



**Conseil économique
et social**

Distr.
GÉNÉRALE

ECE/CES/GE.41/2007/5
22 mars 2007

FRANÇAIS
Original: ANGLAIS

COMMISSION ÉCONOMIQUE POUR L'EUROPE
CONFÉRENCE DES STATISTICIENS EUROPÉENS

COMMISSION DE STATISTIQUE

Groupe d'experts des recensements de la population
et des habitations

Dixième session
Astana, 4-6 juin 2007
Point 3 a) de l'ordre du jour provisoire

PLANIFICATION ET GESTION DES RECENSEMENTS

Gestion de la qualité du recensement en France

Présenté par la France *

La réunion est organisée conjointement avec Eurostat.

Résumé

Le Bureau de la Conférence des statisticiens européens, réuni à Washington les 19 et 20 octobre 2006, a approuvé la reconduction du mandat du Groupe directeur pour les recensements de la population et des habitations et le plan des futures activités de la Conférence en la matière. Il a par ailleurs décidé que le Groupe directeur coordonnerait les travaux entrepris dans le cadre des diverses réunions. Le présent document a été établi à la demande du Groupe directeur pour les recensements de la population et des habitations, et doit être soumis pour discussion à la réunion conjointe CEE/Eurostat sur les recensements de la population et des habitations, qui aura lieu à Astana (Kazakhstan), du 4 au 6 juin 2007. Il renferme des éléments de fond qui serviront de base aux discussions qui auront lieu lors de la séance de la réunion consacrée au thème: «Planification et gestion des recensements».

* Le présent document a été établi par l'Institut national de la statistique et des études économiques (INSEE).

