



Conseil économique et social

Distr. générale
14 juillet 2015
Français
Original : anglais

Commission économique pour l'Europe

Conférence des statisticiens européens

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

Dix-septième session

Genève, 30 septembre-2 octobre 2015

Point 2 b) de l'ordre du jour provisoire

Innovations projetées pour le cycle de recensements de 2020 et résultats des essais

Utilisation conjuguée de multiples sources de données dans le recensement de la population

Note de l'Institut national de statistique de l'Italie

Résumé

La nouvelle stratégie appliquée au recensement italien de la population et des habitations suppose une utilisation généralisée des sources administratives et statistiques disponibles, ainsi qu'un recours limité aux enquêtes par sondage, conformément à l'objectif du plan de modernisation adopté par l'Institut national de statistique de l'Italie (Istat) pour passer à un système intégré de registres. Un processus parallèle plus léger devrait permettre d'améliorer les méthodes et d'acquérir de l'expérience. Des incidences financières positives devraient également être relevées et la demande de financement public diminuerait et s'étalerait dans le temps. De même, la production de données en continu rendrait bien plus efficaces et faciles les relations avec les utilisateurs.

Le présent document traite de certaines des prescriptions les plus importantes du point de vue méthodologique, technologique et organisationnel à remplir pour obtenir les résultats susmentionnés. L'Istat propose de renforcer la coopération entre les pays en ce qui concerne l'exploitation de multiple sources, notamment pour ce qui est des points suivants : i) modèles faisant un usage intégré de ces sources; ii) vérification de la couverture des registres de population pour déterminer si elle est suffisante pour produire des décomptes statistiques; iii) regroupement des données provenant des enquêtes disponibles; iv) mesure de la qualité des données obtenues par intégration des différentes sources.



I. Introduction

1. Pour le prochain cycle de recensements, c'est devenu une nécessité plus qu'un choix que d'associer les sources de données disponibles, y compris les enquêtes, les recensements et les registres administratifs. Il est trop complexe de procéder tous les dix ans à un recensement de la population entièrement fondé sur le dénombrement sur le terrain, une telle opération n'étant plus viable, et les parties prenantes comme les politiciens, les administrateurs et les chercheurs demandent des données démographiques, sociales et économiques périodiquement actualisées et harmonisées, qui doivent aussi être géoréférencées au niveau de détail géographique le plus élevé. Il faut évoluer vers une utilisation plus efficace des sources de données administratives en raison du coût élevé des recensements traditionnels et de la charge de travail qui les accompagne.

2. Si le dernier cycle de recensements se caractérisait par la multimodalité en matière de collecte de données, la caractéristique du prochain cycle sera très probablement le recours à de nombreuses sources administratives et statistiques et, dans ce domaine, il est indispensable de mettre en place une coopération efficace entre les pays.

3. Les individus laissent en permanence des traces dans les registres administratifs, qui renferment potentiellement un énorme volume de données géoréférencées exploitables à des fins statistiques. Ces données peuvent cependant ne pas être à jour et leur couverture peut être erronée en raison de l'absence d'harmonisation entre les diverses classifications et définitions, ce qui pourrait compromettre l'exploitabilité.

4. Dans la version révisée des Principes et recommandations de l'ONU, les autres méthodes de recensement sont représentées dans une matrice où les rangées décrivent la collecte de données par un dénombrement sur le terrain et les colonnes l'utilisation de registres administratifs et/ou de registres de population comme sources de données. Les colonnes contiennent de nombreux types de registres : registres administratifs, créés et utilisés principalement à des fins administratives hors du cadre des autorités statistiques nationales, registres statistiques et registres de base.

5. Dans le nouveau recensement italien, qui est un recensement en continu (*censimento permanente*), le défi consiste à opter pour une utilisation plus intensive des sources administratives et statistiques conjuguée à un recours additionnel mais limité à des enquêtes par sondage ponctuelles afin de produire des résultats chaque année. Le but visé est d'évoluer vers un nouveau type de recensement qui applique des méthodes statistiques et les nouvelles technologies pour produire chaque année des résultats dans une série d'opérations d'intégration bien conçues à partir des sources administratives et statistiques disponibles. Nous étudions à fond des méthodes qui permettront de généraliser l'utilisation d'une seule source, ainsi que les concepts à la base de l'exploitation simultanée de nombreuses sources.

