



Conseil économique et social

Distr. générale
27 mars 2012
Français
Original: anglais

Commission économique pour l'Europe

Conférence des statisticiens européens

Soixantième réunion plénière

Paris, 6-8 juin 2012

Point 3 a) de l'ordre du jour provisoire

Cycle de recensements de 2010 – innovations et enseignements tirés

Difficultés liées à la transition des recensements traditionnels aux recensements à partir des registres en Autriche

Note de l'Office national des statistiques de l'Autriche

Résumé

Le présent document renferme un aperçu des temps forts de la transition du recensement classique au recensement basé sur les registres en Autriche, ainsi que la description d'une approche structurelle de l'évaluation de la qualité des données administratives. Ce processus en trois étapes (données brutes, données combinées et données imputées) permet de produire des indicateurs de qualité qui devraient couvrir l'ensemble des informations disponibles relatives à la qualité pour chaque attribut. Afin de garantir l'applicabilité du cadre d'évaluation de la qualité au recensement de 2011, réalisé à partir des registres, la procédure a été expérimentée sur les statistiques relatives au marché du travail produites en 2009.

Cette approche offre pour avantage de permettre une déconnexion entre évaluation de la qualité et traitement des données, rendant du même coup possible l'application des mêmes méthodes à d'autres ensembles de données produits à partir des registres. De plus, la coopération entre l'Office national des statistiques et les autorités dépositaires des registres se trouve renforcée par le fait que le dépositaire des données est intégré au processus d'évaluation de la qualité.

I. La transition du recensement traditionnel au recensement basé sur les registres en Autriche

1. En mai 2001, le dernier recensement traditionnel a été réalisé en Autriche, de pair avec un recensement des bâtiments et des logements et un recensement des unités locales d'emploi. Ce recensement combiné, réalisé à partir de questionnaires sur papier, a porté sur environ 2 millions de bâtiments, 3,8 millions de logements et 3,3 millions de ménages, pour un total de 8,1 millions de personnes. Le recensement traditionnel représentait donc un effort complexe et coûteux.

2. Au regard des exigences considérables inhérentes aux méthodes traditionnelles de recensement, l'importance des sources de données administratives dans le contexte des statistiques ne cesse de se renforcer depuis quelque temps. Le traitement des données déjà enregistrées par les administrations offre de nombreux avantages par rapport aux données d'enquête (réduction des coûts, suppression de la charge de travail pour les répondants et disponibilité rapide des données). L'utilisation des données tirées de registres existants devrait permettre de refléter pleinement la réalité à un coût raisonnable en combinant les registres par des variables de liaison uniques, d'améliorer la qualité des données et d'harmoniser les définitions. Un autre avantage évident découle du fait que les données des registres sont régulièrement actualisées de façon à prendre en compte le moindre changement intervenant dans les données décrivant les données et leurs attributs (Statistique Finlande, 2004, p. 10).

3. Ainsi, les offices nationaux de statistique sont de plus en plus nombreux à promouvoir le remplacement des recensements traditionnels, coûteux, par des recensements basés sur les registres. En 2000, le Conseil des ministres autrichiens a décidé de mettre en place la nouvelle méthode pour le recensement de 2011, et en 2006, les autorités judiciaires ont prononcé la mise en application du cadre réglementaire. L'Autriche est, avec la Suède, la Finlande, la Norvège, le Danemark et la Slovaquie, un des six pays européens ayant réalisé un recensement basé sur les registres en 2011.

4. Le dénombrement de la population réalisé lors de ce recensement est utilisé pour déterminer les sommes versées dans le cadre du système de péréquation fiscale entre les municipalités et le budget fédéral et le nombre d'électeurs potentiels en vue des prochaines élections. Par ailleurs, les renseignements relatifs aux navetteurs, à l'éducation et à l'emploi offrent des éléments de réflexion importants pour les politiques économique et sociale. Toutefois, les données administratives peuvent reposer sur des définitions différentes qui ne répondent pas aux besoins des offices nationaux de statistique, alors même que les données sont de qualité (voir Nations Unies, 2007, p. 3). Les offices nationaux de statistique doivent par conséquent déterminer si les données conviennent au problème considéré.