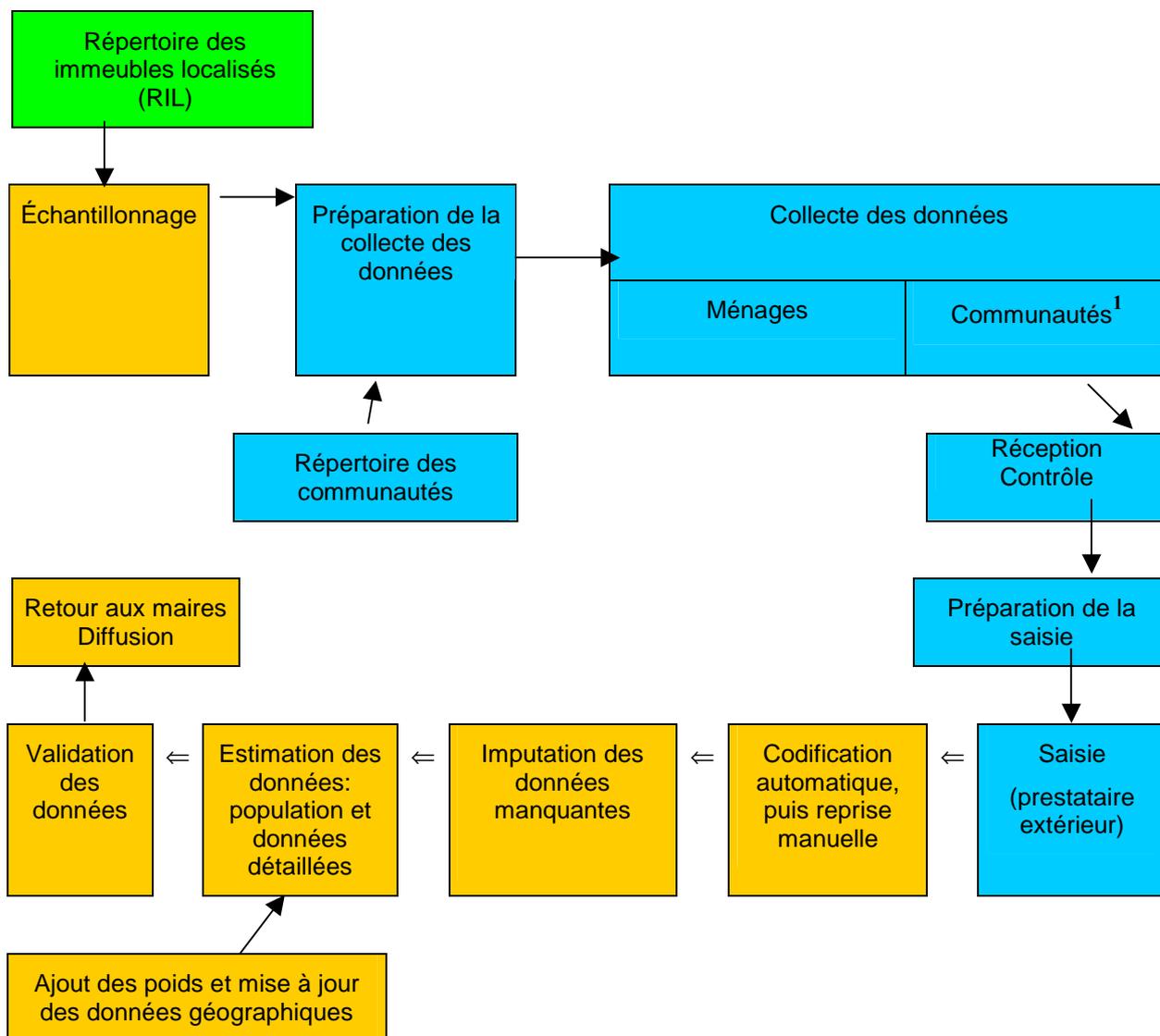
INTRODUCTION

1. Le recensement français repose depuis 2004 sur une nouvelle méthode, alliant deux principes: la collecte tournante sur un cycle de cinq ans et le recours au sondage dans les communes de 10 000 habitants et plus. Le nouveau recensement permettra ainsi de fournir chaque année des résultats détaillés à tous les niveaux géographiques d'intérêt, du pays entier à la commune, voire au quartier. Cette réforme du recensement répond au besoin, de plus en plus marqué, de données démographiques récentes, surtout à des niveaux géographiques fins. Elle permet également de lisser les coûts du recensement, qu'ils soient humains ou budgétaires, et de mieux organiser et contrôler les opérations.
2. La réussite d'une telle méthode repose sur plusieurs préalables, à la fois techniques (disposer de systèmes d'informations géographiques infracommunales et de données administratives pour les mettre à jour et actualiser les données), méthodologiques (techniques d'échantillonnage notamment) et politiques (concertation approfondie avec les communes, partenaires du recensement, et l'ensemble des utilisateurs du recensement).
3. Le recensement tel qu'il se pratique en France répond aux cinq critères essentiels définis par l'ONU, ainsi qu'aux cinq critères généraux de qualité des opérations statistiques (voir le document de fond ECE/CES/2006/24).
4. La qualité est gérée au fur et à mesure du déroulement des opérations de préparation, de collecte, de traitement et de diffusion des résultats du recensement. Elle repose sur des protocoles de contrôle et de mesure de la qualité et sur des bilans mobilisant les différents acteurs du recensement que sont les communes, les acteurs de la collecte, les équipes de l'INSEE et les sous-traitants.
5. Des investigations en cours ou en projet devraient permettre, d'ici à la fin du premier cycle, une analyse quantitative et qualitative de la couverture du recensement.
6. À mi-chemin de son premier cycle quinquennal, le recensement français affiche un bilan positif, mais un travail important reste à accomplir.

I. GESTION DE LA QUALITÉ, PROCESSUS PAR PROCESSUS

7. La qualité du recensement est contrôlée pour chacun des processus: contrôle de la qualité de la base de sondage (Répertoire d'immeubles localisés (RIL) dans les villes de plus de 10 000 habitants); contrôle de la qualité de la collecte des données; contrôle de la saisie et de la codification; contrôle de la qualité des données avant leur diffusion.
8. Pour assurer le contrôle des opérations, l'INSEE a décidé de vérifier ses registres par des enquêtes de terrain et de confronter les données relevées sur le terrain avec les registres. Quant aux opérations de saisie et de codification, elles font l'objet d'une double saisie et d'une double codification sur un échantillon des bulletins, suivies, dans les deux cas, d'une analyse des divergences.
9. Le diagramme ci-dessous montre l'enchaînement des différentes étapes du recensement.

Graphique. LES ÉTAPES DU RECENSEMENT



¹ On entend par «communautés» les ménages institutionnels. Les communautés font l'objet d'une enquête approfondie dont les résultats sont vérifiés à l'aide du «Répertoire des communautés», lequel est mis à jour tous les ans avec le concours de sources extérieures (listes des établissements scolaires, des hôpitaux, etc., fournies par les ministères).

A. Qualité de la base de sondage des grandes communes

10. La mesure de la qualité du Répertoire d'immeubles localisés (RIL) est réalisée grâce à une enquête annuelle.

1. Objectifs et résultats de l'enquête qualité du RIL

11. L'objectif de l'enquête est de mesurer le niveau d'exhaustivité du RIL en termes de déficit et d'excédent d'adresses de catégorie *habitation*, ainsi que de logements associés. Il consiste également à faire en sorte que toutes les adresses relevées sur le terrain soient effectivement inscrites au RIL et, inversement, que toutes les adresses inscrites au RIL existent sur le terrain. Si la base de sondage comportait un déficit trop important, elle ne serait pas véritablement représentative. Si elle comportait un excédent trop important, cela alourdirait inutilement le travail de collecte des données (nombreux déchets). Enfin, un bon équilibre excédent-déficit est nécessaire pour réaliser des estimations sans biais au moment du traitement des données.

