



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****Aspects du recensement intégré en Slovaquie ayant
une incidence sur l'utilisation des sources de données
administratives****Note du Bureau de statistique de la République slovaque****Résumé*

En Slovaquie, le recensement de 2021 repose sur une nouvelle méthode, qui combine des données provenant de multiples sources administratives et des données recueillies sur le terrain. La phase préparatoire du recensement de 2021 a consisté à assurer l'accès aux sources considérées comme pertinentes mais aussi à travailler avec ces dernières, les soumettre à une évaluation qualitative, analyser les erreurs, proposer des corrections et mettre au point les modalités de leur intégration. Le présent document regroupe les principales conclusions et recommandations concernant la procédure à suivre.

* Rédigée par M^{mes} Ludmila Ivančíková et Zuzana Podmanická.



I. Introduction

1. Jusqu'ici, tous les recensements de la population en Slovaquie étaient faits selon la méthode traditionnelle. Le recensement prévu pour 2021 repose sur une nouvelle méthode qui comporte plusieurs éléments novateurs. Il s'agit d'une méthode combinée qui utilise à la fois des sources de données administratives et des données recueillies directement auprès de la population (enquête exhaustive).
2. L'organisation d'un recensement exhaustif nécessite à la fois que l'on dispose de données détaillées à l'échelon régional (municipalités – unités administratives locales de niveau 2) et que l'on obtienne des données fondées sur des mailles utilisables dans le cadre d'Eurostat, conformément au règlement d'application en vigueur. De même, du point de vue de la diffusion des données, la nécessité de pouvoir comparer les nouvelles données recueillies avec celles des recensements précédents et l'élaboration d'hypercubes pour l'Union européenne et de datacubes par le Bureau de statistique de la République slovaque, ont aussi joué un rôle.
3. L'un des aspects primordiaux de cette nouvelle méthode est l'intégration des données statistiques. Elle suppose à la fois l'intégration des données provenant de multiples sources administratives entre elles et l'intégration de celles-ci avec celles recueillies auprès de la population. Il s'agit d'une intégration à plusieurs niveaux, ainsi que nous allons le voir dans les paragraphes ci-après.

II. Cadre législatif de l'utilisation des sources de données administratives

4. Le modèle de recensement présenté au niveau national par le législateur en 2018, qui a ensuite été repris dans le projet de loi sur le recensement de la population et des habitations de 2021, est le premier recensement intégré de l'histoire de la Slovaquie. Ce recensement dispense la population de l'obligation de remplir de complexes formulaires sur papier par autodénombrement. Le Bureau de statistique recueillera directement auprès de la population uniquement les données qui ne pourront pas être fournies par les sources de données administratives. La population sera complètement dispensée de communiquer des données sur les habitations.
5. Bien que l'utilisation de données administratives soit fondée sur les recommandations internationales en matière de recensement, le Bureau de statistique de la République slovaque n'a encore jamais concrètement utilisé ce type de données dans le domaine des statistiques sociales ; dans le domaine des statistiques commerciales, elle l'a déjà fait mais de façon très limitée. La première étape a donc consisté à répertorier les sources de données pertinentes, les administrateurs des registres et les éléments contenus dans ceux-ci, qui pourraient être utiles pour le recensement. En tout, le Bureau de statistique a communiqué avec 17 administrateurs de sources de données administratives entre 2016 et 2018 et il a retenu 12 sources.
6. Plusieurs administrateurs de sources de données administratives ont été sollicités pour communiquer des données dans le cadre d'accords bilatéraux. Dans un premier temps, on a vérifié que les données communiquées étaient pertinentes pour le recensement. Bien que la conclusion de ces accords ait pris du temps et qu'elle ait donné lieu à un certain nombre de tractations, cela a été une réussite, notamment dans le cas des registres contenant les données relatives à la population. En ce qui concerne l'entrée en vigueur de la Réglementation générale sur la protection des données (GDPR), qui est intervenue en Slovaquie au titre d'une loi spéciale sur la protection des données personnelles, certains administrateurs de sources de données administratives ont soulevé la question de la légitimité de cette requête. Les réserves formulées par ceux-ci ont mis en lumière la nécessité d'une définition claire de leur rôle dans le recensement, la pertinence de l'utilisation des données aux fins d'essai et de vérification et l'absence de législation nationale¹.