A. Le principe de redondance

5. Le choix des sources de données appropriées pour produire l'information requise constitue un problème majeur. Le recensement basé sur les registres vise à couvrir l'ensemble des variables pertinentes qui étaient obtenues par la méthode traditionnelle auparavant. À cet égard, le recensement de 2001 a marqué le début de la création de certaines sources de données, telles que le registre central de la population, le registre des logements et des bâtiments ou le registre des performances éducatives. Avant 2001, l'Autriche ne disposait pas de réseau interconnecté de registres sur la population. Chaque municipalité administrait ses propres registres, et les données n'étaient généralement pas saisies électroniquement. Avec la création d'un registre des logements, en 2001, le registre centralisé de la population a évolué, et les municipalités ont dû communiquer leurs

registres. Le dernier recensement traditionnel en Autriche a été réalisé sur la base de l'information obtenue à partir de ce nouveau registre de la population. En 2004, le registre des logements et des habitations a pour la première fois été synchronisé avec le registre de la population. De plus, le registre de la performance éducative a été établi au cours du processus de recensement de 2001.

6. La figure 1 dépeint l'ensemble des registres de référence utilisés aux fins du recensement, ainsi que les liens entre ces registres et les thèmes respectifs. Les sources de données indiquées en rouge sont administrées par l'Office national des statistiques, et les autres sont fournies par des détenteurs de données externes, tels que le service pour l'emploi ou la sécurité sociale. Le registre central de la population représente l'élément principal du recensement, car les unités analysées sont les individus et les résidences principales en Autriche. Pour garantir la qualité des résultats du recensement, les registres de référence sont étayés par sept registres aux fins de comparaison. Ces sept champs d'unités administratives sont soumis par 35 détenteurs de données et sont principalement utilisés pour des vérifications croisées et pour obtenir des renseignements partiellement ou totalement indisponibles dans les registres de référence (Berka *et al.*, 2010, p. 300).

7. Compte tenu de l'autonomie des différents registres et des processus de collecte de données, les sources renferment parfois des valeurs contradictoires pour une même variable. C'est pourquoi le principe de redondance est appliqué pour garantir un niveau de qualité suffisant, l'idée étant d'obtenir à partir d'autant de registres que possible des renseignements concernant le sexe, la nationalité ou l'âge. Une méthode particulière, élaborée par l'Office national des statistiques, vise à définir un registre de référence spécifique pour obtenir l'information requise pour une variable donnée, les registres aux fins de comparaison étant utilisés pour confirmer les valeurs des registres de référence (Lenk, 2008, p. 3). Si les registres renferment des données suffisantes concernant la plupart des caractéristiques requises pour les recensements traditionnels, certaines variables ne peuvent néanmoins pas être intégrées dans les recensements basés sur les registres. Par exemple, la durée des trajets pour se rendre au travail, les langues parlées ou la religion ne figurent dans aucun des registres utilisés.

B. Protection de la confidentialité des données

8. Les noms et numéro de sécurité sociale des individus étant absents des données communiquées, il est nécessaire de disposer d'un numéro d'identification unique pour fusionner les données provenant de sources différentes. Pour garantir la confidentialité des données, il est nécessaire d'entrer un numéro personnel d'identification spécifique à chaque administration aux fins des statistiques officielles avant de pouvoir communiquer les données. Chaque administration en Autriche (sécurité sociale, services fiscaux ou protection sociale) dispose de ses propres numéros personnels d'identification. Ces numéros constitués de 172 chiffres, qui servent à garantir la confidentialité des données lors de la communication électronique entre les différentes administrations publiques, sont produits par la Commission autrichienne de protection des données à partir du registre source des numéros personnels d'identification (Stammzahlregister).

9. La figure 2 montre comment la protection de la confidentialité des données est assurée par l'obtention de numéros d'identification spécifiques par administration et de numéros individuels d'identification aux fins des statistiques officielles. Les détenteurs des données demandent ces numéros individuels d'identification à la Commission de protection des données pour chaque personne en communiquant le nom, le sexe, la date de naissance, le lieu de naissance et l'adresse de l'individu concerné. Le numéro individuel d'identification spécifique à l'administration et le numéro personnel d'identification aux fins des statistiques officielles sont obtenus à partir du registre source des numéros

d'identification à l'aide d'un algorithme spécial très complexe mis au point par la Commission de protection des données. Les numéros individuels d'identification aux fins des statistiques officielles ne sont communiqués que sous forme cryptée aux détenteurs de registres autres que l'Office national des statistiques. Ces derniers doivent envoyer les données accompagnées de ce numéro encrypté à l'Office national des statistiques. De plus, le détenteur du registre fournit son propre numéro d'identification crypté pour chaque individu, afin de pouvoir identifier les données correspondantes en cas de nouvelle demande de l'Office national des statistiques (Lenk, 2008, p. 5).

