



**Conseil économique
et social**

Distr.
GÉNÉRALE

ECE/CES/GE.41/2007/6
22 mars 2007

FRANÇAIS
Original: ANGLAIS

COMMISSION ÉCONOMIQUE POUR L'EUROPE

CONFÉRENCE DES STATISTICIENS EUROPÉENS

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

Dixième session

Astana, 4-6 juin 2007

Point 3 b) de l'ordre du jour provisoire

**TECHNOLOGIE UTILISÉE DANS LE CADRE DES RECENSEMENTS: FAITS
NOUVEAUX ET INCIDENCES SUR LES MÉTHODES DE RECENSEMENT**

Mise en correspondance des fichiers aux fins du recensement aux Pays-Bas

Communication des Pays-Bas*

Cette session est organisée conjointement avec Eurostat.

Résumé

À la réunion qu'il a tenue à Washington, DC (États-Unis) les 19 et 20 octobre 2006, le Bureau de la Conférence des statisticiens européens a approuvé le renouvellement du mandat du Groupe directeur des recensements de la population et des habitations ainsi que le plan des futures activités de la Conférence concernant les recensements de la population et des habitations. Le Bureau est également convenu que le Groupe directeur coordonnerait les travaux liés aux divers types de réunions.

Le présent document a été établi à la demande du Groupe directeur afin d'être présenté et examiné à la Réunion commune CEE/Eurostat sur les recensements de la population et des habitations, qui aura lieu à Astana (Kazakhstan) du 4 au 6 juin 2007. Il constitue une base concrète de discussion pour la séance qui sera consacrée à la «technologie utilisée dans le cadre des recensements: faits nouveaux et incidences sur les méthodes de recensement».

* Document établi par Statistics Netherlands.

RÉCAPITULATIF

1. Aux Pays-Bas, les données provenant de nombreuses sources différentes ont été combinées aux fins de l'établissement des tableaux de recensement de 2001. Depuis 1971, date du dernier recensement réalisé au moyen d'un dénombrement complet, la volonté de participation de la population a décliné de façon spectaculaire. Statistics Netherlands a trouvé le moyen de remédier au problème: le recensement virtuel, qui exploite les registres et résultats d'enquêtes disponibles. Les tableaux qui en résultent sont non seulement comparables avec les recensements précédemment réalisés aux Pays-Bas, mais aussi avec ceux des autres pays qui ont participé à la série de recensements de 2000.

2. Pour le recensement de 2001, davantage d'informations détaillées sont requises par rapport aux recensements antérieurs. L'expérience que Statistics Netherlands a acquise en matière de traitement à des fins statistiques des données provenant de différents registres administratifs lui a permis de constituer une base de données statistiques sociales, qui contient des données démographiques et socioéconomiques détaillées et cohérentes sur les personnes et les ménages. Le Registre de la population en est la pièce maîtresse. La base de données est édifiée en mettant en relation les microdonnées qui proviennent de plusieurs registres administratifs et les résultats des enquêtes par sondage, procédé qui garantit la cohérence et l'exhaustivité des données présentes dans la base. Les enquêtes par sondage demeurent indispensables lorsque les informations ne figurent pas dans les registres.

Mots clefs: recensement, micro-intégration, mise en relation des microdonnées

I. INTRODUCTION

3. Le premier recensement réalisé aux Pays-Bas s'est déroulé en 1795, pour l'établissement des circonscriptions électorales. Les Provinces-Unies qui constituaient alors les Pays-Bas étaient encore une république, et les frontières étaient différentes des frontières actuelles. Après Napoléon, les Pays-Bas sont devenus un royaume, et un recensement a été organisé tous les dix ans. Le premier a eu lieu en 1829. Avant la création de Statistics Netherlands, six autres recensements ont été effectués en 1839, 1849, 1859, 1869, 1879 et 1889, sous la responsabilité du Ministère de l'intérieur. En 1899, Statistics Netherlands a été créé et chargé de procéder au huitième recensement. Au XX^e siècle, six autres recensements traditionnels ont été menés en 1909, 1920, 1930, 1947, 1960 et 1971. Les trois derniers recensements (1981, 1991 et 2001) ont été réalisés non sur la base d'un dénombrement complet de la population mais à partir de registres et de résultats d'enquêtes que Statistics Netherlands a pu utiliser.