¹ En Slovaquie, le recensement de la population et des habitations est le seul qui soit organisé en vertu d'une loi distincte et non pas en vertu de la loi sur les statistiques d'État n° 540/2001.

7. Une solution a été apportée par l'amendement de la loi n° 540/2001 (Recueil des lois sur les statistiques d'État), dont la Section 15 dispose que les recensements de la population et des habitations sont un type particulier d'enquête statistique qui est confiée au Bureau de statistique, organisme d'État ; son contenu, sa portée et les modalités de sa mise en œuvre sont définis par des dispositions spéciales. Pour la préparation et la réalisation du recensement, le Bureau de statistique recourt aussi à des données provenant de sources administratives conformément à la Section 13 . Grâce à cet amendement, le Bureau de statistique avait déjà accès à toutes les sources de données administratives pertinentes, avant même le recensement. Du point de vue de la méthode, cet amendement lui a permis d'analyser les données en détail, de mettre au point et de valider des méthodes d'évaluation de la qualité, d'utiliser des métadonnées, de contrôler le traitement des données et enfin de concevoir et d'éprouver une méthode d'intégration statistique.

8. Le projet de loi sur le recensement de la population et des habitations de 2021 stipule que, à la demande écrite du Bureau de statistique, l'administration publique et le directeur de son système d'information sont tenus de fournir au Bureau de statistique les données, les informations, les métadonnées et les indicateurs démographiques, notamment le numéro d'identification personnel utilisé par le Bureau de statistique afin d'intégrer les données provenant du système d'information de l'administration publique et celles provenant des autres sources de données administratives dont ils ont la gestion.

9. La loi porte aussi sur le cas de deux administrateurs de sources de données administratives qui, aux termes d'un accord, ont communiqué des données à valider dans le cadre de la préparation du recensement, mais une seule fois, et qui ont subordonné la communication de données supplémentaires à la mise en œuvre de la loi sur le recensement. De plus, la loi organise la communication des données administratives pendant le recensement, mais surtout pendant l'enquête postrecensement.

10. La loi énumère les administrateurs des sources de données administratives, qui sont au nombre de neuf : Ministre de l'intérieur, Ministre de l'éducation, Ministre de l'environnement, Autorité de la géodésie, de la cartographie et du cadastre, Bureau d'assurance sociale, Direction des finances, Administration pénitentiaire et judiciaire, Bureau du travail, des affaires sociales et de la famille et Autorité de surveillance des soins de santé. Tous ont l'obligation de communiquer des données au Bureau de statistique au bon moment et en respectant les délais fixés. La loi stipule en outre que les administrateurs des sources de données administratives doivent communiquer leurs données par voie électronique, gratuitement et lorsque le Bureau de statistique leur en fait la demande par écrit, sur la base d'un arrangement négocié à l'avance. Les modalités et les aspects techniques de la communication des données administratives seront définis dans un accord dont les termes seront fixés par le Bureau de statistique par décret.

III. Utilisation des sources de données administratives

11. Le recensement de 2021 exploitera les sources de données administratives suivantes :

- a) Registre des personnes physiques ;
- b) Registre des adresses ;
- c) Registre central des titulaires d'une police d'assurance de santé publique ;
- d) Registre des personnes mises en examen ou condamnées, tenu par l'Administration pénitentiaire et judiciaire ;
- e) Systèmes d'information de l'administration publique gérés par le Ministre de l'éducation ;
- f) Système d'information du Bureau d'assurance sociale ;
- g) Systèmes d'information de l'Autorité de la géodésie, de la cartographie et du cadastre ;
- h) Système d'information de la Direction des finances ;

i) Système d'information du Bureau central du travail, des affaires sociales et de la famille ;

j) Système d'information du Ministère de l'environnement.

