

CONSEIL NATIONAL DE L'INFORMATION STATISTIQUE

Formations Démographie, conditions de vie et Statistiques régionales et locales

Rapport du groupe de travail

**Utilisation des données produites
par le recensement rénové de la population
et leur diffusion**

Président : Jean-Pierre GIBLIN

Rapporteurs : André de LOS SANTOS, Jean-Michel DURR

Décembre 2005

SOMMAIRE

1. INTRODUCTION	4
2. LE PANORAMA DES UTILISATIONS	4
2.1 Les deux grandes catégories de résultats du recensement	5
2.2 Les exemples d'utilisation des résultats du nouveau recensement	6
a. L'utilisation au niveau régional et national	7
b. L'utilisation sur des territoires multi-communaux	7
c. L'utilisation au niveau communal et infra-communal	8
d. Le cas des données bi-localisées	10
e. La construction neuve	11
2.3 Les propositions du groupe de travail	11
3. LA QUALITÉ DES RÉSULTATS	16
4. LES GRANDES LIGNES DE PRODUITS DE DIFFUSION	17
4.1 Les principes de la politique de diffusion de l'Insee pour les prochaines années	17
a. Travailler avec les utilisateurs pour optimiser les produits et services en fonction de leurs besoins	17
b. Faciliter, dans toute la mesure du possible, l'accès aux données statistiques et leur utilisation	18
4.2 Contraintes et opportunités	18
a. L'annualité des résultats	18
b. La nouvelle technique de recensement	19
c. La popularisation des raccords à internet et au haut débit	19
4.3 Les grandes lignes de produits	20
a. Les chiffres clés	20
b. Les analyses et synthèses	22
c. Les résultats détaillés	23
d. Les fichiers-détail	24
e. Les tabulations sur mesure	24
4.4 La communication sur les produits et services de diffusion du recensement	25
5. LA MISE À DISPOSITION DES RÉSULTATS	25
5.1 Le schéma général	25
5.2 Publier une estimation de la population des communes avant la fin de l'année 2004	25
5.3 Produire les premiers résultats nationaux avec quelques indicateurs régionaux avant le démarrage de la campagne 2005	25
5.4 Organiser avant l'été 2005 les premières restitutions statistiques aux communes	25
5.5 Démarrer la diffusion de données régionales avant l'été 2005	26
5.6 Avant la fin mars 2006	26
6. LA RESTITUTION D'INFORMATIONS AUX COMMUNES	26
7. LE RYTHME SOUHAITABLE DE MISE À JOUR DES ZONAGES	29
8. LA DIFFUSION DE RÉSULTATS SUR DES ZONAGES À LA DEMANDE	29
9. PROPOSITIONS CONCERNANT LA DIFFUSION DU RIL	30
10. LISTE DES ANNEXES	33

1. INTRODUCTION

Le groupe de travail sur les données produites par le nouveau recensement et leur diffusion a été mis en place à la demande des formations « Démographie, conditions de vie » et « Statistiques régionales et locales » du CNIS (cf. mandat du groupe de travail en annexe I).

En effet, la rénovation du recensement de la population va entraîner des changements dans l'utilisation des données produites ainsi que dans leur fréquence de mise à disposition. Ces changements vont nécessiter un temps d'apprentissage pour les utilisateurs mais aussi pour les diffuseurs de l'information. Dans ce contexte, il était demandé au groupe de travail que les options et propositions émises intègrent un souci important de pédagogie.

Le groupe a tenu huit réunions :

- le 27 juin 2003,
- le 24 septembre 2003,
- le 17 novembre 2003,
- le 27 janvier 2004,
- le 23 mars 2004,
- le 28 avril 2004,
- le 16 juin 2004,
- le 13 septembre 2004.

Les réunions ont été l'occasion de présenter en détail la méthodologie du nouveau recensement et de construire des exemples d'utilisation tirés de la pratique des utilisateurs. A cet effet, deux sous-groupes ont été constitués. L'un s'est attaché à décrire les problématiques d'utilisation au niveau régional, l'autre au niveau communal et infra-communal.

Le groupe a produit un certain nombre de recommandations présentées dans ce rapport. Ce rapport reprend dans sa structure les thèmes de son mandat.

Remarque : dans le corps du rapport, on parlera des « résultats » du recensement et non pas des « données », de manière à réserver ce dernier mot à la désignation des informations élémentaires - qui ne sont pas communicables parce que couvertes par le secret statistique - dont l'agrégation conduit aux résultats mis à la disposition des utilisateurs.

2. LE PANORAMA DES UTILISATIONS

En premier lieu, le groupe de travail a mis en évidence une large diversité d'utilisation des résultats des recensements de population.

Il est utile de rappeler tout d'abord que l'objet du recensement est, outre la détermination de la population légale de toutes les circonscriptions administratives, la description des caractéristiques démographiques et sociales de la population et des caractéristiques des logements à tous les niveaux géographiques du territoire, du quartier (IRIS) dans les villes jusqu'au pays dans son ensemble. De plus, le recueil simultané d'informations portant sur la population et les logements offre de multiples possibilités de croisement, comme par exemple de mettre en rapport les caractéristiques des personnes avec celles de leur logement. De plus, par le recueil de données de localisation comme le lieu de travail ou le lieu de résidence cinq ans auparavant, le recensement peut fournir une description « au lieu de travail » des actifs travaillant dans une commune ou un territoire ou « au lieu de résidence antérieure » des personnes ayant quitté la commune ou la zone concernée.

Pour les **communes**, le recensement fournit tout d'abord l'effectif officiel de la population, dont l'importance est primordiale de par ses implications, en termes financiers notamment. Au-delà de l'effectif, le sens de la variation de la population est examiné attentivement par les élus municipaux, une augmentation de la population étant souvent connotée positivement. De même, les collectivités locales sont grandement intéressées par les phénomènes touchant à la dynamique démographique, et notamment la mesure des flux migratoires. A cet égard, l'actualisation annuelle de ces données présente un intérêt certain pour la gestion communale, et en particulier il est jugé intéressant de disposer tous les ans d'éléments de structure concernant la pyramide des âges, le taux de chômage ou le taux de nouveaux arrivants. A cette connaissance générale au niveau de la commune, s'ajoute une préoccupation de suivi des quartiers dans le cadre de la gestion infracommunale, sur des territoires dont le périmètre varie sensiblement selon les utilisations (*périmètre scolaire, transports, services de proximité...*). Enfin, les communes sont de plus en plus impliquées dans des **structures intercommunales**, ce qui nécessite des informations homogènes et comparables entre les différentes communes et leurs quartiers. A cet égard, les **agences d'urbanisme** sont particulièrement concernées par les problématiques d'agglomération et de déplacements urbains, mais aussi dans le cadre plus large des aires urbaines.

Les **collectivités régionales** et les organismes ayant une problématique régionale, tels l'IAURIF, sont intéressés par les migrations résidentielles interrégionales, par sexe et âge ou à défaut par des matrices de migration sur des zones plus larges. Les migrations d'emploi et de retraite constituent notamment des enjeux importants pour les régions. Les régions sont demandeuses de statistiques dans le domaine de l'emploi et de la formation, domaines dans lesquels elles ont des compétences étendues. Naturellement, leurs préoccupations ne se limitent pas à l'échelon régional mais portent également sur les niveaux infra-régionaux.

Les **chercheurs** ont des attentes proches de certains grands utilisateurs de la statistique publique. Plusieurs types de problèmes peuvent être énumérés. Les aspects géographiques, tout d'abord, notamment la possibilité de reconstituer des agrégats. Ensuite, la question de la comparaison entre le nouveau recensement et les précédents. Les chercheurs s'intéressent en effet plus spécifiquement à la comparabilité des données obtenues. Enfin, l'accès aux variables sensibles est également un sujet de préoccupation de la part des chercheurs, pour qui elles constituent un thème important d'étude.

Les **administrations** recouvrent un champ très varié d'utilisation des résultats du recensement en fonction de leurs domaines de compétence : territoires ruraux ou urbains, gestion de l'eau, répartition des hommes et des activités sur les territoires, politique de la ville, évaluation des politiques publiques ; etc. Elles se retrouvent en particulier sur le besoin de disposer d'informations homogènes sur l'ensemble du territoire, structurelles et en évolution, et en particulier de résultats bilocalisés de migration ou de déplacement domicile-travail.

Enfin, des **sociétés privées** dans le domaine du **géomarketing** réalisent des produits à partir des résultats détaillés du recensement, le plus souvent au niveau infra-communal, qui s'adressent à différents types de clients : des entreprises, qui utilisent ces informations à des fins de marketing ou de constitution de dossiers administratifs (cas des implantations commerciales), mais également des collectivités locales. Elles expriment un fort besoin d'actualisation des données, mais aussi de disposer de bases de données de format adapté et à faible coût.

2.1. Les deux grandes catégories de résultats du recensement

Le nouveau recensement est basé sur le principe des enquêtes répétées annuellement sur des populations différentes. Ainsi, les communes de moins de 10 000 habitants sont réparties en cinq groupes, de même que les adresses des communes de 10 000 habitants ou plus. Les groupes de communes de moins de 10 000 habitants sont recensés successivement chaque année. Dans les communes de 10 000 habitants ou plus, les groupes d'adresses sont enquêtés successivement chaque année avec un taux de sondage de 40 %. Au bout de cinq ans, tout le territoire a été pris en compte et l'ensemble de données ainsi constitué couvre donc la totalité des communes de moins de 10 000 habitants, et 40 % des communes de 10 000 habitants ou plus.

La construction des résultats du recensement va utiliser ces données de deux manières :

- à **tous les niveaux géographiques**, on utilisera l'ensemble des données collectées au cours des cinq dernières années. Chaque année A, les données les plus anciennes (celles de l'année A-5) seront remplacées par les données collectées en A. L'ancienneté des données est donc toujours la même, de trois ans à quatre ans. En effet, la disponibilité des données interviendra en fin d'année, et elles seront utilisées pendant un an avant d'être renouvelées ;
- pour la **France et les régions**, le fait que les groupes de rotation des communes de moins de 10 000 habitants soient équilibrés au niveau régional va permettre de produire, avec les seules données collectées d'une année, des résultats significatifs à ces niveaux de manière standard. Les résultats auront dans ce cas une ancienneté de 1 à 2 ans, selon la date d'utilisation.

La première catégorie de résultats est dénommée « **résultats du recensement** » car elle constitue l'essentiel de la production du recensement. La seconde est appelée « **résultats de l'enquête annuelle de recensement** ». Les différents niveaux d'utilisation font l'objet du tableau de la proposition 1 (cf. infra).

Selon les niveaux géographiques, il serait envisageable de produire des résultats à partir de deux, trois ou quatre années de collecte, afin de d'accroître leur fraîcheur. C'est le choix du bureau du Census américain pour la diffusion de l'*American Community Survey*, qui remplace le questionnaire long du recensement des Etats-Unis. L'Insee a écarté cette possibilité dans l'immédiat afin de ne pas compliquer l'utilisation des résultats du nouveau recensement. Elle sera cependant utile pendant les quatre premières années, avant d'atteindre le régime de croisière, afin de produire de premiers résultats pour les communes de 10 000 habitants ou plus.

Le cumul des données des cinq dernières années correspond à une moyenne (mobile) des résultats sur la période de cinq ans. On peut éventuellement, à l'aide de modèles statistiques, se rapprocher de la situation du milieu de la période, c'est-à-dire l'année A-2.

