

**Commission économique pour l'Europe**

## Conférence des statisticiens européens

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

Genève, 2-4 octobre 2024

Point 2 l) de l'ordre du jour provisoire

**Révision des recommandations de la Conférence des statisticiens européens  
pour les recensements de la population et des habitations de 2030 :****Diffusion de données de recensement ventilées****Élaboration des recommandations relatives à la diffusion  
des données et à divers sujets connexes (sécurité,  
documentation, métadonnées et archivage)****Note de l'Équipe spéciale de la diffusion de données de recensement  
ventilées de la Conférence des statisticiens européens\****Résumé*

Le présent document contient le projet de chapitre consacré à la diffusion des données et à divers sujets connexes (sécurité, documentation, métadonnées et archivage) qui figurera dans les recommandations de la Conférence des statisticiens européens (CSE) pour les recensements de la population et des habitations de 2030 ainsi qu'un résumé des changements apportés par rapport aux Recommandations pour les recensements de 2020. L'objectif principal est de recueillir les observations et les propositions des experts nationaux sur ce projet afin qu'il soit tenu compte des besoins et des priorités des organismes nationaux de statistique.

\* L'Équipe spéciale de la diffusion de données de recensement ventilées se compose des membres suivants : Fabian Bach (Commission européenne, Eurostat (Président)), Glen Hohlmann (Canada), David Price (Canada), Bernie Casey (Irlande), Klara Shalev (Israël), Alejandra Silva (Commission économique pour l'Amérique latine et les Caraïbes de l'ONU), Baktybek Kainazarov (Fonds des Nations Unies pour la population, Kirghizistan), Andrew Smith (Division de statistique de l'ONU), Ala Negruta (Bureau régional d'ONU-Femmes pour l'Europe et l'Asie centrale), Lara Cleveland (IPUMS). Les opinions exprimées sont purement celles des auteurs et ne peuvent en aucun cas être considérées comme une position officielle de la Commission européenne, à laquelle appartient le Président de l'Équipe spéciale ayant élaboré le présent rapport.

NOTE : Les appellations employées dans le présent document ne reflètent aucune prise de position du Secrétariat de l'Organisation des Nations Unies quant au statut juridique de pays, territoires, villes ou zones quelconques, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.



## I. Introduction

1. Tous les dix ans, la Conférence des statisticiens européens (CSE) publie des recommandations pour aider les pays à effectuer les recensements de la population et des habitations. Ces recommandations sont élaborées par des équipes spéciales composées d'experts et supervisées par le Groupe directeur de la CSE sur les recensements de la population et des habitations.
2. Les travaux de l'Équipe spéciale de la diffusion de données de recensement ventilées s'articulent autour de trois axes thématiques (les aspects généraux de la diffusion, le contrôle de la divulgation des données statistiques et les résultats géoréférencés). À l'issue de l'enquête de la CEE sur les recensements de 2020, l'Équipe spéciale a organisé trois téléconférences et divers échanges écrits. Elle a analysé les résultats de l'enquête et a élaboré des propositions visant à mettre à jour les parties des recommandations de la CSE qui relèvent de son domaine de compétence.
3. Les principaux sujets abordés dans la proposition de mise à jour sont les suivants : la technologie et les outils de diffusion modernes (par exemple, les tableurs et les outils de visualisation des données) ainsi que les progrès réalisés récemment dans le domaine du contrôle de la divulgation des données statistiques (les méthodes basées sur le bruit et leur incidence sur les données liées au maillage ou les outils de diffusion flexibles, entre autres) et dans celui des unités géographiques fonctionnelles (notamment les mailles).
4. On trouvera dans la section II du présent document un résumé des modifications apportées au chapitre consacré à la diffusion des Recommandations pour les recensements de 2020 et dans la section III une vue d'ensemble du projet de chapitre.
5. La section IV présente le projet de chapitre consacré à la diffusion des données (proposé sous la forme d'un chapitre à part entière intitulé « Diffusion des données et sujets connexes : sécurité, documentation, métadonnées et archivage ») qui figurera dans les recommandations pour les recensements de la population et des habitations de 2030.
6. L'objectif principal est de recueillir les observations et les propositions des experts nationaux du recensement sur ce projet afin qu'il soit tenu compte des besoins et des priorités des organismes nationaux de statistique.

## II. Résumé des modifications introduites par rapport aux Recommandations de 2020

7. L'Équipe spéciale a révisé les sections ci-après des Recommandations de 2020 :
  - a) « Confidentialité et sécurité », paragraphes 140 à 148 du chapitre I (Méthodologie) ;
  - b) « Diffusion, documentation, métadonnées et archivage », paragraphes 292 à 318 du chapitre III (Action sur le terrain et autres activités opérationnelles).
8. Sur la base de son analyse des résultats de l'enquête de 2023, l'Équipe spéciale s'est concentrée sur trois principaux sujets qui devraient être abordés dans les recommandations de 2030 :
  - a) *Technologie et outils de diffusion modernes* : l'enquête a permis de confirmer que les modes électroniques de diffusion des résultats étaient de loin la méthode la plus répandue dans les pays de la Commission économique pour l'Europe (CEE). Les outils interactifs en ligne tels que les tableurs et les outils de cartographie élaborés à partir de systèmes d'information géographique (SIG) ne sont utilisés que depuis peu, tandis que les outils qui proposent des fonctions de visualisation des données faciles à utiliser et accessibles sont prisés par de nombreux pays, qui ne cessent d'innover dans ce domaine. L'Équipe spéciale a modifié le texte des Recommandations de 2020 à plusieurs endroits pour mieux rendre compte de cette tendance générale et mettre en évidence les divers avantages qui en découlent, tels que la satisfaction des utilisateurs et une efficacité accrue de la diffusion ;

