



Commission économique pour l'Europe

Conférence des statisticiens européens

**Groupe d'experts des recensements
de la population et des habitations****Vingt et unième réunion**

Genève, 18-20 septembre 2019

Point 4 de l'ordre du jour provisoire

**Travaux de recherche sur l'utilisation des données
administratives pour les recensements****Approches novatrices prévues pour le recensement de 2021 et
au-delà au Canada****Note établie par Statistique Canada****Résumé*

Au Canada, la prochaine édition du recensement quinquennal est prévue pour mai 2021. Le pays utilise encore une approche traditionnelle, principalement fondée sur une méthode de collecte par vagues qui suscite un fort taux de réponse sur Internet. Bien que cette approche ait récemment produit des succès, Statistique Canada étudie la mise en place d'un modèle de recensement mixte, qui s'appuierait si possible d'abord sur des informations provenant de registres statistiques et ne ferait appel à des méthodes de collecte traditionnelles que lorsque les sources administratives disponibles (pour le questionnaire complet) ou la qualité des registres s'avèrent insuffisantes. Le présent document décrit les travaux de recherche menés sur l'élaboration de registres statistiques à l'appui de la mise en place d'une approche mixte, éventuellement pour le cycle de 2026. On y trouve aussi décrites brièvement les méthodes de collecte prévues en 2021, et plus particulièrement la manière dont les travaux sur les registres statistiques pourront être mis à profit, de même que des informations provenant d'autres sources, à l'appui d'opérations de suivi sur le terrain en 2021.

* Document établi par Sylvain Delisle, Marc Hamel et Sophie Lefebvre.



I. Introduction

1. En application de la loi sur la statistique, Statistique Canada doit effectuer tous les cinq ans le recensement de la population, dont la prochaine édition aura lieu en 2021. Le programme de recensement canadien comprend un questionnaire abrégé, avec une dizaine de questions sur les renseignements démographiques de base, et un questionnaire complet qui sert à collecter des informations socioéconomiques plus détaillées sur un échantillon de 25 % des ménages. Le programme s'appuie sur une approche mixte, la majeure partie des données sur les individus et sur les unités de logement à un moment précis étant recueillies auprès des ménages et certaines informations étant obtenues à partir de dossiers administratifs. Par exemple, les données sur les variables relatives aux revenus sont obtenues à l'aide des registres de l'impôt sur le revenu et les informations sur le statut vis-à-vis de l'immigration sont obtenues à partir de la base de données longitudinale gérée par Immigration, Réfugiés et Citoyenneté Canada.

2. Bien que le recensement de 2016 ait été un succès, avec des taux record de réponse sur Internet (68.4 %) et d'autodéclaration (88.8 %), on s'attend à ce que le recensement canadien rencontre certains défis à l'avenir. Premièrement, comme c'est généralement le cas pour toute méthode, l'approche actuelle ne peut sans doute pas être maintenue à long terme, car la société continue d'évoluer. Le succès de la démarche actuelle est fortement tributaire de la méthode de collecte par vagues, qui exige de pouvoir joindre les ménages par la poste¹. En effet, bien qu'un questionnaire électronique serve de principal mode de réponse au recensement, son utilisation exige de faire appel au système postal pour fournir un code d'accès sécurisé à tous les ménages. Or ce mode de communication risque d'être progressivement abandonné par les Canadiens, à mesure que les communications numériques deviennent de plus en plus la norme. Les solutions de remplacement qui ont été mises à l'essai jusqu'à présent pour aviser les répondants, telles que le courrier électronique et des boîtes de messagerie numériques, ne se sont pas avérées concluantes. Deuxièmement, les Canadiens ont exprimé des préoccupations au sujet de la charge de travail pour les répondants imposée par les pouvoirs publics de différents niveaux². Il y a une certaine attente que le programme de recensement s'efforce de tirer parti des données déjà recueillies par des organismes publics, ou même des acteurs privés, avant de procéder à toute collecte directe. Enfin, on ne saurait ignorer les tendances internationales et les travaux de recherche qui accordent une importance croissante aux méthodes de recensement fondées en totalité ou en partie sur des données administratives.

