Nations Unies ECE/CES/GE.41/2010/

Conseil économique et social

Distr. générale
1 juillet 2010

Original : Français

Commission économique pour l'Europe

Conférence des statisticiens européens

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

Treizième réunion
Genève, 7-9 juillet 2010
Point 2 de l'ordre du jour
Communication et diffusion des données de recensement, y compris l'utilisation de systèmes d'information géographique (SIG)

La diffusion des résultats du recensement français depuis 2009

Note établie par l'Institute national de la statistique et des etudes économiques, France

Résumé

La diffusion des résultats du recensement français est annuelle depuis 2009. Cette diffusion a lieu exclusivement sur internet, sous la forme de produits variés adaptés à différents types d'utilisateurs. Une documentation et un accompagnement spécifique visent à rendre les internautes autonomes dans leur usage des données. Les produits de diffusion ont été conçus après des consultations d'utilisateurs, en tenant compte des caractéristiques méthodologiques de l'opération et de l'avis de la CNIL, commission en charge de la surveillance du respect des libertés et de la vie privée. Dès leur deuxième édition, en 2010, ils ont évolué pour tenir compte des premières réactions des utilisateurs.

Merci de recycler

I. Introduction

1. Comme tout recensement, le recensement français permet d'une part d'élaborer des données « officielles » de population à tous les échelons géographiques et d'autre part de construire des statistiques détaillées décrivant la population et la société à des niveaux géographiques fins. Dans le cas de la France, il faut tenir compte de certaines particularités : le nombre important d'unités administratives (36 681 communes – Local Administrative Units (LAU 2)), la grande diversité de leurs tailles (la moitié des communes ont moins de 420 habitant et la moitié de la population vit dans les 930 plus grandes communes) et le mode spécifique de recensement qui permet la mise à jour de l'ensemble des données tous les ans mais nécessite, en contrepartie, certaines précautions d'usage.

2. Ce texte ne porte que sur la diffusion des résultats statistiques, il n'aborde pas celles des populations légales, diffusées annuellement sur internet en fin d'année[1].

3. Il décrira d'abord la stratégie définie par l'Institute national de la statistique et des etudes économiques (L'Insee) pour construire son offre de données, puis la nature des produits de diffusion eux-mêmes, enfin quelques éléments de bilan et les évolutions envisagées.

II. La stratégie de diffusion des données statistiques

A. Tous les résultats mis à jours tous les ans

4. La méthode de recensement mise en place en France depuis 2004 (voir annexe) permet l'élaboration de résultats complets de recensement chaque année. Ces résultats utilisent des données tirées des 5 dernières enquêtes annuelles de recensement, ramenées toutes à la même année de référence. Ainsi en 2009 on a diffusé des données tirées des enquêtes de recensement 2004 à 2008, ayant pour année de référence 2006, puis en 2010 des données issues des collectes 2005 à 2009, correspondant à l'année de référence 2007, etc… Chaque année on produit donc des chiffres nouveaux, mais qui ne correspondent que partiellement à une information nouvelle (seul un cinquième de l'information est vraiment nouvelle). Les données des millésimes successifs ne permettent donc pas de constituer de réelles séries temporelles annuelles. Seuls les résultats issus de deux cycles de 5 années de collecte disjointes peuvent être comparés en toute rigueur car ils sont totalement indépendants.

5. Malgré cette difficulté, l'Insee a décidé de publier, chaque année, une série complète de résultats du recensement : cela correspond à une demande forte des utilisateurs de pouvoir utiliser à chaque instant les informations les plus récentes.

6. Ceci permet également une cohérence entre résultats statistiques et chiffres officiels de population. Ces derniers, conformément à la loi qui régit le recensement, doivent être mis à jour annuellement. Ainsi chaque année, durant les dernières jours de décembre, les populations de toutes les entités géographiques sont publiées et authentifiées. Six mois plus tard, en juillet, toutes les données statistiques sont publiées en parfaite cohérence avec les précédentes. Cette mise en cohérence était une demande forte des utilisateurs, notamment les acteurs publics locaux.