12. Depuis deux ans, le taux de déficit a été sensiblement amélioré, passant de 2,3 % en 2003 à 1,1 % en 2005, et à 0,9 % en 2006. L'effort de mise à niveau en 2004 et 2005 a prioritairement porté sur les communes présentant un risque élevé de déficit. Ce sont donc ces communes qui ont enregistré l'amélioration la plus nette. Ainsi, 12 % d'entre elles connaissent encore un déficit supérieur à 2,5 %, contre 22 % en 2004 et 28 % en 2003. À titre de comparaison, dans un recensement général, on estimait à une sur six la part des communes dont le taux d'omission était supérieur à 3 %.

13. Le taux d'excédent a, lui aussi, été réduit, mais dans une moindre mesure, ce qui a entraîné, en 2005, une dégradation du solde. Cette année (en 2006, et encore en 2007), les opérations d'amélioration de la qualité viseront à réduire le taux d'excédent, afin de parvenir à un solde équilibré d'ici à deux ans. Les premiers résultats pour 2006 sont prometteurs: le solde a diminué de 0,2 %, à la faveur d'une baisse des excédents supérieure à celle des déficits.

14. Remarque: lors des estimations de population communale, le processus d'extrapolation des résultats de l'échantillon à l'ensemble de la commune prend en compte les retours de la collecte (correction de la différence entre le nombre de logements inscrits au RIL et celui effectivement relevé). Le défaut de couverture du RIL est par conséquent neutralisé à ce stade.

Tableau. Résultats pour 2006: déficits et excédents d'adresses d'habitation et de logements du RIL

		2003	2004	2005	2006
Adresses	En déficit	2,3 %	1,9 %	1,3 %	1,3 %
	En excédent	2,0 %	1,9 %	1,8 %	1,7 %
	Solde excédent/déficit	-0,3 %	0 %	0,5 %	0,4 %
Logements	En déficit	2,3 %	1,9 %	1,1 %	0,9 %
	En excédent	2,1 %	2,3 %	1,8 %	1,4 %
	Solde excédent/déficit	-0,2 %	0,4 %	0,7 %	0,5 %

15. Le principal enseignement à tirer est que la qualité du RIL est globalement satisfaisante, mais qu'il existe de fortes disparités entre communes.

16. Le travail accompli par les communes sur les listes d'adresses à confirmer apparaît comme un élément aussi déterminant de la bonne qualité du RIL. Cette enquête a également permis de valider le processus de mise à jour: la contribution des deux grands fichiers de mise à jour, le fichier des permis de construire et le fichier des adresses de La Poste (très utiles pour identifier les adresses disparues), est indéniable lorsqu'elle a pu être couplée avec les réponses des communes.

Enquête qualité du RIL: mode opératoire

L'échantillon est très large, et se compose de 190 000 adresses et 520 000 logements (soit plus de 3 % des logements).

Les enquêteurs disposent de la liste de toutes les adresses telle qu'elle figure dans le RIL: habitations et autres catégories (établissements, équipements et «autres»). Ils contrôlent l'existence de toutes ces adresses, avec des consignes de priorité sur les habitations. Une adresse d'habitation figurant sur la liste mais absente sur le terrain constitue un excédent. Tel est également le cas d'une adresse qui, sur le terrain, s'avère n'être qu'un accès secondaire. Une adresse d'habitation qui existe sur le terrain mais est absente de la liste constitue un déficit. Les enquêteurs vérifient également l'intitulé de l'adresse, le nombre de logements, les confusions de catégories et les adresses comportant plusieurs accès.

2. Amélioration de la qualité du RIL: plans d'action décentralisés, pilotés par les directions régionales

17. L'enquête qualité du RIL permet de fixer des objectifs pour améliorer le répertoire, et ces objectifs sont décentralisés; chaque direction régionale élabore, en concertation avec l'Unité Recensements de la direction générale, un plan d'action avec des objectifs en fonction des points faibles de sa région. La mise en œuvre de ces plans d'action se fait avec la participation des directions régionales de l'INSEE, mais aussi des communes, car la qualité du répertoire conditionne celle du recensement sur leur territoire. Nous avons donc une gestion opérationnelle de la qualité. Les actions correspondantes peuvent être les suivantes: renforcement de la formation des acteurs communaux; ciblage des travaux de terrain (INSEE ou communes) sur certains quartiers ou certains types d'habitat (par exemple les adresses non normalisées telles que les lieux-dits); expertise approfondie de certains fichiers.