3. Pour faire face à ces défis, Statistique Canada a lancé un nouveau programme de recherche appelé « Projet de transformation du Programme de recensement » (PTPR). Ce projet vise à rendre le recensement de la population plus convivial et davantage centré sur les répondants, afin de mieux répondre aux nouveaux besoins en matière d'information, de réduire la charge de travail pour les Canadiens et de limiter les coûts. Les initiatives antérieures telles que le Projet de la stratégie du Recensement de 2016³ ont abouti à la conclusion que l'adoption d'un recensement fondé exclusivement sur les registres administratifs (comme ce qui a été mis en œuvre dans les pays nordiques, par exemple) ne serait pas possible au Canada à court ou à moyen terme, car les conditions nécessaires à la mise en œuvre d'une telle approche n'étaient pas réunies.

4. Le présent document présente une brève description des recherches effectuées sur l'expansion de la méthode mixte pour les futurs cycles de recensement au Canada. Il présente également les possibilités de tirer parti de certains de ces travaux pour appuyer les opérations de collecte en vue du prochain recensement, en 2021.

¹ Hamel (2012).

² Statistique Canada (2019).

³ Howatson-Leo et Trépanier (2011).

II. Élaboration de registres statistiques

5. Depuis 2010, Statistique Canada examine comment les données administratives peuvent être exploitées davantage dans le cadre du recensement de la population canadienne. Certains renseignements produits pour le recensement ont été obtenus de sources administratives au cours des trois derniers cycles (par exemple, revenu et catégorie au regard de l'immigration, en 2016). Avant l'introduction d'une approche multimodale, les effets modaux des différents modes ou sources de réponse au recensement avaient été étudiés, et il avait été constaté que la qualité de l'information n'avait pas été sensiblement modifiée⁴. En fait, les réponses provenant de sources administratives et des questionnaires en ligne sont généralement considérées comme étant de meilleure qualité.

6. Le recensement administratif tel qu'il est réalisé, par exemple, dans les pays nordiques (Danemark depuis 1981, Finlande depuis 1990, Norvège et Suède depuis 2011) ou aux Pays-Bas, est souvent cité comme le moyen le plus efficace de tirer parti des données administratives ou des registres pour obtenir les dénombrements de la population et des logements. Toutefois, ce modèle ne convient pas dans le contexte canadien, puisque le Canada ne dispose pas des éléments fondamentaux que sont un système unifié de registres administratifs et un numéro d'identification personnel pour chaque résident.

7. La recherche initiale sur l'utilisation des sources administratives dans le contexte du recensement s'est concentrée sur la construction d'une population statistique (Base de données statistique et démographique du Canada) par la compilation de renseignements provenant de diverses sources administratives, principalement des données fiscales et d'immigration, ainsi que de renseignements provenant des registres des naissances et décès. Les premiers résultats, prometteurs, des travaux sur la Base de données statistique et démographique ont ensuite amené Statistique Canada à entreprendre l'élaboration d'un Registre statistique de la population (RSP) et à transformer le Registre des adresses (RA) en un registre statistique des bâtiments (RSB). Ces registres sont des registres statistiques, c'est-à-dire qu'ils sont construits à partir de diverses sources de données administratives et ne serviront qu'à appuyer le mandat de Statistique Canada consistant à fournir des produits statistiques dépersonnalisés et protégés. Le RSP fournira une liste continuellement mise à jour des personnes vivant au Canada à un moment donné, ainsi que les caractéristiques démographiques de base comme l'âge, le sexe à la naissance et le lieu de résidence.

8. Le RSB sera une évolution du Registre des adresses de Statistique Canada et fournira une liste exacte de tous les bâtiments au Canada, y compris les logements collectifs, les logements privés occupés et inoccupés, et les bâtiments non résidentiels.

9. Le RSP et le RSB, jumelés au Registre des entreprises existant et à un éventuel registre des activités, fourniront une infrastructure de registres statistiques de base interconnectés qui seront accessibles à tous les programmes de Statistique Canada. Par exemple, le RSP pourrait être utilisé par le programme d'estimations démographiques pour les estimations intercensitaires ou comme base d'échantillonnage pour les programmes de statistiques sociales.