10. Chaque détenteur de données est en mesure de décrypter ses propres numéros individuels d'identification et d'obtenir un numéro à 28 chiffres, de sorte que seul l'Office national des statistiques est en mesure de décrypter les numéros individuels d'identifications aux fins de statistiques et de les utiliser comme variable commune de liaison pour les nouvelles données de registres. La procédure de cryptage des données personnelles a été utilisée pour la première fois en 2006 lors d'un recensement expérimental.

Figure 1
Registres et thèmes du recensement basé sur les registres en Autriche

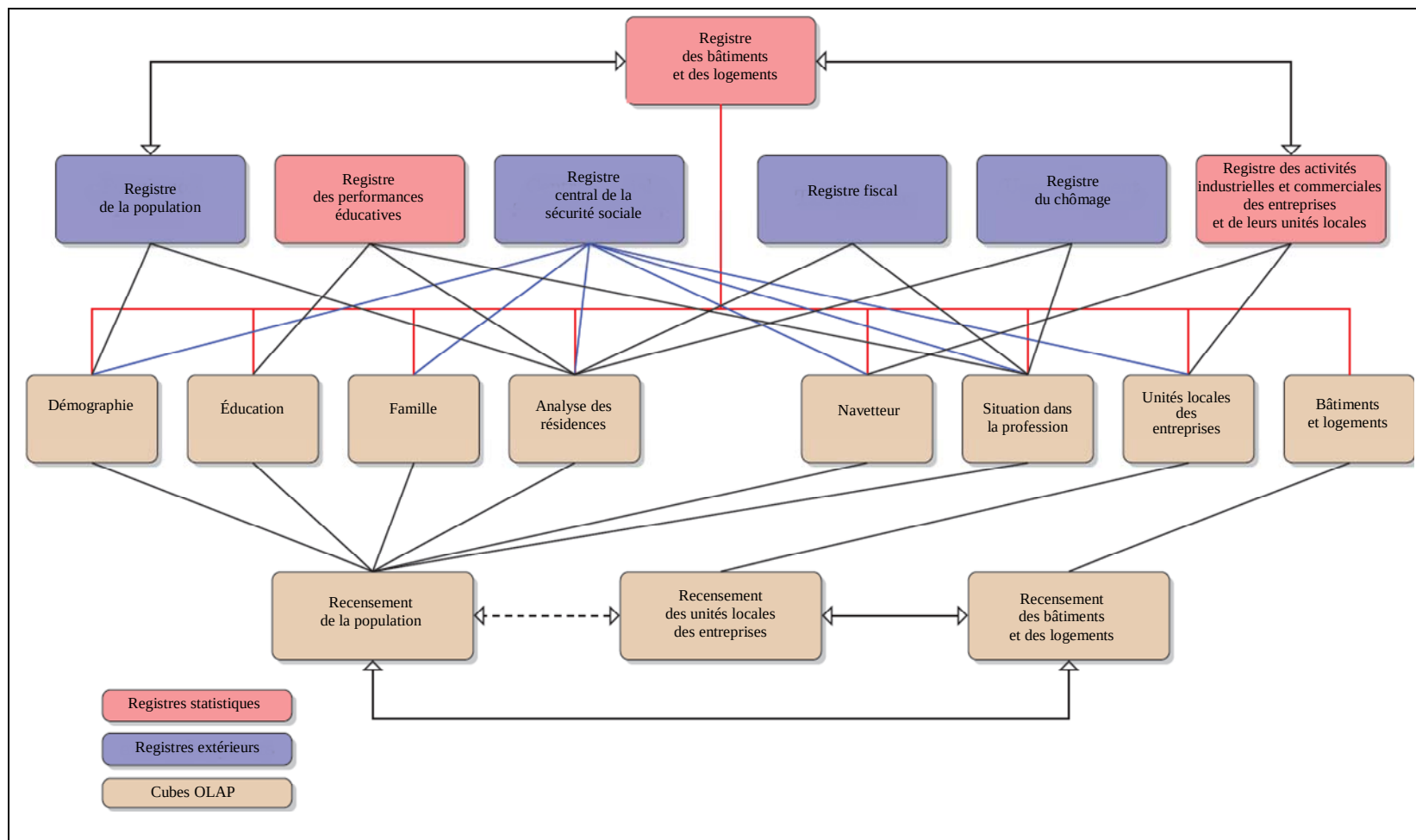
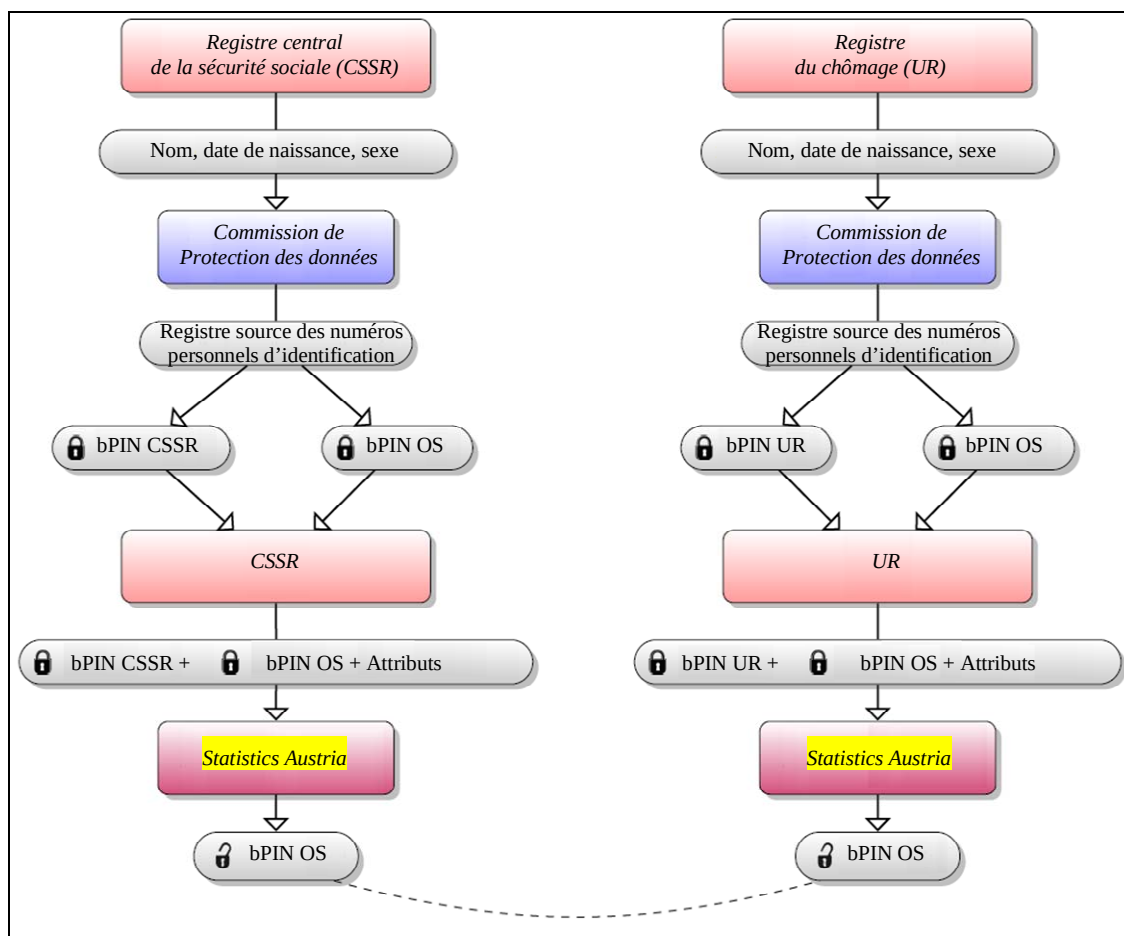


Figure 2

Protection de la confidentialité des données dans le cadre du recensement basé sur les registres

bPIN CSSR = Numéro individuel d'identification aux fins du CSSR.

bPIN OS = Numéro individuel d'identification aux fins des statistiques officielles.

bPIN UR = Numéro individuel d'identification aux fins de l'UR.