De plus, les utilisateurs de résultats de recensement se souviennent certainement que l'exploitation des données n'est pas nécessairement exhaustive : ce n'est le cas que pour les variables dites "principales". Les variables les plus complexes, comme la profession détaillée, la structure des ménages, etc., sont exploitées à partir d'un échantillon d'un quart de la base globale. En effet, la détermination de la profession détaillée, par exemple, est assez lourde. Les algorithmes de codification automatique générant des exceptions, une reprise manuelle est nécessaire. L'Insee ne serait pas en mesure de mener à bien l'exploitation exhaustive des données en une année.

Avec le nouveau recensement, le principe du sondage au quart pour ces variables est conservé pour les communes de moins de 10 000 habitants. En revanche, pour les communes de 10 000 habitants ou plus, qui font déjà l'objet d'un sondage à 40 % en cinq ans, l'intégralité des données sera exploitée. Ce traitement différencié répond au caractère plus important, dans ces dernières, des problématiques liées à la profession et à l'activité économique.

2.2 Les exemples d'utilisation des résultats du nouveau recensement

Afin de concrétiser le plus possible les utilisations des résultats du nouveau recensement, le groupe de travail a essayé de dégager des exemples précis d'utilisation sur lesquels ont été réalisés des simulations afin de mettre en évidence le type de résultats qui seraient disponibles avec le nouveau recensement et la précision correspondante.

Les travaux du groupe et des sous-groupes ont ainsi permis de constituer un ensemble de **fiches présentant des exemples d'utilisation** des résultats du nouveau recensement pour les différentes problématiques auxquelles les utilisateurs sont confrontés et en mettant en évidence les différences par rapport aux recensements généraux.

Une première grille d'analyse naturelle a consisté à **distinguer les utilisations selon le niveau géographique**, en partant du niveau régional pour atteindre le niveau infra-communal. Une attention plus particulière a été portée à certaines variables dont l'utilisation peut paraître complexe ou problématique aux utilisateurs dans le cadre du nouveau recensement, telles que l'âge, en raison du caractère pluriannuel de la collecte, mais aussi les informations bilocalisées, à cause de la non-simultanéité du recueil de l'information. Enfin, l'utilisation des résultats portant sur les logements, en lien avec le traitement particulier de la **construction neuve**, a fait l'objet d'un approfondissement.

a. L'utilisation au niveau régional et national

Les résultats détaillés issus des cinq enquêtes de recensement les plus récentes seront utilisables sans précaution particulière, si ce n'est de bien mesurer qu'ils sont représentatifs, en moyenne, de l'année médiane de ces cinq dernières années. Publiés en fin d'année N, ils porteront donc sur le début de l'année N-2. Ils seront donc directement comparables aux résultats que produisait un recensement général.

Toutefois, rappelons que le plan de sondage (*cf. fiche en annexe IV-4*) des enquêtes annuelles est optimisé au niveau de la région (les groupes de rotation des communes de moins de 10 000 habitants sont définis régionalement) et que les variables d'équilibrage sont principalement de nature démographique.

Les résultats des enquêtes annuelles fourniront donc de très bons résultats pour la pyramide des âges ou la population active par exemple. Ces résultats seront disponibles avant la fin de l'année de l'enquête et fourniront une information « avancée » très récente avec un bon niveau de précision pour la mise à jour des principales grandeurs statistiques régionales. La précision des résultats issus de l'échantillon annuel dépend en effet de la répartition par taille des communes de moins de 10 000 habitants, mais aussi de la part des communes de 10 000 habitants ou plus dans la région. A cet égard, des abaques régionaux de précision seront publiés. De plus, les résultats de l'enquête annuelle correspondent à des données collectées à la même date, ce qui peut permettre de mesurer des effets « conjoncturels » qui seront lissés par le cumul quinquennal des résultats détaillés.

Il est aussi naturel de s'intéresser aux migrations interrégionales même si celles-ci ne font pas partie de l'équilibrage en tant que tel. Le calcul d'une matrice régions x régions risque d'être peu significatif pour les régions avec lesquelles les flux sont très faibles. Etablir des matrices annuelles interrégionales (22x22) est donc clairement du ressort des résultats détaillés incorporant un cycle complet d'enquêtes. Les enquêtes annuelles peuvent cependant fournir des résultats récents et fiables, pour peu que l'on considère des sous-ensembles suffisamment vastes et pertinents. Dans beaucoup de régions, l'essentiel des migrations a lieu entre la région et les régions limitrophes, et bien sûr avec l'Ile-de-France.

L'exemple de la région Aquitaine (*cf. fiche migrations régionales en annexe III-1*), montre que l'analyse annuelle du bilan migratoire pour une région, à partir des enquêtes annuelles de recensement, peut se décliner de la façon suivante :

- les entrées et les sorties, le solde apparent ;
- déclinaison par tranches d'âge, en niveau et en structure, des groupes les plus concernées par les flux migratoires ;
- bilan selon les régions limitrophes, l'Île-de-France, le « reste », en niveau et en structure ;
- éventuellement, bilan selon des types de territoire (par exemple, taille de communes).

Dans chaque cas, les résultats sont accompagnés d'indicateurs de leur marge de précision.

b. L'utilisation sur des territoires multi-communaux

Les résultats détaillés portent sur les cinq dernières enquêtes de recensement. Pour le territoire multicommunal, toutes les communes de moins de 10 000 habitants qui le composent ont été recensées au cours de ces cinq dernières enquêtes, et on dispose d'un échantillon de 40 % de la population et des logements de ses communes de 10 000 habitants ou plus. Les données ainsi recueillies sont donc représentatives en moyenne du milieu de la période, à savoir l'année N-2. Pour la majorité des comportements démographiques, assez inertes, le décalage temporel de collecte selon les communes entre les cinq années n'apporte pas de perturbation significative. De plus, par comparaison avec les résultats d'un recensement général, utilisés de nombreuses années après sa collecte, l'utilisateur n'est pas conduit à des conclusions erronées. En revanche, le caractère annuel de la mise à jour des données peut légitimement inciter à examiner les résultats pour des variables nettement plus conjoncturelles comme l'emploi ou le chômage. Dans ce cas, le décalage temporel, même mineur au sein d'une période de cinq années, peut induire en erreur l'utilisateur qui n'aurait pas pris quelques précautions.

On montre cependant (*cf. fiche sur la zone d'emploi de Longwy en annexe III-2*) que l'on peut en déduire une vision, certes lissée, de l'évolution de l'emploi dans une zone d'emploi, et ainsi suivre par exemple l'impact d'une restructuration industrielle. De plus, en tenant compte de l'année de recensement des différentes communes, il est possible de réaliser des comparaisons spatiales entre communes de niveaux d'emploi et de chômage. Pour ce faire, il suffit de comparer la commune à sa référence régionale de l'année, connue au moyen de l'enquête annuelle, en calculant par exemple un indice de sur- ou sous-exposition au chômage. Les indices ainsi calculés sur la période des cinq dernières années peuvent être comparés entre eux en s'affranchissant ainsi des différences conjoncturelles de situation de l'emploi au cours de la période. Dans tous les cas, l'information est nettement plus riche qu'avec un recensement général qui ne produisait une photographie que tous les huit ou neuf ans, sans considération de cycle conjoncturel.

c. L'utilisation au niveau communal et infra-communal

C'est au niveau communal que la différence de traitement entre les communes au-dessous ou au-dessus du seuil de 10 000 habitants a le plus d'implications. En effet, les communes de moins de 10 000 habitants, recensées exhaustivement tous les cinq ans, connaîtront une situation assez proche de celle des recensements généraux. L'intervalle intercensitaire étant réduit à cinq ans, elles pourront cependant bénéficier d'une actualisation plus fréquente, mais ponctuelle. En revanche, dans les communes de 10 000 habitants ou plus, la collecte et donc le rafraîchissement sont annuels. Toutefois, les résultats qui vont en découler sont d'une nature différente de ceux, ponctuels, des recensements généraux et vont donc nécessiter une certaine appropriation pour donner leur pleine mesure.

En particulier, le recours à l'échantillonnage, même à un taux élevé de 40 % sur cinq ans, marque une rupture majeure dans l'esprit des utilisateurs. Il faut tout d'abord rappeler que les recensements généraux n'étaient jamais exploités exhaustivement pour toutes les variables et que certaines variables, comme la profession et la catégorie socioprofessionnelle, mais aussi l'activité économique ou les structures familiales, n'étaient exploitées que pour un quart des individus. La raison en est la forte charge de traitement qu'elles engendrent, ce qui aurait conduit à un délai trop grand de publication, accentuant encore le décalage avec la réalité. Dans le nouveau recensement, les variables profession, activité économique et structures familiales seront toujours exploitées sur un quart des questionnaires des communes de moins de 10 000 habitants, mais sur la totalité des questionnaires collectés dans les communes de 10 000 habitants ou plus. Pour ces communes, on atteindra donc un taux de 40 % pour ces informations au lieu de 25 % par le passé.

En second lieu, les données du recensement général, portant sur une même date, présentaient un caractère plus concret et rassurant que l'accumulation de données sur une période pluriannuelle. Il incombe donc aux statisticiens de prouver aux utilisateurs l'utilité, pour leurs besoins habituels, de résultats construits à partir de ces données collectées au cours de cinq années et notamment de montrer que, par rapport à des résultats ponctuels mais utilisés plusieurs années après, des informations pluriannuelles présentent un caractère lissé permettant de mieux saisir les tendances de fond au-delà des bruits de fond conjoncturels et de suivre les mouvements plutôt que de les constater *ex post*.

S'agissant du calcul de la population légale, un membre du groupe a fait part de ses interrogations quant à la publication en fin d'année A, de chiffres portant sur l'année A-2. Il craint que les communes ayant réalisé l'enquête de recensement l'année A ne comprennent pas qu'on « rétopole » ce chiffre pour le ramener à l'année A-2 (*cf. contribution de M. Ousset en annexe V-7*). La présentation des raisons tant techniques que d'égalité de traitement entre les communes, conformes à l'avis du Conseil d'Etat (*cf. réponse d'Alain Godinot en annexe V-7bis*) a conduit les membres du groupe à recommander à l'Insee de développer une communication adaptée envers les communes de moins de 10 000 habitants et à recommander le démarrage précoce des travaux du groupe de travail piloté par la DGCL, chargé d'examiner les conséquences de l'actualisation annuelle de la population légale à partir de 2008 sur les différents textes qui en dépendent.

Les membres du groupe de travail, et notamment les représentants de communes, ont exprimé leur volonté d'être clairement informés de la nature et de la précision des résultats que produira le nouveau recensement au niveau communal et infra-communal, et d'être accompagnés par l'Insee dans la maîtrise de ce nouveau corpus de résultats.

A cet égard, plusieurs travaux ont été menés par l'équipe du recensement.

Illustrer la précision au niveau communal et infra-communal

Le tableau suivant montre la précision obtenue par les résultats détaillés pour différentes tranches d'effectifs (cf. *fiche sur la précision à l'Iris en annexe III-4*). Elle est mesurée ici par le coefficient de variation, rapport de l'écart type à la moyenne, qui indique l'écart moyen entre la valeur estimée et la vraie valeur en pourcentage de cette dernière.