12. Lorsqu'il a arrêté les détails de la procédure d'application du projet de recensement 2021, le Bureau de statistique a supposé avec raison que les préparatifs du recensement dureraient plus longtemps que d'habitude et qu'ils nécessiteraient, en plus d'un certain nombre d'études, l'analyse de données administratives.

13. Pendant la phase préparatoire, l'analyse de chacune des sources de données est conduite de façon uniforme. Étant donné que certaines sources sont analysées à plusieurs reprises, on en profite pour vérifier qu'elles s'améliorent. Néanmoins, on est obligé de constater que, bien que les administrateurs des sources de données administratives soient conscients d'un certain nombre d'incohérences et qu'ils les acceptent, il arrive souvent que la nature des registres et des sources de données administratives elles-mêmes (ainsi que la législation correspondante) ne permette pas de corriger les ensembles de données, même aux fins du recensement.

14. Les analyses portent sur les caractéristiques techniques des ensembles de données communiquées, l'exploitabilité des données, l'exhaustivité de certaines données, les listes de codes et la possibilité de les convertir en listes statistiques imposées, les liens logiques entre les variables, la détection des erreurs et les possibilités de correction optimale de l'absence de données ou d'erreurs logiques.

15. Un système préétabli de vérifications, de conversions et d'édits d'ordre logique sera vérifié, modifié et installé dans le système de traitement électronique des informations.

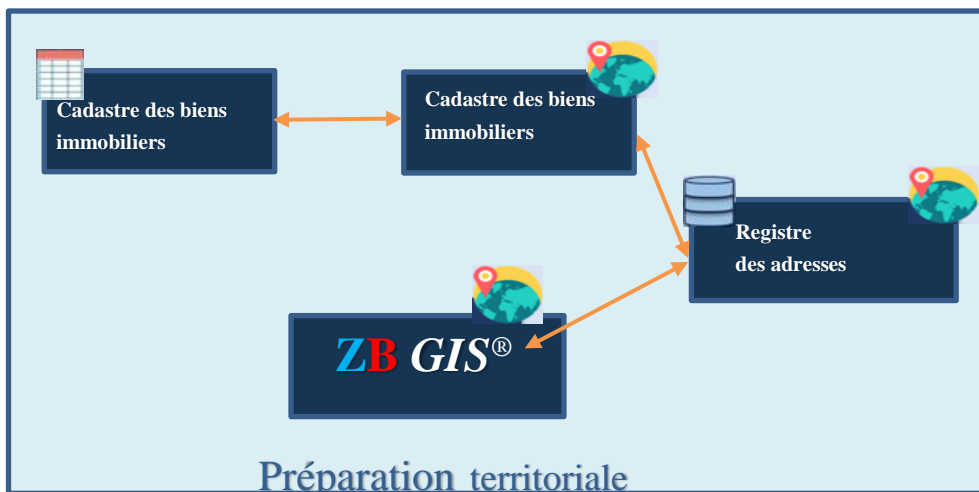
16. Les sources de données administratives seront utilisées au moment de la préparation territoriale, du recensement des habitations, de la collecte électronique des données concernant la population et aussi au moment du traitement.

17. Afin que le recensement porte sur la totalité du territoire, on procède à ce qu'il est convenu d'appeler une « préparation territoriale ». Dans le cas du recensement de 2021, il s'agit de la reconnaissance dans l'espace et dans la base de données de tous les objets contenus dans l'unité territoriale statistique dans lesquels la population peut vivre au moment du recensement, qu'il s'agisse ou non d'une résidence officielle.

18. Pendant la préparation territoriale on utilisera les sources de données administratives suivantes : le Registre des adresses administré par le Ministère de l'intérieur, les systèmes d'information du Bureau de la géodésie, de la cartographie et du cadastre (Cadastre des biens immobiliers et ZBGIS ®) et le système d'information du Ministère de l'environnement. La préparation territoriale permettra de dresser une liste d'adresses donnant le plus grand ensemble d'objets habités ou habitables (maisons individuelles, immeubles d'appartements, cabanes, abris de jardin, écoles, etc.) obtenues grâce aux sources de données administratives ainsi que la localisation géographique de ces objets sur les cartes de base.