II. Mise en œuvre des recensements préliminaires

A. Le recensement expérimental de 2006

11. En 2006, l'Office national des statistiques a réalisé un recensement expérimental à l'échelle d'un recensement complet de la population, des bâtiments et des logements, ainsi que des unités d'emploi locales, afin d'évaluer la transition du recensement traditionnel au recensement basé sur les registres. Pour la première fois, les divers détenteurs de données ont été invités à communiquer des renseignements afin de produire des statistiques administratives combinées. Malgré des problèmes mineurs en ce qui concerne les sources de données, le recensement expérimental a été un succès d'ensemble. En particulier, les données relatives à la démographie, à l'éducation et à l'emploi se sont révélées être d'excellente qualité et d'une grande plausibilité, alors que des problèmes se sont posés pour identifier l'emplacement du lieu de travail des navetteurs et pour associer les individus aux logements (Statistics Austria, 2009, p. 7). L'Office national des statistiques a validé la

qualité des résultats en les comparant avec une enquête par échantillonnage pour les mêmes dates de référence (Lenk, 2008, p. 14).

12. En 2008, l'entrée en application de la loi de péréquation a marqué un tournant décisif dans l'histoire des recensements en Autriche. En vertu de cette loi, les dotations budgétaires fédérales aux municipalités ont été déterminées sur la base du dénombrement réalisé dans le cadre du recensement expérimental basé sur les registres, ce recensement, qui n'était initialement pas destiné à être contraignant, ayant finalement acquis la valeur d'un instrument politique pertinent.

B. Analyse des résidences

13. Pour garantir la qualité du recensement expérimental amélioré, il était important de détecter les valeurs manquantes, les entrées multiples ou les erreurs de mesure, et de les éliminer dans l'optique du dénombrement. Cette procédure, connue sous le nom d'analyse des logements, énumère cinq raisons d'exclure des individus du recensement:

- a) La personne est décédée avant la date de référence, mais elle apparaît encore dans le registre central de la population;
- b) La personne apparaît plusieurs fois dans le registre central de la population, à la date de référence, avec, par exemple, deux résidences principales ou davantage;
- c) La personne a été enregistrée moins de quatre-vingt-dix jours avec une résidence principale en Autriche, ce qui établit un lien entre le nombre d'habitants et une durée de séjour minimale;
- d) La détection de ce qu'on appelle le «tourisme de recensement», c'est-à-dire des personnes qui sont enregistrées dans une municipalité pendant une durée n'excédant pas cent quatre-vingts jours, mais dans une autre municipalité avant et après la période de référence;
- e) Les écarts possibles en l'absence de résidence principale, c'est-à-dire les personnes qui n'apparaissent que dans le registre central de la population et dans aucune autre source administrative.

14. Le dernier cas de figure appelle des précisions de la part de l'Office national des statistiques, car les signes de vie ne sont pas satisfaisants (voir le tableau 1). Les individus qui n'apparaissent que dans le registre central de la population requièrent des précisions, et ils reçoivent une lettre à cet effet¹, ce qui serait le cas pour les ID3459 dans cet exemple. À la suite du recensement expérimental, quelque 45 000 lettres ont été envoyées aux personnes dont la situation n'était pas claire, et 9 000 des personnes concernées ont indiqué avoir leur résidence principale en Autriche. Enfin, environ 0,5 % de la population n'ont pas été dénombrés, dont environ 10 000 suppressions pour raisons techniques (voir les paragraphes 1 à 3). Les municipalités autrichiennes doivent être informées des individus non dénombrés, ce qui permet aux services d'enregistrement de prouver la résidence des personnes en question, voire de les supprimer du registre des résidences. En raison des résultats du recensement expérimental de 2006, environ 80 % des cas non dénombrés ont été éliminés des registres des résidences par les municipalités.

15. Pour le recensement de 2011 basé sur les registres, l'Office national des statistiques applique la même procédure que lors du recensement expérimental de 2006. L'analyse des résidences a commencé en janvier 2012. Dans un premier temps, environ 54 000 ont été

¹ La lettre, qui est adressée chaque fois qu'il y a des imprécisions, ne renferme qu'une question: «Avez-vous eu une résidence principale en Autriche à la date de référence? Oui ou non.»

envoyées pour préciser les résidences principales. La deuxième phase est prévue pour septembre 2012, c'est-à-dire après achèvement de la transmission des données par les détenteurs externes.