Tranche	Précision (CV)
>50 000	<1 %
20 000-50 000	1,50 %
10 000-20 000	2,00 %
6 000-10 000	2,50 %
3 000-6 000	3,00 %
2 000-3 000	3,50 %
1 000-2 000	4,50 %
500-1 000	6 %
250-500	8 %
<250	>8 %

A partir de ces indications, une réflexion a été entamée pour reconsidérer le détail d'information à la fois utile et fiable au niveau infra-communal. En effet, les tableaux des précédents recensements offraient un détail de données au niveau de l'IRIS et même de l'îlot qui, s'ils étaient principalement utilisés pour la reconstitution de territoires plus vastes, n'en conduisaient pas moins certains utilisateurs à être abusés par l'illusion de précision d'un nombre absolu. En effet, même exhaustif, le recensement n'était pas « parfait » et pouvait souffrir certaines incertitudes, tant en ce qui concerne le dénombrement de la population que plus encore la déclaration précise de ses caractéristiques. Ainsi, il arrive qu'un agent recenseur « oublie » un logement de sa zone de collecte ou que des habitants ne renseignent pas complètement leur questionnaire. Autant ces imperfections se compensent sur des effectifs d'une certaine importance, grâce à la loi des grands nombres, autant le poids d'une erreur individuelle peut peser lourd dans un effectif très faible. De plus, la stabilité temporelle de données n'est naturellement pas assurée au fil des ans sur des données très fines et donc très sensibles aux comportements individuels.

Illustrer la mesure des mouvements migratoires

La connaissance des mouvements migratoires au niveau communal étant une préoccupation majeure des communes, une fiche a été construite pour illustrer l'apport du nouveau recensement en ce domaine. Le recensement général ne permettait cette mesure que tous les huit ou neuf ans, donc à des intervalles variables, et induisait une analyse fortement normative de ces mouvements, en les considérant stables sur la période et en occultant par exemple les phénomènes de retournement.

Le nouveau recensement mesure les migrations de la même façon qu'auparavant en déduisant le solde migratoire dit « apparent » de l'évolution de la population de laquelle est retranché l'accroissement naturel (les naissances moins les décès) tel que le fournissent les statistiques de l'état civil. Le fait de pouvoir réaliser ces calculs à intervalles rapprochés permettra de repérer plus rapidement les inflexions éventuelles des tendances. Compte tenu du caractère cyclique du nouveau recensement, il est conseillé de calculer ces soldes sur un empan de cinq ans, mais on pourra l'actualiser tous les ans.

Pour intégrer le fait de la non-simultanéité de la collecte des données, on a besoin de faire l'hypothèse supplémentaire que les probabilités de migration ont une certaine inertie dont les évolutions ne sont significativement décelables que sur des périodes d'une certaine longueur. La conséquence de cette hypothèse est que, comme pour les autres variables du recensement, les résultats produits donneront une vision « moyenne » de la période et que les évolutions seront lissées. Cette façon de modéliser est toutefois usuelle en démographie, par exemple dans l'analyse de la fécondité.

Typologie d'IRIS

Une des premières utilisations des résultats du recensement au sein d'une commune est la cartographie sociale de ses quartiers. Il s'agit le plus souvent d'établir une typologie des différents quartiers au regard de critères comme la structure sociale de la population, l'exposition au chômage, les caractéristiques du parc de logements. L'interrogation légitime des membres du groupe de travail a porté sur la robustesse d'une telle typologie à l'échantillonnage dans les grandes villes. En effet, quel serait l'intérêt d'une actualisation annuelle si l'aléa engendré par le sondage masquait les évolutions réelles des quartiers ? Une fiche a été réalisée sur ce thème et a montré, à partir de simulations réalisées sur des Iris du recensement de 1999, que le classement des Iris n'était pas affecté par le recours au sondage. De plus, cette robustesse n'est pas affectée, bien au contraire, par la complexité des critères pris en compte pour réaliser la typologie. En revanche, ce type d'analyse pourra être mené à tout instant, sur la période des cinq dernières années, sans considération de la date du dernier recensement.

Zones à la demande

Le nouveau recensement, en s'appuyant non plus sur la brique de base que constituaient les îlots en ville, mais sur le répertoire des adresses, ouvre un champ tout à fait prometteur. En effet, la possibilité de recomposer des territoires infra-urbains, non plus en termes d'îlots, mais en termes d'adresses, peut permettre de produire des informations nouvelles, comme par exemple, dans le cas d'un projet de tramway, le nombre de personnes résidant le long d'un axe de circulation.

Un exemple a été construit à partir de la directive européenne sur le bruit, qui fait obligation aux pays membres d'établir des cartes d'exposition au bruit de la population dans les agglomérations de plus de 250 000 habitants à partir de 2007, puis de plus de 100 000 habitants à partir de 2012. A partir de la caractérisation des voies en termes d'exposition au bruit, il est ainsi possible d'estimer la population résidente le long des axes concernés, avec une précision tout à fait acceptable.

Par ailleurs, l'analyse des caractéristiques de la zone de chalandise d'un projet d'implantation commerciale, est nécessaire pour l'examen du projet présenté en commission d'équipement commercial (CDEC). Dans le cas d'un équipement urbain, la zone de chalandise peut-être limitée à la proximité immédiate du commerce. Faute d'avoir la possibilité d'obtenir les résultats sur cette zone spécifique, il peut être utile de disposer au minimum d'une clé de répartition des résultats des IRIS englobant, par exemple le nombre de logements ou de personnes au niveau très fin de l'îlot ou même de l'adresse.

d. Le cas des données bi-localisées

Le cas des données de flux est peut-être celui qui interpelle le plus les utilisateurs. Comment obtenir la même information qu'au recensement général où les flux, par exemple de déplacements de travail de commune à commune, étaient saisis au même instant ? Les communes étant recensées à des dates différentes, les déplacements d'une commune A vers une commune B seront observés une année donnée, alors que les flux réciproques, à savoir de la commune B vers la commune A seront peut-être observés l'année N+4 !

Une simulation a été réalisée sur l'agglomération de Montpellier. La typologie des déplacements domicile-travail au sein de cette agglomération montre une très forte polarisation autour de la ville-centre. Par ailleurs, dans une agglomération de cette importance, la répartition des communes de moins de 10 000 habitants entre les groupes de rotation de communes présente un caractère tout à fait équilibré. Ainsi, tous les ans, 45 000 habitants seront recensés en moyenne dans l'agglomération, actualisant régulièrement les informations.

Pour fixer les idées sur la précision, considérons à titre d'exemple le nombre d'habitants de Lattes qui se déplacent pour aller travailler à Montpellier. Il était de l'ordre de 2 700 personnes en 1999. Si ce niveau reste inchangé, l'incertitude sera de l'ordre de 3,5 % (en termes de coefficient de variation). Ce qui signifie qu'une évolution annuelle moyenne de 1,5 % du nombre de personnes allant travailler à Montpellier sera jugée significative au bout de 5 ans. L'actualisation des informations relatives aux déplacements domicile-travail sera plus rapide que les 7 à 9 ans qu'il y avait entre les anciens recensements généraux. Le recensement général apportait certes la connaissance « anecdotique » des tout petits flux, mais ce sont précisément ces flux qu'il fallait souvent filtrer dans les tableaux afin de ne pas en compliquer la lecture !

Toutefois, le groupe n'a pas eu le temps d'approfondir l'illustration des résultats bilocalisés et il serait souhaitable de poursuivre les travaux méthodologiques et pédagogiques en ce sens.

e. La construction neuve

Dans les communes de 10 000 habitants ou plus, le plan de sondage tient compte spécifiquement des immeubles neufs. En effet, il est important de connaître le nombre exact de logements des adresses nouvelles afin de réaliser les extrapolations. A cet effet, la construction neuve sera enquêtée exhaustivement dans les trois ans au maximum qui suivent l'entrée des logements neufs dans le RIL. Au démarrage du nouveau recensement, le délai sera toutefois de cinq ans afin d'absorber le stock de constructions neuves depuis le recensement de 1999.

Ceci a pour conséquence que toute la construction neuve présente dans le RIL l'année médiane du cycle d'estimation (soit 2006 pour le premier cycle 2004-2008 par exemple) sera connue au moment où on calculera cette estimation (en 2008 pour le premier cycle).

Chaque année, l'Insee disposera sur les constructions neuves de données exhaustives datées de trois ans en moyenne. Dès lors, une caractérisation statistique fine et annuelle des nouveaux habitants sera envisageable au bénéfice des utilisateurs. C'est donc un apport non négligeable du nouveau recensement pour l'étude de la construction neuve, qui reste le plus souvent un phénomène conjoncturel (un programme immobilier concentre généralement la construction neuve sur quelques années). Les recensements généraux qui avaient lieu tous les 7 à 9 ans ne permettaient pas d'appréhender de manière totalement satisfaisante la construction neuve.

De plus, une des composantes nécessaires aux projections de population scolaire est l'estimation du nombre d'enfants à scolariser par niveau dans les constructions neuves. Les résultats du nouveau recensement apporteront une information fine pour l'estimation des taux de peuplement et de scolarisation.

2.3 Les propositions du groupe de travail

Les deux types de résultats du recensement, ceux issus de cinq années d'enquête, et ceux issus de la seule enquête annuelle, ont fait l'objet de plusieurs présentations méthodologiques et sont illustrés par plusieurs fiches. On l'a vu, chaque année A, l'enquête annuelle permettra d'obtenir des résultats avancés pour la France et chacune de ses régions, mais aussi pour un cinquième des communes de moins de 10 000 habitants, celles enquêtées exhaustivement dans l'année. Les résultats détaillés seront, eux, produits l'année A à tous les niveaux géographiques, jusqu'à l'IRIS en infracommunal, et porteront sur l'année médiane des cinq dernières années, soit A-2.

Proposition 1 : Faciliter l'utilisation des outils de diffusion en ligne en publiant un tableau précisant les résultats à utiliser et les précautions d'emploi en fonction du niveau géographique (cf. ci-dessous).

Tableau de synthèse des types de résultats en fonction du niveau territorial

Niveau géographique	Résultats produits fin année N	Date de référence
France, région	Résultats de l'enquête annuelle	Début de l'année N
	Résultats détaillés issus des cinq dernières enquêtes	Début de l'année N-2
Commune de moins de 10 000 habitants	Résultats détaillés de l'enquête de recensement	Début de l'année de l'enquête
	Résultats détaillés : pour la commune, il s'agit d'actualiser chaque année les données de la dernière enquête de recensement pour les caler sur le niveau de population estimé en N-2.	Début de l'année N-2
Commune de 10 000 habitants ou plus	Résultats détaillés issus des cinq dernières enquêtes	Début de l'année N-2
Ensemble multi-communal	Résultats détaillés issus des cinq dernières enquêtes : résultats détaillés N-2 pour les communes de 10 000 hab. ou plus de la zone résultats ramenés à N-2 pour les communes de moins de 10 000 hab. de la zone	Début de l'année N-2

Propositions d'utilisation des résultats portant sur l'année N en fonction du niveau territorial

Niveau géographique – thèmes	Source	Précautions	Disponibilité
Région			
Population	Résultats de l'enquête annuelle	A comparer aux autres régions et à la France seulement	Fin de l'année N
	Résultats détaillés	Aucune	Fin de l'année N+2
Structures par âge	Résultats de l'enquête annuelle	Groupes d'âges	Fin de l'année N
	Résultats détaillés	Age détaillé	Fin de l'année N+2
Analyse des migrations interrégionales	Résultats de l'enquête annuelle	Se limiter aux principales régions d'échanges	Fin de l'année N
	Résultats détaillés	Aucune	Fin de l'année N+2
Population active - chômage	Résultats de l'enquête annuelle	Groupes d'âges agrégés	Fin de l'année N
	Résultats détaillés	Aucune	Fin de l'année N+2
Profession - CS	Résultats de l'enquête annuelle	Se limiter à la CS	Fin de l'année N
	Résultats détaillés	Profession détaillée	Fin de l'année N+2