4. Initialement, les recensements avaient un objectif double: corriger les erreurs des registres municipaux de la population, et obtenir des informations complémentaires sur des phénomènes socioéconomiques dans le pays. Depuis que les recensements sont effectués à partir de registres, le premier objectif est devenu caduque. Par ailleurs, la qualité du Registre central de la population, qui regroupe tous les registres municipaux de la population, s'est considérablement améliorée au fil du temps. Les municipalités ont en effet intérêt à tenir à jour leurs registres de la population car elles reçoivent de l'administration centrale une allocation qui est généralement calculée en fonction de la taille de la population d'après les registres locaux. Une autre raison tient au fait qu'il est dans la pratique très difficile de vivre au sein de la société néerlandaise sans être inscrit sur le Registre de la population. Ainsi, tant les municipalités que les citoyens sont

suffisamment encouragés à préserver la qualité de ces registres. Les mesures prises récemment par la municipalité de Rotterdam pour améliorer la qualité du registre municipal de la population relatif à certains vieux quartiers en est la preuve. Le second objectif demeure valable aujourd'hui, et bon nombre de résultats de recensement sont publiés dans un contexte historique ou international. Actuellement, les données du recensement sont très utilisées pour les comparaisons entre les pays.

5. Le premier recensement traditionnel de la population réalisé dans le Royaume des Pays-Bas, tel que prescrit par un décret royal, s'est déroulé en 1829. Il a marqué le début d'une période de cent quarante années dans l'histoire du pays, pendant laquelle 14 recensements traditionnels ont été réalisés; le dernier remonte à 1971. Depuis lors, les objectifs d'un gouvernement qui réunit toutes sortes d'informations sur les citoyens ont suscité une méfiance de plus en plus grande. La société s'est montrée moins coopérative, obligeant Statistics Netherlands à trouver des solutions de remplacement au recensement traditionnel de la population. Plutôt que de dénombrer la population sur le terrain, Statistics Netherlands a étudié la possibilité d'utiliser les registres administratifs et les enquêtes par sondage comme nouvelles sources de données, en vue de se procurer les informations nécessaires pour le recensement.

6. Lors des recensements de 1981 et 1991, les données démographiques ont été extraites du Registre de la population. Celles relatives aux caractéristiques socioéconomiques, telles que l'emploi et l'éducation, ont été fournies par l'Enquête sur les forces de travail. Cependant, ces deux sources ont été exploitées séparément, et il n'a pas été prêté une attention particulière à la cohérence des informations au niveau des microdonnées. De plus, il existait parfois des divergences entre les totaux des tableaux fournis par une source et les totaux correspondants de l'autre source. Pour corriger ce problème de cohérence, les résultats des tableaux ont été repondérés au niveau des totaux du Registre de la population (pour ce qui est des résultats obtenus dans le cadre du recensement de 1981, voir Vliegen et Van de Stadt, 1988; pour la compilation des résultats du recensement de 1991, voir Corbey, 1994). Les données recueillies pour les recensements de 1981 et 1991 étaient bien moins détaillées que la série de tableaux correspondant à celui de 2001. Contrairement à ce qui s'est passé pour les deux recensements précédents, Statistics Netherlands a publié les données du recensement de 2001 au niveau des municipalités.

7. Dans le cadre du Programme de recensement de 2001, Statistics Netherlands a adopté une nouvelle démarche, unique en Europe. L'un des acquis les plus importants des années 90 dans le domaine des statistiques sociales est la possibilité d'acquérir une quantité croissante d'informations statistiques socioéconomiques à partir des registres administratifs. Il est maintenant possible de disposer d'informations complètes et détaillées sur l'emploi et la sécurité sociale. En reliant et en intégrant les microdonnées démographiques et socioéconomiques provenant d'une grande diversité de registres administratifs et d'enquêtes par sondage, Statistics Netherlands a constitué une base de données statistiques sociales, qui renferme des informations cohérentes et détaillées sur les personnes, les ménages, les emplois et les prestations (sociales), et constitue donc une source de données appropriée pour le recensement de la population de 2001. La cohérence entre les totaux (partiels) fournis par les enquêtes par sondage et les registres est obtenue grâce à la pondération séquentielle, nouvelle technique de pointe en la matière.

8. La présente communication s'appuie largement sur les chapitres 1 et 13 de l'ouvrage de Schulte Nordholt *et al.* (2004). La section II ci-après analyse plus en détail le recensement de 2001. La méthode employée pour compiler les données de ce recensement est exposée à la section III. L'association des sources de données, en particulier le couplage des microdonnées et leur micro-intégration sont présentés à la section IV. La section V est consacrée à la création de la Base de données statistiques sociales. Enfin, la section VI est réservée aux conclusions.