II. Innovations relatives à la collecte de données multimodale dans le recensement de 2011

6. Le recensement de 2011 a été appréhendé d'une façon totalement différente par rapport aux recensements précédents. Les registres de population administratifs (« Anagrafi ») ont été utilisés pour envoyer par la poste des questionnaires aux familles. Les répondants avaient la possibilité de choisir entre plusieurs solutions pour remplir et retourner le questionnaire, selon une approche permettant plusieurs modes de renvoi :

- En ligne, à l'aide du mot de passe fourni avec le questionnaire;

- À n'importe quel bureau de poste en Italie;
- Après d'un des centres municipaux de recensement, avec la possibilité de bénéficier de l'assistance de spécialistes;
- Directement auprès d'un des agents recenseurs municipaux.

7. Des techniques d'échantillonnage ont été utilisées pour mesurer certaines variables intéressantes. Dans le cadre de cette démarche, on a eu recours à deux versions du questionnaire, une version courte pour deux tiers des familles des grandes municipalités et une version complète pour les autres familles.

8. Un système en ligne a été mis au point pour gérer les réponses multimodales et aider les enquêteurs à récupérer les non-réponses. Conçu de façon à permettre une automatisation de tout le travail administratif, il a permis de mettre à jour pratiquement en temps réel l'état de chaque questionnaire. Il a également permis de produire des rapports intermédiaires sur l'état d'avancement du recensement, d'envoyer la localisation des zones aux agents recenseurs et de surveiller leurs travaux, de récupérer de façon ciblée les non-réponses et les individus non enregistrés, de comparer les données de recensement et les registres de population locaux, et de produire des comptes relatifs concernant les personnes trouvées et non trouvées. La disponibilité de données actualisées en permanence concernant l'état de chaque questionnaire a permis aux agents recenseurs d'être orientés uniquement vers les ménages auxquels le questionnaire avait été envoyé, mais qui ne l'avaient pas encore retourné, et ainsi de récupérer systématiquement les non-réponses.

9. Dans les grandes municipalités a été utilisée une liste supplémentaire basée sur les relevés des numéros de logements antérieurs au recensement et renfermant des renseignements concernant les logements potentiellement inhabités pour lesquels il n'existe aucune famille correspondante dans les registres municipaux. L'utilisation d'autres listes auxiliaires provenant de sources administratives centrales et locales, qui renferment des renseignements concernant la présence d'individus non enregistrés dans les registres municipaux, a permis de récupérer ces individus de façon ciblée et systématique.

10. Pour signaler les personnes non inscrites dans les registres municipaux et rendre les informations spatiales disponibles au niveau unitaire, qu'il s'agisse de personnes physiques ou morales, des sources administratives ou statistiques ont été incluses dans un système intégré de microdonnées (SIM). Des données de nombreuses sources telles que les registres des entreprises, des impôts, de l'éducation, de l'emploi et autres registres pertinents ont été corrélées au moyen de codes uniques individuels. L'une des principales sources utilisées au cours de ce processus était la liste des permis de séjour accordés à des non-ressortissants des pays communautaires. Les thèmes ci-après figurent dans le système intégré de microdonnées : caractéristiques des ménages, lieu de résidence habituel, lieu de travail, lieu d'étude (école, collège ou université), situation dans la profession, caractéristiques de l'instruction, type de logement et de locaux d'habitation, etc. Pour gérer un éventuel sous-dénombrement, des données provenant de sources telles que l'administration des impôts ou les archives des permis de séjour des étrangers ont été utilisées pour dresser une liste de personnes ne figurant pas dans les registres de population municipaux mais susceptibles de résider dans chaque municipalité.

11. Les importantes innovations apportées au recensement de 2011 n'ont pas encore permis d'établir un équilibre stable et durable entre les coûts et les avantages d'un recensement. De fait, les coûts restent élevés et trop concentrés dans le temps alors que le recours à des données administratives en l'état n'est pas encore adéquat pour satisfaire à la demande de données de type censitaire. De plus, les données des recensements deviennent vite obsolètes, et il n'y a une offre de données géographiques

très détaillées que tous les dix ans, d'où la nécessité d'élaborer une approche différente.