B. Collecte de données

18. La collecte de données représente souvent le maillon le plus fragile du processus statistique. Sa qualité repose sur des procédures précises et rigoureuses, mais aussi sur l'adhésion des répondants (qui conditionne la sincérité de leurs réponses) et l'engagement de ceux qui recueillent l'information, à savoir les agents recenseurs. Une préparation rigoureuse (repérages, formations) et des contrôles stricts permettent d'en assurer la qualité. Des contrôles sont réalisés par les communes et l'INSEE en cours de collecte. Une fois les données collectées, le contrôle

est du ressort de l'INSEE; il permet de vérifier les données collectées et, le cas échéant, de les corriger.

1. Organisation des enquêtes de recensement

19. Les enquêtes du recensement sont menées en France sur la base d'un partenariat entre l'INSEE et les communes. Ce partenariat est défini dans la loi qui institue le recensement et précisé dans plusieurs textes d'application, dont l'élaboration a reposé sur une concertation approfondie avec les communes.

20. L'INSEE supervise et organise les enquêtes, alors que les communes les préparent et les effectuent. Les communes sont chargées, en particulier, de recruter et rémunérer les agents recenseurs, de les encadrer (par le biais d'une personne spécialement désignée par leurs soins et appelée «coordonnateur communal»), de mener les opérations de collecte (repérage des logements, dépôt et retrait des questionnaires par les agents recenseurs) et de mener un premier contrôle de la qualité (notamment de l'exhaustivité de la collecte).

21. Pour encadrer et surveiller les opérations de terrain, l'INSEE s'appuie sur des superviseurs, des agents de l'INSEE responsables d'un secteur géographique donné et d'un ensemble de communes. Ces superviseurs doivent notamment veiller à la mise en place, dans les communes, des moyens nécessaires aux enquêtes de recensement, former les coordonnateurs communaux et les agents recenseurs (pour ces derniers, la formation est assurée conjointement avec la commune), de suivre, avec les coordonnateurs communaux, le déroulement de la collecte, d'aider ceux-ci à résoudre les problèmes rencontrés et, enfin, de procéder avec les communes à la clôture de la collecte. On dénombre en tout 550 superviseurs, dont 50 sont recrutés spécifiquement par l'INSEE aux fins de cette opération et pour toute la durée de l'enquête.

2. Préparation de la collecte et contrôles effectués au fil de la collecte

22. Le mode opératoire de collecte repose sur un protocole précis, dont les agents recenseurs sont responsables, sous le contrôle du responsable communal et du superviseur. Le manuel de l'agent recenseur explicite les tâches à accomplir. Une documentation détaillée est remise aux responsables communaux et aux superviseurs. Pendant la collecte et à la fin de celle-ci, la commune vérifie que l'agent recenseur est bien passé dans chaque logement à recenser. Aucune adresse, aucun logement ne doit avoir été oublié.

23. Les superviseurs effectuent par ailleurs des contrôles lorsqu'ils se rendent dans les communes. Ils vérifient que les procédures ont bien été respectées (confidentialité, respect du calendrier) et que les questionnaires déjà collectés ont bien été remplis, afin de déceler les erreurs de collecte les plus fréquentes. En fin de collecte, lorsqu'ils transmettent les questionnaires à l'INSEE, ils fournissent une évaluation de la qualité de la collecte sur la commune et font part des éventuelles difficultés rencontrées.

24. Outre ce dispositif de base, les communes, qui sont responsables de la collecte, mettent en place un dispositif de suivi et de contrôle plus précis.

25. En amont de la collecte, les adresses de l'échantillon à recenser sont vérifiées en deux temps: validation de l'adresse et du nombre de logements par l'équipe centrale de la commune,

puis repérage sur le terrain par l'agent recenseur, qui doit authentifier les adresses et comptabiliser les logements (cette phase permet également à la commune de tester l'agent recenseur). Un deuxième repérage sur le terrain est réalisé en cas de désaccord entre l'agent recenseur et l'équipe centrale (un seuil d'écart peut être toléré, en fonction des moyens dont dispose l'équipe communale). Ainsi, lorsque le travail de collecte s'engage, il y a accord entre l'INSEE, le coordonnateur communal et l'agent recenseur sur l'existence d'une adresse, son classement en adresse habitée et le nombre de logements à y recenser. La rigueur de cette vérification en amont facilite le travail de collecte de l'agent recenseur.

26. Le suivi de la collecte, au fil des opérations et pour chaque agent recenseur, permet de vérifier l'avancement de la distribution et du retour des imprimés et de détecter certaines difficultés rencontrées par des agents recenseurs, ou des carences. Pour cela, l'INSEE fournit aux communes un outil de contrôle et de suivi de la collecte, mais les communes peuvent aussi, si elles souhaitent des fonctionnalités plus riches, développer ou faire développer leur propre outil. L'examen des premiers documents retournés par l'agent recenseur permet, par l'analyse de la qualité des réponses, d'évaluer la compréhension des consignes par l'agent recenseur et son niveau d'engagement dans la collecte.