[1] http://www.insee.fr/fr/ppp/bases-de-donnees/recensement/populations-legales/

B. Internet, et seulement Internet

7. Jusqu'au recensement de 1999, les résultats détaillés des recensements de la population faisaient l'objet d'une importante diffusion sous la forme de volumes papier, dont l'addition représentait plusieurs m^3. Cette diffusion était complétée d'une diffusion sur internet, plus tardive. Conséquence logique de la décision de mettre à jour chaque année l'ensemble des résultats, il n'était plus possible d'envisager le maintien de publications sous forme papier : outre les arguments de coût et de développement durable, compte tenu des délais d'impression et de distribution, ils auraient été « périmées » presque dès leur édition.

8. Depuis 2009, les résultats détaillés du recensement ne sont donc disponibles que sur internet. Cette offre est entièrement gratuite, comme toute l'offre de l'Insee qui ne nécessite par de traitement « à la demande ». La gratuité a été étendue, par rapport aux recensements précédents, à des types de produits autrefois payants et/ou réservés à certains publics, notamment les fichiers de micro-données.

9. Notons que même les chiffres de population officielle des collectivités locales, qui sont authentifiés par décret, ne sont pas imprimés : le texte qui les authentifie stipule que « Les chiffres de la population [..] des communes, […] sont arrêtés aux valeurs figurant dans les tableaux consultables sur le site Internet de l'Institut national de la statistique et des études économiques (www.insee.fr). »

C. Souhaits des utilisateurs et respect du secret statistique

10. Les utilisateurs du recensement ont été consultés grâce au CNIS (Conseil national de l'information statistique), organisme de concertation entre utilisateurs et producteurs de la statistique publique, où siègent des représentants des administrations, des partenaires sociaux et des associations. Un groupe de travail spécifique sur l'utilisation et la diffusion des données du recensement a fonctionné en 2003 et 2004, rassemblant plus de 70 utilisateurs d'horizons très variés. Ses recommandations portaient notamment sur la nécessité de tenir compte de la diversité des publics (de l'élu au chercheur) et des niveaux géographiques (du niveau national au quartier), et sur l'exigence de bien informer sur les précautions particulières à prendre du fait de la nouvelle méthode. L'Insee s'est efforcé de prendre en compte l'ensemble de ces demandes lors des travaux de définition des produits.

11. Bien entendu, la définition des produits de diffusion du recensement a fait l'objet de discussions avec la CNIL (Commission nationale Informatique et libertés). De ces discussions ont résulté un arrêté de diffusion qui précise la nature des produits de diffusion autorisés et les niveaux de détail des données (précision géographique ou détail de nomenclature) selon les produits. Ce texte précise également les limites spécifiques aux variables jugées « sensibles » , à savoir les pays de naissance, les pays de résidence antérieure et la date d'arrivée en France. Tous les produits ont été conçus conformément à ces exigences.

D. Rendre l'utilisateur autonome

12. Les données du recensement sont traditionnellement, parmi celles diffusées par les INS, celles qui sont le plus utilisées, notamment par des acteurs locaux. Grâce à Internet et du fait de l'élargissement de l'offre, on pouvait s'attendre à une augmentation le nombre d'utilisateurs. De plus, la nouvelle méthode de recensement (sondage en grandes communes, étalement sur 5 années de la collecte) rend nécessaire certaines précautions d'usage. Enfin, comme à chaque recensement, certaines évolutions du questionnement et des concepts rendent les comparaisons avec les recensements précédents délicates.

13. Or l'Insee ne peut pas assurer une formation, ni même un accompagnement personnalisé à chaque utilisateur. Même si il est organisé pour répondre aux questions des internautes, le risque d'un trop grand nombre d'interrogations par rapport aux moyens de réponses personnalisées était fort. Un grand effort a donc été fourni, dans la conception du site, pour rendre autant que possible l'utilisateur autonome.

14. L'essentiel des résultats statistiques figurent dans un sous-site spécifique du site insee.fr, http://www.recensement.insee.fr, qui a ouvert le 1er juillet 2009, soit 15 mois seulement après la fin de la dernière collecte prise en compte. Cet espace est conçu pour faciliter la consultation grâce à un mode de navigation intuitif.