Tableau 1

Signes de vie dans les analyses des résidences

bPIN OS	CPR	CSSR	TR	UR	SWR	CAR ...
⋮	⋮	⋮	⋮	⋮	⋮	⋮
ID3457	✓	✓	✓	⚡	⚡	⚡
ID3458	✓	✓	⚡	✓	⚡	⚡
ID3459	✓	⚡	⚡	⚡	⚡	⚡
ID3460	✓	⚡	⚡	⚡	✓	⚡
⋮	⋮	⋮	⋮	⋮	⋮	⋮

CPR = Registre central de la population, CSSR = Registre central de la sécurité sociale, TR = Registre fiscal, UR = Registre du chômage, SWR = Registre des bénéficiaires de la protection sociale, CAR = Registre des allocations familiales.

C. Statistiques sur le marché du travail basées sur les registres

16. Les statistiques annuelles du marché du travail basées sur les registres offrent des possibilités intéressantes pour les statistiques en temps réel. À partir de 2009, les données relatives à la démographie, à l'éducation, aux navetteurs et aux marchés du travail sont produites annuellement et permettent d'obtenir des informations de fond, y compris au niveau des municipalités, sur les questions économiques et sociales en Autriche. Par exemple, les demandes fréquentes de données administratives permettent d'analyser en profondeur l'évolution du chômage des jeunes entre deux années spécifiques ou l'évolution des différences sexospécifiques en matière de performances éducatives. Les sources de données utilisées et les processus employés pour produire ces attributs sont identiques aux méthodes de recensement basées sur les registres. Quelque 35 sources de données administratives et statistiques sont reliées pour obtenir les variables souhaitées. Avec la mise en œuvre des statistiques du marché du travail basées sur les registres, l'Office national des statistiques répond aux normes contemporaines relatives aux statistiques et à l'analyse des données. Les données sont même rendues publiques via la base de statistiques sur le site Web de l'Office national des statistiques (www.statistik.at).

III. Considérations relatives à un cadre d'évaluation de la qualité

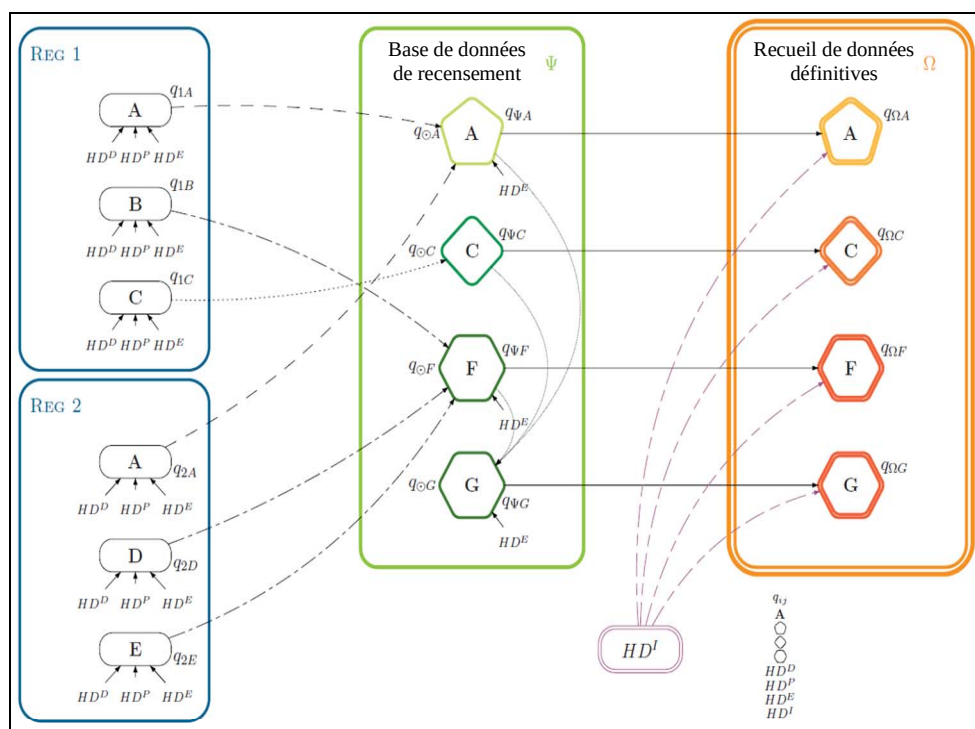
17. La brièveté de la période de transition du recensement traditionnel au recensement basé sur les registres a été une tâche exigeante. La période transitoire au cours de laquelle les données d'enquête ont peu à peu été remplacées par les données administratives a duré une vingtaine d'années dans certains pays européens (voir Ruotsalainen (2008) pour l'exemple finlandais). Les offices nationaux de statistique et les détenteurs de données ont ainsi eu tout le loisir d'examiner en profondeur la question de l'évaluation de la qualité dans certains pays, alors que le calendrier de la transition a été particulièrement ambitieux