27. Certaines communes procèdent à un bilan hebdomadaire avec chaque agent recenseur, d'une durée d'environ deux heures, dont le but est de s'assurer que toutes les adresses, tous les logements qui y figurent et toutes les personnes qui y résident sont bien recensés. Chaque semaine, l'agent recenseur est tenu de se conformer à des objectifs chiffrés (imprimés distribués, documents collectés en retour). À la fin de cette période, les dernières difficultés de collecte sont analysées. Plusieurs indicateurs sont suivis: nombre et pourcentage de Fiches de logement non enquêté (FLNE) (ces fiches doivent être établies le plus tard possible, en phase de clôture des opérations), taux de vacance des logements, qualité du carnet de tournée.

3. Après la collecte, les contrôles de l'INSEE

28. Après réception des documents transmis par les communes, des vérifications sont menées en bureau, puis sur le terrain. Le protocole de contrôle est adapté à la nouvelle configuration des enquêtes de recensement. Par rapport à ce qui était le cas pour les recensements généraux, il est désormais possible de réaliser des vérifications plus approfondies pour chaque commune, en raison du moindre volume de données collectées chaque année et du nombre moins élevé de communes visées. Le fichier de la taxe d'habitation joue un rôle important dans les contrôles en bureau et permet de vérifier l'exhaustivité de la collecte au niveau des communes ou des zones de collecte données. Ces contrôles, en bureau ou sur le terrain, sont menés «au plus près de la collecte». Les erreurs qu'ils permettent de détecter sont corrigées.

29. Ils sont sélectifs, car ils résultent de «tamis» successifs:

a) Un premier contrôle effectué lors de la réception et de l'enregistrement des questionnaires permet d'évaluer la qualité pour chaque commune.

Cette étape concerne toutes les communes visées aux fins de l'enquête de l'année. Elle constitue la phase dite «réception-enregistrement des questionnaires». Les agents de l'INSEE procèdent en bureau au décompte des questionnaires par enregistrement de codes à barres. Ils vérifient les adresses, les bulletins vierges. Ils établissent des indicateurs par

commune sur l'évolution du nombre de logements et de personnes, la structure du parc de logements (résidences principales, résidences secondaires, logements occasionnels, logements vacants), et la proportion de FLNE. À l'issue de cette étape, ils dressent un état des lieux pour chaque commune en tenant compte de ces indicateurs quantitatifs et des indicateurs qualitatifs sur le déroulement de la collecte fournis par les superviseurs et de leur appréciation sur la réception des documents. Ils attribuent à chaque commune, sur la base d'une échelle de notation, une note «qualité», et soumettent à un contrôle approfondi celles dont les notes sont les plus faibles (environ 10 % des communes).

b) Cette appréciation débouche, pour environ 10 % des communes, sur un contrôle approfondi en bureau.

Ce contrôle approfondi peut porter soit sur la commune entière, soit sur une portion de commune. Il s'appuie majoritairement sur les fichiers de taxe d'habitation et est orienté en fonction du problème mis en évidence par les indicateurs

- i) Sur l'exhaustivité du nombre de logements:
 - a. Présence ou absence de l'adresse,
 - b. Comparaison du nombre de logements,
 - c. Comparaison des noms des occupants (communes de plus de 10 000 habitants);
- ii) Sur le pourcentage de fiches de logement non enquêté (FLNE):
 - a. Vérification des FLNE à partir des informations (catégorie de logement) contenues dans le fichier de la taxe d'habitation,
 - b. Nombre supposé de personnes résidant dans le logement;
- iii) Sur la structure par catégorie de logement (principal, secondaire, vacant): confrontation avec le fichier de la taxe d'habitation.

Il peut également être procédé à des enquêtes de contrôle par téléphone auprès des habitants.

Ces contrôles permettent soit de confirmer les informations relevées, soit de mettre en évidence des erreurs, qui sont alors corrigées.

c) Enfin, les cas non résolus en bureau font l'objet de vérifications sur le terrain menées par l'INSEE (environ 30 000 vérifications effectuées sur le terrain en 2006).

Si les divergences sont trop nombreuses et s'il est impossible de les régler en bureau, les communes font entièrement l'objet de contrôles sur le terrain. La commune est informée de cette décision, mais elle ne reçoit pas la liste des adresses ou des logements concernés.

Les contrôles sur le terrain portent sur des communes pour lesquelles la qualité est jugée insuffisante au moment du contrôle en bureau. Dans la grande majorité des cas, les contrôles confirment le résultat de la collecte. Ils permettent aussi de déceler et corriger certaines erreurs. Ces erreurs n'induisent pas de biais systématique, ni à la hausse ni à la baisse. Ils donnent lieu, le cas échéant, à des corrections de questionnaires, et les communes en sont informées. Ils permettent aussi de procéder aux ajustements requis et de modifier les protocoles de collecte ou de vérification pour l'année suivante.