II. LE RECENSEMENT DE 2001

9. En 2003, les données ont été combinées en vue d'établir les tableaux du recensement néerlandais de 2001. Aux Pays-Bas, on a utilisé à cet effet des données dont Statistics Netherlands disposait déjà au lieu d'interroger les habitants en procédant à un dénombrement complet de la population. De la sorte, la «facture de recensement» du contribuable néerlandais a considérablement diminué. Le coût d'un recensement traditionnel serait d'environ 300 millions d'euros, alors qu'avec cette méthode il n'a été «que» de 3 millions. Ces estimations prennent en compte le coût de tous les travaux préparatoires (mise au point d'une nouvelle méthode et du logiciel correspondant, par exemple). Le coût des registres n'est pas inclus (ils existent déjà), mais celui de l'analyse des résultats l'est. Les registres ne sont pas mis à jour pour les recensements, mais ils le sont à d'autres fins. Réaliser des économies sur les dépenses induites par le recensement n'est possible que dans les pays disposant de suffisamment d'informations issues des registres.

10. Le recensement de 2001 porte sur 40 grands tableaux, dont 28 se rapportent aux Pays-Bas en général, 9 aux régions COROP (NUTS 3) et 3 aux municipalités (NUTS 5). Les 40 tableaux se décomposent comme suit: 8 ont trait au logement, 2 aux déplacements pendulaires et les 30 autres sont des tableaux démographiques se rapportant à l'activité professionnelle, au niveau d'instruction et à l'activité économique. En outre, des données démographiques et sur le logement et l'emploi sont recueillies au niveau des districts suburbains pour 10 grandes villes participant à l'Audit urbain II (Statistics Netherlands, 2003).

11. Le recensement virtuel aux Pays-Bas a débuté plus tard que dans les autres pays qui ont effectué un recensement traditionnel. Il aurait été absurde de lancer le projet de recensement 2001 tant qu'on ne disposait pas de toutes les sources voulues: certains registres n'ont été disponibles que relativement tard. Toutefois, les Pays-Bas ont construit les 40 tableaux de recensement plus rapidement que la plupart des autres pays qui ont participé à la série de recensements de 2000. En fait, ils ont été l'un des premiers à envoyer le jeu complet des 40 tableaux à Eurostat, qui était chargé de coordonner les contributions de tous les États membres de l'Union européenne, des pays en voie d'adhésion et des États membres de l'Association européenne de libre-échange (AELE). Ils ont bénéficié du fait qu'il n'était pas nécessaire de vérifier et rectifier les formulaires de recensement envoyés en retour. Il convient néanmoins de relever que seules des données d'échantillons sont disponibles pour certaines variables, ce qui signifie qu'il a été impossible d'atteindre le niveau de détail requis dans certains tableaux concernant les Pays-Bas.

12. Dans les pays nordiques (Danemark, Finlande, Islande, Norvège et Suède), les registres comportent actuellement plus de variables que ceux des Pays-Bas. Le problème dû à l'insuffisance du degré de détail dans les résultats n'y joue donc pas un rôle majeur. Toutefois, certains de ces pays ont procédé à un dénombrement (restreint) pour les variables absentes de

leurs registres. La plupart des autres pays sont dans la même situation que les Pays-Bas, où certaines variables qui interviennent dans le recensement se trouvent dans les registres, tandis que d'autres ne peuvent être obtenues que par sondage. C'est la raison pour laquelle, dans leur démarche, les Pays-Bas portent un grand intérêt à l'association des registres et des enquêtes ainsi qu'à l'utilisation de techniques statistiques modernes et de logiciels pour construire les tableaux. Il est évidemment d'une importance capitale que les bureaux de statistique soient en mesure d'utiliser les registres appropriés pour le recensement. Pour Statistics Netherlands, la nouvelle loi sur les statistiques adoptée en 2003 en a accru la possibilité. Néanmoins, dans les années à venir, il lui faudra veiller à entretenir de bons contacts avec les détenteurs des registres. La communication en temps voulu à Statistics Netherlands des variables qui lui sont utiles pour la production de statistiques est essentielle.