b) *Progrès réalisés récemment dans le domaine du contrôle de la divulgation des données statistiques* : l'enquête a montré que de nombreux pays de la CEE ont procédé à d'importantes mises à niveau de leurs systèmes de contrôle de la divulgation des données statistiques, principalement en raison de l'apparition de méthodes plus efficaces basées sur le bruit, qui permettent notamment d'éviter la suppression de cellules, mais aussi de faire face aux nouveaux risques associés à l'utilisation de produits de diffusion plus précis et plus souples. Il est à noter que de nombreux pays de la CEE s'exposent à des risques supplémentaires en utilisant des outils de diffusion interactifs ayant directement accès aux microdonnées ainsi qu'au problème de la différenciation géographique (lié à l'adoption de mailles). L'Équipe spéciale a modifié le texte des Recommandations de 2020 à plusieurs endroits et propose d'ajouter de nouveaux paragraphes pour rendre compte de cette évolution et mettre en évidence les liens qui existent entre les technologies modernes de diffusion et les nouveaux risques de divulgation ;

c) *Progrès réalisés récemment dans le domaine des unités géographiques fonctionnelles* : la collecte d'informations géoréférencées précises (coordonnées) sur la population et les habitations est en passe de devenir la norme dans la région de la CEE, ce qui présente divers avantages en termes de diffusion, tels que la possibilité de choisir librement ou de personnaliser les unités géographiques de sortie. Aussi les mailles sont-elles devenues un niveau de découpage géographique essentiel dans la région de la CEE. Les utilisateurs, quant à eux, profitent d'outils de cartographie interactifs très flexibles. L'Équipe spéciale a modifié le texte des Recommandations de 2020 à plusieurs endroits pour rendre compte de cette évolution et mettre en évidence les principaux avantages qui en découlent en termes de comparabilité entre les pays et de satisfaction des utilisateurs.

9. Enfin, dans le cadre de la démarche générale entreprise par plusieurs équipes spéciales pour réorganiser de manière plus cohérente les différentes sections du chapitre III des Recommandations de 2020, l'Équipe spéciale propose de fusionner les sections susmentionnées en un nouveau chapitre intitulé « Diffusion et sujets connexes : sécurité, documentation, métadonnées et archivage ». Cela permettrait notamment de donner plus de visibilité à la confidentialité et à la sécurité et de regrouper les questions liées à la sécurité et à l'archivage des données confidentielles du recensement.

### III. Vue d'ensemble du chapitre

10. La diffusion des résultats est l'un des aspects cruciaux du recensement, car c'est l'opération qui permet de passer des fichiers de microdonnées complets compilés par les autorités chargées du recensement aux données avec lesquels les utilisateurs peuvent travailler. Ainsi, il ne s'agit pas seulement de la diffusion des produits finaux, mais aussi de la manière dont ces produits sont obtenus à partir des microdonnées complètes. Comme indiqué dans la section II, du fait de l'évolution récente des technologies de diffusion, les produits modernes présentent des caractéristiques (en termes d'interactivité, de flexibilité et de précision géographique) qui ont des incidences sur la sécurité des données personnelles. Le présent chapitre est donc consacré à la diffusion des résultats et à diverses questions connexes, à savoir la sécurité (notamment le contrôle de la divulgation des statistiques), les documents relatifs au recensement et aux résultats, les métadonnées accompagnant les résultats du recensement et l'archivage à long terme des données de recensement (accès aux relevés historiques des recensements). On trouvera ci-après une brève description des différentes sections.

11. La section « Confidentialité et sécurité » est subdivisée en deux parties :

a) La partie intitulée « Principes de confidentialité » présente les principes clefs régissant les mesures visant à garantir la sécurité physique et logique des données personnelles traitées dans le cadre d'une opération de recensement. Il s'agit notamment de placer la vie privée de l'individu au centre des préoccupations et de garantir au public que les informations personnelles traitées par les autorités chargées du recensement seront protégées à toutes les étapes et ne seront utilisées qu'aux fins du recensement statistique. Des mesures suffisantes doivent être prises en interne à cette fin, par exemple en ce qui concerne l'infrastructure informatique et la gestion de l'accès du personnel aux données ;

b) La partie intitulée « Contrôle de la divulgation des données statistiques » décrit le processus visant à protéger les données statistiques de façon à ce qu'elles puissent être publiées sans que des informations confidentielles concernant une personne ou une entité soient communiquées. L'objectif principal de toute méthode de contrôle de la divulgation des statistiques devrait être de garantir un niveau de protection suffisant tout en assurant une déperdition minimale d'informations et donc une utilité maximale des données. On trouvera dans cette partie une description des méthodes permettant de prévenir la divulgation des données tabulaires et des microdonnées ainsi qu'une présentation des autres risques de divulgation associés aux nouvelles technologies de diffusion.

12. La section intitulée « Diffusion » traite des aspects essentiels de la stratégie de diffusion et des produits qui en découlent, notamment en ce qui concerne :

a) Le principe de la diffusion simultanée à tous les utilisateurs, selon un calendrier établi ;

b) Les différentes manières de mettre les résultats d'un recensement à la disposition de l'utilisateur, par exemple sous la forme de rapports ou de tableaux statiques publiés, de produits interactifs (éventuellement accompagnés d'outils de visualisation des données ou de fonctions cartographiques), de produits fournis sur commande ou adaptés à des besoins particuliers ou de microdonnées ;

c) Le rapport coût-efficacité de la stratégie de diffusion, avec une description des gains d'efficacité qui pourraient être réalisés grâce à l'adoption de nouvelles solutions technologiques (par exemple, l'automatisation de méthodes de contrôle de la divulgation des statistiques et d'outils interactifs qui peuvent partiellement remplacer les services à la demande) ;

d) L'accès facile et gratuit aux produits, avec une description des normes techniques applicables aux interfaces et les canaux d'information modernes tels que les médias sociaux ;

e) Les progrès technologiques réalisés grâce à des outils de diffusion puissants tels que les SIG, qui facilitent la cartographie, et les outils de visualisation géographique.