Niveau géographique – thèmes	Source	Précautions	Disponibilité
Zone multi-communale (département, arrondissement, canton...)			
Tous thèmes – derniers résultats	Résultats détaillés	Aucune	Fin de l'année N+2
Mesure des évolutions : données non conjoncturelles	Résultats détaillés	Privilégier l'évolution sur 5 ans (renouvellement complet)	Fin de l'année N+2
Mesure des évolutions : données conjoncturelles : ex : suivi de l'emploi dans une zone d'emploi	Résultats détaillés	Pour un suivi annuel : vérifier les dates de recensement des communes de moins de 10 000 hab.	Fin de l'année N+2
Comparaisons entre communes de la zone : données non conjoncturelles	Résultats détaillés	Vérifier le caractère non conjoncturel	
Comparaisons entre communes de la zone : données conjoncturelles : ex : taux de chômage des communes de la zone	Résultats détaillés pour les communes de plus de 10 000 hab., dernière enquête de recensement pour les moins de 10 000 hab.	Calculer le rapport du taux de chômage de chaque commune de moins de 10 000 hab. à celui de la région de la même année. Comparer les résultats obtenus entre les communes.	
Commune de moins de 10 000 habitants			
Tous thèmes – derniers résultats (y compris IRIS)	Résultats de la dernière enquête de recensement	Aucune	Fin de l'année de l'enquête
Mesure des évolutions : tous thèmes (y compris IRIS)	Résultats de la dernière enquête de recensement	Evolution sur 5 ans	Fin de l'année de l'enquête
Communes de 10 000 habitants ou plus (en régime permanent)			
Tous thèmes – derniers résultats	Résultats détaillés	Aucune	Fin de l'année N+2
Logements neufs et population correspondante	Résultats détaillés	Aucune (exhaustif)	Fin de l'année N+2
Mesure des évolutions : données non conjoncturelles	Résultats détaillés	Privilégier l'évolution sur 5 ans (renouvellement complet)	Fin de l'année N+2
Mesure des évolutions : données conjoncturelles : ex : suivi de l'emploi dans une commune	Résultats détaillés	Mesure des évolutions annuelles délicates : seul 1/5 étant rafraîchi, privilégier les évolutions quinquennales plus robustes (renouvellement complet).	Fin de l'année N+2
Comparaisons entre communes de 10 000 hab. ou plus	Résultats détaillés	Aucune	Fin de l'année N+2
Comparaisons avec une commune de moins de 10 000 hab.	Résultats détaillés	Si la donnée est de nature « conjoncturelle », vérifier la date d'enquête de la commune de moins de 10 000 hab.	Fin de l'année N+2

Niveau géographique – thèmes	Source	Précautions	Disponibilité
IRIS 2000 (communes de 10 000 hab. ou plus)			
Tous thèmes – derniers résultats – effectifs absolus	Résultats détaillés	Vérifier la précision en fonction notamment de l'effectif	Fin de l'année N+2
Tous thèmes – derniers résultats – proportions	Résultats détaillés	Plus robuste que les effectifs. Vérifier cependant la précision	Fin de l'année N+2
Logements neufs et population correspondante	Résultats détaillés	Aucune (exhaustif)	Fin de l'année N+2
Mesure des évolutions : données non conjoncturelles	Résultats détaillés	Evolution sur 5 ans (renouvellement complet)	Fin de l'année N+2
Mesure des évolutions : données conjoncturelles : ex : suivi de l'emploi annuel	Résultats détaillés	Les résultats détaillés ne permettent pas un suivi annuel en effectif absolu mais en revanche, ils offrent un suivi lissé en moyenne mobile quinquennale	Fin de l'année N+2
Comparaisons entre IRIS de la commune ou d'autres communes.	Résultats détaillés	Vérifier la précision en fonction notamment de l'effectif. Meilleur en proportion (par exemple : typologies, comparer des taux de moins de 20 ans...)	Fin de l'année N+2
Comparaisons entre l'IRIS et sa commune.	Résultats détaillés	Vérifier la précision en fonction notamment de l'effectif. Meilleur en proportion (par exemple comparer le taux de moins de 20 ans...)	Fin de l'année N+2

L'annualisation des résultats reste évidemment un point de progrès majeur du recensement rénové. Jusqu'à présent, les résultats étaient mis à disposition au fur et à mesure sur une période d'environ 3 ans. Chaque phase de livraison (dénombrement, évolutions démographiques, exploitation principale, exploitation complémentaire) enrichissait la phase précédente, voire dans certains cas apportait des résultats qui se "substituaient" à des résultats produits antérieurement (cf. par exemple la PCS issue de l'exploitation complémentaire remplaçant la position professionnelle de l'exploitation principale).

Proposition 2 : livrer en une seule fois les résultats relatifs à une année, et ne pas modifier les résultats de l'exploitation principale par ceux de l'exploitation complémentaire.

Niveaux et structures

A partir des recensements généraux, les résultats publiés étaient beaucoup plus des effectifs que des structures. Or, si les niveaux sont utiles pour mesurer par exemple l'effectif d'une population concernée par une action publique, ils ne permettent pas facilement les comparaisons entre communes ou territoires, sauf à recalculer des proportions. De plus, s'agissant d'enquêtes effectuées pour partie par sondage, la précision obtenue sur une proportion est supérieure à celle sur une donnée en niveau absolu. Il est donc souhaitable d'introduire plus d'informations de structure dans les résultats produits par le nouveau recensement, sans pour autant négliger la diffusion des effectifs les plus pertinents. Les indicateurs en valeur absolue sont en effet fondamentaux lorsqu'il s'agit de mettre en œuvre ou d'orienter l'action publique. Un travail a été engagé pour repérer de "grands indicateurs" jugés essentiels à toutes les échelles géographiques tant communales qu'infra-communales.

Proposition 3 : développer la diffusion de résultats en structure, surtout au niveau infra-communal, sans pour autant négliger la diffusion des niveaux les plus pertinents. Mettre en évidence les niveaux absolus présentant une bonne fiabilité.

Proposition 4 : identifier de "grands indicateurs" significatifs à tous les échelons géographiques et principalement au niveau de l'IRIS.

Les comparaisons spatiales

L'année de collecte est utile pour effectuer des comparaisons entre les communes de moins de 10 000 habitants. En effet, afin de pouvoir les comparer, les résultats des communes de moins de 10 000 habitants seront calés sur le niveau de population estimé en A-2 de la commune. Pour la majorité des critères, cette procédure sera suffisante. En revanche, pour des variables à caractère conjoncturel comme l'emploi ou le chômage, comparer des résultats issus de données recueillies à des dates différentes peut être hasardeux. Dans ce cas, il convient de passer par l'intermédiaire d'une référence régionale ou nationale. Par exemple, le taux de chômage de chaque commune pourra être rapporté au taux de chômage régional de l'année de collecte, puis cet indicateur de sur- ou sous-exposition au chômage pourra être comparé entre les communes. En revanche, les communes de 10 000 habitants ou plus ne posent pas ce problème car elles sont enquêtées tous les ans dans les mêmes conditions ; leurs résultats sont donc strictement comparables entre eux.

Proposition 5 : fournir systématiquement l'année de collecte des résultats concernant les communes de moins de 10 000 habitants dans les fichiers-détail qui seront diffusés.

Proposition 6 : dans les tableaux concernant une zone multi-communale, préciser systématiquement l'année de collecte des résultats concernant les communes de moins de 10 000 habitants la composant.

Proposition 7 : diffuser les méthodes à employer pour comparer les résultats à caractère conjoncturel entre communes de moins de 10 000 habitants.

Proposition 8 : afin de faciliter une lecture territoriale des résultats, introduire dans les produits de diffusion des indicateurs de comparaison territoriale "poids", "indice de représentativité", "médiane ou quartiles".

Les comparaisons temporelles

La comparaison de résultats publiés chaque année peut conduire l'utilisateur à comparer abusivement une année à la précédente. S'agissant des résultats des enquêtes annuelles, il faut s'assurer que l'incertitude liée au sondage n'est pas supérieure à l'ampleur de l'évolution. Pour les résultats détaillés issus des cinq années, le rafraîchissement d'une année sur l'autre ne porte que sur un cinquième des communes de moins de 10 000 habitants et sur un cinquième de l'échantillon des communes de 10 000 habitants ou plus. Il est donc plus judicieux d'effectuer les comparaisons à cinq ans, sur un ensemble complètement renouvelé.

Proposition 9 : faire figurer autant que possible les résultats comparables d'il y a cinq ans dans les tableaux, ou l'évolution correspondante. Par exemple, le tableau « population par tranches d'âge » pourrait être produit avec les colonnes correspondant à l'année et à cinq ans auparavant.

Proposition 10 : introduire dans les produits de diffusion des indicateurs d'évolution donnant le sens de la tendance (non chiffrée), ou l'évolution en indice.

Mise en cohérence avec d'autres sources

Le nouveau recensement, pas plus que l'ancien, n'aura la capacité à répondre à l'ensemble des besoins d'information exprimés. Le recensement général fournissait tous les huit à neuf ans une base sur laquelle des estimations étaient élaborées au cours de la période intercensitaire à partir d'autres sources, le plus souvent administratives. Avec le nouveau recensement, ce dispositif est appelé à changer sensiblement. Le recensement quitte son rôle de référence décennale immobile pour offrir une actualisation régulière. Bien adapté au suivi des phénomènes à évolution relativement lente, comme le sont la plupart des phénomènes sociodémographiques, il ne peut évidemment se substituer au suivi conjoncturel que permettent des sources telles que les demandes d'emploi de l'ANPE pour la mesure du chômage ou les DADS pour celle de l'emploi salarié. Il faut donc construire un nouveau paysage statistique local en donnant à chaque source sa place en cohérence avec les autres.

Proposition 11 : développer une approche multi-sources, en particulier pour les problématiques conjoncturelles telles que l'emploi ou le chômage.

3. LA QUALITÉ DES RÉSULTATS

Les recensements généraux fournissaient les indications sur la précision des résultats, par exemple pour les données issues du sondage au quart, sous une forme très fruste en annexe aux publications. La généralisation de la production de résultats à partir d'échantillons rend nécessaire un affichage plus clair de la précision des informations diffusées. C'est surtout au niveau de l'infra-communal que le besoin d'information sur la fiabilité des données est le plus fort.

Les utilisateurs ont souhaité un affichage de la précision en termes de coefficient de variation, qui donne une précision en pourcentage, plutôt que d'intervalle de confiance. Cependant, il est rappelé que l'on passe aisément de l'un à l'autre. En effet, le coefficient de variation est le rapport de l'écart-type à la moyenne, tandis que l'intervalle de confiance à 95 % est égal au résultat trouvé plus ou moins deux écarts-types. Il s'interprète de la façon suivante : pour un résultat donné, la « vraie » valeur a 95 % de chances de se trouver dans l'intervalle compris entre le résultat moins deux écarts types et le résultat plus deux écarts types.

Proposition 12 : privilégier l'utilisation de la précision en pourcentage (*coefficient de variation*).

Au-delà de la valeur numérique elle-même, il est intéressant d'essayer de bâtir des règles à partir de cet indicateur permettant d'apprécier la qualité des estimations.