Figure 1

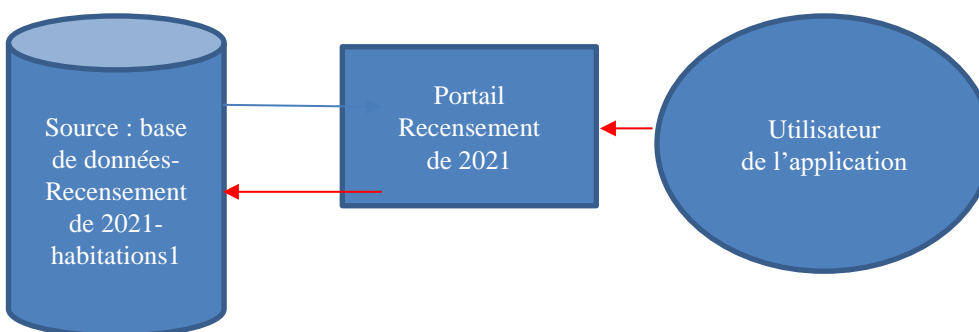
Sources de données administratives pour la préparation territoriale du Recensement de 2021



19. Le recensement des habitations fera l'objet d'une préparation territoriale. Comme cela a déjà été indiqué, les autorités s'efforceront d'alléger la charge que les enquêtes font peser sur la population. À cette fin, lors du recensement des habitations, les données administratives communiquées au Bureau de statistique (plus précisément le cadastre des biens immobiliers) seront utilisées pour préremplir le formulaire de chaque objet habitable conformément à la préparation territoriale, après quoi les municipalités ajouteront les données manquantes en se servant des registres existants, principalement les registres municipaux, les registres de gestionnaires d'immeubles d'appartements et les registres des sociétés de gestion. De la sorte, les habitants seront dispensés de participer à la collecte des données concernant les caractéristiques de leurs habitations et le recensement des habitations sera effectué par les municipalités (unités administratives locales de niveau 2).

Figure II

Collecte électronique des données concernant les habitations dans le Recensement de 2021.



20. La personne chargée du recensement des habitations sera désignée par la municipalité. Le recensement des habitations sera effectué exclusivement de façon électronique, via le portail Web du Bureau de statistique.

21. La collecte des données commencera le 1^{er} juin 2020, lorsque le module baptisé « Collecte-Recensement des habitations 2021 » (« ZBER SDB 2021 »), qui contient les formulaires électroniques concernant les habitations, qui ont été préremplis dans toute la mesure possible, sera mis à la disposition de toutes les municipalités par l'intermédiaire du portail du Bureau de statistique. L'accès de chaque municipalité sera protégé par un mot de passe et un nom qui sera constitué du code statistique de la municipalité conformément à la liste des codes imposée. Le code et le nom d'accès seront distribués aux municipalités par courrier électronique, conformément à la loi n° 305/2013, Recueil § 12. Les personnes autorisées par la municipalité vérifieront les listes des habitations sur la base des sources de données recommandées. Les données seront collectées jusqu'au 12 février 2021 ; la date de référence du recensement est fixée au vendredi 1^{er} janvier 2021 à 0 heure.