III. Au-delà de 2011

12. Faisant le bilan du précédent cycle de recensement de la population et des habitations, nous projetons d'appliquer de nouvelles méthodes pour produire des données censitaires. Il est de la plus haute importance d'appliquer le principe essentiel qui consiste à fournir des statistiques détaillées au niveau géographique le plus bas. Il est envisagé d'utiliser des registres – principalement des registres de population – conjointement avec d'autres sources pour produire des statistiques aréolaires détaillées sur la population et les habitations, et d'appliquer aux mêmes fins la méthode des enquêtes en continu.

13. Les techniques, méthodes et solutions organisationnelles mises en œuvre pour le cycle de 2011 sont réexaminées et associées à de nouvelles autres dans un cadre innovant qui les rend compatibles avec des objectifs stratégiques plus avancés.

14. Il faut étudier cette nouvelle approche pour de nombreuses raisons, à savoir: i) la nécessité de produire des statistiques plus fréquentes et à jour; ii) le budget restreint alloué au recensement; iii) le manque d'empressement de la population à participer au recensement; iv) le renforcement des capacités techniques nécessaires pour manipuler les sources de données.

15. Les renseignements concernant les individus et les ménages sont recueillis auprès des sources administratives ou statistiques disponibles, à savoir les différents types de registres, dont les suivants sont de première importance : registres des personnes, des ménages et des logements ainsi que données d'enquête existantes corrélées au niveau individuel. Des enquêtes par sondage ponctuelles fournissent des informations sur les thèmes de recensement non disponibles auprès des sources administratives ou pour ajuster les données de qualité médiocre issues des registres.

16. Des méthodes statistiques seront utilisées pour atteindre les deux principaux objectifs du recensement de la population, à savoir :

- Dénombrement des résidents habituels et production de données essentielles sur la structure démographique de la population et des ménages (objectif C);
- Production de données socioéconomiques issues du recensement (objectif D).

17. S'agissant de l'objectif C, c'est-à-dire les estimations annuelles de la population ventilée par structure démographique (sexe, âge, situation matrimoniale, citoyenneté), les questions les plus graves auxquelles il faudra faire face au cours des prochaines années sera le dénombrement de la population vieillissante et des migrants.

18. Concernant l'objectif D, pour satisfaire à l'obligation nationale et internationale d'estimer les caractéristiques socioéconomiques des ménages et des individus, le principal objectif est de collecter dans le cadre des enquêtes uniquement les variables qui doivent l'être conformément aux règlements de l'UE et qu'il n'est pas possible d'estimer en associant les informations issues des registres de population, des sources administratives et des enquêtes statistiques. Une étude de substituabilité visant utiliser les données socioéconomiques provenant des sources administratives est en cours. Le but est de remplacer les données traditionnellement collectée par voie de recensement par des estimations obtenues d'informations disponibles : recours à des sources administratives et regroupement d'enquêtes par sondage déjà existantes. Il faut pour cela renforcer les études actuelles ayant trait aux sources administratives locales et centrales. Nous répondrions aux besoins des utilisateurs en fournissant des mises à jour plus fréquentes, évitant ainsi une dégradation de l'exactitude au cours de la

décennie. En particulier, les individus et ménages retireront un avantage du fait que les bureaux de statistique utiliseront des informations déjà disponibles qui devraient normalement leur être de nouveau demandées lors d'enquêtes onéreuses et redondantes. Les données administratives et les données massives peuvent aussi aider à multiplier les données aréolaires et permettre la construction de nouveaux indicateurs aux fins de l'élaboration des politiques. Enfin, une utilisation accrue des données administratives renforcera les capacités des organismes publics à forger des éléments d'appréciation leur permettant d'évaluer l'efficacité de leurs programmes et politiques. Des mesures seront prises pour déterminer les zones qu'il sera plus constructif de couvrir avec de nouvelles informations.

