

**Commission économique pour l'Europe**

Conférence des statisticiens européens

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

Genève, 20-22 septembre 2023

Point 4 l) de l'ordre du jour provisoire

**Révision des recommandations de la Conférence des statisticiens européens
pour les recensements de la population et des logements pour le cycle de 2030 :****Diffusion de données de recensement ventilées****Rapport d'activité préliminaire de l'Équipe spéciale
de la diffusion de données de recensement ventilées
de la Conférence des statisticiens européens****Note de l'Équipe spéciale de la diffusion de données de recensement
ventilées de la Conférence des statisticiens européens*.¹***Résumé*

Après le lancement du processus d'examen des recommandations de la Conférence des statisticiens européens, cette équipe spéciale a organisé trois téléconférences et divers échanges écrits. Ses travaux s'articulent autour de trois axes (diffusion générale, contrôle de la divulgation des données statistiques, résultats géoréférencés). L'Équipe spéciale s'est penchée sur le questionnaire utilisé pour l'enquête de 2013, a proposé des mises à jour de diverses questions et a élaboré de nouvelles questions en fonction des thèmes qui entrent dans son champ d'étude.

Les principaux sujets que la mise à jour des questions permet de mieux aborder sont les suivants : la technologie et les outils de diffusion modernes, y compris les tableurs et les outils de visualisation des données, et les progrès récents dans le domaine du contrôle de la divulgation des données statistiques, y compris les méthodes basées sur le bruit et leur incidence sur les données liées au maillage ou les outils de diffusion flexibles, ainsi que dans celui des unités géographiques fonctionnelles, y compris les mailles.

* Document établi par Fabian Bach, Commission européenne – Eurostat. Les opinions exprimées sont purement celles de l'auteur et ne peuvent en aucun cas être considérées comme une position officielle de la Commission européenne.

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.

¹ Le présent document a été soumis avec retard en raison de sa transmission tardive par l'Équipe spéciale.



I. Introduction : Champ d'étude et plan de travail de l'Équipe spéciale

1. La principale partie des recommandations de 2020 de la Conférence des statisticiens européens² que l'Équipe spéciale examinera correspond au sous-titre « Diffusion », paragraphes 292 à 307 de la section « Diffusion, documentation, métadonnées et archivage » du chapitre III (Activités sur le terrain et autres activités opérationnelles). L'examen pourra aussi porter sur les questions abordées dans la section « Confidentialité et sécurité », paragraphes 140 à 148 du chapitre I (Méthodologie).

2. Le champ d'étude de l'Équipe spéciale comprend les éléments suivants :

- a) Formats et produits de diffusion, avec leurs justifications respectives, leurs publics cibles et leurs considérations spécifiques ;
- b) Assurance de la qualité de diffusion à des niveaux de détail élevés ;
- c) Incidence des différentes méthodes de diffusion des données de recensement sur le plan de la confidentialité et du contrôle de la divulgation ;
- d) Planification de la diffusion ;
- e) Diffusion des résultats géoréférencés (l'Équipe spéciale devra examiner cette question en coordination avec l'Équipe spéciale chargée de l'information géospatiale) ;
- f) Documentation et métadonnées ;
- g) Archivage.

3. L'Équipe spéciale se compose de neuf membres provenant des organismes et pays suivants : Bureau régional d'ONU-Femmes pour l'Europe et l'Asie centrale, Canada, CEPALC, Eurostat (présidence), Division de statistique de l'ONU, FNUAP Kirghizistan, Integrated Public Use Microdata Series (IPUMS), Irlande et Israël.

4. L'Équipe spéciale est convenue d'un plan de travail répartissant ses activités en deux grandes étapes : 1) examen des questions pertinentes du questionnaire utilisé pour l'enquête de 2013 (jusqu'en avril 2023) ; 2) examen des questions figurant dans les recommandations de la Conférence mentionnées plus haut, sur la base de l'analyse des résultats de l'enquête de 2023 (à partir de l'automne 2023, après la publication des résultats de l'enquête). Tout au long de ces deux étapes, les membres de l'Équipe spéciale seront répartis en trois sous-groupes thématiques chargés de formuler des propositions initiales qui seront discutées par l'ensemble du groupe :

- a) Diffusion générale, métadonnées et archivage (sous la direction de l'Irlande) ;
- b) Contrôle de la divulgation des données statistiques (sous la direction d'Eurostat) ;
- c) Résultats géoréférencés (sous la direction du Canada).