en Autriche. L'Office national des statistiques n'étant pas responsable de la gestion des données externes, il convient d'accorder une attention particulière à l'évaluation de la qualité dans le contexte des recensements basés sur les registres. L'analyse de la qualité des données issues des registres doit satisfaire à plusieurs conditions, telles que la transparence, la précision et la faisabilité. Notre approche contribue à la création d'un cadre qualité pour l'analyse des données administratives, qui repose sur l'utilisation de trois hyperdimensions différentes pour l'obtention des indices de qualité (voir Berka *et al.*, 2012). Ce cadre est étroitement lié au flux de données, mais se distingue du traitement des données, ce qui permet de faire en sorte que le traitement ne soit pas influencé, mais évalué.

18. Le flux des données des recensements basés sur les registres en Autriche comporte trois niveaux: les données brutes (c'est-à-dire les registres), les données combinées (base de données de recensement) et les données imputées (données finales). La figure 3 dépeint le traitement des données, qui commence par la réception des données brutes soumises par les diverses administrations. Reliée via le numéro individuel unique d'identification aux fins de statistiques officielles, l'information est fusionnée en cubes dans la base de données de recensement, laquelle est ensuite enrichie des imputations de non-réponses venant compléter les données finales renfermant les données effectives et les estimations.

Figure 3

Cadre qualité pour les recensements basés sur les registres



A. Évaluation de la qualité au niveau des registres

19. L'information concernant la qualité des données brutes est obtenue à travers trois hyperdimensions: documentation, prétraitement et source externe. Avant de s'intéresser aux données, la documentation passe en revue les processus relatifs à la qualité mis en place par les services en charge des registres, ainsi que la documentation des données (métadonnées). Les niveaux de confiance et de fiabilité des détenteurs de données sont examinés à l'aide d'un questionnaire comprenant 16 questions ouvertes et 9 questions à échelle (par exemple,

fins administratives ou traitement des données). L'Office national des statistiques est donc en mesure de vérifier les méthodes de collecte des données ou les dispositions juridiques applicables à leur enregistrement qui risquent d'avoir une incidence importante sur leur qualité. Les réponses aux questionnaires, qui sont produites par les experts des administrations concernées, doivent apporter des résultats suffisamment probants.

20. La deuxième hyperdimension, le prétraitement, s'intéresse aux erreurs de forme des données brutes. Les écarts, les non-réponses et les clefs primaires manquantes sont détectés à ce stade de l'évaluation de la qualité. Le résultat final de cette hyperdimension est constitué du rapport des entrées utilisables au nombre total d'entrées. Cette procédure est, elle aussi, exécutée pour chaque attribut de chaque registre. Si la proportion d'entrées utilisables pour un attribut d'un registre donné est inférieure à celle du même attribut pour un autre registre, la qualité mesurée sera elle aussi inférieure.

21. Enfin, la troisième hyperdimension (source externe) consiste à établir une comparaison entre des données issues de registres et une source externe. En Autriche, le microrecensement est fréquemment utilisé pour les enquêtes représentatives, et il est considéré comme le meilleur ensemble de données comparatives disponible. La vérification de la cohérence avec la source externe forme une troisième mesure de la qualité, constituée du rapport du nombre de valeurs concordantes au nombre total d'entrées reliées. Si l'attribut n'est pas couvert par l'enquête de référence, nous nous fions à l'avis des experts locaux.