Ils portent soit sur des adresses, soit sur des logements.

Dans le cas des adresses, on contrôle les adresses qui ne semblent pas correspondre à celles demandées dans l'échantillon (communes de plus de 10 000 habitants), ou qui risquent d'avoir été oubliées (communes de moins de 10 000 habitants). Il y a alors aussi vérification du nombre de logements.

Sept mille cinq cents contrôles d'adresse ont été menés sur le terrain en 2005, ce qui représente 1,6 adresse pour 1 000 pour les communes de moins de 10 000 habitants, et 8,6 pour 1 000 pour les communes de plus de 10 000 habitants.

Ils portent également sur des logements lorsqu'il y a doute sur leur catégorie (principal, secondaire, vacant ou occasionnel). Il y a alors aussi vérification du nombre de personnes. Environ 14 400 contrôles sur les logements ont été menés en 2005, soit environ un contrôle pour 300 logements. Une grande majorité porte sur les logements ayant fait l'objet de FLNE.

En 2006, l'INSEE a développé ses efforts sur les contrôles de logements: 23 500 contrôles de logements ont été réalisés par les directions régionales (soit plus d'un pour 200). Comme en 2005, une proportion importante de ces contrôles a porté sur des logements pour lesquels une FLNE avait été établie.

Après ces contrôles, le taux définitif des FLNE était de 3 % environ. Il a décliné régulièrement depuis 2004. Les contrôles nous permettent de confirmer la pertinence du système de la FLNE dans la prise en compte, pour les résultats statistiques (chiffres de la population et résultats détaillés), des ménages qui n'ont pas pu être contactés directement: ils confirment que des FLNE ont été établies pour des résidences principales, et non pour des résidences secondaires ou des logements vacants, et que l'estimation du nombre de personnes vivant dans ces logements est fiable.

30. En conclusion de cette revue des contrôles de la collecte de données, il faut souligner que cet aspect représente un des principaux avantages de la nouvelle méthode: en raison de la moindre taille de chaque enquête de recensement, et aussi grâce au système des tamis successifs, les contrôles sont beaucoup plus efficaces que dans un recensement classique. Les communes en sont conscientes, ce qui a un effet dissuasif par rapport à d'éventuelles tentatives de fraude.

C. Traitements

31. La **saisie** est effectuée par un prestataire externe. Il y a saisie optique des questionnaires, qui sont scannés, puis soumis à un logiciel de reconnaissance des formes. Une saisie des libellés mentionnés en clair complète le processus. Un cahier des charges précis définit ce qu'on attend

du prestataire en termes de qualité. Les spécifications concernent à la fois la sécurité des données, pendant les transports et chez le prestataire (sécurité physique et informatique, engagement à ne pas divulguer) et la qualité des données (nombre de documents saisis par zone de collecte, taux d'erreur maximal par variable). Il y a comptage des documents et contrôle des écarts avec le premier comptage réalisé à l'INSEE au moment de l'enregistrement des codes à barres (voir par. 29 a)).

32. Les erreurs de saisie sont évaluées par double saisie sur un échantillon de bulletins, chez un second prestataire, qui travaille à partir des images scannées en utilisant les mêmes règles de saisie. Les divergences sont analysées à l'INSEE, de façon à établir les taux d'erreurs sur le codage de chacune des variables. La double saisie est réalisée au fil de la campagne de saisie, ce qui permet, éventuellement, de corriger les protocoles pour les lots suivants.

33. Des contacts et des réunions régulières sont organisés avec le prestataire, afin d'anticiper les problèmes ou de les résoudre au plus vite. Il est demandé au prestataire de procéder aux ajustements ou aux améliorations éventuels, en cours de campagne ou entre deux campagnes, s'il s'agit d'améliorations plus lourdes, suite aux enseignements de la double saisie. Par exemple, l'optimisation des outils de reconnaissance de forme a permis de réduire sensiblement le taux d'erreur de saisie des dates de naissance. À ce sujet, il convient de noter que, même sur des données aussi simples que des dates de naissance, il est illusoire de s'attendre à un taux d'erreur zéro: l'information étant manuscrite au départ, des erreurs d'interprétation sont inévitables.

34. Le **codage** comprend d'abord une phase automatique de codification selon des nomenclatures d'activité et de profession, notamment, puis une phase de reprise par des opérateurs en direction régionale des cas non codés. On développe actuellement un dispositif de contrôle de qualité sur ce processus. Le principe est de procéder à une seconde codification sur un échantillon de bulletins représentatifs des différents modes de codification (automatique et reprise), puis d'analyser les divergences.

35. L'objectif est de mesurer la qualité du codage automatique et celle de la reprise manuelle (pour obtenir dans un premier temps un pourcentage global des cas bien codés, puis, à terme, des pourcentages par grandes modalités des variables, tels que le pourcentage de cadres supérieurs bien codés).