5. Le présent document rend compte de la première phase de travail, qui s'est achevée avec succès, à savoir l'examen du questionnaire. L'Équipe spéciale a proposé au Groupe directeur des mises à jour et de nouvelles questions dans les sections suivantes du questionnaire de 2013 : Méthodologie ; Technologie ; Communication et publicité ; Documentation, métadonnées et archivage ; Sécurité, confidentialité et contrôle de la divulgation ; Diffusion ; Innovations ; Qualité et couverture ; Caractéristiques géographiques. Elle a également formulé des commentaires et des considérations (sans avoir pu proposer de mise à jour des questions) sur certains éléments des sections suivantes : Problèmes rencontrés, succès obtenus et enseignements tirés ; Coûts et avantages.

² https://unece.org/DAM/stats/publications/2015/ECECES41_EN.pdf.

II. Échange de données d'expérience entre les membres de l'Équipe spéciale en ce qui concerne la diffusion

6. Lors de la deuxième réunion, tous les membres de l'Équipe spéciale ont eu l'occasion de présenter des produits ou des aspects spécifiques de la diffusion issus de leur propre expérience. Sept membres ont pris la parole pour faire les interventions suivantes :

a) Eurostat a présenté les principaux canaux de diffusion des résultats du recensement de 2021 de l'UE, à savoir le Census Hub³, qui offre un accès complet à tous les résultats via une interface de création de tableaux, Eurobase⁴, qui présente une sélection d'hypercubes via un navigateur en ligne, et les articles de Statistics Explained⁵. Enfin, les résultats comprennent également des indicateurs clefs sur une maille européenne de 1 kilomètre carré (la population totale a déjà été publiée⁶) ;

b) L'Irlande a donné un aperçu de ses produits de diffusion du recensement national, qui comprennent des tableaux PX Stat (plus de 1 000 ayant jusqu'à quatre dimensions)⁷, des publications électroniques⁸, des statistiques sur la population de petites régions, y compris une cartographie du recensement et des cartes interactives⁹, des micro-fichiers de recherche¹⁰ et des mailles statistiques ;

c) La CEPALC a présenté sa plateforme régionale de diffusion de données REDATAM¹¹ en mettant l'accent sur deux produits dans le contexte du recensement : les microdonnées de recensement harmonisées et les tables de migration concernant la région de l'Amérique latine ;

d) Le Canada a décrit ses produits de diffusion du recensement¹², y compris un certain nombre de profils¹³ et de cartes¹⁴ de recensement, avant de présenter plusieurs projets, notamment un nouveau tableur hybride en ligne ;

e) L'IPUMS¹⁵ a présenté ses activités et ses produits axés sur la diffusion de microdonnées de recensement et d'enquête comparables à l'échelle internationale et provenant de toutes les régions du monde ;

f) Israël a décrit la situation du pays et les difficultés que celui-ci rencontre actuellement en ce qui concerne principalement la modernisation en profondeur de l'ensemble de son système de diffusion des données de recensement. L'un des plus gros problèmes est de savoir comment diffuser les données en toute sécurité ;

g) Le FNUAP Kirghizistan a présenté un résumé de ses activités régionales passées et en cours, y compris le soutien à la publication du recensement kirghize (notamment à partir du cycle de 2010) et les efforts actuels visant à mieux prendre en compte les questions du Groupe de Washington¹⁶ sur les statistiques relatives au handicap et sur la migration.

³ <https://ec.europa.eu/CensusHub2> (pour l'instant, seuls les résultats du recensement de 2011 de l'UE peuvent être consultés ; les résultats de 2021 seront ajoutés dès qu'ils seront disponibles).

⁴ <https://ec.europa.eu/eurostat/data/database> (l'onglet relatif au recensement de 2021 donne accès à certains ensembles de données fournis volontairement par les membres de l'UE, à savoir la population par grande tranche d'âge et au niveau NUTS 3, ainsi que divers ensembles de données spécifiques sur la population de citoyenneté ukrainienne – dans le contexte de l'invasion non provoquée de l'Ukraine par la Russie).

⁵ <https://ec.europa.eu/eurostat/statistics-explained/index.php?search=census>.

⁶ https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_and_housing_census_2021_-_population_grids.

⁷ <https://data.cso.ie>.

⁸ <https://www.cso.ie/en/census>.