22. À partir de ces trois mesures de la qualité, un indicateur global de qualité pour chaque attribut et pour chaque registre peut être obtenu par calcul d'une moyenne pondérée. Cet indicateur permet par conséquent de déterminer les effets sur la qualité de la production de données aux données finales brutes contenues dans les registres.

B. Évaluation de la qualité de la base de données de recensement

23. Toute l'information issue des registres est combinée dans la base de données de recensement, qui couvre l'ensemble des attributs requis aux fins du recensement. Un même attribut étant susceptible d'être fourni par plusieurs sources de données, un ensemble de règles prédéfinies par l'Office national des statistiques permet de recueillir l'information appropriée dans les registres considérés. S'agissant de l'évaluation de la qualité de la base de données du recensement, nous distinguons trois types d'attributs en fonction de leur origine:

a) Des attributs uniques existent dans un seul registre, à savoir celui du niveau d'éducation (voir l'attribut C dans la figure 3);

b) Des attributs multiples apparaissent dans plusieurs registres, par exemple dans le registre du sexe (voir l'attribut A dans la figure 3). Les informations provenant de sources multiples sont combinées par l'application d'un ensemble de règles, afin d'obtenir la valeur la plus appropriée pour l'attribut correspondant dans la base de données du recensement;

c) Les attributs obtenus sont créés sur la base d'attributs différents, par exemple l'emploi (voir les attributs F et G dans la figure 3). Les registres ne renferment aucune information correspondant à ces spécifications pour ces attributs.

24. Une description détaillée de l'évaluation de la qualité pour les trois types d'attributs de la base de données de recensement et proposée par Berka *et al.* (2010) et Berka *et al.* (2012).

25. Les recherches en cours portent sur l'Ensemble de données finales, qui correspond à la base de données de recensement après application des imputations. La quantité de non-

réponses est effectivement réduite par les imputations, mais le processus d'imputation lui-même doit être surveillé, une tâche réalisée en utilisant les informations obtenues de façon constante par l'hyperdimension «Imputation».

C. Conclusion et perspectives

26. Le présent document décrit les temps forts de la transition du recensement traditionnel au recensement basé sur les registres en Autriche. Il décrit par ailleurs une approche structurée pour l'évaluation de la qualité des données administratives. Un processus en trois étapes (données brutes, données combinées et données imputées) permet d'obtenir les indicateurs de qualité pour trois hyperdimensions. Ces mesures doivent couvrir l'ensemble des informations sur la qualité pour chaque attribut. Pour garantir l'applicabilité du cadre de qualité au recensement de 2011 basé sur les registres, la procédure est expérimentée en utilisant des données issues des statistiques sur le marché du travail de 2009, elles-mêmes produites sur la base des registres.

27. Le cadre qualité actuellement disponible offre pour avantage considérable de dissocier l'évaluation de la qualité et le traitement des données. La séparation de la procédure de traitement est requise pour évaluer le processus sans l'influencer, ce qui permet d'appliquer les mêmes méthodes aux autres ensembles de données basés sur les registres. Qui plus est, la coopération entre l'Office national des statistiques et les administrations en charge des registres est renforcée par le fait que l'administration est intégrée au processus d'évaluation de la qualité.

IV. Références

- Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., & Schwerer, E. (2010). A quality framework for statistics based on administrative data sources using the example of the Austrian census 2011. *Austrian Journal of Statistics*, volume 39, n° 4, 299-308.
- Berka, C., Humer, S., Lenk, M., Moser, M., Rechta, H., & Schwerer, E. (2012). Combination of evidence from multiple administrative data sources: quality assessment of the Austrian register-based census 2011. *Statistica Neerlandica*, volume 66, Issue 1, 18-33.
- Lenk, M. (2008). *Methods of Register-based Census in Austria* (Tech. Rep.). Statistics Austria, Wien.
- Ruotsalainen, K. (2008). *Finnish Register-based Census System* (Tech. Rep.). Statistics Finland.
- Statistics Austria (2009). *Bericht "über die Probezählung 2006 – Ergebnisse und Evaluierung"* (Tech. Rep.). Statistics Austria.
- Statistics Finland (2004). *Use of registers and administrative data sources for statistical purposes* (Tech. Rep.). Statistics Finland.
- Nations Unies (2007). *Register-based statistics in the Nordic countries*. Commission économique pour l'Europe, New York et Genève.