La première règle qui semble se dégager vient de l'effectif considéré pour le croisement d'intérêt. Il est intéressant de garder en tête le tableau de décision qui suit :

Effectif	Règle
Au-dessous de 50	Danger
Entre 50 et 200	Vigilance
Entre 200 et 500	Un peu de prudence
Au-dessus de 500	Normalement pas de problème

L'appréciation du précédent tableau peut être modulée par des éléments propres. Par exemple, si les variables concernées sont fortement corrélées à l'âge, la règle est probablement pessimiste.

Proposition 13 : la précision des résultats doit être indiquée à l'utilisateur de la manière aussi simple et intuitive que possible, par exemple en affectant une couleur aux cases de tableau dont la significativité serait inférieure à un certain seuil.

Si on raisonne dans le cadre d'un IRIS standard de 2 000 habitants, alors une 2^e règle, qui découle en partie de la première, serait de se limiter si possible à des croisements avec 10 modalités au maximum pour les variables qui concernent la population et les logements. Même si, bien entendu, il y a peu de chances que toutes les modalités du croisement aient le même effectif, cela permet d'assurer, en moyenne, un effectif par case suffisant.

Concrètement on peut imaginer :

- décliner une variable en 10 modalités au plus ;
- croiser une variable à 2 modalités avec une variable à 5 modalités au plus ;
- croiser une variable avec 3 modalités avec une variable ayant aussi 3 modalités au plus ;

Proposition 14 : fournir à l'utilisateur des éléments simples lui permettant de déterminer les conditions minimales de significativité des résultats : effectif minimum, nombre de modalités de la variable, catégorie de variable...

L'étape suivante consiste à essayer de construire, à partir de ces deux règles, des tableaux pour lesquels est garantie une fiabilité des résultats à tous les niveaux. Pour ce faire, on raisonnera en distinguant les variables relatives au logement de celles relatives à la population.

Proposition 15 : élaborer des tableaux à l'Iris offrant une bonne garantie de précision des résultats sur les croisements proposés. Alléger les tableaux Analyse en privilégiant les croisements les plus pertinents.

4. LES GRANDES LIGNES DE PRODUITS DE DIFFUSION

Le groupe de travail a cerné le contexte de la mise en place de la future gamme de produits de diffusion du recensement avant d'en esquisser les principaux traits, au moins à titre exploratoire. Les éléments de contexte abordés ont été les principes de la politique de diffusion de l'Insee pour les prochaines années, puis les contraintes et les opportunités.

Comme l'indique son titre, ce chapitre n'aborde pas la question des travaux sur mesure. Il ne traite pas non plus, du moins directement, de la restitution d'informations aux communes, le chapitre V étant consacré à cette question.

Enfin la gamme de produits décrite ici ne pourrait être réalisée complètement qu'au terme d'un premier cycle de cinq années de collecte. Pendant cette période de montée en charge les réalisations seront partielles.

4.1 Les principes de la politique de diffusion de l'Insee pour les prochaines années

a. Travailler avec les utilisateurs pour optimiser les produits et services en fonction de leurs besoins

De nombreux enseignements doivent être tirés de l'évaluation faite par le CNIS de la diffusion des résultats du RP 1999. Notamment des principales conclusions du rapport : besoin d'améliorations significatives dans la présentation de la gamme de produits ; dans les délais de mise à disposition et dans la gestion des annonces de dates de disponibilité ; dans le nombre parfois excessif des versions successives des produits et enfin dans la stabilité des principes de tarification.

Le groupe de travail du CNIS sur la diffusion des résultats du nouveau recensement est un des lieux privilégiés de cette concertation.

b. Faciliter, dans toute la mesure du possible, l'accès aux données statistiques et leur utilisation

Par une mise à disposition systématique sur internet

La priorité sera donnée à la diffusion par internet de la plus grande part possible de l'information produite par l'Insee, y compris les résultats détaillés, sous réserve qu'ils ne soient pas couverts par le secret statistique. La commodité d'accès à cet outil, non seulement pour consulter mais aussi pour télécharger des résultats, devrait faciliter considérablement l'accès aux informations et leur utilisation.

Par une nouvelle politique tarifaire

La nouvelle politique de tarification de l'Insee aboutit en moyenne à une baisse des prix des produits de diffusion, et même à des prix systématiquement nuls pour tout ce qui est ou sera diffusé par internet.

En effet, les droits d'accès et redevances de rediffusion seront supprimés.

En conséquence l'accès aux informations sur internet sera gratuit. Quant aux cédéroms, leur prix sera fonction, comme pour les précédents recensements, des frais de réalisation, mais il ne dépendra plus du nombre de données qu'ils offriront. On peut donc imaginer de produire des cédéroms beaucoup plus « pleins » qu'au RP 1999 sans augmenter les prix. Autrement dit, le catalogue de cédéroms sera beaucoup plus ramassé et l'achat d'une collection complète pour couvrir tout le territoire reviendra nettement moins cher.

Enfin la rediffusion, commerciale ou non, du contenu de tout produit gratuit ou payant sera totalement libre de droits, sous réserve de la mention de la source et du respect de l'intégrité des données.

La combinaison de cette nouvelle tarification avec la volonté d'offrir le maximum de résultats sur internet sera déterminante pour la diffusion des résultats du nouveau recensement : à tous les niveaux géographiques standard, supra et infra-communaux, ils pourront être gratuitement consultés et/ou téléchargés sur le site de l'Insee.

Par le rapprochement de sources sur un même support

La politique de facilitation des approches multi-sources s'appliquera particulièrement aux résultats locaux. Des investissements sont actuellement en cours pour structurer de façon souple les gisements d'informations finement localisées, diffusés par l'Insee. Ils aboutiront en 2005. Ces bases seront mises en ligne sur le site de l'Insee.

4.2 Contraintes et opportunités

a. L'annualité des résultats

L'annualité des résultats limitera fortement les gammes de cédéroms et d'ouvrages sur papier.

Jusqu'ici l'Insee étalait sur trois ans, pour un recensement donné, son programme de publications sur supports matériels. Après le recensement de 1999, par exemple, sont parus : trois cents fascicules et quatre cents cédéroms départementaux, plus de vingt fascicules régionaux (dits « verts »), une vingtaine d'ouvrages et de cédéroms divers... La tentative de réaliser tous les ans le même catalogue de produits aboutirait donc à une accumulation explosive de retards.

Bien au contraire il faut viser, pour une source annuelle, de mettre à disposition la totalité des résultats dans les trois mois qui suivent la fin de la phase de production et de validation. Cela ne signifie pas mettre à disposition tous les produits dans ce délai, mais cela implique d'offrir un accès avant trois mois à tous les résultats sur au moins un support.

Si on peut espérer arriver à des délais aussi resserrés pour la mise à disposition en ligne, le problème est tout autre pour les produits sur support matériel. On ne peut en effet attendre beaucoup d'une plus grande automatisation de la réalisation des imprimés et cédéroms : les chaînes d'édition étaient déjà largement automatisées aux recensements de 1990 et de 1999. On ne peut non plus mobiliser en permanence plus de moyens que ceux qui étaient rassemblés sur une période limitée pour les recensements généraux de la population.

Il faudra faire des choix. Pour éclairer ces choix, il faut se demander quel doit être le rôle des imprimés et cédéroms dans ce nouveau contexte, quelles statistiques voire quelles autres informations, il est nécessaire de diffuser sur ces supports, pour quels publics ? à quel niveau géographique ? avec quelle périodicité ?

b. La nouvelle technique de recensement

Le recensement produira des résultats chaque année. Le prix en sera le recours à une technique d'observation plus complexe. D'une part il recourra aux sondages. Ce n'est pas une nouveauté absolue puisque les précédents recensements établissaient les valeurs d'un certain nombre de variables particulièrement importantes à partir d'un échantillon. Le recours à des observations sur un intervalle pluriannuel, avec calcul des résultats en moyenne sur cette période, est une innovation importante.

L'interprétation des résultats ne pourra donc pas reposer sur les mêmes précautions méthodologiques que lors des derniers recensements. Plusieurs notes ont exposé au groupe de travail la portée et les limites des résultats qui seront produits. Le groupe souhaite que ces informations soient systématiquement portées à la connaissance des futurs utilisateurs.

Cette information pourra se faire au moyen d'outils et de documents adaptés, voire de présentations orales, mais elle devra aussi se trouver sur les produits de diffusion.

Proposition 16 : Les produits de diffusion du recensement devront, dans la mesure du possible et sans que cela nuise à la clarté des documents offerts, fournir aux utilisateurs des éléments d'aide à l'interprétation des résultats, notamment sur l'appréciation de leur précision en niveau comme en évolution.

c. La popularisation des raccordements à internet et au haut débit

Le nombre d'abonnements à internet a atteint un niveau significatif : 23 millions d'internautes de plus de dix ans en mars 2004 d'après la société Médiamétrie, soit près de la moitié de la population d'âge correspondant. Qui plus est, la croissance de ce nombre reste très vive (+15 % en 2003) et on peut espérer atteindre 70 à 80 % de la population générale avant la diffusion des premiers résultats complets du recensement. De plus, chez les professionnels publics ou privés qui sont de loin les plus gros utilisateurs du recensement, le taux de connexion est déjà très élevé. Des sondages amènent à penser que les trois quarts au moins des personnes qui s'adressent à l'Insee sont déjà connectées, soit chez elles, soit à titre professionnel. Il est donc légitime de penser à s'appuyer largement sur ce média comme vecteur de la mise à disposition du recensement.

Cette opinion est confortée par le succès rapide, particulièrement en France, des liaisons à haut débit. Les internautes disposant de ce type de connexion peuvent non seulement consulter des données sur un site, mais encore télécharger des lots de données importants afin de les traiter de manière autonome sur leur propre système. Cette utilisation hors ligne correspond à de nombreuses utilisations des résultats du recensement, particulièrement chez les professionnels.

Cela ne signifie pas qu'il soit possible de s'en remettre au seul support internet :

- le support imprimé reste le plus commode quand on a fréquemment besoin de consulter - et non de traiter - un petit nombre d'informations, toujours les mêmes : il peut donc être utile pour mettre à disposition les indicateurs les plus recherchés sur les zonages les plus utilisés ;
- les résultats présentés sous forme « littéraire » comme les analyses et les synthèses se lisent toujours plus facilement sur papier qu'à l'écran ;
- certains utilisateurs comme les relais de l'Insee, les rediffuseurs, les entreprises de marketing, des banques, des entreprises à réseaux, voudront se procurer de vastes bases de résultats du recensement : il est prévu que ces bases de données pourront être téléchargées par lots. Mais au-delà d'un certain volume, il sera plus pratique pour ces utilisateurs comme pour l'Insee de procéder à un transfert de ces bases sur cédérom.

4.3 Les grandes lignes de produits

Les produits sont distingués par les types d'utilisations auxquelles ils se prêtent. En allant de ceux qui offrent des chiffres clés, accessibles au public le plus vaste, jusqu'à ceux qui fournissent les informations les plus spécialisées ou les plus détaillées, destinées à des professionnels, on peut citer :

- produits rassemblant sur chaque territoire un certain nombre de chiffres clés ou d'indicateurs, sous forme soit de séries chronologiques, soit de fiches thématiques articulant des résultats de sources diverses, soit de cartes thématiques ;
- analyses et synthèses ;
- résultats détaillés sous forme de tableaux à plusieurs dimensions comme de bases de données ;
- fichiers-détail.

a. Les chiffres clés

On peut distinguer deux grands types de produits rassemblant des indicateurs ou « chiffres clés » : les « séries chronologiques » et les « profils ».