III. Élargir l'approche mixte pour les futurs recensements

10. Des pays comme le Royaume-Uni ou la Nouvelle-Zélande envisagent d'adopter une méthode de recensement fondée en tout ou en partie sur des données administratives, même en l'absence d'un solide système unifié de registres gouvernementaux et de numéros d'identification personnels. Depuis 2016, Statistique Canada étudie l'utilisation de l'information stockée dans un éventuel RSP comme éventuelle nouvelle modalité de réponse au recensement de la population.

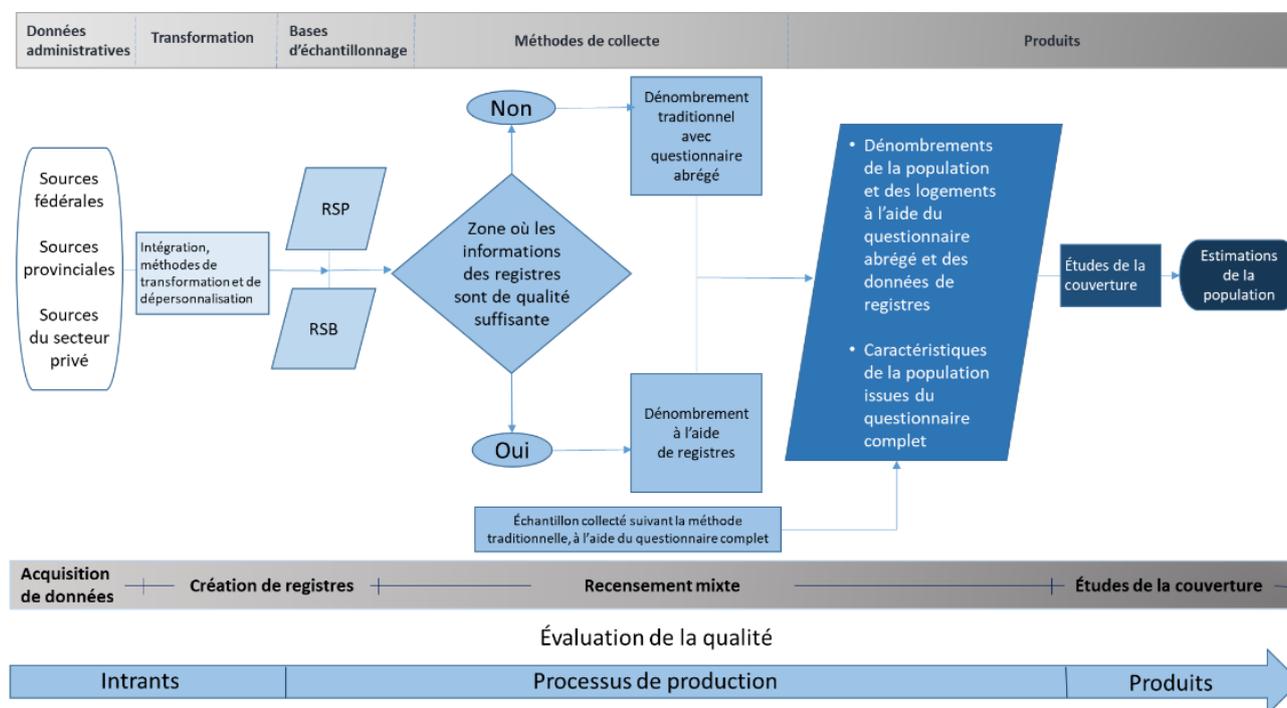
11. La principale conclusion des recherches menée à ce jour est que les données administratives brutes actuellement en possession de Statistique Canada sont très prometteuses pour ce qui est de fournir une bonne couverture de la population canadienne

⁴ Grondin et Sun (2008).

et pourraient éventuellement être utilisées pour produire des dénombrements de la population. Il faut poursuivre les travaux, car les recherches démontrent également que la couverture infraprovinciale doit être améliorée par le perfectionnement des dossiers administratifs actuels ou par l'acquisition de nouveaux dossiers. Malgré cela, les informations contenues dans les registres statistiques pourraient très probablement être jugées de qualité suffisante dans de nombreux domaines pour être utilisées dans le programme de recensement dans un avenir prévisible. On pourrait y parvenir en élargissant la méthode actuelle de recensement mixte.

