



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**Utilisation des données administratives pour établir une liste de la population visée par le recensement****Note de l'Office de statistique de Pologne****Résumé*

On trouvera dans le présent document une description de la méthode suivie pour élaborer une liste de la population visée par le recensement à partir de données issues de sources administratives. En outre, une nouvelle approche novatrice est présentée, qui, en plus de définir la population couverte par le recensement, permettra de générer un ensemble de données qui intègre d'éventuelles coupes transversales et se positionne au niveau le plus bas possible de la division territoriale (mailles de 1 km² comprises). L'approche tient compte de la spécificité des données des registres, notamment en termes de qualité et de complexité. Le document présente des questions tant méthodologiques que techniques en lien avec la logique de l'élaboration de la liste de la population visée. Il traite des sources de données et de leurs méthodes d'exploitation, ainsi que de la production de rapports de qualité.

* Note établie par Pawel Murawski.



I. Introduction

1. Le recensement est la seule enquête permettant d'obtenir des informations détaillées sur la population, sa répartition territoriale, sa structure démographique et sociale et les migrations professionnelles (en particulier des salariés), les ménages et les familles, ainsi que sur l'état quantitatif et qualitatif des ressources en logements et en bâtiments.
2. Ces informations sont collectées à tous les niveaux de la division territoriale du pays : national, régional et local. Cette approche, fondée sur les données recueillies lors du recensement, permet de prendre un instantané de la société pour faire en quelque sorte l'état des lieux de la société au moment du recensement. Dans le cadre de la mise en œuvre de mesures visant à accroître leur utilisation, la liste de la population visée sera établie sur la base exclusive de données issues de sources administratives. Ce document expose la méthode d'élaboration de la liste, passe en revue les sources de données et traite de questions tant méthodologiques que techniques en lien avec la logique de la création de la liste de la population visée.

II. Élaboration de la liste de la population visée

3. L'élaboration de la liste de la population visée a pour but de fournir un descriptif global et exhaustif de la population en vue du recensement de 2021 et des recensements annuels.
4. Cette mesure s'est appuyée sur l'intégration de données issues de nombreuses sources hétérogènes aux fins d'obtenir une représentation cohérente des données recueillies.
5. À ce stade, il a été supposé que le « dénominateur commun » utilisé pour combiner les sources de données serait le numéro d'identification personnel (PESEL), car il s'agit du principal identifiant auquel la plupart des sources administratives ont recours.
6. Le tableau 1 présente les sources administratives sélectionnées en vue de la création de la liste de la population visée.

Tableau 1

Description des sources de données administratives exploitées pour élaborer les listes de la population visée

N°	Désignation des données	Fournisseur des données/source des données	Contenu
1	PESEL	Ministère des affaires numériques/Registre des numéros d'identification personnels	Personnes inscrites comme résidents à titre temporaire ou à titre permanent en Pologne
2	KEP	Ministère des finances/Liste nationale des contribuables	Personnes physiques (contribuables) Personnes physiques exerçant une activité économique (contribuables)
3	ZUS	Institut polonais de sécurité sociale/Registre central des assurés	Personnes physiques (assurés)
		Institut polonais de sécurité sociale/Registre central des contribuables	Personnes physiques (contribuables)
4	KRUS	Caisse d'assurance sociale des agriculteurs	Agriculteurs, conjoints et membres du ménage qui travaillent avec eux (assurés)

N°	Désignation des données	Fournisseur des données/source des données	Contenu
		Caisse d'assurance sociale des agriculteurs	Personnes bénéficiant de pensions vieillesse et de pensions d'invalidité de la KRUS pour les agriculteurs
5	NFZ	Fonds national de santé	Personnes physiques (bénéficiaires)
6	ARIMR	Agence pour la restructuration et la modernisation de l'agriculture	Personnes physiques (agriculteurs)
7	PESEL_ WYMELD	Ministère des affaires numériques/Registre des numéros d'identification personnels	Personnes qui ont été rayées du registre sans destination connue et qui, jusque-là, avaient un lieu de résidence permanent en Pologne.

A. Création d'un ensemble de données concernant les personnes dotées de numéros d'identification personnels uniques (numéro PESEL)

7. La première étape de l'élaboration de la liste de la population visée consiste à collecter et à fusionner les numéros PESEL uniques des registres administratifs repris dans le tableau 1. Tous les numéros d'identification personnels attribués sont soumis à un contrôle d'exactitude.

8. La validation du numéro PESEL consiste à vérifier le nombre de caractères de l'identifiant, la conformité du chiffre de contrôle et la possibilité d'occurrence de la date de naissance dans le calendrier codé dans le numéro d'identification. L'algorithme de validation des numéros PESEL intègre la vérification automatique des dates écoulées et à venir, en tenant compte des années bissextiles.

9. Les variables sexe et âge doivent être prises en compte dans la liste des personnes titulaires d'un numéro d'identification personnel en raison de la nécessité d'une validation supplémentaire en vue de son élaboration.