13. La section intitulée « Documentation et métadonnées » présente des recommandations en ce qui concerne :

a) L'élaboration d'un dossier complet de documents de référence et de métadonnées expliquant les produits statistiques, précisant leur utilité et en facilitant l'utilisation ;

b) La mise en place d'un système de métadonnées bien structuré, qui repose sur des normes internationales tout en répondant aux besoins particuliers du pays ;

c) Les éléments standard d'un système de métadonnées ;

d) L'établissement de rapports méthodologiques lorsque la méthode de recensement utilisée a changé depuis le recensement précédent ;

e) La consultation des parties prenantes de divers horizons aux fins de l'amélioration des métadonnées.

14. La section intitulée « Archivage et accès à des données confidentielles du recensement » comprend notamment des recommandations sur :

a) Le recensement en tant que source de données unique en son genre, ce qui suppose de confier à l'autorité chargée du recensement l'importante responsabilité de préserver, pour l'avenir, cette photographie de la société à un instant donné ;

b) Les mesures à prendre concernant les finalités, les contextes et les responsabilités en matière de sécurité des données pour que les autorités chargées du recensement puissent rendre publiques les données de recensement après une période définie (qu'il est recommandé de fixer à au moins cent ans) ;

c) Lorsque les relevés du recensement sont archivés, l'importance qu'il y a à préserver l'ensemble des métadonnées et des documents exposant les procédures/opérations et d'y faciliter l'accès.

## **IV. Diffusion et sujets connexes : sécurité, documentation, métadonnées et archivage<sup>1</sup>**

### **A. Confidentialité et sécurité**

#### **1. Principes de confidentialité**

15. Le recensement permet de réunir des informations sur chaque personne et ménage du pays. Il peut prendre la forme d'un dénombrement direct, d'une compilation d'informations provenant de registres administratifs ou d'une combinaison entre la collecte de données provenant de registres et celle de données tirées d'enquêtes. Il ne s'agit pas de rechercher des détails particuliers à propos d'individus, mais plutôt d'obtenir des statistiques sur la collectivité dans son ensemble et sur des groupes qui en font partie. En conséquence, la population a le droit de s'attendre à ce que les informations personnelles fournies à titre confidentiel soient protégées, et elle doit être assurée qu'il en est bien ainsi. Les noms, adresses et numéros d'identification personnels doivent être séparés des autres données dès que possible et ne doivent pas être publiés afin que les données obtenues ne contiennent aucun élément d'identification personnelle. Le principe de confidentialité, généralement consacré par des lois relatives à la protection des données statistiques et personnelles, s'applique à l'ensemble des données de recensement, allant des renseignements personnels (tels que ceux figurant dans les questionnaires ou ceux provenant des registres de la population ou de sources de données administratives) aux informations produites et rendues publiques.

16. Il faut certifier à la population que toutes les informations personnelles (telles que les noms, les adresses complètes, les numéros d'identification personnels ou les clefs de couplage internes) seront traitées de manière strictement confidentielle par l'organisme de recensement et par toute personne qu'il emploie ou qui lui fournit des services aux fins du recensement. De nombreux pays sont dotés d'une législation visant à protéger ces informations, qu'elle se rapporte spécifiquement au recensement ou qu'elle porte plus généralement sur la confidentialité des statistiques ou la protection des données personnelles et la liberté de l'information.

17. Le traitement des informations, quelle que ce soit l'étape du recensement à laquelle elles ont été obtenues, doit obéir aux principes suivants :

a) Seules les personnes qui relèvent de l'organisme de recensement ou agissent en son nom doivent avoir accès aux informations personnelles obtenues dans le cadre du recensement ;

b) Il est essentiel de protéger la vie privée de l'individu et donc de traiter les informations personnelles obtenues dans le cadre du recensement de telle manière qu'elles ne soient pas révélées au public. En outre, les membres d'un ménage doivent pouvoir, s'ils le souhaitent, remplir chacun un questionnaire distinct afin que les informations personnelles qui y figurent ne soient pas révélées aux autres membres de leur ménage ou au recenseur. Dans les recensements réalisés à partir des données provenant de registres administratifs, les informations personnelles doivent être stockées séparément des clefs de couplage des pouvoirs publics ; elles doivent rester confidentielles et ne pas être divulguées au moment de la diffusion des résultats ;

<sup>1</sup> Les renvois internes à ce chapitre sont donnés selon la numérotation utilisée dans le présent document. Pour les renvois à d'autres chapitres, les numéros des paragraphes des Recommandations de 2020 sont indiqués entre crochets. Dans les deux cas, les numéros des paragraphes faisant l'objet d'un renvoi seront différents dans la version finale publiée des Recommandations de 2030.

c) Tous les fonctionnaires de l'organisme de recensement et les agents extérieurs qui lui fournissent des services liés au recensement doivent recevoir des instructions strictes concernant la confidentialité et signer un engagement juridiquement contraignant à ce sujet. Ils doivent être passibles de poursuites pour toute violation de la loi ;

d) La sécurité physique des documents ou des données numériques qui sont stockés aux fins du recensement et qui contiennent des informations personnelles détenues par l'organisme de recensement, le personnel de terrain ou des agents autorisés doit être assurée de façon rigoureuse et, si cela est jugé nécessaire, être vérifiée de façon indépendante ;

e) Les modes d'accès aux données personnelles issues du recensement et les modes de traitement de ces données, notamment les systèmes informatiques et les applications téléphoniques, doivent être strictement protégés pour empêcher tout accès non autorisé ;

f) Au moment de la publication de statistiques issues du recensement, il convient de prendre les mesures qui s'imposent pour empêcher que des informations sur des particuliers ou des ménages susceptibles d'être identifiés ne soient divulguées par inadvertance. Des précautions particulières peuvent être de rigueur en ce qui concerne les résultats statistiques relatifs à des zones de taille réduite.