13. Plus de 50 pays de la région de la Commission économique des Nations Unies pour l'Europe ont pris part à la série de recensements de 2000. Un grand nombre d'entre eux ont choisi un jour de l'année 2001 comme jour de référence, même s'ils ont opté pour des jours très différents. Comme il faut beaucoup de temps avant que tous les pays n'achèvent les tableaux requis par les organisations internationales, les Pays-Bas ont pris l'initiative d'établir des comparaisons simples entre neuf pays européens qui avaient construit la série de tableaux pour Eurostat relativement vite et qui étaient disposés à s'associer aux analyses comparatives (voir le chapitre 12 de l'ouvrage de Schulte Nordholt *et al.*, 2004). Ils ont également comparé les résultats de leur recensement de 2001 avec ceux de leurs recensements précédents (Schulte Nordholt, 2005).

III. MÉTHODE DE COMPILATION DES DONNÉES DU RECENSEMENT DE 2001

14. Le recensement virtuel actuel se rapporte à 2001. Le Registre central de la population, qui regroupe tous les registres municipaux de la population, en est la pièce maîtresse. Il contient les informations démographiques concernant chacun des habitants des Pays-Bas (Prins, 2000).

15. Même si le Registre de la population a idéalement pour but de consigner toutes les personnes qui composent la population, il ne saurait être parfait: il arrive ainsi que des personnes déménageant dans une autre municipalité «oublient» de le signaler aux autorités. Les registres municipaux ne sont donc pas toujours à jour. Il arrive également que deux personnes enregistrées à des adresses distinctes vivent en fait ensemble. Il est de leur intérêt d'être déclarées à deux adresses différentes lorsque l'une d'elles bénéficie d'une aide sociale et que l'autre est salariée: la première risquerait de perdre son droit aux prestations si l'organisme de sécurité sociale venait à découvrir que les deux personnes vivent ensemble.

16. Un groupe important de population qui échappe au Registre de la population est celui des personnes qui vivent dans le pays à l'insu des autorités, le plus souvent dans l'illégalité. Ce groupe de population ne figure pas parmi les personnes recensées. Les résidents illégaux posent problème pour les services de statistique car, d'une part, ils participent à la vie économique et, en tant que tels, sont inclus dans les statistiques économiques mais, d'autre part, ne sont pas pris en compte dans les statistiques démographiques. Il est très peu probable qu'ils soient comptabilisés dans un recensement traditionnel, de toute façon. Statistics Netherlands a tenté d'évaluer la taille de cette population illégale, mais cette tâche s'est avérée particulièrement difficile du fait de la quasi-absence d'informations. L'estimation officielle du

nombre de résidents illégaux établie par Statistics Netherlands au 1^{er} janvier 2001 s'inscrit dans une marge relativement large: entre 46 000 et 116 000 personnes (Hoogteijling, 2002).

17. Un certain nombre d'enquêtes et de registres intégrés ont été mis en relation avec le Registre de la population. Pour ce faire, seules les correspondances exactes ont été prises en compte sur la base de ce que l'on appelle le numéro de sécurité sociale et de fiscalité (numéro SoFi). L'ensemble du système intégré s'appelle la Base de données statistiques sociales. Mise au point initialement en vue de réaliser des recensements virtuels, elle est désormais utilisée pour l'établissement de nombreuses statistiques sociales.

18. Les données du Registre de la population au 1^{er} janvier 2001 ont servi de base pour construire la série de tableaux. Différentes variables telles que l'activité professionnelle et le niveau d'instruction ont été extraites de l'Enquête sur les forces de travail. L'importance variable de chaque emploi a été obtenue à partir de la grande Étude sur l'emploi et les revenus. Pour obtenir un nombre suffisant d'enregistrements, on a combiné les informations sur les personnes ayant répondu à l'Enquête sur les forces de travail de 2000 avec celles obtenues dans le cadre de la même enquête en 2001. En ce qui concerne les tableaux sur le logement, nous avons utilisé les données du Registre de la population au 1^{er} janvier 2001, le Registre du logement de 2001 et l'Étude sur les conditions de logement de 2000.

19. Certaines variables du Registre de la population et des séries de données de la Base de données statistiques sociales – âge, sexe, situation matrimoniale et situation au regard de l'emploi, par exemple – sont disponibles dans leur intégralité. Les variables provenant des enquêtes ne sont disponibles que pour une partie de la population; il s'agit notamment du niveau d'instruction atteint (Enquête sur les forces de travail) ou du mode d'occupation du logement – locataire ou propriétaire – (Étude sur les conditions de logement).

20. Pour procéder à l'estimation la plus fiable possible pour un tableau, chaque estimation est fondée sur le plus grand nombre possible d'enregistrements. Les tableaux qui ne contiennent que des variables issues de registres sont comptés à partir des registres. Ceux qui contiennent au moins une variable issue d'une enquête sont estimés sur la base de la plus grande combinaison possible de registres et d'enquêtes.