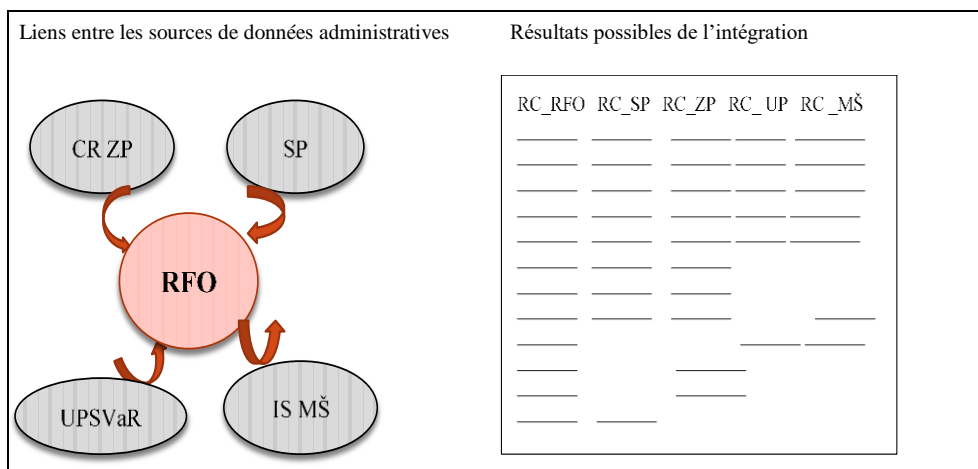
22. Pendant la collecte des données, les bases de données seront alimentées à la fois avec des variables calculées (par exemple le nombre de personnes vivant dans une habitation) et des variables dérivées (par exemple le propriétaire de l’habitation). Cet ensemble de données traitées concernant les habitations, qui est composé à la fois de variables calculées et de variables dérivées, sera communiqué à chaque municipalité pour son cadastre, une fois que le recensement de la population et des habitations sera complètement terminé. En d’autres termes, les ensembles de données municipales seront communiqués à chaque municipalité, sous différents formats correspondant aux particularités du système d’information de chaque municipalité. Dans le même temps, cette base de données permettra aux municipalités de mettre à jour ces données en continu et ainsi d’assurer la durabilité du système et son utilisation pour répondre aux besoins locaux des municipalités ainsi qu’aux exigences nationales qui interviendront après le recensement de 2021.

23. Pour le recensement de la population et la collecte de données la concernant, on utilisera des données provenant du Registre des personnes physiques (RFO), du Registre des adresses, du Registre central des titulaires de polices d’assurance de santé publique (CR ZP), du Registre des personnes mises en examen ou condamnées, des systèmes d’information de l’administration publique administrés par le Ministère de l’éducation (IS MŠ), le système d’information de l’organisme d’assurance sociale (SP), le système d’information de la Direction des finances et le système d’information du Bureau central du travail, des affaires sociales et de la famille (UPSVaR). Les quelques données qui ne figureront pas dans les sources de données administratives seront recueillies directement auprès de la population, exclusivement par voie électronique.

24. Dans le cadre de la préparation et du traitement des données, l’intégration des données administratives se déroulera en trois phases :

- a) Première phase : transfert et stockage des lots de données ;
- b) Deuxième phase : vérification et traitement initial des données, source par source ; dans le cas du Registre des personnes physiques dont les lots contiennent plusieurs tableaux, on recherchera les liens entre le numéro personnel d’identification et la base de données résultant des données provenant du Registre. Dans le même temps, les sources éventuelles d’erreur relevées lors des analyses initiales seront supprimées ;
- c) Troisième phase : intégration des sources de données administratives entre elles, en utilisant un identifiant unique et la liste des codes d’intégration.

Figure III
Exemple d’intégration de sources de données administratives concernant la population



25. La collecte de données auprès de la population sera entièrement électronique. Les habitants seront dénombrés au moyen d’un formulaire électronique. Le formulaire sera disponible en ligne et hors ligne mais les formulaires hors ligne ne seront utilisés qu’à titre exceptionnel. En ce qui concerne la nature des données recueillies, l’autodénombrement sera préféré pour remplir les formulaires électroniques. Dans le cas des personnes mineures et des habitants dans l’incapacité de fournir eux-mêmes des données, le formulaire

électronique sera rempli par leur représentant(e) légal(e), un proche ou une personne vivant sous le même toit.

26. La collecte de données auprès de la population se fera entre le 15 février 2021 et le 31 mars 2021. Sur l'ensemble du territoire de la République slovaque, la date de référence du recensement (c'est-à-dire la date à laquelle se rapporteront les données) a été fixée au vendredi 1^{er} janvier 2021 à 0 heure. Cette date a été choisie compte tenu de l'utilisation des registres et des autres sources de données administratives.

27. La collecte des données se fera de telle façon que les habitants puissent s'autodénombrer pendant toute la période du recensement, soit du 15 février 2021 au 31 mars 2021. En outre, pendant toute cette période, ils pourront aussi s'adresser à des points de contact où ils pourront s'autodénombrer ou être dénombrés par un agent recenseur.