## 2. Contrôle de la divulgation des statistiques

18. Le contrôle de la divulgation des statistiques vise à protéger les données statistiques de façon à ce qu'elles puissent être publiées sans que des informations confidentielles susceptibles d'être liées à des personnes physiques ou morales particulières soient communiquées. L'objectif principal de toute méthode de contrôle de la divulgation des statistiques devrait être de garantir un niveau de protection suffisant tout en assurant une déperdition minimale d'informations et donc une utilité maximale des données. Il existe une variété d'indicateurs types permettant d'évaluer systématiquement le rapport utilité-risque et de choisir la méthode de contrôle de la divulgation des statistiques qui convient le mieux au programme de diffusion des données du pays.

19. Les mesures visant à empêcher la divulgation des données tabulaires peuvent comprendre tout ou partie des actions suivantes :

a) La limitation du nombre de catégories dans lesquelles une variable peut être classée, par exemple par le recours à des tranches d'âge plutôt qu'à des âges précis, en particulier lorsqu'il s'agit d'âges avancés (« recodage global ») ;

b) Lorsque le nombre de personnes ou de ménages d'une zone se situe au-dessous d'un seuil déterminé, l'élimination des résultats statistiques (« suppression locale ») – sauf peut-être aux fins d'un dénombrement élémentaire – ou leur inclusion dans ceux d'une zone voisine suffisamment peuplée ;

c) L'ajout de « bruit » aux enregistrements de microdonnées avant la production des tableaux (« perturbation pré-tabulation »), par exemple en permutant certaines caractéristiques des enregistrements les plus confidentiels, en cherchant une correspondance dans les microdonnées à partir d'un ensemble de variables prédéterminées et en permutant tout ou partie des autres variables des enregistrements appariés (« permutation ciblée d'enregistrements ») ;

d) L'ajout de « bruit » aux tableaux produits (« perturbation post-tabulation »), par exemple en arrondissant les chiffres des cellules à la valeur supérieure ou inférieure du multiple le plus proche de la base prédéfinie (arrondissement classique) ou en ajoutant un bruit d'ampleur limitée de manière contrôlée et cohérente dans les tableaux (ajout de bruit contrôlé, par exemple avec la méthode des clefs de cellule).

20. En cas de divulgation de microdonnées provenant d'un recensement (comme les microdonnées issues de fichiers à usage public ou sous contrat), il importe de supprimer, dans la base de données, toutes les informations relatives au nom, à l'adresse et à des caractéristiques particulières susceptibles de permettre l'identification des répondants. Grâce aux ensembles de microdonnées à usage scientifique, il est possible de réaliser des analyses

bien plus complexes que celles que permettent les tableaux, les outils de visualisation et les indicateurs publiés. La constitution d'un échantillon de microdonnées, qui ne donne accès qu'à une fraction de la population totale, renforce la protection des données tout en préservant les informations sur les caractéristiques de la population. En outre, il est possible de recourir à un recodage global ou à une suppression locale des microdonnées pour diminuer le risque de divulgation. La perturbation des microdonnées ou l'utilisation ciblée de méthodes synthétiques peuvent aussi contribuer à la protection des informations confidentielles, mais ces méthodes doivent être utilisées de façon contrôlée et rigoureusement mises à l'essai, car elles peuvent réduire la précision et l'utilité des microdonnées.

21. À l'issue d'une évaluation utilité-risque rigoureuse, les autorités chargées du recensement peuvent conclure qu'il est préférable d'appliquer une combinaison de plusieurs mesures de protection à un programme de diffusion des données. En outre, il est important que la méthode de contrôle de la divulgation des statistiques retenue soit appliquée de manière cohérente à l'ensemble des résultats, car les incohérences entraînent généralement de nouveaux risques de divulgation. Certaines des méthodes basées sur le bruit mentionnées au paragraphe 19 ont été mises au point précisément pour compenser les lacunes des méthodes traditionnelles (perte excessive d'informations et manque de cohérence découlant d'une suppression locale dans les grands ensembles de tableaux).

22. Des outils interactifs sont de plus en plus utilisés pour la diffusion des statistiques (voir par. 26). Il est important d'y intégrer des procédures automatisées de contrôle. L'utilisation d'outils très souples qui permettent aux utilisateurs d'exécuter des requêtes hautement personnalisées directement à partir des microdonnées comporte des risques de divulgation supplémentaires, par exemple en cas d'attaques massives et systématiques par falsification de requête. Dans de tels cas, il est généralement nécessaire de mettre en place des méthodes ciblées de contrôle de la divulgation des statistiques.

23. Dans de nombreux pays, les recensements se caractérisent par un niveau élevé de détail géographique, comme en témoigne le nombre croissant de données diffusées au niveau des plus petites unités géographiques, qui sont désormais souvent complétées par des données maillées (voir par. [428 à 433]). Toutefois, la publication de données concernant des unités géographiques très petites et ne se chevauchant pas (limites administratives et limites déterminées par les mailles) entraîne des risques particuliers en matière de divulgation (« différenciation géographique ») qui doivent être pris en compte dans la méthode de contrôle de la divulgation des statistiques. Ainsi, plusieurs pays ont jugé que les méthodes basées sur le bruit post-tabulaire étaient efficaces à cet égard lors de la préparation du dernier recensement.

24. Quelle que soit la méthode de contrôle de la divulgation des statistiques adoptée, il est essentiel d'en présenter les caractéristiques générales aux utilisateurs. Cela vaut en particulier pour les méthodes dans lesquelles les données publiées ressemblent à des données non traitées, ou pour les méthodes qui peuvent nuire à l'additivité interne des tableaux.