⁹ <https://visual.cso.ie/?body=entity/ima/cop/2016&boundary=C03736V04484>.

¹⁰ <https://www.cso.ie/en/census/census2016reports/powscar>.

¹¹ <https://www.cepal.org/en/subtopics/redatam>.

¹² <https://www12.statcan.gc.ca/census-recensement/index-fra.cfm?DGUID=2021A000011124>.

¹³ <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=F>.

¹⁴ <https://www12.statcan.gc.ca/census-recensement/2021/geo/maps-cartes/index-fra.cfm>.

¹⁵ www.ipums.org.

¹⁶ <https://www.washingtongroup-disability.com>.

III. Examen du questionnaire

7. Les principales considérations de l'Équipe spéciale et les mises à jour du questionnaire que celle-ci a proposées en conséquence sont présentées ci-après, selon les trois volets thématiques décrits au paragraphe 4.

A. Diffusion générale, métadonnées et archivage

8. L'Équipe spéciale a constaté une tendance générale à utiliser des outils de diffusion plus interactifs, basés sur le Web, depuis la dernière mise à jour des recommandations de la Conférence des statisticiens européens. Par exemple, plusieurs pays ont mis en place des tableurs interactifs en tant qu'outils de diffusion permettant aux utilisateurs de consulter des tableaux croisés personnalisés. Ces outils peuvent ensuite être reliés aux microdonnées (comme l'Australian Table Builder¹⁷ et le système mis en place par le Bureau de statistique du Bangladesh¹⁸) ou à un ensemble prédéfini de tableaux croisés très détaillés (comme le Census Hub de l'UE (voir la note de bas de page 3)). Bien que la phase finale et certains autres choix de conception puissent avoir une incidence sur les scénarios de contrôle de la divulgation (voir le paragraphe 16), l'Équipe spéciale s'est généralement accordée pour reconnaître que les phases initiales interactives et conviviales de ces outils représentaient une amélioration importante de la diffusion des données de recensement. Ces outils sont souvent associés à des outils interactifs de visualisation des données.

9. L'Équipe spéciale a noté par ailleurs que les modes de collecte et de stockage des données de recensement continuaient d'évoluer rapidement avec la numérisation de la société dans un grand nombre de pays. D'une part, de nombreux pays réduisaient les contacts directs avec la population en s'appuyant de plus en plus sur des sources de données administratives et autres. D'autre part, les recensements basés entièrement ou partiellement sur une enquête directe s'effectuaient de plus en plus à l'aide d'outils intelligents et numériques.

10. Enfin, l'Équipe spéciale a noté que la communication avec les utilisateurs jouait un rôle de plus en plus important dans la réussite des recensements. Cela se vérifiait à la fois avant le recensement (la consultation des parties prenantes permettant par exemple de planifier le recensement) et après la publication des résultats (types de documentation, de métadonnées et d'informations sur la qualité fournies, par exemple).

11. Par conséquent, l'Équipe spéciale a proposé diverses mises à jour du questionnaire en vue d'améliorer dans ce sens le retour d'informations. Plus particulièrement, plusieurs questions ont été complétées ou ajoutées afin de mieux comprendre si et comment des tableurs seraient proposés dans le cadre de la diffusion ; une question a été ajoutée afin de savoir si l'organisme de recensement avait spécialement consacré des moyens humains aux technologies de visualisation des données ; des questions ont été complétées ou ajoutées par souci d'adaptation aux nouveaux modes de recensement (utilisation de données administratives, outils de réponse intelligents) ; enfin, certaines questions ont été proposées sur la consultation des parties prenantes, les types de documentation et les types d'indicateurs de qualité fournis avec les données.

B. Contrôle de la divulgation des données statistiques

12. L'Équipe spéciale a reconnu que ce domaine avait connu de profondes évolutions depuis la dernière mise à jour des recommandations de la Conférence des statisticiens européens. Notamment, de nombreuses activités avaient été menées dans le monde entier pour étudier et mettre au point de nouvelles méthodes de contrôle de la divulgation des données statistiques basées sur le bruit, à savoir l'introduction de petites perturbations dans les microdonnées ou les tableaux à multiples entrées afin de protéger les informations confidentielles.

¹⁷ <https://www.abs.gov.au/websitedbs/censushome.nsf/home/tablebuilder>.

¹⁸ <http://redatam.bbs.gov.bd/redbin/RpWebEngine.exe/Portal>.