15. Le site s'articule autour de trois rubriques consacrées aux données statistiques (repérables en rouge) et de deux autres destinées à favoriser l'autonomie de l'internaute (repérables en bleu) : rubrique *Documentation* et rubrique *Aide* (pour la recherche géographique ou l'utilisation des bases téléchargeables).

16. La rubrique « Documentation » est particulièrement riche : elle comprend d'abord une série de conseils d'utilisation des données, portant à la fois sur les spécificités liées à la méthode (précision, temporalité) et sur l'utilisation des données thème par thème (avec notamment des précisions sur la comparabilité avec les RP précédents). On y trouve aussi une liste des variables utilisées avec leurs définitions, une présentation de l'organisation des niveaux géographiques dans les données du recensement et une présentation des grands types de produits présents sur le site.

17. De plus, sur chaque page du site qui présente un produit (fiche, tableau ou base de données) figurent également des fichiers de documentation en format .pdf, facilement téléchargeables, adaptés au thème et à la nature des données concernées.

18. Même si l'Insee n'a pas les moyens de former tous les utilisateurs du recensement, un module d'accompagnement à l'utilisation des résultats a été conçu en juin 2009. Ce module a été mis en œuvre dans la plupart des régions auprès de 5 000 correspondants de l'Insee dans les services de l'État, les principales collectivités territoriales et des organismes d'étude dans les domaines de l'urbanisme et de l'aménagement local. Ce module rappelle les principes de collecte, présente les produits de diffusion et développe les principaux conseils d'utilisation des résultats sur des aspects de précision et pour les thèmes qui font l'objet de changements de concepts par rapport à 1999.

III. Des produits adaptés aux différents publics

19. Même si tous les produits présents sur le site sont dans les faits accessibles à tous les internautes, ils ont été conçus en fonction de « cibles » différentes que l'on peut qualifier ainsi :

(a) Le public en général : qu'il s'agisse du grand public, des décideurs, des élus, des utilisateurs pressés, des médias…, chacun doit disposer de produits immédiatement utilisables;

(b) Les professionnels : ils peuvent construire en toute autonomie une information adaptée à leurs besoins propres ou à ceux de leurs commanditaires en exploitant les bases de données mises en ligne.

En dehors de ces deux catégories une offre spécifique, payante, existe également.

A. Des produits « grand public », immédiatement utilisables.

20. Pour ces utilisateurs, les produits doivent bénéficier d'une ergonomie adaptée : consultables à l'écran, faciles à imprimer et pouvant être utilisés et valorisés directement, sans retraitement. Les données ont été choisies pour pouvoir être utilisées sans précautions d'usage particulières, en particulier en terme de significativité. L'utilisateur peut choisir, parmi les 40 000 territoires proposés, la zone géographique sur laquelle il cherche des données (un moteur de recherche spécifique l'y aide), puis le thème de sa recherche.

21. Trois type de produits appartiennent à cette catégorie : des fiches de chiffres clés, des tableaux détaillés et des cartes d'indicateurs:

(a) Les chiffres clés rassemblent les résultats essentiels du recensement de la population. Ils sont présentés en six fiches thématiques proposant tableaux et graphiques simples. La plupart des données sont présentées en évolution, essentiellement par la comparaison avec les résultats du recensement de 1999. Ils sont disponibles pour chaque commune quelle que soit sa taille et pour les territoires habituels, administratifs ou d'étude, constitués d'un ensemble de communes entières. En 2009, chaque fiche est exportable en format .pdf;