Les séries chronologiques

Ces produits privilégient la dimension temporelle plutôt que le nombre et la diversité des indicateurs. Au dernier recensement, ils comprenaient :

- les « Évolutions démographiques 1982-1990-1999 » (cent fascicules départementaux dits « orange » ; une récapitulation complète sur un cédérom comprenant aussi la base de données communales mise à jour à la date du RP ; un volume « national » reprenant les mêmes informations mais seulement pour les principaux échelons) ;
- et les volets « Indicateurs références » (1982-1990-1999) des volumes départementaux et régionaux des exploitations principale et complémentaire.

Des produits équivalents devront être mis à disposition sur internet :

- en consultation ;
- sous forme de bases téléchargeables.

Par ailleurs, il sera impossible d'éditer chaque année ces centaines de fascicules et de volumes. Y aurait-il une justification à les publier une fois par cycle de cinq ans ? Cela semble peu probable : le réflexe du public sera probablement de préférer une information gratuite et à jour sur internet à des publications payantes et dépassées au bout d'un an.

Faudra-t-il publier chaque année un volume national fournissant ces indicateurs aux principaux échelons administratifs et d'étude : France, régions, départements, unités urbaines, aires urbaines, etc. ? La question mérite d'être approfondie. S'agissant d'un volume annuel le décalage temporel entre la mise à disposition sur internet et sous forme imprimée serait limité. Les principaux échelons de référence faisant l'objet de recherches fréquentes, il pourrait être commode d'en disposer sous forme d'un ouvrage imprimé.

Ceux qui ne pourront ou ne voudront pas consulter ou télécharger les résultats sur internet devront pouvoir commander des bases entières sur cédérom.

Proposition 17 :

Les séries chronologiques :

- seront mises à disposition sur internet : en consultation dans un format aisément lisible par les principaux tableurs ou pour téléchargement dans un format acceptable par les systèmes de gestion de bases de données courants ;
- la possibilité de commander des bases d'indicateurs sur cédérom sera maintenue ;
- une étude sera menée sur l'intérêt d'une publication imprimée annuelle comportant des séries chronologiques d'indicateurs sur la population et les logements aux principaux échelons de référence.

Les « portraits de territoire »

Ces produits fournissent sur une zone géographique donnée des ensembles d'indicateurs complémentaires sur des thèmes variés. Un ensemble de fiches de ce genre peut constituer ce qu'on appelle un « tableau de bord ». Elles apportent souvent une dimension chronologique à certains indicateurs, mais pas de façon systématique. Ces fiches permettent aussi, toujours pour certains indicateurs, de comparer le territoire en question à un territoire de référence (par exemple, dans le portrait d'une commune, comparaison avec le département englobant).

La nécessité de composer des faisceaux d'indicateurs conduit à recourir à des sources variées : ces produits dépassent le cadre de la diffusion du recensement. Si une fiche porte des résultats de plusieurs sources différentes, elle peut être mise à jour plusieurs fois dans la même année.

Enfin, ces fiches allient tableaux, graphiques et cartes.

L'Insee diffuse actuellement deux collections de produits de ce genre :

- les fiches « profils » du recensement de 1999 : c'est un ensemble limité à cette seule source ; elles sont disponibles sur le site du recensement de 1999 ;
- les fiches « portraits de territoire » : au nombre d'une vingtaine, elles combinent des sources variées et sont actuellement imprimées à la demande dans les services d'information de l'Insee.

Les « portraits de territoire » seront mis en ligne avant la fin de l'année 2005 et régulièrement mis à jour, notamment à partir des résultats du recensement ; comme ils intégreront toute l'information figurant sur les fiches « profils » ces dernières ne seront plus diffusées. *Cette gamme de produits sera ainsi unifiée.*

Ces portraits seront disponibles en ligne sous trois formats :

- des fiches regroupant des indicateurs chiffrés et des graphiques sur chaque commune, canton, etc. ;
- des cartes thématiques consacrées à la visualisation spatiale d'une sélection d'indicateurs ; ces cartes seront dotées d'un certain degré d'interactivité (consultation des noms de cellules, choix du nombre de classes...),
- une base de données téléchargeables fournissant par échelon géographique tous les indicateurs contenus dans les fiches : ce format sera particulièrement utile à ceux qui veulent se constituer leurs propres tableaux de bord ; ils pourront télécharger en une seule opération les données relatives à de très nombreux échelons plutôt que de transférer cette information fiche par fiche.

Proposition 18 : Les chiffres clés des territoires :

- feront l'objet au cours de 2005 d'une mise à disposition en ligne systématique, puis seront régulièrement mis à jour ; ils remplaceront les fiches « profils » ;
- ils seront mis à disposition sous trois formats : fiches de tableaux et graphiques, cartes thématiques pour une sélection d'entre eux, bases de données téléchargeables ;
- ils faciliteront les comparaisons entre zones ;
- la possibilité de commander des bases d'indicateurs sur cédérom sera maintenue ;
- ces fiches n'ont jamais fait l'objet de publications imprimées : le recensement en cours ne changera rien à cette pratique.

b. Les analyses et synthèses

On peut distinguer la synthèse générale des résultats du recensement des analyses et synthèses thématiques.

Synthèse générale

Le recensement de 1999 avait donné lieu à la publication de l'ouvrage « Portrait de la France », qui décrivait les principales caractéristiques de la population et des logements. Cette publication d'un faible volume (32 p.) et s'appuyant sur de nombreux graphiques et cartes était destinée à un large public.

L'intérêt du sujet, l'ampleur du public touché par les opérations de recensement, l'accueil fait au « Portrait de la France » de 1999, font penser qu'il pourrait être utile, ne serait-ce qu'au titre de « restitution aux citoyens », de publier un ouvrage de ce genre dans la collection « Références » de l'Insee. La lenteur des mouvements démographiques ne justifierait sans doute pas une mise à jour annuelle. De toute façon, d'autres ouvrages de référence comme les « Tableaux de l'économie française » sont là pour mettre à disposition les mises à jour annuelles des principales variables.

Il serait préférable de s'en tenir à un rythme quinquennal.

Analyses et synthèses thématiques

La fourniture de résultats standard à tous les échelons géographiques par le recensement ne doit pas faire oublier qu'il permet aussi de faire, à des niveaux comme la France ou les régions, des analyses approfondies de la démographie et de l'emploi. Le dernier recensement a été l'occasion de publier sur ces thèmes des volumes combinant analyses et fournitures de tableaux détaillés.

L'intérêt du renouvellement tous les cinq ans de telles études thématiques ne fait guère de doute. En revanche, la formule des ouvrages doit faire l'objet de réflexions plus approfondies : les publications imprimées demeurent un vecteur privilégié de diffusion des études et analyses ; alors que la place des tableaux détaillés serait plutôt sur le site de diffusion de l'Insee.

Proposition 19 :**Les analyses et synthèses :**

Seront publiés tous les cinq ans :

- un ouvrage destiné au grand public fournissant une synthèse générale des résultats des recensements ;
- des volumes d'analyses et de synthèses thématiques.

c. Les résultats détaillés

Pour fixer les idées, rappelons que les résultats détaillés standard du dernier recensement correspondaient aux tableaux « Analyse », soit une quarantaine de tableaux comprenant 18 000 nombres, pour tout niveau géographique, de la France à l'IRIS¹. On approche donc le milliard d'informations diffusées.

Supports électroniques : internet et cédéroms

A l'avenir, les résultats détaillés diffusibles seront systématiquement mis à disposition sur internet, soit pour consultation, soit pour téléchargement. Plutôt que de répartir ces résultats en une quantité énorme de tableaux à deux dimensions comme sur le site du RP 1999, ils seront présentés en tableaux à trois dimensions ou plus, la géographie étant une de ces dimensions.

Ces tableaux multidimensionnels seront mis à disposition sous deux formats :

- un format qui adoptera en ligne une présentation sous forme de tableaux : il sera adapté à la consultation ainsi qu'à l'extraction vers les principaux tableurs de quelques comptages ou de quelques tableaux ;
- un format adapté à la réutilisation dans les gestionnaires de bases de données (SGBD) des utilisateurs qui voudront travailler dans leurs propres systèmes sur les informations les plus détaillées : uniquement destiné au téléchargement en un ou quelques lots, il ne permettra pas la consultation en ligne.

Cette diffusion en ligne n'éliminera pas forcément tout besoin de cédérom. A partir d'une certaine taille des fichiers à transférer, il pourra être préférable d'avoir recours au cédérom, malgré la croissance exceptionnellement rapide des liaisons haut débit. Pour autant rien n'indique qu'il faille prévoir une publication systématique et a priori sur cédérom : ce dernier semble plutôt devoir devenir un simple intermédiaire de transfert, gravé à la demande.

Les publications imprimées

Les résultats détaillés du RP 99 (Tableaux Références et Analyses) se trouvaient essentiellement sur deux collections d'imprimés :

- les cent volumes départementaux « Tableaux Références et Analyses » de l'exploitation principale (+ 1 volume France) ;
- la trentaine de volumes régionaux « Tableaux Références et Analyses » de l'exploitation complémentaire (+ 1 volume France).

Même si la livraison simultanée ou quasi-simultanée des deux exploitations permettait de ne pas distinguer deux séries (principale et complémentaire), la parution annuelle d'une centaine de ces publications assez volumineuses n'est pas envisageable.

Faudrait-il passer à une parution quinquennale ? Les publications imprimées de grandes masses de chiffres sont des produits très contestables. Contrairement aux résultats de type « profil », « portrait » ou autres tableaux de bord, les résultats détaillés sont acquis pour être retravaillés plutôt que simplement consultés. Les éditer sur papier, c'est obliger l'acquéreur à en ressaisir un grand nombre. Pour toutes les autres sources que le recensement, l'Insee a renoncé à publier des résultats détaillés sur papier : d'où la mutation de la série « Insee Résultats », composée autrefois de gros volumes imprimés, en cédéroms maintenant.

¹ Avec des restrictions en fonction du niveau géographique pour les croisements avec les variables dites sensibles.

Proposition 20 :**Les résultats détaillés :**

- seront mis à disposition sur internet : soit pour consultation ou extraction vers des tableurs, soit pour téléchargement, dans un format aisément lisible par les SGBD courants ;
- la possibilité de commander des bases de résultats détaillés sur cédérom sera maintenue ;
- l'intérêt de publier sur papier de gros volumes de données, ne serait-ce que tous les cinq ans, alors que l'information correspondante sera disponible sous forme électronique, apparaît trop faible. Ces publications ne seront donc pas en principe entreprises.

d. Les fichiers-détail

Il s'agit des fichiers de données individuelles, appelées aussi micro-données. Seule la diffusion de données individuelles ni directement nominatives (anonymisées) ni indirectement nominatives, peut être envisagée.

Au recensement de 1999, il avait été possible de diffuser :

- le fichier-détail des logements avec un repérage géographique fin ;
- le fichier-détail des logements et des individus avec un repérage géographique ne descendant pas en dessous de zones d'au moins cinquante mille habitants ;
- un échantillon d'un vingtième des logements et des individus avec un repérage géographique fin.

Proposition 21 :**Fichiers-détail :**

- Les membres du groupe de travail souhaitent qu'en termes d'accès aux données détail des possibilités analogues à celles du recensement de 1999 soient offertes par le nouveau recensement.
- Les formats de diffusion de ces fichiers devront être adaptés aux besoins les plus fréquents des utilisateurs.

e. Les tabulations sur mesure

La mise à disposition gratuite et sur internet d'un grand nombre de résultats jusqu'au niveau de la commune et de l'IRIS devrait répondre aux besoins de la majorité du public.