## B. Diffusion

25. Un recensement n'est pas terminé tant que les informations recueillies ne sont pas mises à la disposition des utilisateurs sous une forme et selon un calendrier qui répondent à leurs besoins en constante évolution. En conséquence, pour la diffusion des résultats du recensement, il est important d'être à l'écoute des besoins des utilisateurs et de respecter des normes de qualité exigeantes dans la production des statistiques. Les résultats d'un recensement doivent être diffusés simultanément auprès de tous les utilisateurs, et il convient de veiller avec le plus grand soin à éviter de divulguer par inadvertance des informations sur des particuliers susceptibles d'être identifiés. Diverses mesures statistiques peuvent être appliquées pour protéger la confidentialité (voir par. 18 à 24).

26. Il existe plusieurs manières de mettre les résultats d'un recensement à la disposition des utilisateurs :

a) Sous la forme de rapports publiés (soit sur papier soit, plus communément, sur un support numérique) présentant des tableaux normalisés convenus à l'avance,

habituellement établis aux niveaux national, régional ou local et qu'il est possible de se procurer auprès des administrations publiques ou directement auprès d'autres sources ;

b) Sous la forme de rapports/résumés ad hoc non publiés contenant les données normalisées du recensement, mais ventilées selon d'autres limites géographiques ou sous-groupes qui n'avaient pas été publiés précédemment (le demandeur peut participer au coût de production) ;

c) Sous la forme de données publiées sur les sites Web des organismes nationaux de statistique ou par d'autres moyens électroniques. Celles-ci peuvent se présenter sous la forme de bases de données agrégées ou de microbases de données, mises à la disposition des usagers pour être traitées en ligne ou téléchargées gratuitement et parfois accompagnées d'outils de visualisation dynamiques ou interactifs, notamment de fonctions de cartographie augmentant la valeur des statistiques ;

d) Sous la forme de produits fournis sur commande ou adaptés à des besoins particuliers à partir d'une base de données, d'un tableur automatique ou d'un service de données statistiques, et présentant des tableaux à double entrée portant sur des variables n'apparaissant pas dans les rapports publiés ni dans les résumés, mais qui doivent se conformer aux mêmes mesures de contrôle relatives à la divulgation de l'information statistique que celles appliquées aux produits normalisés ;

e) Sous la forme de microdonnées (souvent appelés fichiers à usage public) portant généralement seulement sur une fraction échantillonnée de la population totale (souvent au niveau du ménage, pour inclure les individus qui le composent), en utilisant des méthodes de contrôle de la divulgation des statistiques (voir par. 18 à 24). Pour compenser les risques accrus découlant de la publication de microdonnées, une combinaison de mesures de contrôle supplémentaires peut être appliquée : les utilisateurs potentiels de données sont souvent vérifiés ; les données ne font l'objet que d'une diffusion restreinte ; les données généralement fournies ou auxquelles il est donné accès dans des conditions de sécurité sont soumises à un contrôle strict et à des mesures drastiques de protection de la confidentialité.

27. Lorsqu'il n'existe pas d'outils de diffusion personnalisables ou que les outils existants ne permettent pas de fournir les tableaux qui ne sont demandés que par quelques utilisateurs, comme certaines administrations publiques ou certains organismes de recherche spécialisés, ces tableaux peuvent être fournis sur demande. Néanmoins, dès lors qu'ils sont produits, l'organisme national de statistique ne devrait imposer aucune restriction à leur mise à la disposition du grand public par ses équipes chargées de l'assistance aux clients et utilisateurs.

28. Lorsque les résultats sont publiés sur papier, les coûts d'impression deviennent un problème, car les exemplaires physiques ne peuvent pas atteindre autant d'utilisateurs qu'une publication en ligne. Les publications traditionnelles, en particulier sous forme de document imprimé, sont sur le point de devenir complètement obsolètes. S'il est vrai que les publications imprimées permettent de communiquer des informations cohérentes sur des sujets particuliers et peuvent donc convenir à un utilisateur ou à un marché donné, les utilisateurs souhaitent désormais voir des formes de diffusion numériques plus dynamiques et plus interactives.

29. Lorsque des données sont fournies sous forme électronique, il faut donner aux utilisateurs les moyens de les extraire facilement. Ces données doivent être accessibles dans des formats normalisés ou au moyen d'une interface de programmation d'applications (API) et complétées, idéalement, par des fonctionnalités interactives faciles d'utilisation qui permettent de personnaliser les résultats. Pour que les données de recensement soient faciles à trouver et pertinentes, il faut avant tout que les tableaux soient immédiatement exploitables. La base de données de sortie doit être formatée conformément aux normes internationales relatives aux métadonnées, telles que la norme SDMX d'échange de données et de métadonnées statistiques. Les stratégies de diffusion doivent également être harmonisées avec les règles fixées par les autorités nationales en matière de données ouvertes.

30. Il convient de mettre en place, chaque fois que possible, des outils permettant aux utilisateurs de commander, de définir, de personnaliser et de recevoir en ligne des résultats du recensement et des échantillons (métadonnées) à usage public, en prenant des mesures appropriées pour protéger la confidentialité statistique des données et la sécurité de la

transmission. Lors de la conception des produits du recensement, il faut prendre en compte toutes les formes de technologie dont se servent les utilisateurs, comme les smartphones et autres appareils portatifs. Il est également important de veiller à la convivialité de l'accès aux données. Les tendances évoluent rapidement et il est essentiel, si l'on veut que les données de recensement conservent leur pertinence, que les utilisateurs disposent d'un ensemble complet de tableaux ou de produits permettant d'accéder rapidement aux données. Les tableaux en ligne permettent aux utilisateurs de personnaliser rapidement leurs requêtes et d'interroger directement la base de données pour obtenir les résultats immédiatement. L'existence d'un tel ensemble éviterait d'avoir à élaborer de grandes quantités de tableaux que les utilisateurs devraient parcourir pour trouver si les données recherchées ont été publiées.