(b) Les tableaux détaillés regroupent un ensemble de plus de 80 tableaux. Ils fournissent des résultats détaillés couvrant l'ensemble des thèmes du recensement de la population. Plus riches en détail d'information que les chiffres clés, les tableaux détaillés fournissent les effectifs correspondant au croisement de deux, trois ou quatre variables statistiques portent uniquement sur le recensement le plus récent. Ils ne proposent pas de données comparatives dans le temps, mais ils sont organisés selon la même structure thématique que les résultats du recensement de la population de 1999. Ils ne sont disponibles que pour des zones de plus de 2000 habitants. Selon la taille du territoire concerné (2 à 10 000 habitants, 10 000 à 50 000, plus de 50 000 habitants), certains tableaux sont plus ou moins détaillés. Par exemple, le secteur d'activité économique sera disponible en 5, 16 ou 36 postes selon la taille de la zone. Chaque tableau est exportable en format .xls ou .txt;

(c) Les cartes d'indicateurs **:** Destinées à un public large, ces cartes thématiques sont publiées sur le site Statistiques locales de l'Insee (lien direct depuis le site du recensement). Elles sont aussi à l'usage des professionnels grâce aux fonctionnalités de l'outil cartographique. En effet, il s'agit de cartes dynamiques sur la France métropolitaine et d'outre-mer par région, département, zone d'emploi, canton ou commune. Les indicateurs cartographiés, sont, pour la plupart, présents dans les fiches chiffres clés. Dans la mesure du possible, un historique sur plusieurs années ou plusieurs périodes est proposé pour ces indicateurs. L'ensemble des données sont exportables.

B. Pour les professionnels, des bases de données téléchargeables

22. En plus des produits immédiatement exploitables, les professionnels ont besoin de données pouvant être retravaillées. Plus sensibilisés aux questions de significativité statistique, de comparabilité entre millésimes ou d'évolutions des concepts, ils devront consulter la documentation et utiliser leurs propres outils de traitement et d'édition.

23. L'Insee met à leur disposition des données téléchargeables de deux types : le premier type, appelé « bases » réclame seulement une bonne connaissance des fonctionnalités d'un tableur. Elles sont le plus souvent disponibles sous deux formats, .xls et .txt. Les « fichiers de micro données » qui constituent le second type, doivent être lus et traités avec des logiciels de traitement de bases de données ou des logiciels statistiques.

Une documentation complète et des conseils d'utilisation des produits sont présentés sous forme de fiches .pdf associées aux différents fichiers.

24. La première série de bases téléchargeables reprend tous les indicateurs des «chiffres clés » ou des « tableaux détaillés » dans des tables permettant d'additionner à volonté des données communales. Elles comprennent donc des données détaillées même pour les petites communes, afin de ne pas contraindre les usages par des experts.

25. Les communes de plus de 5 000 habitants sont découpées en quartiers fixes d'environ 2 000 habitants : les Iris. Ce découpage a été réalisé en 1999 pour permettre la diffusion de statistiques publiques infra-comunales en respectant le secret statistique. Cinq bases thématiques proposent des données essentielles au niveau des quartiers Iris, afin de pouvoir effectuer des regroupements sur des zones infra-communales composées d'un ensemble d'Iris. Ces bases sont accompagnées d'indicateurs de robustesse statistique (pour chaque zone) et de coefficients de variation (pour toutes les modalités des variables diffusées), ce qui permet à l'utilisateur expert d'adapter l'usage qui en est fait à la précision de la donnée.

26. Les Bases sur les flux de mobilité fournissent, pour les migrations résidentielles, les déplacements domicile-lieu de travail et les déplacements domicile-lieu d'études, les flux totaux de mobilité entre communes qui permettent tous les regroupements souhaités.

27. Les bases téléchargeables permetent aux utilisateurs toutes les agrégations communales d'indicateurs déjà calculés. Pour réaliser leurs propres tableaux, des fichiers de micro données sont proposés aux professionnels avertis (organismes publics, collectivités territoriales, grandes entreprises, bureaux d'études et de conseil, chercheurs…). Ces fichiers donnent des informations au niveau des personnes ou des logements. Ils permettent de faire des analyses exploratoires de données, de modéliser des comportements, ou simplement de réaliser une tabulation sur une sous-population particulière. Du fait de la méthode, chaque donnée individuelle doit être utilisée avec une pondération, comme dans les enquêtes par sondage. Mais les tailles de fichier ne sont pas du même ordre. Ces fichiers volumineux, disponibles en format .txt, nécessitent des gestionnaires de bases de données ou des logiciels statistiques. Ils sont de trois types :