36. Le deuxième objectif du contrôle qualité du codage est de mesurer un taux de bulletins individuels non codables (par exemple un intitulé vague de type fonctionnaire, sans autre indication), ce qui permettra de se donner une cible réaliste en matière de qualité et d'éviter la «surqualité».

37. Ce dispositif de mesure et de contrôle permettra également d'enrichir les fichiers d'apprentissage qui servent à la codification automatique, afin d'en améliorer la performance, de mieux cibler la formation des codeurs de la reprise et d'assurer finalement une bonne gestion de la qualité.

38. **Redressements.** Il s'agit d'analyser les non-réponses, totales ou partielles, afin de les imputer, mais aussi d'analyser les bulletins qui présentent des incohérences (par exemple une personne de 95 ans qui déclarerait être active). La procédure la plus fréquente est la méthode hot-deck, qui consiste à compléter l'information manquante en s'aidant de la réponse d'un

«donneur» qui lui ressemble. Là encore, les redressements s'affinent d'année en année à partir des résultats précédents, ce qui constitue un moyen efficace d'amélioration de la qualité.

39. En particulier, le redressement des FLNE permet d'imputer, pour le calcul de la population des communes, la population des résidences principales pour lesquelles le contact avec l'agent recenseur n'a pas eu lieu (absences de longue durée, personnes injoignables, refus). Pour les résidences principales dont le nombre de personnes est indiqué (dans plus de 8 cas sur 10), on «impute» pour le logement autant d'individus que la FLNE en indique dans le ménage. Pour les autres FLNE, il est également procédé à une imputation, suivant une méthode visant à reproduire la structure, par taille de ménage, observée en moyenne nationale pour les FLNE comportant des données. Pour tenir compte du fait qu'un nombre restreint de FLNE sont remplies à tort pour des résidences non principales, ce redressement est fait à concurrence de la part des résidences principales dans la commune. En 2005, l'impact du redressement des FLNE était de 1 025 000 personnes dans les communes de plus de 10 000 habitants, et de 441 000 personnes dans celles de moins de 10 000 habitants. L'impact de l'«abattement» décrit plus haut est d'environ 90 000 personnes.

D. Diffusion

40. La qualité de cette étape doit beaucoup à la préparation, qui en a été faite avec les utilisateurs dans le cadre des groupes du Conseil national de l'information statistique (CNIS). Deux groupes ont ainsi travaillé, un sur la diffusion générale, l'autre sur la diffusion en direction des communes. La diversité des utilisations du recensement selon les demandeurs (communes, chercheurs, administrations, sociétés privées) a été mise en évidence, ainsi que la pédagogie nécessaire pour présenter les données. En effet, ce seront des moyennes ou des données lissées, et pas des données brutes, et leur précision ne permettra pas tous les croisements théoriquement possibles du fait du recours au sondage.

41. Après la diffusion, des enquêtes seront menées pour vérifier l'adéquation aux besoins, dans les groupes du CNIS et par une enquête de satisfaction auprès d'utilisateurs.

42. La qualité des chiffres diffusés repose aussi sur une étape de validation systématique avant la diffusion, qui permet de voir s'il ne reste pas de résultats aberrants.

43. De cette étape de validation on a aussi des enseignements à tirer pour la diffusion des résultats des années suivantes.

II. PERSPECTIVES POUR L'ANALYSE DE LA COUVERTURE DU RECENSEMENT

44. On l'a vu, un travail important de contrôle *in itinere* est mené au moment de la collecte du recensement et dans les trois mois qui suivent la réception des questionnaires. Ces contrôles donnent lieu à correction immédiate de la feuille de logement et du ou des bulletins individuels concernés.

45. Au niveau global, l'analyse de la couverture du recensement prendra plusieurs formes:

a) Une analyse du surdénombrement, fondée sur la détection de personnes recensées en double. Pendant la collecte des données, on peut identifier les personnes qui ont deux lieux de

résidence, les étudiants en particulier, et leur réserver un traitement particulier; le recensement français a retenu à cet effet la notion de «double compte»: les étudiants de moins de 25 ans² sont comptés avec la population des deux communes concernées, celle de leurs parents et celle où ils font leurs études. Aux fins du décompte de la population des communes ou de l'établissement de statistiques structurelles, les étudiants de moins de 25 ans ne sont pris en compte que dans une seule commune:

- i) La commune de leurs parents s'ils ont moins de 18 ans;
- ii) La commune où ils résident pour faire leurs études s'ils sont âgés de 18 à 25 ans;
- iii) Ce système existe en France depuis le recensement de 1962;
- iv) Pour les autres personnes dont on ne pouvait déterminer qu'elles avaient une «double résidence» au moment de la collecte de données, on s'appuiera sur l'échantillon démographique permanent (EDP), échantillon au 1/100 de la population française, alimenté par les données d'état civil et les formulaires de recensement. La procédure d'alimentation de cet échantillon permet de détecter si plusieurs bulletins sont établis pour une même personne. Compte tenu de la méthode de recensement, il convient d'avoir les cinq campagnes pour en tirer des enseignements fiables. Ces enseignements porteront sur une estimation du volume global de surdénombrement et permettront de caractériser les populations concernées, notamment en termes de résidence (grande commune, petite commune, communauté) ou de sexe et âge. Du fait du sondage, du recensement et surtout de l'EDP, les résultats n'auront qu'une portée nationale.

b) Une analyse des omissions, qui pourra s'appuyer sur des fichiers administratifs de couverture universelle tels que les fichiers de type «assurance maladie». Une première maquette sera proposée en 2007.

46. Ces investigations, centrées sur la collecte, compléteront celles, décrites à la section IA (qualité du RIL).

47. Ces pistes sont plus prometteuses que celles offertes par une enquête postcensitaire du type de celles qui ont été menées en France en 1962 et 1990. Il paraît en effet plus pertinent de contrôler la qualité d'une enquête en la confrontant à une source administrative plutôt qu'à une autre enquête, qui présenterait les mêmes imperfections (en particulier les biais dus aux difficultés à contacter les ménages).

III. À MI-CHEMIN DU PREMIER CYCLE DE COLLECTE, UN PREMIER BILAN POSITIF

48. Après trois collectes, il est possible de dresser un premier bilan positif de la mise en place du nouveau recensement. L'adhésion des communes et des habitants à la nouvelle méthode est très large. Toutes les communes ont participé au recensement en 2005 et 2006 et les deux

² Au-dessus de 25 ans, ils ne sont comptés que dans la commune où ils font leurs études.

communes qui avaient refusé de participer en 2004 ont pris part au recensement les années suivantes.

49. Quant aux habitants, on pouvait craindre que le recours au sondage n'entraîne une moindre mobilisation que lors des recensements généraux. Il n'en a rien été, grâce à une communication très efficace, au niveau tant national que local. Cette communication a fortement insisté sur le changement de méthode et sur l'utilité du recensement. Elle a été portée par la mise à disposition de résultats des premières enquêtes annuelles, qui a permis de montrer par l'exemple que la nouvelle méthode fonctionnait et quels types de résultats elle pouvait fournir. Pour les trois années 2007 à 2009, le changement de méthode étant supposé assimilé par le public, la communication se situe désormais sur le registre de l'utilité, avec également un message renforcé sur le côté pratique: comment le recensement se passe concrètement, où se faire recenser en cas de double résidence...

50. La qualité des bases de sondage et l'exhaustivité de la collecte se sont améliorées au fil des campagnes. Le taux de déficit du RIL a ainsi été ramené à 0,9 % des logements. La part des ménages qui n'ont pu être recensés directement, et donc pour lesquels a été établie une fiche de logement non enquêté, est très faible, de l'ordre de 3 %. Dans la plupart des cas, il s'agit de ménages impossibles à joindre (personnes en congés ou hospitalisées, ou ayant des horaires très décalés rendant difficile un contact avec l'agent recenseur...). Les effectifs correspondant à ces ménages sont pris en compte pour l'établissement des statistiques.

51. Enfin, dès la fin de 2004, l'INSEE a pu restituer de premières estimations de population à toutes les communes de moins de 10 000 habitants qui avaient été enquêtées, ainsi qu'à une centaine de grandes villes. Ces résultats ont ensuite été diffusés largement via Internet, deux jours avant le début de la collecte 2005.

52. Fin 2005, l'INSEE a étendu ce dispositif aux communes de moins de 10 000 habitants enquêtées en 2006 ainsi qu'à une centaine de grandes villes supplémentaires. Ces résultats ont été rendus publics le 17 janvier 2006. Le mouvement s'est poursuivi fin 2006 et début 2007, avec les communes enquêtées en 2006 et environ 200 grandes communes supplémentaires.

53. Cette diffusion de premiers résultats a très largement contribué à la conviction des utilisateurs quant à l'intérêt de la nouvelle méthode.

54. Sur le plan organisationnel, les effets positifs de l'annualisation de la collecte commencent à se faire sentir: amélioration des organisations, des applications informatiques, professionnalisation des différents acteurs, à l'INSEE comme dans les communes.

55. Un important travail de consolidation, permis par l'expérience des premières collectes, reste à mener avant la fin du premier cycle, pour atteindre le régime de croisière. Pour la suite, la réflexion en matière d'améliorations s'orientera dans deux directions: l'ouverture à la collecte par Internet et l'évolution des questionnements, pour les adapter aux nouvelles attentes des utilisateurs et à la disponibilité de nouvelles sources administratives, notamment en matière d'emploi.