Il reste que certains utilisateurs, particulièrement parmi les professionnels, ne pourront pas se satisfaire complètement de résultats standard. Ils devront donc avoir recours à des tabulations sur mesure. Si des règles de diffusion des fichiers-détail comparables à celles de 1999 sont reconduites, ces utilisateurs pourront réaliser leurs propres tabulations sur mesure en acquérant un extrait de fichier-détail ne comportant qu'un repérage géographique grossier (zones de 50 000 habitants ?).

Comme il faut prévoir des limites à la précision géographique des fichiers-détail dont la diffusion sera autorisée, seul l'Insee aura la possibilité de réaliser des tableaux sur mesure sur des zones géographiques fines qui seront vraisemblablement des équivalents de l'Iris en termes de taille de population.

Proposition 22 :**Les tabulations sur mesure :**

- l'offre de l'Insee en matière de tabulation sur mesure devra être claire, explicite, expliquée, largement portée à la connaissance des intéressés et divulguée suffisamment tôt pour que les utilisateurs l'intègrent dans leurs programmes d'analyse des résultats ;
- les procédures de commande et de fourniture de ces travaux devront être aussi simples et rapides que possible.
- le service de tabulation devra être rendu dans des conditions techniques et tarifaires adaptées.

4.4 La communication sur les produits et services de diffusion du recensement

Le recensement de 1999 a donné lieu à une offre de produits d'information standard divers et très nombreux, et à une double gamme de services sur mesure. Cette richesse n'allait pas sans une certaine complexité et des utilisateurs ont pu avoir des difficultés à repérer le produit qui convenait à leurs besoins. D'autres ne l'ont pas repéré.

La gamme de produits proposée pour le recensement actuel est beaucoup plus simple et la difficulté de choix en sera sensiblement diminuée pour le public. Néanmoins la quantité d'information apportée par le recensement sera considérable, elle portera sur un très grand nombre d'échelons géographiques, et sera mise à disposition sur différents supports et formats. La communication sur l'offre de résultats du recensement et l'aide au choix des utilisateurs restent donc indispensables.

Proposition 23 : Le plan de diffusion du recensement devra faire l'objet suffisamment tôt d'une très large communication. Ses modes de présentation devront être adaptés à la diversité des utilisateurs et permettre à ces derniers de se déterminer à partir d'une claire connaissance du choix qui leur sera offert.

5. LA MISE À DISPOSITION DES RÉSULTATS

5.1 Le schéma général

Le schéma consiste à commencer à publier très tôt, en augmentant régulièrement le volume d'informations produites.

5.2 Publier une estimation de la population des communes avant la fin de l'année 2004

L'effectif recensé lors de la collecte est retourné à la commune à la fin de la phase de réception, d'enregistrement et de contrôle. Ce chiffre correspond pour les communes de moins de 10 000 habitants à l'effectif de la population, alors que pour les communes de 10 000 hab. ou plus, il représente la population de l'échantillon enquêté.

Pour les communes, le premier résultat du recensement est bien l'effectif de la population. Même si ce chiffre ne sera authentifié qu'en 2008, il peut être diffusé sous forme provisoire à l'instar des recensements généraux. Au RP 99, la population estimée des communes, établie à partir de la saisie des bordereaux communaux, avait eu lieu début juillet, en même temps qu'une conférence de presse de présentation d'une première estimation de la population française. Pour les communes de moins de 10 000 habitants, rien n'interdit de reproduire un pareil schéma. En revanche, l'exercice est plus délicat pour les communes de 10 000 habitants ou plus, à partir de seulement 8 % de la population des ménages. Il sera donc nécessaire de définir un seuil de population, qui serait abaissé au fur et à mesure des années.

Le calcul des chiffres de population pour les communes concernées pourrait intervenir au plus tard avant la fin de l'année. La mise en ligne pourrait se faire par le site *internet* habituel *insee.fr*. Avant la mise en ligne, une communication aux communes sera nécessaire, surtout pour les communes de plus de 10 000 habitants, car le chiffre ne se déduit pas immédiatement des comptages.

5.3 Produire les premiers résultats nationaux avec quelques indicateurs régionaux avant le démarrage de la campagne 2005

A partir des premiers résultats statistiques sur la population et les logements, il est envisagé de publier avant le lancement de la campagne 2005, début janvier, un *Insee Première* de cadrage.

Le ou les tableaux publiés dans l'*Insee Première* seraient également mis en ligne sur *insee.fr*.

5.4 Organiser avant l'été 2005 les premières restitutions statistiques aux communes

Après les différentes phases de validation, c'est-à-dire à partir du printemps 2005, il est proposé de démarrer la fabrication de produits de restitution d'information statistique aux communes, sous la forme de fiches « portrait de territoires ». Ces fiches présenteraient de manière synthétique les principaux résultats pour la commune, avec une comparaison régionale, l'indication des principales évolutions depuis 1999 et une mise en évidence graphique.

5.5 Démarrer la diffusion de données régionales avant l'été 2005

Après validation régionale des données de la campagne 2004, il sera possible de diffuser les résultats aux niveaux national, régional et communal pour les communes de moins de 10 000 habitants sur un premier sous-ensemble des variables exploitées.

En complément des fiches communales de restitution, il est proposé de mettre en ligne de premiers tableaux aux niveaux France et régions. Il est même proposé d'aller plus loin et de commencer à mettre en ligne de premiers « cubes » permettant à l'utilisateur de réaliser lui-même ses propres tabulations à l'intérieur d'un ensemble déjà constitué de croisements de plusieurs variables. Cela suppose cependant la disponibilité des outils informatiques correspondants. De plus, il est nécessaire d'avoir consulté la CNIL.

5.6 Avant la fin mars 2006 :

- Publier l'ensemble des résultats France et régions de la campagne 2005
 - o Au début de l'année 2006, on devrait avoir codifié la profession, redressé l'ensemble des variables des deux campagnes 2004 et 2005. On pourra alors diffuser l'ensemble des tableaux de la campagne 2005 aux niveaux France et régions.
- Augmenter le volume de la restitution aux communes
 - o En cumulant les campagnes 2004 et 2005, descendre le seuil de restitution aux communes.

6. LA RESTITUTION D'INFORMATIONS AUX COMMUNES

Le contenu détaillé des fiches de restitution évoquées ci-dessus sera établi après concertation avec les représentants des communes au sein d'un sous-groupe du groupe de travail du CNIS. Leurs propositions vont dans le sens d'élaborer des indicateurs susceptibles de satisfaire les attentes des communes tout en présentant de bonnes garanties de fiabilité.

Les communes de moins de 10 000 habitants ne posent pas de problème particulier. En revanche, il est difficile de prévoir un dispositif uniforme pour les communes de 10 000 habitants ou plus aux cours des premières années, d'autant que seule la population des ménages aura été enquêtée. Pour les communes de plus de 80 000 habitants, il est envisageable de produire quelques résultats synthétiques comme la répartition en grands groupes d'âges ou quelques caractéristiques des logements. Pour les communes proches du seuil de 10 000 habitants, il ne sera pas possible de produire des résultats avant d'avoir cumulé deux ou trois années. En effet, le poids des adresses de grande taille dans l'échantillon annuel peut perturber la significativité des résultats en introduisant des effets de grappe.

Proposition 24 : s'agissant du mode de restitution, il est proposé d'adresser les fiches à la commune par courrier et de rendre publics ces documents via le site *internet* de données locales sous forme d'un produit de diffusion en ligne.

Proposition 25 : organiser à l'égard des communes des présentations plus ou moins personnalisées selon la configuration des régions.

Le calendrier de restitution serait le suivant :

Le dispositif envisagé pour les communes de 10 000 habitants ou plus :

Année de diffusion	2004		2005		2006		2007		2008		2009
	2 ^e semes- tre	1 ^{er} semes- tre	2 ^e semes- tre	2 ^e semestre							
Source de données	2004	2004	2004+2005		2004+2005 +2006		2004+2005 +2006+2007		2004 +2005 +2006 +2007 +2008	2005 +2006 +2007 +2008 +2009	
Communes de 80 000 habitants ou plus	Estima- tion de popula- tion	N2	Estima- tion de popula- tion	N3	Estima- tion de popula- tion	N3	Estima- tion de popula- tion	N3	Popula- tion légale +N4	Population légale +N4	
Communes de 50 000 à 79 999 habitants	Estima- tion de popula- tion	N1	Estima- tion de popula- tion	N2	Estima- tion de popula- tion	N3	Estima- tion de popula- tion	N3	Popula- tion légale + N4	Population légale + N4	
Communes de 20 000 à 49 999 habitants			Estima- tion de popula- tion	N1	Estima- tion de popula- tion	N2	Estima- tion de popula- tion	N3	Popula- tion légale +N4	Population légale +N4	
Communes de 10 000 à 19 999 habitants					Estima- tion de popula- tion	N1	Estima- tion de popula- tion	N2	Popula- tion légale +N4	Population légale + N4	

Avec :

- **N1** : niveau 1 présentant l'estimation de la population de la commune accompagnée de grands ratios et des principales composantes de cette population ;
- **N2** : niveau 2 présentant l'estimation de la population de la commune accompagnée des principales caractéristiques regroupées de cette population ;
- **N3** : niveau 3 présentant l'estimation de la population de la commune accompagnée des principales caractéristiques détaillées de cette population ;
- **N4** : produit final de restitution présentant les caractéristiques détaillées de la population.

Il est à noter aussi que les restitutions issues de la seule enquête de recensement de 2004 rendront compte de la situation au 1^{er} janvier 2004 ; celles issues des deux enquêtes de recensement de 2004 et 2005 rendront compte de la situation au 1^{er} juillet 2004 (point médian des deux enquêtes de recensement) ; celles issues des trois enquêtes de recensement de 2004, 2005 et 2006 rendront compte de la situation au 1^{er} janvier 2005 (point médian des trois enquêtes de recensement) ; et ainsi de suite jusqu'à rejoindre, par sauts semestriels, la date du 1^{er} janvier 2006 qui sera la référence des premières productions annuelles détaillées en régime de croisière.

Après quoi, le nombre d'enquêtes de recensement utilisées demeurant égal à cinq, la date de référence se décalera d'un an tous les ans.

Le dispositif pour les communes de moins de 10 000 habitants

La collecte dans les communes de moins de 10 000 habitants étant exhaustive, les caractéristiques socio-démographiques des communes enquêtées peuvent être diffusées sous le format final dès le premier cycle. Cependant, les traitements de codification et de redressement feront l'objet d'une montée en charge progressive pour atteindre leur régime courant en 2006. Il est donc nécessaire de prévoir, d'ici 2006, un produit ne présentant que les résultats des variables ayant fait l'objet d'un traitement de vérification et/ou de redressement post-collecte.

Année de diffusion	2004		2005		2006		2007		2008		2009
	2 ^e semes- tre	1 ^{er} semes- tre	2 ^e semes- tre	2 ^e semes- tre							
Date de diffusion											
Source de données	2004		2005		2006		2007		2008	2009	
Communes de moins de 10 000 habitants	Estima- tion de popula- tion	N3b	Estima- tion de popula- tion	N3b	Estima- tion de popula- tion	N4	Estima- tion de popula- tion	N4	Popula- tion légale + N4	Population légale + N4	

Avec :

- **N3b** : niveau 3 présentant la population des résidences principales de la commune accompagnée des principales caractéristiques détaillées de cette population (variables intégrées dans le processus de codification-redressement provisoire). Ce produit est différent du produit nommé N3 dans le cas des communes de 10 000 habitants ou plus.
- **N4** : produit final de restitution présentant les caractéristiques détaillées de la population.