(a) Un fichier de logements avec des données localisées à la commune décrivant les caractéristiques des logements et celles des ménages qui les occupant;

(b) Deux fichiers de personnes :

(i) L'un avec des données localisées à la région portant sur les caractéristiques de chaque personne recensée, celles de son ménage, ainsi que celles de sa résidence principale. Les variables y sont disponibles avec des modalités détaillées (y compris nationalité, pays de naissance, ancienneté d'arrivée en France) ;

(ii) L'autre avec des données localisées au canton-ou-ville et département, avec des modalités moins détaillées pour des raisons de confidentialité;

(c) Cinq fichiers de personnes qui permettent d'étudier, de commune à commune, les déplacements domicile - lieu de travail, les déplacements domicile - lieu d'études et les migrations résidentielles. Le nombre de variables descriptives des individus y est limité.

C. Quelques prestations spécifiques payantes

28. Pour des demandes spécifiques qui ne trouvent pas satisfaction par l'intermédiaire des produits précédents, l'Insee propose des « produits sur mesure » qui font l'objet d'une tarification. Par ailleurs, pour certains organismes ayant une mission de service public, un service complémentaire payant pour la « diffusion infra communale de zones à façon » sera

proposé à partir de 2011 : les communes pourront passer commande à l'Insee de chiffres-clés sur des zones d'au moins 1 000 logements définies "à façon", en fonction des besoins liés à leurs politiques (zones de bruit, restructuration urbaine, impact d'un équipement…).

IV. Premier bilan, premières perspectives

A. Le coût du dispositif

29. La mise en place de l'ensemble de ce dispositif de diffusion a eu un coût important. Il a fallu définir précisément le contenu de chaque produit (fiche, tableau ou base), construire les variables ou les modalités sous-jacentes, concevoir l'organisation du site internet, le développer, puis tester l'ensemble.

30. Les forces de travail mobilisées au niveau des statisticiens, des concepteurs de produits et des informaticiens, entre 2007 et 2009, sont évaluées à 11 années de travail de cadres de haut niveau. Le développement du site lui-même, sous-traité à une société spécialisée, a coûté à l'Insee 280 000 euros. Au total, la mise en place de la diffusion des résultats statistiques du recensement français a coûté 1,4 million d'euros.

31. Sa mise à jour annuelle, qui comprend le renouvellement total des données, la correction d'erreurs et quelques évolutions, mobilise environ 4 années/personnes et des ressources en sous-traitance, pour un coût de l'ordre de 400 000 euros.

B. Succès et nouvelles demandes

32. Depuis son ouverture, le site consacré au recensement est visité par près de 60 000 internautes par mois. Le test réalisé à l'été 2009 et les réactions recueillies depuis montrent que le site est très apprécié. Fortement attendu en termes de contenu, il apparaît accessible, facile d'utilisation grâce à une présentation sobre avec des textes introductifs jugés assez explicites et l'impression d'un accès facilité aux données principales. La nouvelle interface apparaît classique, familière et adaptée, son appropriation est immédiate. Le site a globalement séduit l'ensemble de ses publics, y compris celui des professionnels de l'information statistique qui apprécient sa richesse.

33. Mais, comme toujours, cette mise à disposition a suscité des demandes supplémentaires, qui s'expriment essentiellement à travers plusieurs canaux :

(a) Des internautes interrogent directement l'Insee, par mail ou par téléphone, et font part de leurs souhaits ou de leurs suggestions;

(b) Un groupe du Conseil national de l'information statistique (CNIS) rassemblant divers publics - collectivités locales, entreprises, chercheurs,… - a en charge les propositions d'amélioration des produits de la diffusion du recensement et de son accompagnement;

(c) A l'occasion des travaux d'étude ou d'expertise menées par les directions régionales de l'Insee avec les acteurs publics locaux, ces derniers font part de leurs remarques et demandes;

(d) Lors de l'ouverture du site en juillet 2010, une enquête par internet est organisée pour compléter ces remontées.