B. Contrôle de groupes donnés

10. La deuxième étape de l'élaboration de la liste concerne le contrôle des groupes de personnes suivants.

1. Personnes décédées

11. La présence, dans la liste, de personnes décédées est contrôlée sur la base des informations reçues pour l'année en cours et l'année précédente. Cette procédure permet d'identifier clairement toutes les personnes décédées figurant sur la liste.

2. Personnes âgées de 90 ans et plus

12. Les personnes âgées de 90 ans et plus font l'objet d'une analyse (observation d'un ensemble de données issues des dossiers de personnes dont le décès a peu de chances d'être confirmé par un certificat de décès) fondée sur la variable WIEK calculée. Pour que le contrôle s'effectue en bonne et due forme, le numéro d'identification personnel (PESEL) de ces personnes doit apparaître dans les ensembles de données PESEL ou WYMELD et dans au moins un des autres ensembles de données.

3. Personnes qui ont quitté la Pologne

13. Pour identifier ces personnes, on s'appuie sur une catégorie de données obtenues dans le cadre de travaux de recherche en cours : les résidents permanents d'autres pays,

enregistrés en Pologne dans le cadre d'un séjour temporaire, qui ont signalé quitter le territoire ou dont la durée de séjour déclarée a expiré.

4. Résidents et bénéficiaires d'une pension en dehors de la Pologne

14. Les codes d'assurance du Registre central des assurés permettent d'identifier les personnes résidant et bénéficiant de prestations dans un État membre de l'Union européenne autre que la Pologne ou dans un État membre de l'Association européenne de libre-échange (AELE). Ces cas sont examinés selon que ces personnes séjournent ou non sur le territoire polonais.

5. Personnes n'apparaissant que dans un seul registre

15. Personnes pour lesquelles le numéro d'identification personnel (ou une partie de celui-ci) ne figure que dans un seul registre.

6. Personnes possédant une adresse de résidence en dehors de la Pologne

16. Sur la base des données de la Liste nationale des contribuables, les personnes possédant une adresse à l'étranger sont marquées au moyen du code de leur pays de résidence.

17. Le contrôle des différents groupes figurant sur la liste de la population visée conduit au marquage des cas examinés (points 1 à 6 ci-dessus), ce qui permet l'identification des dossiers, facilite leur vérification approfondie dans l'inventaire et permet d'apporter les corrections nécessaires dans le processus d'estimation de la population.

III. Association des données personnelles et détermination des adresses

18. La troisième étape de l'élaboration d'une liste de la population visée consiste à ajouter les noms et prénoms des personnes et à déterminer leur adresse. Il s'agit d'une étape importante de l'élaboration de la liste complète nécessaire au recensement, eu égard à la personnalisation des données et à la réalisation du recensement au moyen d'un questionnaire de recensement (authentification). En outre, les données personnelles comptent au nombre des différents dénominateurs communs possibles.

A. Nom et prénom

19. L'attribution du nom et du prénom est précédée d'une recherche d'occurrences de noms et de prénoms différents associés à un même numéro d'identification personnel ; cela inclut la recherche de valeurs itératives dans des registres administratifs isolés et de différences entre registres. Il arrive que des registres isolés comportent ce genre d'erreurs, principalement des erreurs dans des lettres individuelles, qu'il est facile de contrôler par recoupement avec un autre registre.

B. Adresse

20. Établir l'adresse consiste à attribuer des adresses aux personnes figurant sur la liste de la population visée et à vérifier l'exactitude des résultats et des algorithmes relatifs à la sélection de l'adresse de résidence.

21. Pour déterminer les adresses des personnes figurant sur la liste, quatre types d'adresses sont pris en compte : résidence permanente, résidence temporaire, adresse légale et adresse de correspondance, collectées dans les registres figurant dans le tableau 2.

Tableau 2
Ordre dans lequel les adresses sont attribuées

<i>N°</i>	<i>Source</i>	<i>Type d'adresse</i>
1	KEP	adresse de résidence (adresse légale pour la dernière étape)
2	ZUS	adresse de résidence (adresse légale, adresse de correspondance pour la dernière étape)
3	PESEL	adresse temporaire
4	PESEL	adresse permanente
5	KRUS	adresse de résidence (adresse de correspondance pour la dernière étape)
6	ARiMR	adresse de résidence (adresse de correspondance pour la dernière étape)
7	NFZ	adresse de résidence

22. La pertinence des données et le recoupement des adresses déterminent l'ordre de sélection des adresses à partir d'un ensemble de données spécifique. Le degré de cohérence est pris en compte pour toutes les adresses.