Une information synthétique pour une lecture territoriale

Si l'analyse ville entière est évidemment fondamentale, elle ne constitue souvent qu'une première étape dans les communes. En effet, pour pouvoir être appropriés par les différents services municipaux, les résultats du recensement doivent être "territorialisés" sur la base des découpages infra-communaux propres à chaque service : quartiers, périmètres scolaires, ZAC, unités d'action sociale...

Si les périmètres sont à "géométrie variable" selon les services, la "lecture territoriale" est généralement la même et s'articule autour de deux dimensions : l'espace et le temps. Concrètement, elle consiste à observer un "territoire d'étude" par rapport à un "territoire de référence" pour mettre en évidence les points "remarquables" propres au territoire étudié : positionner par exemple un quartier dans la ville. Elle consiste également à observer un "territoire d'étude" dans son évolution dans le temps : étudier l'évolution d'un quartier entre le recensement de 1990 et celui de 1999. Pour permettre une bonne utilisation des résultats annuels issus du recensement rénové, une réflexion autour d'une "information synthétique" paraît nécessaire.

Proposition 26 : concevoir des produits tenant compte des besoins spécifiques aux communes, notamment à l'infra-communal, et permettant la comparaison avec un territoire de référence et présentant les évolutions (cf. [proposition 18](#)).

La formation des utilisateurs

Plus encore que par le passé, l'utilisation des résultats du nouveau recensement nécessitera quelques précautions. On l'a vu avec les recensements généraux, la publication d'informations à des niveaux infra-communaux très fins, dans le but de faciliter les recompositions sur des zones spécifiques, conduisait cependant certains utilisateurs à les utiliser directement et sans précaution, négligeant la précision initiale des données et sa dégradation avec le temps. Il convient donc d'aider les utilisateurs, au moyen de documents pédagogiques mais aussi de formations. Des formations spécifiques aux utilisateurs communaux pourraient être envisagées avec le CNFPT ou des organismes spécialisés.

Proposition 27 : développer des outils pédagogiques permettant aux utilisateurs de se familiariser avec les résultats du nouveau recensement.

Proposition 28 : construire des formations spécifiquement destinées aux utilisateurs communaux, en développant l'aspect infra-communal.

Proposition 29 : constituer un groupe de travail avec des représentants de communes afin d'avancer dans la mise en œuvre des propositions 25 à 27.

7. LE RYTHME SOUHAITABLE DE MISE À JOUR DES ZONAGES

La mise en place du nouveau recensement conduit en théorie à la possibilité de revoir les zonages tous les ans. Naturellement, l'utilisation d'un zonage d'étude suppose une certaine pérennité afin notamment de mesurer des évolutions. On peut donc envisager deux options :

- se caler sur le rythme quinquennal de recensement au cours duquel on a pris en compte l'ensemble du territoire et revoir les principaux zonages tous les cinq ans,
- se caler sur le rythme des recensements généraux et revoir ces zonages tous les dix ans, afin de laisser une plus grande durée d'utilisation à périmètre inchangé.

Le groupe de travail préconise une certaine pérennité dans les zonages afin de faciliter les comparaisons temporelles. Un rythme de l'ordre de dix ans paraît satisfaisant pour la plupart des zonages.

En ce qui concerne les IRIS : un toilettage, sur la base des principes définis à l'occasion du recensement de 1999 devrait permettre de corriger les quelques anomalies observées dans certaines communes ou de prendre en compte les évolutions de l'urbanisation de certains quartiers. Ce toilettage devrait avoir lieu en 2006-2007 afin de pouvoir être mis en œuvre pour la diffusion des premiers résultats infra-communaux en 2008. L'avis de la CNIL est nécessaire pour engager une telle opération mais la Commission ne devrait pas y être opposée dès lors qu'elle restera strictement dans les limites d'un « toilettage ».

Proposition 30 : Envisager un rythme de mise à jour des zonages offrant un bon compromis entre prise en compte de l'évolution des territoires et capacité de comparaison temporelle.

Proposition 31 : Envisager un toilettage des IRIS en 2006-2007, selon les mêmes principes que ceux établis pour le RP 99, afin de pouvoir diffuser dès 2008 sur les IRIS mis à jour.

8. LA DIFFUSION DE RÉSULTATS SUR DES ZONAGES À LA DEMANDE

Jusqu'à présent les communes reconstituaient leurs territoires infra-communaux à partir des îlots, qui étaient dans le recensement traditionnel la plus petite brique territoriale. Avec les nouvelles modalités du recensement, l'îlot disparaît et est remplacé par l'IRIS 2000. Toutefois, le découpage en IRIS ne permet pas toujours de recomposer les différents zonages de la gestion communale utilisés par les services municipaux.

D'où le besoin exprimé par les communes d'obtenir avec le recensement des données sur des « agrégats d'adresses » autres que l'IRIS 2000. Cette possibilité fait partie des préconisations de la commission spéciale mise en place en application de l'article 158 de la loi n° 2002-276 du 27 février 2002, relative à la démocratie de proximité, qui recommande « à l'Insee d'examiner avec la CNIL la possibilité de produire, au bénéfice des utilisateurs qui viennent d'être cités, et s'agissant de communes découpées en IRIS, des tableaux statistiques portant sur des zones à la demande ».

Cette possibilité constituerait un véritable progrès par rapport aux recensements précédents et permettrait de travailler sur des territoires « en couloir » ce qui n'était pas possible avec une diffusion à l'îlot : étude d'impact de bruit (couloir aérien ou axe routier) ou de desserte d'une nouvelle ligne de transport (tramway, bus, ...).

Proposition 32 : mettre en place, avec l'accord de la CNIL, un dispositif permettant aux communes et aux aménageurs de disposer de résultats du recensement sur des agrégats d'adresses de taille semblable à celle des IRIS 2000.

9. PROPOSITIONS CONCERNANT LA DIFFUSION DU RIL

Le RIL est un **répertoire d'immeubles localisés**, en fait un répertoire associant **adresses localisées** et immeubles, constituant ainsi un « référentiel adresse métier », localisé par des coordonnées géographiques. Dans un premier temps, il comprend les adresses des immeubles d'habitation (au sens du recensement) celles des communautés et celles des établissements du répertoire SIRENE et ultérieurement celles des équipements au service du public. Il a été constitué pour les besoins du recensement de la population, pour lequel il constitue la base de sondage des adresses dans les communes de 10 000 habitants ou plus.

Le RIL ne concerne pour le moment que les communes de 10 000 habitants ou plus. Cela représente environ 900 communes regroupant :

- 4 % du territoire français,
- 30 millions d'habitants, soit environ la moitié de la population française,
- 5 millions d'adresses d'habitation,
- 2 millions d'adresses d'entreprises du fichier SIRENE, dont un bon nombre est commun aux adresses d'habitation.

Ce répertoire a été créé par arrêté du 19 juillet 2000 (modifié par arrêté du 9 octobre 2002). Il est réalisé par l'Insee au sein des directions régionales, en association avec les communes dont l'expertise est sollicitée.

A l'occasion des travaux du groupe, une présentation du projet de référentiel à grande échelle (RGE) a été faite par le secrétaire général du Conseil national de l'information géographique (CNIG). Le concept de RGE est né du rapport du député Langagne, qui a été remis le 30 septembre 1999 ; il portait sur l'Institut géographique national (IGN) et sa place dans la société nationale de l'information. A l'issue de ce rapport, il avait été préconisé de mettre en place le plus rapidement possible un ensemble de données géographiques cohérentes, permettant une couverture totale du territoire national, avec une précision métrique. Par ailleurs, toujours à la suite de ce rapport, une instruction a été conduite par le ministère de l'Équipement qui a décidé en février 2001 de constituer sous la maîtrise d'œuvre de l'IGN le RGE de précision métrique.

Le RGE est non seulement une orthophotographie (une photographie aérienne « redressée » pour la rendre superposable à la carte) ; mais également un ensemble de données topographiques, qui vont décrire la situation topographique du territoire ; il contient également une composante parcellaire, dans la mesure où il décrit la propriété en s'appuyant sur le plan cadastral de la Direction générale des Impôts. Enfin, la quatrième composante est la composante adresse. Il convient de signaler que les quatre composantes sont superposables les unes aux autres.

Suite à la décision interministérielle survenue au cours de l'année 2001, le CNIG a conduit une réflexion sur la composante adresse, et a précisé quels étaient les enjeux nécessaires à l'existence des adresses géo-référencées au niveau de l'administration française. Les enjeux concernent des gains en matière de distribution du courrier, une réduction des coûts de fonctionnement des services

publics, tant au niveau de l'État que des collectivités territoriales, la modernisation des services publics. Des gains étaient attendus pour les gestionnaires de réseau, dans la mesure où ces organismes sont confrontés à des problèmes de géo-référencement au niveau des équipements mis en place et de leurs clients. Enfin, de nouvelles activités économiques peuvent surgir de l'utilisation d'adresses géo-référencées : à partir d'un point précis, il peut être intéressant de connaître, dans un voisinage proche, tous les établissements présentant un intérêt pour une activité économique donnée.

Les membres du groupe ont fait part d'une certaine perplexité devant l'apparente hétérogénéité induite par la perspective de la coexistence du RIL et du RGE et ont souhaité que l'IGN et l'Insee travaillent en étroite collaboration à aboutir à une convergence. On doit comprendre cependant qu'il y a une complémentarité entre ces deux approches, le RGE ayant vocation à satisfaire l'ensemble des besoins liés à l'adressage alors que le RIL doit être vu comme une approche métier de cet élément de référence.

Par ailleurs, la diffusion du RIL, non encore autorisée par les arrêtés de constitution du RIL, a fait l'objet de propositions de la part du groupe. En effet, le RIL présente l'intérêt d'un répertoire exhaustif des adresses d'une commune. Les informations rattachées, comme le nombre de logements de chaque adresse, sont initialisées à partir du recensement de 1999 et mis à jour exhaustivement pour les logements neufs et avec une probabilité de 40 % en cinq ans pour les logements existants. En tant que telles, ces données seraient très utiles à la connaissance fine de la répartition territoriale du parc de logements d'une commune.

De plus, l'utilisation du RIL comme clé de répartition, permettant de ventiler approximativement des informations générales diffusées à l'IRIS, sur des sous-ensembles particuliers définis en termes d'adresses, serait très précieuse pour de nombreux utilisateurs. A titre d'exemple, la ventilation pourrait se faire à partir du nombre de logements de chaque adresse. Ainsi, les publics non habilités à obtenir des résultats du recensement sur des zones infracommunales à la demande, pourraient cependant calculer une clé de répartition des résultats à l'Iris afin de réaliser une approximation des caractéristiques de territoires infracommunaux. Cette possibilité est particulièrement souhaitée par les entreprises de conseil chargées de constituer des dossiers administratifs de projets d'implantation commerciale. Il est rappelé que lors du recensement de 1999, 14 données (sexe et âge) étaient diffusibles à l'îlot pour tous les publics.

Proposition 33 : favoriser la convergence des projets RGE et RIL afin que les communes ne soient pas confrontées à des normes différentes et/ou des investissements redondants.

Proposition 34 : charger l'Insee d'instruire un dossier de demande d'autorisation de la CNIL pour la diffusion du RIL afin de valoriser cette source de connaissance de la répartition territoriale du parc de logements et de « localisation » des résultats du recensement à l'IRIS.