12. Un recensement mixte élargi au Canada intégrerait l'information des registres statistiques à la collecte traditionnelle. Globalement, cela signifie que la base d'échantillonnage du recensement des logements canadiens serait divisée en deux parties, en utilisant diverses caractéristiques (couverture et exactitude du registre, type de logement, zone géographique, considérations opérationnelles, coûts, etc.) D'une part, les résultats du recensement seraient produits à l'aide de données administratives de qualité suffisante provenant du RSP et du RSB au lieu d'un dénombrement direct. Ainsi, le RSP serait utilisé pour contribuer des informations démographiques de base, telles que la date de naissance, le sexe à la naissance et le lieu géographique, au dénombrement de la population. Le dénombrement traditionnel à l'aide du questionnaire abrégé serait utilisé pour la deuxième partie, lorsque les informations contenues dans les registres ne sont pas jugées de qualité suffisante ou sont trop peu précises. Parallèlement à cela, on continuerait à sélectionner un échantillon représentatif de logements pour fournir des caractéristiques détaillées au moyen du questionnaire complet, comme c'est le cas dans le modèle actuel. La figure I montre les différentes étapes du cheminement pour aller des sources de données administratives aux estimations démographiques.

Figure I
Passer des sources de données administratives aux estimations de la population au moyen d'un recensement mixte



13. Statistique Canada effectue actuellement plusieurs simulations de cette approche, dans le cadre desquelles les premières itérations du RSP sont comparées à la base de données du recensement de 2016 et aux résultats du Programme d'estimations démographiques, à la fois pour valider la qualité des registres statistiques et pour déterminer quelle serait l'utilisation optimale de ces registres dans une approche mixte. Des simulations semblables seront effectuées parallèlement au recensement de 2021. C'est à l'issue de ces recherches que l'on se prononcera sur la mise en œuvre éventuelle d'une approche mixte élargie en 2026.

IV. Utilisation des registres statistiques pour le recensement de 2021

14. Pour le programme du recensement de la population de 2021, Statistique Canada réutilisera et répétera en grande partie les approches mises en œuvre avec succès pour le programme de 2016. Toutefois, les possibilités d'amélioration de l'approche actuelle, qui est axée sur un taux maximal d'autodéclaration, sont maintenant plus limitées. Le principal gain d'efficacité potentiel pour la méthode de recensement actuelle consiste à tirer parti de l'information facilement accessible pour réduire la nécessité d'effectuer le suivi des logements où aucune réponse n'a été obtenue. En utilisant l'information provenant de la collecte ainsi que des données issues des nouveaux registres statistiques et des nouvelles sources de données administratives dont dispose Statistique Canada, il pourrait être possible de réduire les coûts et la charge de travail pour les répondants, tout en maintenant la qualité élevée du recensement. La stratégie proposée pour 2021 comporte trois grands volets :

a) Réduire le surdénombrement de la base de recensement : faire passer le nombre de logements de la base d'échantillonnage devant être annulés de 4,5 % du total en 2016 à 3,5 % en 2021, grâce à une validation à l'aide d'informations provenant de sources administratives ;

b) Prévoir l'état d'occupation du logement : classer comme à annuler, inoccupés ou occupés les logements pour lesquels aucune réponse n'a été obtenue ;

c) Arrêter le suivi de certains cas de non-réponse et utiliser les données du RSP comme réponse indirecte.

15. La réduction du surdénombrement sera en grande partie réalisée grâce à l'appariement avec divers dossiers administratifs. Cela fait plusieurs cycles de recensement que Statistique Canada utilise le Registre des adresses comme base d'échantillonnage. La majeure partie des actualisations du Registre des adresses sont effectuées à l'aide de sources administratives. Le fichier principal utilisé à cette fin est le fichier des points de remise tenu à jour par Postes Canada. Moins de 20 % des adresses sont validées dans les bureaux de Statistique Canada ou sur le terrain.