21. Nous avons assuré la cohérence entre les tableaux en procédant à une pondération séquentielle qui produit un nouvel ensemble de facteurs de pondération pour chaque tableau faisant l'objet de l'estimation en appliquant de façon répétée l'estimateur par régression. Avec cette méthode, les poids des enregistrements dans les microdonnées sont adaptés de telle sorte que les nouvelles estimations de tableau sont conformes à toutes les estimations de tableau antérieures.

22. Les chiffres du recensement de 2001 se rapportent aux personnes vivant aux Pays-Bas au 1^{er} janvier 2001 (dénombrement de chaque individu en tant qu'unité). Les personnes qui, d'après le Registre de la population, vivaient aux Pays-Bas au début de cette journée ont été «comptées» dans le recensement virtuel. La plus grande partie de la population du pays se compose de ménages privés, le reste étant rattaché à un ménage institutionnel. Le nombre de salariés présenté dans les tableaux correspond à la fin de l'année 2000, le 22 décembre ayant été pris comme date de référence afin de déterminer le nombre d'emplois occupés par des salariés aux Pays-Bas. Il a été impossible d'utiliser une date de référence pour le nombre de salariés en 2001, les ensembles

de données de la Base de données statistiques sociales de 2001 n'étant pas disponibles en temps voulu pour le recensement de 2001. Les données de la Base de données statistiques sociales ont été extraites des informations contenues dans les registres sur les emplois occupés par les salariés. Si un salarié occupe plusieurs emplois à la fois, il peut figurer à plusieurs reprises dans le registre des salariés. Ce sont les caractéristiques de l'emploi principal qui sont alors utilisées dans l'ensemble des tableaux. L'emploi principal d'un salarié est par définition celui qui est associé au salaire brut le plus élevé pris en compte par les assurances sociales.

IV. ASSOCIATION DES SOURCES DE DONNÉES: COUPLAGE DES MICRODONNÉES ET MICRO-INTÉGRATION

A. Couplage des microdonnées

23. La plupart des registres administratifs actuels sont fournis avec une clef unique. Il s'agit du numéro de sécurité sociale et de fiscalité (numéro SoFi), identifiant unique attribué à chacun des habitants des Pays-Bas (inscrit dans le registre) et des Néerlandais vivant à l'étranger qui perçoivent un revenu aux Pays-Bas, pour lequel ils doivent s'acquitter d'un impôt auprès des autorités fiscales néerlandaises.

24. Pour empêcher toute utilisation abusive du numéro SoFi, Statistics Netherlands le retranscrit à des fins de traitement statistique en ce qu'il est convenu d'appeler le numéro d'identification personnel (personne-RIN). Les identifiants personnels tels que la date de naissance et l'adresse sont remplacés par l'âge à la date de référence et l'adresse-RIN, conformément aux règles édictées par l'Autorité néerlandaise de protection des données, afin de protéger la vie privée des citoyens.

25. Le numéro SoFi étant utilisé par les administrations chargées de la sécurité sociale et les autorités fiscales, on peut supposer qu'il s'agit d'un identifiant d'excellente qualité. Il se peut que, pour un nombre limité de ces numéros, des valeurs erronées ont été enregistrées dans les fichiers de données, auquel cas il ne sera pas possible de les coupler à d'autres fichiers; cela dit, d'une manière générale, le taux de concordance est proche de 100 %. Il peut y avoir, dans certains cas, utilisation abusive des numéros SoFi, par des travailleurs non déclarés par exemple, ce qui entraîne une erreur d'appariement. Parfois, on dispose d'indices de non-concordance: lorsque l'on rapproche le Registre des emplois et celui de la population, par exemple, et que l'employé s'avère être un enfant en bas âge, ou bien lorsque la base de l'administration fiscale (FiBase) montre un revenu exceptionnellement élevé pour un employé, alors qu'il s'agit en réalité de l'accumulation des revenus de toutes les personnes qui utilisent le même numéro SoFi.

26. Tous les fichiers de données statistiques sociales peuvent être mis en relation avec le Registre de la population. Concrètement, cela signifie que ces fichiers sont tous indirectement reliés les uns aux autres via le Registre de la population. Ce dernier peut donc être considéré comme la pièce maîtresse de l'ensemble des sources de données sociales. Lorsque l'on met le Registre de la population en relation avec le Registre des emplois, ou avec un registre des prestations sociales, on relie différentes unités statistiques (personnes, emplois, prestations). Il peut alors y avoir de multiples couplages, une personne pouvant avoir plusieurs emplois ou bénéficier de plusieurs formes d'aide sociale.