IV. Projets lancés par l'Istat

19. Dans le plan de modernisation de l'Istat, le défi que devra relever le système statistique italien sera d'adopter un système intégré, qui mettra l'accent sur les nouvelles méthodes statistiques et les nouvelles technologies qui renforceront l'efficacité d'utilisation des sources de données administratives et statistiques pour réduire au minimum les coûts ainsi que la charge de travail imposée aux répondants. L'élément central de ce système intégré qui englobera les domaines économique, environnemental, démographique et social sera constitué de registres et cadres de base, intégrés et harmonisés, de la population et des entreprises. Un système coordonné d'échantillons d'unités statistiques pour les différentes enquêtes sera mis au point ainsi que des questionnaires personnalisés visant des groupes de population. L'unique code d'identification utilisé pour intégrer les registres dans le système intégré de microdonnées (SIM) permettra aussi, au sein de l'organisme de statistique national, d'associer plus facilement des données issues de différentes sources, de procéder à une comparaison dans le temps et de stimuler les usages multiples des données. Des méthodes statistiques seront appliquées pour corriger les données. Celles-ci ne devraient être collectées auprès des individus, des ménages ou des entreprises que s'il est impossible de les obtenir dans les registres de base et auprès des sources administratives et statistiques disponibles.

20. Un rôle primordial est attribué à l'ANNCSU (acronyme italien du registre national des rues et des adresses) et à son système de géoréférencement des rues et des numéros de logement. Cette base de données géographique est de fait essentielle pour localiser des unités dans le système d'informations, qui est un outil capital pour améliorer le géocodage des données issues de sources administratives. Chaque numéro de logement sera lié par géocodage aux zones de dénombrement créées par l'Istat.

21. Les piliers du programme numérique italien sont énoncés dans une loi datant de la fin de 2012. Outre le *censimento permanente* (recensement en continu) et l'ANNCSU, les archives nationales de la population résidente (ANPR) seront l'unique registre de population national dans lesquels tous les registres municipaux seront transférés. Il est primordial d'intégrer ces trois piliers pour produire des données aréolaires à partir du registre de population.

22. Pour que les sources administratives servent à des fins statistiques, il est essentiel de faire en sorte qu'elles répondent aux critères de qualité en appliquant une stratégie de surveillance et de correction permanentes des données. L'Istat a mis en place le projet ARCOLAIO, dont le but est de contrôler la qualité de l'utilisation statistique des sources administratives. Nous étudions également à fond les méthodes qui peuvent être utilisées pour traiter ces aspects.

23. Selon le règlement européen concernant les recensements de la population et du logement (Parlement européen, 2008), la population visée par un recensement est l'ensemble des résidents habituels définis comme étant ceux qui habitent sur leur lieu

de résidence habituelle depuis une période de plus de douze mois ou qui ont l'intention de le faire. Comme les intentions ne sont pas enregistrées dans les sources administratives et qu'il est très difficile d'identifier les résidents qui ont vécu au même endroit pendant plus d'une année, cette définition ne peut pas être aisément appliquée dans la nouvelle approche. De plus, la population résidente habituelle « réelle » peut différer de la population enregistrée à cause des migrations non documentées et des faux enregistrements effectués pour des raisons de commodité fiscale.

24. Le système ARCHIMEDE (archive de microdonnées économiques, démographiques et sociales) est une infrastructure innovante qui produira des microdonnées provenant de sources administratives et d'enquêtes statistiques après vérification de leur qualité et traitement. Les données, liées par géocodage aux zones de dénombrement, seront diffusées dans le respect de la confidentialité et de manière efficace et transparente. Les résultats du système ARCHIMEDE seront définis en tenant rigoureusement compte des besoins des utilisateurs pour fournir aux administrations publiques, aux chercheurs et aux utilisateurs des ensembles de données à usage général se rapportant à des unités statistiques et géographiques de base (descendant jusqu'aux zones de dénombrement).

25. Du point de vue méthodologique, nous étudions les capacités de prévision des modèles dont la résidence habituelle est la variable-réaction, les données provenant de multiples sources administratives et statistiques étant des variables indépendantes. Nous étudions, entre autres, les modèles de structure latente (Biemer Paul P., 2011, p. 258) et plusieurs généralisations de méthodes de double saisie pour l'utilisation de listes multiples. Les capacités de prévision de ces modèles peuvent être considérablement accrues par l'adjonction de nouvelles sources au système de microdonnées intégrées telles que les listes des contrats des compagnies de gaz et d'électricité, ainsi que les listes d'usagers de services locaux.