31. Les médias sociaux, qui sont devenus de plus en plus appréciés et efficaces, permettent de communiquer de petites quantités de produits – en particulier à l'intention d'utilisateurs non spécialistes – ou de diffuser en temps voulu des résultats en lien avec des événements ou des dates importantes. Compte tenu de la diversité des plateformes actuellement disponibles, les images contenant du texte et les infographies peuvent être utilisées dans le cadre de campagnes menées sur les médias sociaux pour mettre en évidence l'utilisabilité des résultats du recensement. L'utilisation de ces médias témoignera souvent de la volonté de l'organisme national de statistique de s'ouvrir aux utilisateurs et d'établir un dialogue avec eux afin de répondre à leurs questions et préoccupations (voir également par. [297]).

32. Si l'accès aux bases de micro ou macrodonnées ou la diffusion de ces bases sur les médias sociaux peut largement contribuer à augmenter le nombre d'utilisateurs et donc à accroître la demande de données de recensement, il convient toutefois d'être attentif à ce qui suit :

a) Certains tableaux à double entrée peuvent présenter des problèmes de qualité notables compte tenu des non-réponses, des erreurs d'échantillonnage ou de traitement, ou encore des méthodes de traitement ou d'imputation. Les autorités chargées du recensement devraient mettre en place des procédures de mise en garde des utilisateurs potentiels à cet égard afin de ne pas décrédibiliser le recensement dans son ensemble. Certains organismes nationaux de recensement refusent d'autoriser la publication de certains de ces tableaux pour des raisons liées à leur qualité sur le fond, même si cela risque de mécontenter certains utilisateurs. D'autres ne les publieront que s'ils peuvent s'appuyer sur une politique claire couvrant à la fois les considérations de fond et les aspects techniques ;

b) S'agissant de certains tableaux à double entrée détaillés et de tous les fichiers contenant des données individuelles, des problèmes pourraient se poser en cas de diffusion d'informations sur des répondants identifiables en violation des règles sur la confidentialité des données du recensement. Ce point est examiné plus avant aux paragraphes 15 à 24.

Les problèmes que sont la qualité sur le fond et la confidentialité doivent être traités et des garde-fous établis, mais ils ne constituent pas un obstacle pour la diffusion d'un large éventail de produits de recensement.

33. Divers produits doivent être disponibles pour répondre aux besoins en constante évolution des utilisateurs. Il est probablement nécessaire de fournir :

a) Des résumés nationaux, régionaux et locaux ;

b) Des rapports sur les principales constatations relatives à certaines caractéristiques, complétés par des résultats et des analyses détaillés, soit sous une forme standard pour les zones allant jusqu'au niveau local, soit sous la forme de statistiques ventilées détaillées sur des caractéristiques particulières et des populations ciblées ;

c) Des profils de population ou des résumés statistiques essentiels pour les zones de petite taille et les petits groupes de population ;

d) Des analyses spatiales et graphiques, notamment un atlas du recensement, éventuellement complété par une analyse graphique interactive ou des outils cartographiques permettant aux utilisateurs d'effectuer une analyse personnalisée ;

e) Des produits à forte valeur ajoutée tels que des classifications des zones ou des ménages ;

f) Des métadonnées supplémentaires portant sur les définitions, les classifications et des évaluations de la couverture et de la qualité. Les métadonnées peuvent varier en fonction du type de recensement (dénombrement, recensement mixte ou recensement fondé sur des registres), mais leur niveau de détail doit rester le même.

34. Les résultats du recensement doivent être diffusés selon un plan structuré, assorti d'un calendrier annoncé à l'avance. Cela suppose que les données nationales soient publiées rapidement, dans le respect des délais fixés dans le calendrier, et que les données désagrégées soient publiées après une phase de traitement, le tout conformément au plan établi. Néanmoins, il est important que les délais soient respectés pendant la phase de diffusion, que l'on devrait chercher à raccourcir tout en garantissant l'intégrité des données.

35. La première publication des chiffres de la population est généralement très attendue par les utilisateurs, qui vont du grand public aux administrateurs de programmes et responsables des politiques. C'est pourquoi, le dénombrement à peine terminé, certains pays publient des chiffres provisoires qui donnent une idée générale de l'évolution démographique, même s'ils sont susceptibles d'être modifiés une fois l'ensemble des opérations de traitement et de vérification menées à terme. Il faut toutefois appeler l'attention des utilisateurs sur ce qu'implique l'utilisation de chiffres provisoires de la population, qui peuvent différer sensiblement des chiffres finaux .

36. Très tôt dans le processus, le public doit être informé du calendrier et du contenu des publications à venir, si l'on veut continuer à susciter son intérêt pour le recensement. Les parutions peuvent être progressives : on peut commencer par un simple résumé descriptif portant sur les grandes divisions géographiques d'un pays puis finir par des tableaux à double entrée plus complets et des rapports thématiques descriptifs.

37. Les données devraient en principe être mises à disposition gratuitement au point d'accès ou de livraison, mais le montant des redevances, lorsqu'il est nécessaire d'en percevoir (par exemple en cas de produits personnalisés ou commandés), doit être fixé de manière à rendre l'accès aux résultats abordable pour toutes les catégories d'utilisateurs. Des outils de publication interactifs flexibles qui traitent automatiquement les demandes de données personnalisées (voir par. 26) peuvent contribuer à répondre efficacement aux besoins des utilisateurs et à accroître ainsi la valeur des données diffusées gratuitement. La possibilité de diffuser des informations en ligne a permis de réduire la dépendance à l'égard des documents papier. Toutefois, les organismes nationaux de statistique peuvent toujours proposer un service payant d'impression à la demande pour fournir des documents de recensement aux utilisateurs qui ne sont pas en mesure d'accéder à des copies numériques ou de les recevoir. Ces utilisateurs ne doivent pas être désavantagés par l'absence de produits imprimés.