27. Dans les enquêtes par sondage auprès des ménages (Enquête sur les forces de travail, par exemple), on ne dispose pas du numéro SoFi. Dans ce cas, on utilise une variable (ou clef) différente, qui est souvent constituée à partir des identifiants personnels suivants:

- a) Sexe;
- b) Date de naissance;
- c) Adresse¹.

28. Cette sorte de clef permettra généralement d'établir une distinction entre les personnes. Toutefois, elle ne constitue pas une combinaison d'identifiants unique à 100 %. Elle peut aboutir à un défaut de concordance dans le cas de jumeaux de même sexe. Il peut également y avoir erreur de concordance lorsqu'une partie de la date de naissance ou le code postal et le numéro du logement sont inconnus ou erronés. Un autre inconvénient tient au fait que la clef n'est pas liée à la personne mais à l'adresse, ce qui pose un problème en cas de déménagement récent. Lorsque l'on utilise cette clef pour coupler le Registre de la population et l'Enquête sur les forces de travail et que l'on tolère un écart entre les sources pour l'une des variables – sexe, année de naissance, mois de naissance ou jour de naissance – tout au plus, on obtient un couplage des enregistrements de l'Enquête sur les forces de travail proche de 100 %.

29. Avec la stratégie de couplage ainsi adoptée, Statistics Netherlands tente d'obtenir le plus grand nombre possible de correspondances et de réduire autant que faire se peut le nombre de défauts de concordance. Dès lors, pour obtenir un taux de concordance plus élevé, on s'efforce encore de relier les enregistrements restants non liés au moyen de différentes variantes de la clef, en omettant par exemple le numéro du logement et en tolérant des écarts dans les caractères numériques du code postal. Pour maintenir le risque d'une non-concordance à son niveau le plus bas possible, certains dispositifs de sûreté sont intégrés dans le processus. Cette dernière tentative permet d'obtenir 1 % de concordance en plus.

30. Finalement, environ 2 à 3 % des enregistrements de l'Enquête sur les forces de travail n'ont pu être reliés au Registre de la population, ce qui est dans l'ensemble un résultat satisfaisant, mais l'on ne peut écarter la sélectivité dans le couplage des microdonnées. Si les enregistrements qui n'ont pu être reliés correspondent à une catégorie donnée de la population, les estimations fondées sur les enregistrements qui l'ont été risquent donc d'être faussées puisqu'elles ne correspondent pas à l'ensemble de la population. Les études réalisées par le passé ont montré que, dans les enquêtes par sondage sur les ménages, le taux de correspondance était moins élevé dans le cas des jeunes de 15 à 24 ans que dans celui des autres tranches d'âge. La raison en est qu'ils déménagent plus fréquemment, et sont donc souvent enregistrés à la mauvaise adresse. Le taux est moins élevé pour les personnes vivant dans les quatre grandes villes du pays – Amsterdam, Rotterdam, La Haye et Utrecht – que pour le reste de la population. Il l'est également dans le cas des minorités ethniques, en raison

¹ En fait, la combinaison d'un code postal (déterminé le plus souvent en fonction de la rue) et du numéro du logement vient remplacer l'adresse. Le code postal aux Pays-Bas se compose de quatre chiffres, suivis de deux lettres.

notamment du fait que les dates de naissance sont souvent moins bien enregistrées (Arts *et al.*, 2000).

31. De nos jours, le Registre de la population sert de base d'échantillonnage pour l'Enquête sur les forces de travail. Le taux de concordance est donc proche de 100 %, et l'on ne rencontre plus aucun problème de sélectivité en matière de couplage.

B. Micro-intégration

32. Parvenir à coupler le Registre de la population avec toutes les autres sources de données mentionnées permet d'obtenir des informations bien plus cohérentes sur les différentes composantes démographiques et socioéconomiques de la vie de chaque individu. Il faut cependant garder à l'esprit que certaines sources sont plus fiables que d'autres. Certaines présentent une meilleure couverture que d'autres, et il se peut même que les informations se contredisent selon la source. Il est donc important de bien cerner les points forts et les points faibles de toutes les sources de données utilisées.