Tableau 3
Degré de cohérence des adresses

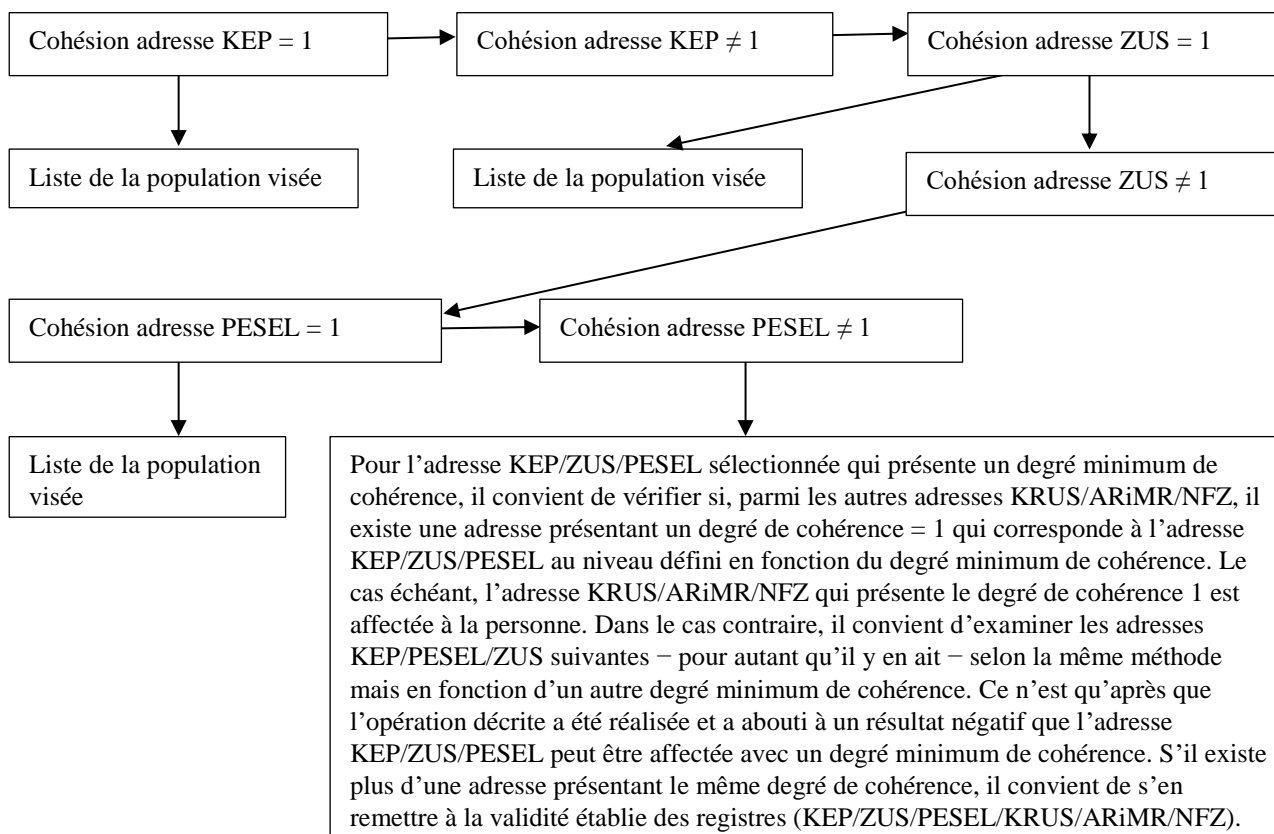
<i>N°</i>	<i>Degré de cohérence</i>
1	cohérence au niveau de la rue
2	cohérence au niveau du lieu (ville)
3	cohérence au niveau de la commune
4	absence de cohérence

23. La procédure d'attribution des adresses selon l'algorithme a été divisée en quatre étapes :

- a) Premièrement, sélectionner la cohérence de l'adresse au niveau de la rue à partir des ensembles de données KEP, ZUS et PESEL successivement ;
- b) Deuxièmement, à partir des ensembles de données KEP, ZUS et PESEL, sélectionner les adresses présentant un degré minimum de cohérence (autre que 1) et les corriger/compléter sur la base de l'adresse issue des ensembles de données KRUS, ARiMR et NFZ qui recoupe l'adresse KEP, ZUS et PESEL, au niveau défini par le degré minimum de cohérence. Les adresses pour lesquelles il n'a pas été possible de trouver un dossier de résidence plus complet sont incluses dans la liste de la population visée ;
- c) Troisièmement, attribuer une adresse présentant un degré minimum de cohérence à partir des ensembles de données KRUS, ARiMR, et NFZ ;
- d) Quatrièmement, ajouter les adresses dites complémentaires.

Graphique 1

Schéma simplifié de l'algorithme de sélection de l'adresse



IV. Conclusion

24. La méthode présentée pour l'élaboration d'une liste de la population visée uniquement à partir de données issues de sources administratives permettra de couvrir l'ensemble de la population appelée à faire l'objet du recensement. Il s'agit d'une procédure entièrement automatique et reproductible qui permet une pleine intégration dans le cadre d'autres mécanismes et garantit la qualité des données.