

Anonymisation des microdonnées du Recensement Général de la Population et de l'Habitat de 2014

Le Haut Commissariat au Plan (HCP) s'est engagé dans un processus d'ouverture des microdonnées issues de ses principales opérations statistiques, recensements et enquêtes aussi bien auprès des ménages que des entreprises.

Ce processus devrait s'accompagner de mesures adéquates permettant d'éviter tout risque de divulgation des données à caractère personnel que pourraient contenir ces microdonnées, et ce conformément aux "principes fondamentaux de la statistique officielle" des Nations Unies et aux dispositions du Décret Royal portant loi n° 370-67 du 10 Joumada I 1388 (5 août 1968) relatif aux études statistiques et notamment son article 8 et de la loi n°09-08 relative à "la protection des personnes physiques à l'égard du traitement des données à caractère personnel".

Pour cette raison, le HCP a initié un programme d'anonymisation des microdonnées issues de ses opérations statistiques avec une priorité accordée aux données du Recensement Général de la Population et de l'Habitat de 2014 (RGPH 2014) et de l'Enquête nationale sur la consommation et les dépenses des ménages de 2013-2014.

L'objectif de cette opération d'anonymisation est de produire des fichiers de microdonnées anonymisées portant sur les caractéristiques démographiques et socio-économiques relatives aux ménages et aux personnes observés lors de ces deux opérations d'envergure. Ces fichiers, une fois élaborés, seront mis, en ligne, à la disposition des différents utilisateurs.

Cette note présente un aperçu portant sur les aspects méthodologiques ayant servi pour anonymiser les microdonnées du RGPH 2014.

Il est à souligner que la diffusion des microdonnées de ce recensement, 4 ans après la réalisation de sa phase de collecte, constitue en soi une mesure de protection des données individuelles. En effet, plus on s'éloigne de la date de collecte des données, plus faible est la probabilité d'identifier les données

personnelles des unités observées. C'est une démarche adoptée par plusieurs institutions statistiques nationales en publiant un échantillon des données individuelles anonymisées issues de leurs recensements, en général, 3 à 5 ans après la date de collecte. C'est le cas notamment de Statistique Canada qui a publié les microdonnées du recensement de 2016 en 2019 (3 ans après), du Bureau du recensement des États-Unis (4 ans après le recensement de 2010), de l'Insee (3 ans après le recensement de 2015) et du Bureau de la statistique nationale de la Grande-Bretagne (3 ans après le recensement de 2011).

L'approche méthodologique adoptée par le HCP pour l'anonymisation des microdonnées du RGPH 2014 s'inspire des méthodes préconisées par la Banque mondiale et des bonnes pratiques internationales dans ce domaine, notamment celles convenues avec les institutions statistiques officielles partenaires.

Cette approche prévoit la mise en œuvre d'une série de mesures et de traitements visant à assurer l'équilibre entre la préservation de l'intégrité des données individuelles tout en réduisant au minimum le risque d'identification des données personnelles relatives aux unités observées lors du RGPH à savoir les ménages et les individus.

Les grandes lignes de l'approche méthodologique adoptée pour l'anonymisation des données du RGPH 2014 sont comme suit :

1. le fichier de microdonnées anonymisées du RGPH 2014 comporte un échantillon de 10% de l'ensemble des unités observées lors du dernier recensement à l'exception des ménages nomades, des personnes sans abri et de la population comptée à part. Cet échantillon est jugé largement suffisant pour assurer une représentativité aux niveaux national, régional et provincial avec une déclinaison des indicateurs selon les principales caractéristiques démographiques et socio-économiques (sexe, milieu de résidence, groupes d'âge, type de logement, etc.).
2. Les unités de tirage sont constituées des ménages. Le mode de sélection de ces unités obéit aux principes de tirage aléatoire stratifié où les critères de stratification sont les subdivisions administratives et géographiques (régions, provinces, communes et centres urbains) et le type de logement

occupé par le ménage. Pour chaque ménage sélectionné, tous les membres sont retenus dans l'échantillon, et ce pour permettre aux utilisateurs la possibilité d'analyser les structures intra-ménages.

3. la constitution d'une clé d'identification comportant l'ensemble des variables jugées permettant, une fois combinées, d'identifier les ménages et les individus recensés.
4. le calcul du risque d'identification initial, qui représente la probabilité de reconnaître les ménages ou les individus à partir de la clé d'identification.
5. Pour réduire au minimum ce risque d'identification et de divulgation des données personnelles des ménages et de la population, plusieurs mesures techniques ont été prises, dont:
 - la limitation du niveau de désagrégation géographique des données à diffuser aux niveaux régional et provincial.
 - l'application d'un seuil de diffusion de 1‰ de la population municipale (soit environ 30 000 habitants) pour réduire le risque d'identification des ménages et des personnes qui résident dans les provinces les moins peuplées.
 - la suppression des observations aberrantes permettant d'identifier avec une forte probabilité les ménages. cette mesure a concerné une proportion minimale de 2‰ des ménages.
 - le changement aléatoire des numéros d'ordre des ménages au sein de chaque préfecture ou province, pour perturber leur classement établi initialement selon l'ordre adopté par le code géographique national.
 - le recodage des modalités de réponse pour certaines variables pour dissimuler les cas rares et identifiants.
 - la suppression de certaines informations. Cette mesure a touché une proportion de 1,6‰ des informations relatives aux ménages et 0,6‰ de celles relatives aux individus.

- la modification de certaines données en utilisant la méthode PRAM (Post Randomization Method). Cette mesure a touché 0,4‰ des personnes.

L'opération d'anonymisation des données du RGPH 2014 a donné naissance à un fichier comportant les microdonnées anonymisées de 730 099 ménages et de 3 341 426 personnes portant sur 85 variables.

Ce fichier offre l'opportunité aux différents utilisateurs de confectionner des indicateurs et des tableaux statistiques personnalisés aux niveaux national, régional et provincial, notamment pour analyser les relations inter-variables que les informations contenues dans les bases de données déjà publiées par le HCP sur les résultats du RGPH 2014 ne couvrent pas.

Pour mettre les microdonnées anonymisées du RGPH 2014 à la disposition de tous les utilisateurs, ce fichier sera mis en ligne sur le site web du HCP (www.hcp.ma). Il sera accompagné d'une documentation technique présentant les métadonnées et les conditions d'utilisation pour en assurer un meilleur usage. La documentation relative aux métadonnées comporte notamment le dictionnaire des variables et des modalités ainsi que les concepts et définitions utilisés.