28. L'intégration des données administratives et des données démographiques repose sur le regroupement des variables en quatre groupes, comme suit :

a) Variables du premier type : variables collectées électroniquement, qui peuvent être éventuellement soumises à des vérifications (numéro personnel d'identification, nom et prénom, date de naissance, sexe et adresse) ;

b) Variables du deuxième type : variables provenant des sources de données administratives mais qui ne seront pas utilisées pour la collecte électronique (par exemple nombre d'enfants ou activité économique) ;

c) Variables du troisième type : variables provenant exclusivement des sources de données administratives (par exemple âge, sexe, activité économique ou catégorie d'emploi) ;

d) Variables du quatrième type : variables recueillies exclusivement sur le terrain (par exemple religion, nationalité ou niveau d'éducation).

29. La collecte des variables du premier type provenant de toutes les sources de données administratives donnera le départ de la collecte électronique.

30. L'intégration des données administratives avec celles provenant de la collecte électronique se déroulera en deux phases. Pendant la première phase, seules seront utilisées les variables du premier type qui assureront la fonction d'identification de la collecte électronique. Pendant la seconde phase, une fois terminée la collecte électronique, les données provenant de celle-ci seront mises en lien avec les données traitées provenant des sources administratives.

31. Les intégrations destinées à obtenir des données supplémentaires (par exemple des données sur les ménages ou des données sur la population vivant dans les habitations) subiront un traitement statistique une fois terminée la collecte électronique des données sur la population.

32. Les données provenant des sources administratives qui seront considérées comme de bonne qualité et pertinentes à l'issue du recensement pourront servir de point de départ à l'après-recensement de 2021.

IV. Conclusion

33. Comme le montrent les pays qui effectuent des recensements combinés et des recensements à partir de registres, ce type de recensement repose sur une législation favorable, l'appui du public, une coopération avec les administrateurs des sources de données administratives ainsi qu'avec les autres organismes participant au recensement, l'existence d'un système approprié de registre (ce qui n'était pas le cas en Slovaquie jusqu'en 2011) et l'existence d'un numéro personnel d'identification unique². Si ces conditions sont réunies, alors la Slovaquie pourra passer des recensements traditionnels aux recensements « intégrés ».

² Schulte Nordholt, Eric (2018) The usability of administrative data for register-based censuses, *Statistical Journal of the IAOS* 34:4, pp.487–498.

34. La nécessité d'analyser de nouvelles sources, l'évaluation de leur qualité et des méthodes de leur intégration ainsi que l'utilisation de ces méthodes vont entraîner une modification des méthodes de travail. Dans le même temps, elles favorisent la création de processus qui permettront de répondre aux exigences (même si nous restons sur un plan plutôt idéologique) du postrecensement (et la production plus fréquente d'une partie des données du recensement) et d'une éventuelle transition vers un recensement reposant exclusivement sur les registres.

35. Le recensement intégré présente à la fois des avantages et des inconvénients. En effet, on risque moins de collecter des données qui ne figurent pas dans les registres, sont d'une qualité insuffisante ou ne correspondent pas à la méthode employée pour le recensement. De nombreux pays renoncent à certaines variables dont le recensement était la seule source. Des registres sont créés à des fins non statistiques. Par nature, les registres ne rendent pas toujours les nuances des changements sociaux, tels que la composition de ménages atypiques, contrairement au recensement traditionnel. C'est la raison pour laquelle il faudra procéder à davantage d'analyses en profondeur et peut-être chercher de nouvelles sources, étant donné que le renforcement des enquêtes par échantillonnage semble difficile. En revanche, la participation décroissante de la population au recensement entraîne le recours aux seules sources disponibles ou que l'on peut mettre à la disposition des fournisseurs de statistiques. Leur utilisation permet d'enrichir le contenu des statistiques et de mettre au point de nouvelles méthodes et rend en outre les bureaux de statistique quelque peu dépendants d'autres organismes comme les sources de données administratives.