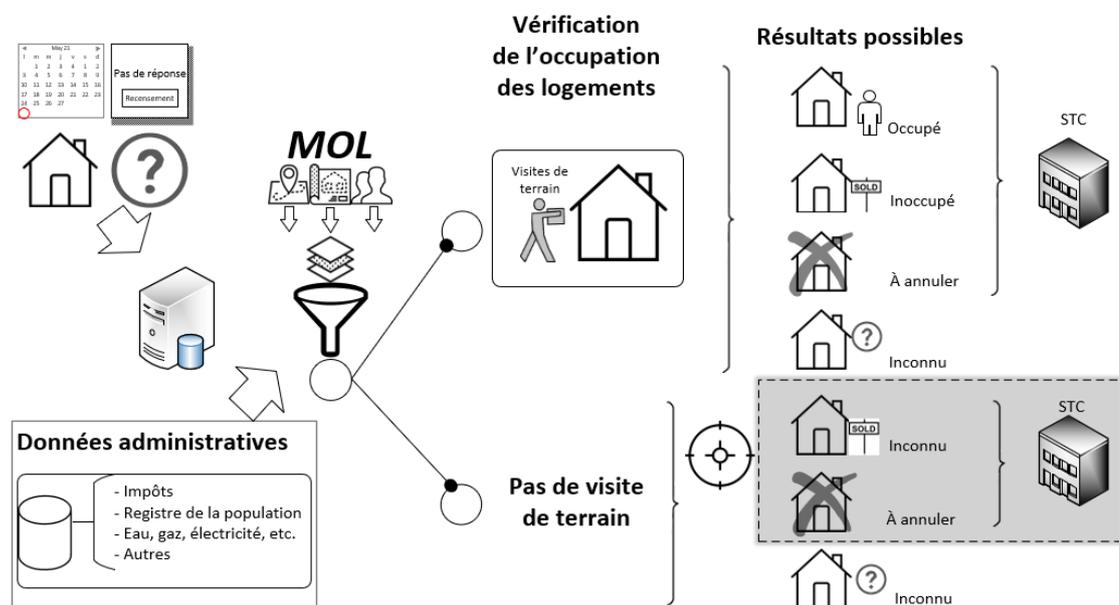
16. Selon la procédure appliquée jusqu'à présent, il y aurait après le recensement une nouvelle actualisation du Registre des adresses qui inclurait les éléments annulés pour la première fois au cours du recensement précédent et ces logements feraient alors partie de la base d'échantillonnage pour le prochain recensement, en prévision du fait qu'ils pourraient être des logements valides à ce moment-là. Cette procédure a été modifiée avec la première mise à jour du Registre des adresses après le recensement de 2016, qui ne portait que sur les logements privés valides du recensement de 2016. Les logements annulés en 2016 ne seront réactualisés que s'ils figurent dans les dossiers administratifs utilisés pour mettre à jour le Registre des adresses. Une méthode en cours d'élaboration devrait permettre de détecter les doublons du fichier (enregistrements distincts se rapportant au même logement). D'autres méthodes ont été mises au point en interne pour affiner les processus d'appariement et de rapprochement afin de réduire au minimum les doublons dus à différents formats d'adresse.

17. Une des méthodes de collecte actuellement employée pour le recensement sera améliorée pour 2021 en utilisant de nouvelles sources de données administratives et en élargissant ses objectifs. L'opération de vérification de l'occupation des logements, qui vise à relever, aussi près que possible du jour du recensement, les logements vacants et les logements à annuler, reposera sur une méthode d'échantillonnage plus précise fondée sur de nouvelles sources de données administratives.

18. Un modèle d'occupation des logements (MOL), en cours d'élaboration, devrait permettre de prédire le statut d'occupation de chaque logement. Sur la base d'informations provenant de diverses sources administratives telles que les dossiers des administrations fiscales, les données relatives aux services collectifs de distribution (eau, gaz, électricité etc.) et les registres statistiques, on attribuera à chaque logement une probabilité d'être occupé, inoccupé et à annuler avant les opérations sur le terrain. La figure II ci-après illustre l'application de ce modèle. Les logements ayant de fortes probabilités d'être inoccupés ou à annuler (c'est-à-dire une plus faible probabilité d'être occupés) seront pris

en compte dans la sélection de l'échantillon pour la vérification. Cet échantillon sera également sélectionné sur la base de critères opérationnels, comme le regroupement des logements à contrôler, afin de maîtriser le coût de la collecte. Pendant l'opération de vérification, l'état d'occupation est vérifié par du personnel de terrain.