38. Pour répondre à des demandes très diverses, il faut concevoir des produits qui permettent de fournir des informations statistiques et géographiques au moyen de SIG ou d'autres API, de la manière la plus souple et la plus large possible et dans le plus grand respect possible de l'obligation de confidentialité (voir par. 23). Parmi les caractéristiques que devraient présenter les produits, on peut citer les suivantes :

a) Les utilisateurs doivent être en mesure de trouver les informations rapidement et simplement, dans un format ouvert ou des formats multiples ;

b) Les organismes nationaux de statistique augmentent considérablement l'utilité de leurs données de recensement lorsqu'ils intègrent des fonctions graphiques et cartographiques dans leurs produits. L'idéal serait que les utilisateurs aient eux-mêmes la possibilité de produire facilement des graphiques ou des cartes puis de les imprimer, de les tracer ou de rendre les images disponibles pour d'autres usages. Plusieurs pays produisent désormais ces types de produits de recensement, parfois en coopération avec des organismes commerciaux ;

c) Les formats de données ouverts, en particulier ceux qui facilitent le partage, doivent être privilégiés, car ils favorisent la mise en réseau des données du recensement avec d'autres bases de données, ce qui permettra une utilisation encore plus large des données.

39. La cartographie thématique et la visualisation des données, qui sont devenues des éléments importants de la diffusion des résultats, intéressent les organismes nationaux de statistique en raison de leur capacité à stimuler des interactions avec les utilisateurs et à en toucher un nombre croissant. Toutefois, la visualisation des données est un vaste domaine, qui va de simples infographies à des outils sophistiqués permettant des analyses multidimensionnelles des données. En outre, elle peut poser des problèmes à certains organismes de recensement en raison des coûts élevés, des exigences en matière d'infrastructure et de la nécessité de se conformer aux prescriptions techniques des différents pays (par exemple, en termes d'accessibilité et d'apparence commune). Les compétences nécessaires pour produire des visualisations efficaces peuvent être insuffisantes, et il peut être difficile de consacrer des ressources adéquates à leur élaboration, en particulier compte tenu des contraintes budgétaires que connaissent de nombreux organismes au cours de la période décennale actuelle. Toutefois, les utilisateurs souhaitent que les contenus en ligne soient bien présentés, attrayants et individualisés, c'est pourquoi le renforcement des capacités de production de contenus visuels devrait faire partie des priorités pour de nombreux organismes nationaux de statistique.

40. Les organismes nationaux de statistique doivent tenter de trouver la juste proportion entre la production de données tabulaires traditionnelles et la création de visualisations de données, qui implique des coûts élevés. Ils doivent pour cela bien cerner les différents types d'utilisateurs et leurs besoins particuliers en menant des recherches pour mieux comprendre les besoins en matière de données des utilisateurs finaux (on trouvera de plus amples informations sur la répartition des utilisateurs au chapitre [x] sur la communication et la sensibilisation). Il n'existe pas de solution passe-partout, ce qui explique pourquoi de nombreux organismes nationaux de statistique créent des unités ou des équipes spécialisées dans les technologies de visualisation des données.

### C. Documentation et métadonnées

41. Pour tout pays, un élément important du programme de diffusion des résultats de son recensement est un dossier complet de documents de référence et de métadonnées expliquant les produits statistiques, précisant leur utilité et en facilitant l'utilisation, notamment en vue de comparaisons avec les recensements précédents et d'autres sources de données.

42. Un système de métadonnées fournit des informations supplémentaires sur les caractéristiques des données recueillies et publiées. Chaque organisme national de statistique utilise son propre système de métadonnées, qui repose sur des normes internationales tout en répondant aux besoins particuliers du pays. Étant donné qu'un recensement et ses résultats sont souvent étroitement liés à d'autres domaines d'activité statistiques, il est recommandé que le système de métadonnées relatif à un recensement utilise les mêmes éléments que le système général de métadonnées de l'organisme national de statistique du pays concerné. Cependant, il est habituellement nécessaire aussi que les métadonnées d'un recensement comportent certains éléments propres au recensement ; le système de métadonnées doit en outre assurer la comparabilité internationale la plus large possible des données.

43. Dans le cadre des recensements de la population de 2030, il faut aussi assurer la comparabilité des données recueillies avec celles des recensements antérieurs tout en intégrant de nouveaux éléments correspondant aux évolutions qui ont eu lieu depuis le recensement précédent. Le système de métadonnées doit donc être également comparable à celui du recensement précédent, mais il doit être mis à jour pour répondre aux nouveaux besoins. Les systèmes de métadonnées des différents organismes nationaux de statistique devraient également faire ressortir la mesure dans laquelle ils utilisent des données provenant d'un dénombrement direct ou de sources administratives.

44. Un système de métadonnées doit comprendre au minimum les éléments suivants :

- a) Des définitions des termes et notions employés ;
- b) Un dictionnaire des données ou un glossaire des termes ;
- c) Des légendes explicatives des tableaux ;

- d) Des classifications et nomenclatures ;
- e) Les questions du recensement (dans le cas où les renseignements sont recueillis au moyen d'un processus traditionnel de dénombrement sur le terrain) ;
- f) Les fins auxquelles les renseignements sont recueillis, en particulier dans le cas des données administratives ;
- g) Les sources de données utilisées, en particulier lorsque les données sont tirées de registres administratifs.