33. Comme il existe des différences entre les sources, il faut procéder à une micro-intégration des données afin de les vérifier et de corriger celles qui sont inexactes. On pense que des données intégrées fourniront des résultats bien plus fiables, puisqu'elles reposent sur une quantité optimale d'informations. La couverture des (sous-) populations s'en trouvera elle aussi améliorée puisqu'il est possible, si des données manquent dans une source, d'utiliser une autre source. L'intégration présente un autre avantage: les utilisateurs des données statistiques obtiennent un chiffre pour chaque phénomène social, et non plus une multitude troublante de chiffres différents selon la source utilisée.

34. Le processus de micro-intégration des sources de données comporte les étapes suivantes (Van der Laan, 2000):

- a) Harmonisation des unités statistiques;
- b) Harmonisation des périodes de référence;
- c) Vérification de la population couverte (couverture);
- d) Harmonisation des variables, en cas de divergence dans les définitions;
- e) Harmonisation des classifications;
- f) Rectification des erreurs de mesure, lorsque des variables correspondantes ne présentent toujours pas la même valeur après harmonisation des définitions;
- g) Imputation, en cas de non-réponse à une question;
- h) Calcul de (nouvelles) variables; création de variables à partir de sources de données différentes;
- i) Vérification de la cohérence d'ensemble.

Toutes ces étapes sont soumises à un ensemble de règles d'intégration et sont entièrement automatisées.

35. La façon dont fonctionne la micro-intégration est illustrée ci-après, avec un cas où les données provenant du Registre des emplois sont confrontées à celles du Registre des prestations sociales. Tant les emplois que les prestations sont enregistrés quantitativement, de telle sorte que les informations sur leur état sont stockées en permanence tout au long de l'année et non à une date de référence donnée. Les analystes du Registre des emplois savent que les dates de début et de cessation d'emploi ne sont pas consignées avec grande précision. Il est toutefois important de savoir si, à la date de référence, il y a ou non emploi, autrement dit si la personne est ou non salariée. À l'aide du registre des prestations sociales, il est parfois possible de déterminer la période d'emploi avec plus de précision.

36. Supposons qu'une personne perde son emploi fin novembre et commence à percevoir une allocation chômage début décembre. Le Registre des emplois indiquera peut-être qu'elle a perdu son emploi à la fin de l'année, soit en raison de lenteurs administratives soit parce qu'elle a reçu des versements après la cessation de son emploi. L'enregistrement des prestations est réputé plus fiable. Lorsqu'il rapproche ces faits, l'agent chargé de l'intégration pourrait décider de rectifier la date de cessation d'emploi pour fin novembre parce qu'il est peu probable que la personne en question travaille en décembre et, parallèlement, touche une allocation chômage. De telles décisions sont prises avec la plus grande prudence. Dès que des indications convaincantes provenant d'autres variables du registre des emplois font apparaître que l'emploi était toujours occupé en décembre, la date de cessation d'emploi ne sera en principe pas rectifiée.

V. LA BASE DE DONNÉES STATISTIQUES SOCIALES

37. Le couplage des microdonnées et la micro-intégration de toutes les sources de données disponibles aboutissent à la constitution de la Base de données statistiques sociales, ensemble complet de fichiers de microdonnées intégrées à leur stade définitif. Cette base renferme des informations statistiques démographiques et socioéconomiques cohérentes et détaillées sur les personnes, les ménages, les emplois et les prestations (sociales). La plus grande partie des informations statistiques est disponible en volume. On trouvera dans Arts et Hoogteijling (2002) une présentation détaillée de la Base de données statistiques sociales.

38. Il ne faut pas se représenter cette Base de données comme un gigantesque fichier contenant des millions d'enregistrements et des milliers de variables: il serait totalement inefficace de stocker les données intégrées telles quelles. De plus, aux fins de la protection des données, Statistics Netherlands ne conserve pas autant d'informations ensemble. Tous les fichiers intégrés sont donc, à leur stade final, conservés séparément. Un seul élément les relie: la clef personne-RIN, présente dans chacun des fichiers intégrés. Ainsi, dès que des utilisateurs souhaitent obtenir une série de variables à partir de la Base de données statistiques sociales, seuls les fichiers présentant les variables demandées seront fournis. Il est aisé de les extraire de la base et de les relier au moyen de la clef.