26. Dans le cadre du recensement en continu, une enquête pilote a débuté en mars 2015 avec la participation de 150 municipalités et de 160 000 ménages. Outre les capacités de prévision des modèles statistiques susmentionnés, son but est de tester l'organisation des enquêtes selon certaines variantes, l'accent étant mis sur les nouvelles technologies de saisie de données et les architectures TI, notamment en ce qui concerne :

- L'architecture TI et l'élaboration d'un système de gestion intégré pour la collecte de données multimodale, qui permet de recourir à de nombreux outils différents (tablettes, mobiles, PC, etc.) avec ou sans connexion au réseau et indépendamment du matériel et du système d'exploitation utilisés;
- La suppression des questionnaires sur papier dans la collecte de données sur le terrain pour mettre en place différentes solutions organisationnelles, en recourant à des systèmes locaux de renvoi des questionnaires et sur la base de différences dans l'utilisation des TIC;
- La vérification des noms de rues et des numéros de logement dans le registre ANNCSU.

V. Conclusion

27. Jusqu'à présent, y compris 2011, les recensements ont lieu en Italie tous les dix ans. La nouvelle stratégie en matière de recensement de la population et des habitations généralisera l'exploitation des sources administratives et statistiques disponibles tout en faisant un usage limité des enquêtes par sondage.

28. La continuité devrait apporter des gains d'efficacité importants dans le travail sur le terrain, ainsi qu'une amélioration de la qualité. Un travail local constant devrait permettre de préserver et développer le savoir-faire au fil du temps; un processus parallèle plus léger devrait permettre d'améliorer les méthodes et d'acquérir de l'expérience.

29. Des incidences financières positives devraient également être relevées : la demande de financement public diminuerait et s'étalerait dans le temps, et les prestations seraient plus intéressantes, voire meilleur marché, que dans le cas d'un recensement ponctuel. De même, la production de données en continu rendrait bien plus efficaces et faciles les relations avec les utilisateurs.

30. Le présent document traite de certaines des prescriptions méthodologiques, technologiques et organisationnelles les plus importantes à remplir pour obtenir les résultats susmentionnés.

31. Nous proposons de renforcer la coopération entre les pays en ce qui concerne l'exploitation de multiples sources, notamment pour ce qui est des points suivants :

- Modèles faisant un usage intégré de ces sources;
- Vérification de la couverture des registres de population pour déterminer si elle est suffisante pour produire des décomptes statistiques;
- Regroupement des données provenant des enquêtes disponibles;
- Mesure de la qualité des données obtenues par intégration des différentes sources.

Références

- Baffour, B., J.J. Brown, P.W.F Smith, 2013, An investigation of triple system estimators in censuses. *Statistical Journal of the International Association for Official Statistics*, 29, 53-68
- Bakker, B.F.M., Rooijen, J. van & Toor, L. van (2014). The system of social statistical datasets of Statistics Netherlands : an integral approach to the production of register-based social statistics. *Statistical journal of the United Nations ECE*, 30(4), 411-424
- Biemer Paul P., 2011 *Latent class analysis of survey errors*, Wiley Series in Survey Methodologies
- CEE (2007), *Register-based statistics in the Nordic countries : review of best practices with focus on population and social statistics*, publication des Nations Unies, ISBN 978-92-1-116963-8
- Parlement européen, 2008, Règlement (CE) n° 763/2008 du Parlement européen et du Conseil du 9 juillet 2008 concernant les recensements de la population et du logement. Dans: *Journal officiel de l'Union européenne*, 13.8.2008, p. L 218/14-L 218/20
- 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* 22, 25-37
- Wallgren, A. et Wallgren, B. (2007), *Register-based statistics – administrative data for statistical purposes*, John Wiley and Sons, Chichester
- Zhang, L-C. (2012), *Topics of statistical theory for register-based statistics and data integration*, *Statistica Neerlandica*