34. Les demandes des utilisateurs portent parfois sur la forme des produits de diffusion, parfois sur les données elles-mêmes, parfois sur la documentation du site. L'Insee analyse

les demandes les plus récurrentes et s'efforce, quand cela est possible, d'y donner suite dans les versions ultérieures du site.

C. Des évolutions dès 2010

35. Dès la deuxième édition du site, plusieurs améliorations ont été apportées. En voici quelques exemples :

(a) Il sera désormais possible d'éditer et d'exporter, en un seul fichier .pdf l'ensemble des fiches « chiffres clés » d'un territoire, qui étaient réparties en 6 fiches thématiques en 1999. L'internaute pourra ainsi d'un seul clic obtenir le portrait de son territoire au recensement;

(b) Les données des fiches « chiffres clés » d'un territoire pourront être exportées en format .xls, au lieu du seul format .pdf;

(c) Des niveaux géographiques sont rajoutés dans les fichiers de micro donnée. Dans le fichier logement, on trouve notamment le code du quartier IRIS. Dans ce domaine, l'Insee est allé aux limites de ce qui a été autorisé par la CNIL;

(d) Des bases de données téléchargeables nouvelles ont été construites pour permettre directement les comparaisons entre les 26 régions (nomenclature d'unités territoriales statistiques (NUTS) 2) et les 100 départements (NUTS 3) sur tous les indicateurs. Dans la version de 2009, l'usager devait lui-même reconstituer ces zones à partir des bases communales.

V. Conclusion

36. La diffusion des résultats statistiques détaillés du recensement peut être considérée comme un succès : une grand nombre de données, correspondant aux attentes d'utilisateurs variés, ont été mises à disposition dans le calendrier prévu. Ils ont été bien accueillis et sont déjà largement utilisés.

37. Pour l'avenir, l'Insee doit envisager un mode de mise à disposition moins coûteux en moyens humains, pour pouvoir à la fois répondre à l'impératif des mises à jours annuelles et répondre à de nouvelles demandes. Ces nouvelles modalités devront également permettre d'intégrer les résultats du recensement dans l'organisation globale du site internet de l'Insee : il faudra passer d'une logique de source à une logique de thème. Le calendrier de cette refonte n'est pas défini.

Annexe : La nouvelle méthode de recensement

1. La nouvelle méthode de recensement substitue au comptage traditionnel organisé tous les huit ou neuf ans une technique d'enquête annuelle permettant aux communes de disposer de données fraîches, actualisées chaque année. La méthode d'enquête varie selon un seuil de population fixé par la loi à 10 000 habitants.

2. **Les communes de moins de 10 000 habitants sont recensées une fois tous les cinq ans par roulement.** Elles ont été réparties en cinq groupes, selon des règles précises qui assurent un nombre d'habitants équivalent dans chaque groupe. Chaque année, l'enquête de recensement porte sur la totalité de la population et des logements des communes du groupe concerné. Au bout de cinq ans, l'ensemble de la population des communes de moins de 10 000 habitants aura été recensé.

3. **Dans les communes de 10 000 habitants ou plus, un échantillon de la population est recensé chaque année.** La collecte annuelle porte sur un échantillon d'adresses représentant environ 8 % de la population. Au bout de 5 ans, l'ensemble du territoire de chaque commune aura été pris en compte et 40 % environ des habitants de ces communes auront été recensés.

4. La fréquence de la collecte est donc quinquennale pour les communes de moins de 10 000 habitants et annuelle pour les communes de 10 000 habitants ou plus. L'enquête de recensement est exhaustive dans le premier cas, c'est une enquête par échantillon dans le second cas. Concrètement, environ 9 millions de personnes sont recensées chaque année, soit 14 % de la population vivant en France.

5. **Les données diffusées sont issues des 5 dernières enquêtes réalisées** : elles portent donc sur 45 millions de personnes soit 70% de la population. Les données collectées au cours de 5 années sont ramenées à une même année de référence (l'année médiane) par des méthodes d'interpolation, d'extrapolation et de moyennes mobiles.