VI. CONCLUSIONS

39. Statistics Netherlands a adopté de nouvelles méthodes de collecte et de traitement des données en vue de la compilation du Programme de tableaux de recensement de 2001. Les éléments les plus importants en sont l'utilisation d'une combinaison de registres administratifs et d'enquêtes par sondage comme source de données, et l'application d'une pondération séquentielle, nouvelle méthode permettant d'établir des estimations de tableaux cohérents sur le plan numérique à partir de cette source de données. On obtient un recensement virtuel, puisque les résultats concernant certaines caractéristiques de la population sont obtenus à partir d'estimations et non d'un dénombrement.

40. La nouvelle méthode utilisée pour produire des tableaux de recensement s'est avérée une bonne solution, bien moins onéreuse que les opérations de recensement coûteuses réalisées par le passé. Aucun travail particulier n'avait été nécessaire pour réunir les données, les sources de données utilisées pour le recensement de 2001 faisant déjà partie du programme de statistique ordinaire de Statistics Netherlands. La plupart des données provenaient des registres, et seules quelques informations complémentaires ont dû être obtenues au moyen d'enquêtes par sondage. La réalisation du Programme de recensement de 2001 n'a donc entraîné aucune charge de travail supplémentaire. En outre, le temps de traitement des données pour ce recensement n'a représenté qu'une fraction de celui qui aurait été nécessaire pour un recensement traditionnel.

41. La nouvelle démarche adoptée présente un inconvénient: la taille de l'échantillon utilisé pour les enquêtes ne suffit pas toujours à garantir la fiabilité des estimations pour de petites sous-populations telles que les municipalités. Ce problème ne se poserait pas dans le cas d'un recensement traditionnel.

42. Lorsque l'on compare la méthode actuelle de compilation des tableaux de recensement avec celles utilisées pour les recensements de 1981 et 1991, on constate que le temps de production a été plus long pour le Programme de recensement de 2001, mais que les estimations dans les tableaux croisés avec des informations provenant de registres et d'enquêtes sont plus fiables. La première raison en est que les informations statistiques ont beaucoup gagné en cohérence grâce à la combinaison des sources de données; la deuxième est qu'il a été possible d'utiliser dans les méthodes d'estimation plus de données auxiliaires qu'auparavant, les registres disponibles étant plus nombreux; enfin, un gros travail a été réalisé en vue de parvenir à une cohérence numérique d'ensemble.

Bibliographie

- Arts, C.H. and E.M.J. Hoogteijling. 2002. 'The Social Statistical Database of 1998 and 1999'. *Monthly Bulletin of Socio-economic Statistics*. Vol. 2002/12 (December 2002), pp. 13-21, 2002. [in Dutch]
- Corbey, P.I., 1994. 'Exit the population census'. *Netherlands Official Statistics*, Vol. 9 (Summer 1994): pp. 41-44, 1994.
- Hoogteijling, E.M.J. 'Illegal people in the Netherlands'. *Monthly Bulletin of Population Statistics*. Vol. 2002/03 (March 2002), page 21. [in Dutch]
- Laan, P. van der, 2000. 'Integrating Administrative Registers and Household Surveys'. *Netherlands Official Statistics*, Vol. 15 (Summer 2000): Special Issue, *Integrating Administrative Registers and Household Surveys*, ed. P.G. Al and B.F.M. Bakker, pp. 7-15.
- Prins, C.J.M., 2000. 'Dutch population statistics based on population register data'. *Monthly Bulletin of Population Statistics*. Vol. 2000/02 (February 2000), pp. 9-15.
- Schulte Nordholt, E., M. Hartgers and R. Gircour (Eds.), 2004. 'The Dutch Virtual Census of 2001, Analysis and Methodology'. Statistics Netherlands, Voorburg / Heerlen, July, 2004. <http://www.cbs.nl/NR/rdonlyres/D1716A60-0D13-4281-BED6-3607514888AD/0/b572001.pdf>.
- Schulte Nordholt, E., 2005. 'The Dutch virtual Census 2001: A new approach by combining different sources'. *Statistical Journal of the United Nations Economic Commission for Europe*, Volume 22, Number 1, 2005, pp. 25-37.
- Statistics Netherlands, 2003. 'Urban Audit II, the implementation in the Netherlands'. Report, BPA no. 2192-03-SAV/II, Statistics Netherlands, Voorburg. <http://www.cbs.nl/NR/rdonlyres/8C6E4C9D-4338-4E32-848B-8D43B9B3242D/0/urbanauditIINetherlands.pdf>.
- Vliegen, J.M. and H. van de Stadt, 1988. 'Is a Census still necessary? Experiences and alternatives'. *Netherlands Official Statistics*, Volume 3, No. 3, pp. 27-34.