45. Lorsque des indicateurs sont prévus par une classification internationale type, celle-ci doit être utilisée. Lorsqu'ils ne peuvent pas être classés selon une norme internationale, il peut être nécessaire de créer une nouvelle nomenclature. Les documents de référence peuvent porter sur un large éventail de questions, par exemple la méthode de base utilisée, la couverture, le taux de réponse, les sources des données, les enquêtes pilotes et les mises à l'essai, les variables dérivées, le taux de réponse par Internet, l'imputation et les enquêtes postérieures au dénombrement, ainsi que les rapports portant sur des descriptions plus générales de l'opération de recensement dans son ensemble et de la qualité des données. L'ampleur de la documentation de référence traitant particulièrement des problèmes de couverture et de qualité est abordée au chapitre [IV], mais il est ici recommandé que les pays intègrent au moins, dans les métadonnées de leur recensement, des mesures de la qualité des données et de la couverture, notamment les taux de réponse (aux niveaux national et local) et les taux d'imputation des données (pour la source de données dans son ensemble et pour chaque thème).

46. Les rapports méthodologiques sont particulièrement importants lorsque la méthode de base utilisée a changé depuis le recensement précédent (par exemple en cas de passage d'un dénombrement sur le terrain traditionnel à une méthode entièrement ou partiellement fondée sur des registres). De tels changements sont susceptibles de modifier les définitions et notions employées et donc la comparabilité entre les recensements.

47. Pour obtenir des métadonnées complètes, les autorités chargées du recensement doivent consulter diverses parties prenantes (notamment des organismes publics, des chercheurs, des décideurs et des membres du public) en utilisant, par exemple, des enquêtes structurées, des groupes de discussion ou d'autres mécanismes de consultation (voir par. [255 à 265]). Les observations peuvent être recueillies pour affiner les descriptions des métadonnées et les rendre plus intuitives et plus complètes.

## **D. Archivage et accès à des données confidentielles du recensement**

48. Le recensement est une source de données statistiques particulière, dont on peut en permanence extraire des informations comparables pour une période allant jusqu'à cent à cent cinquante ans. Il est possible d'en faire une source de données unique, mais cela suppose de confier à l'organisme national de statistique (ou à l'organisme national chargé des archives) l'importante responsabilité de préserver, pour l'avenir, cette photographie de la société à un instant donné. De la sorte, les données de recensement ne sont pas uniquement utiles aux décideurs et utilisateurs d'aujourd'hui, mais le seront aussi aux générations futures. C'est donc à l'organisme national de statistique qu'il incombe de traiter, d'archiver et de conserver ce « trésor ».

49. De nombreux pays ne conservent les données du recensement concernant les individus et les ménages que pendant le temps nécessaire au traitement des données et à la production des résultats statistiques, ou jusqu'à la fin du recensement. Toutefois, lorsqu'il s'agit d'évaluer l'ensemble des coûts et des avantages d'un recensement, il ne faut pas sous-estimer la valeur scientifique, sociohistorique et généalogique des données individuelles. Les organismes nationaux de statistique peuvent donc souhaiter autoriser l'accès (public ou restreint) à l'ensemble des relevés de recensement après un certain temps. Les pays ayant l'intention de conserver les données en vue d'une telle recherche doivent s'assurer qu'il existe un cadre juridique et une infrastructure physique solides pour en protéger la sécurité et la confidentialité jusqu'à ce qu'elles deviennent accessibles au public. Les pouvoirs publics doivent être conscients du fait que les organismes nationaux de statistique risqueraient

d'avoir du mal à recueillir des informations auprès de la population si les assurances concernant la confidentialité des informations recueillies n'étaient pas suivies d'effet. La confiance de la population dans la sécurité et la confidentialité des informations personnelles traitées dans le cadre d'un recensement doit par conséquent être considérée comme primordiale.

50. La période de non-divulgence des relevés de recensement doit être suffisamment longue pour protéger la confidentialité des données concernant des personnes vivantes, notamment les données ayant un caractère sensible, ou au moins réduire au minimum le risque d'atteinte à cette confidentialité. Dans de nombreux pays, cette période est fixée par la loi, mais elle peut varier d'un pays à l'autre. D'autres pays appliquent des dispositions plus générales relatives à la protection des données ou des dispositions législatives concernant la liberté de l'information pour ne pas divulguer des relevés confidentiels jusqu'à ce que le risque de divulgation d'informations personnelles concernant des personnes vivantes ait disparu. Une période de cent ans est donc recommandée même si, en raison de l'augmentation constante de l'espérance de vie, les pays peuvent envisager d'augmenter ce seuil en fonction des circonstances nationales.

51. Outre les demandes d'accès aux données de recensement émanant du public, les organismes nationaux de statistique peuvent recevoir des demandes de la part d'autres organismes publics qui souhaitent vérifier ou confirmer des informations existantes lorsque les données historiques sont rares ou inexistantes. Ils doivent accéder à ces demandes si la loi l'exige ou si cela sert clairement l'intérêt public.

52. Outre l'archivage des relevés de recensement (pour les pays qui y procèdent), il est tout aussi important – voire plus important – que chaque pays veille à préserver l'ensemble des métadonnées et des documents exposant les procédures/opérations produits tout au long du recensement, notamment la totalité des documents se rapportant à la gestion du projet, ainsi qu'à y faciliter l'accès. Il s'agit non seulement de disposer d'éléments utiles pour évaluer, avec le recul, toute l'efficacité du recensement, mais aussi, pour les futurs responsables de la planification de recensements, de pouvoir tirer des enseignements des résultats obtenus et des difficultés rencontrées par leurs prédécesseurs.

53. Ce faisant, les pays doivent s'assurer que face à l'évolution rapide de la technologie, les supports et les médias utilisés pour l'archivage de ces précieuses informations sont régulièrement réévalués de façon à permettre l'extraction rapide de données à tout moment pendant les vingt à cinquante années à venir.

## V. Conclusion

54. Le projet de recommandations portant sur la diffusion des données et divers sujets connexes (sécurité, documentation, métadonnées et archivage) devant figurer dans les recommandations pour les recensements de la population et des habitations de 2030 est présenté pour observations et discussion.