Figure II
Processus dérivé du Modèle d'occupation des logements



19. Selon les performances du Modèle d'occupation des logements, de nouveaux traitements pourraient être appliqués avant de définir l'échantillon pour la vérification, permettant de supprimer complètement les visites sur le terrain ou de les limiter aux opérations de suivi des non-réponses :

a) Les logements ayant une probabilité élevée d'être inoccupés et une faible probabilité d'être à annuler ou occupés seraient classés comme étant inoccupés.

b) Les logements ayant une probabilité élevée d'être à annuler et une faible probabilité d'être inoccupés ou occupés seraient classés comme étant à annuler.

20. Le troisième volet de la stratégie permettra de mettre fin au suivi de certains logements non répondants pendant les opérations de suivi des non-réponses, en utilisant les données du Registre statistique de la population comme réponse indirecte. La stratégie actuellement appliquée au Canada vise à obtenir un taux de réponse de 98 % au minimum pour chacune des quelque 45 000 unités de collecte définies pour les opérations de suivi des non-réponses. Dans ce contexte, le Registre statistique de la population pourrait être utilisé pour un nombre maximal prédéterminé de logements par unité de collecte dans les zones où des efforts importants ont déjà été faits sans succès en vue d'obtenir une réponse des logements non répondants, ou même pour accroître encore le taux de réponse dans les unités de collecte qui sont déjà à 98 % ou au-delà.

21. Parmi les critères à utiliser pour dresser la liste des logements à dénombrer à l'aide du Registre statistique de la population figurent les suivants : le Modèle d'occupation des logements indique une forte probabilité que le logement soit occupé ; le logement ne fait pas partie de l'échantillon du questionnaire complet ; les données issues du Registre statistique de la population sont d'une qualité relativement bonne ; et la limite fixée pour ce type de remplacement n'a pas été dépassée pour l'unité de collecte en question. Des limites doivent être imposées à l'utilisation de ce type de réponse pour des ménages entiers afin d'éviter d'influer sur la variance des résultats des études de couverture du recensement, qui sont eux-mêmes fondés sur des sources administratives.

V. Conclusion

22. Statistique Canada a progressé rapidement dans ses travaux de recherche sur les solutions administratives susceptibles de remplacer un recensement par dénombrement. Cependant, toute mise en œuvre de telles approches à grande échelle devra être examinée très attentivement. En effet, les chiffres de population sont la base du système électoral canadien, et des milliards de dollars de transferts annuels du gouvernement fédéral aux provinces et aux territoires, et par extension aux municipalités, reposent sur les résultats du recensement. Il faudra pouvoir garantir la qualité avant toute décision d'aller de l'avant avec une approche mixte beaucoup plus étendue.

23. La mise en œuvre d'une approche mixte impliquera également certains compromis. Par exemple, certains des résultats du recensement actuellement produits à partir du questionnaire abrégé (100 % de la population), comme la situation matrimoniale et la formation de la famille, devraient être transférés au questionnaire détaillé (échantillon de 25 %) parce que ces renseignements ne sont généralement pas disponibles dans les dossiers administratifs. Cela ne sera peut-être pas un problème majeur, car la valeur de ces variables est beaucoup plus grande lorsqu'elles sont combinées à d'autres variables du questionnaire complet. Il faudra toutefois mener de vastes consultations auprès des utilisateurs de données et des autres parties prenantes avant que l'approche combinée ne puisse être élargie.

Références

- Hamel, Collecte de données par Internet dans le cadre du recensement de la population canadienne (document soumis à la quatorzième réunion du Groupe d'experts des recensements de la population et des habitations de la CEE – ECE/CES/GE.41/2012/6), Genève (2012)
- Statistique Canada (2019) La modernisation : au cœur des efforts de réduction du fardeau de réponse de Statistique Canada, https://www.statcan.gc.ca/fra/apercu/smr09/smr09_100
- L. Howatson-Leo et J. Trépanier (2011) Rapport final sur les options du Recensement de 2016 : cadre proposé de détermination du contenu et options méthodologiques (Statistique Canada). https://www12.statcan.gc.ca/census-recensement/fc-rf/reports-rapports/r2_index-fra.cfm
- C. Grondin, L. Sun, 2006 Census Internet Mode Effect Study, Joint Statistical Meeting of the American Statistical Association, août 2008
-