13. Par exemple, après le recensement de 2011 de l'UE qui avait vu une multitude de cellules de données supprimées dans des tableaux croisés plus complexes et à la suite de premiers travaux menés notamment en Australie¹⁹ et dans d'autres pays, Eurostat avait élaboré un projet avec des experts nationaux afin de mettre au point des méthodes de contrôle de la divulgation des données statistiques plus efficaces en vue du recensement de 2021 de l'UE. Ce projet avait débouché sur des recommandations d'experts²⁰ et des outils logiciels publics²¹ axés sur deux méthodes basées sur le bruit : la permutation ciblée des entrées avant l'établissement des tableaux et l'utilisation d'une clef cellulaire après l'établissement des tableaux. Ces méthodes avaient été choisies non seulement pour réduire la quantité de données supprimées, mais aussi pour faire face à certains risques supplémentaires liés à la publication de données portant sur des zones géographiques non imbriquées (par exemple, les mailles de 1 kilomètre carré et les circonscriptions municipales).

14. Pour le recensement de 2021, un nombre important de pays de l'UE devraient appliquer un système de contrôle de la divulgation des données statistiques fondé sur ces recommandations. Dans le contexte de la modernisation en cours des statistiques démographiques au sein de l'UE, ces recommandations seront évaluées et probablement affinées ou éventuellement complétées par d'autres propositions ; par exemple, la Belgique et la France ont mis au point leurs propres méthodes basées sur le bruit, en particulier pour protéger les données liées au maillage.

15. Le Bureau du recensement des États-Unis a publié pour la première fois les résultats du recensement de 2020 protégés par une méthode de confidentialité différentielle²² basée sur le bruit²³. Cette méthode a suscité des réactions mitigées de la part des experts et des utilisateurs de données, en raison de graves préoccupations quant à son utilité²⁴ et des débats qui s'en sont suivis²⁵.

16. Enfin, l'Équipe spéciale a pris note de certains problèmes de contrôle de la divulgation des données statistiques lorsque des données détaillées sont diffusées au moyen d'outils interactifs permettant aux utilisateurs de personnaliser les recherches de données²⁶. Lorsqu'il est envisagé de fournir de tels outils dans le cadre du programme de diffusion, il faut que la stratégie de contrôle de la divulgation tienne compte des différents scénarios d'attaque.

17. Par conséquent, l'Équipe spéciale a proposé diverses mises à jour du questionnaire en vue d'améliorer dans ce sens le retour d'informations, notamment par l'application de méthodes de permutation (ciblée) des entrées, de clef cellulaire et de confidentialité différentielle en tant qu'options de réponse et par l'ajout de questions sur l'utilisation et les propriétés des outils de production interactifs et leur incidence sur le contrôle de la divulgation des données statistiques, ainsi que d'une question sur les mesures spécifiques de protection contre la différenciation géographique des zones géographiques de production non imbriquées.

¹⁹ Thompson G., Broadfoot S. et Elazar D. (2013), *Methodology for the Automatic Confidentialisation of Statistical Outputs from Remote Servers at the Australian Bureau of Statistics*, [Session de travail conjointe CEE/Eurostat sur la confidentialité des données statistiques](#).

²⁰ https://cros-legacy.ec.europa.eu/content/harmonised-protection-census-data_en.

²¹ <https://github.com/sdcTools/CensusProtection>.

²² Dwork C., McSherry F., et al. (2006), *Calibrating Noise to Sensitivity in Private Data Analysis*, in *Theory of Cryptography*, eds. Halevi S., Rabin T., p. 265 à 284, Berlin, Heidelberg: Springer Berlin Heidelberg.

²³ Abowd J. M. (2018), *The U.S. Census Bureau Adopts Differential Privacy*, ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (actes de la 24^e conférence), KDD '18. London, UK: Association for Computing Machinery, p. 2867. DOI:10.1145/3219819.3226070.

²⁴ Voir par exemple Ruggles S., Fitch C., Magnuson D. et Schroeder J. (2019), *Differential Privacy and Census Data: Implications for Social and Economic Research*, AEA Papers and Proceedings, 109, 403 à 408.08.

²⁵ Voir par exemple Muralidhar et Domingo-Ferrer (2023), *Legacy Statistical Disclosure Limitation Techniques for Protecting 2020 Decennial US Census: Still a Viable Option*, *Journal of Official Statistics*, à paraître, consultable en ligne à l'adresse <https://ssrn.com/abstract=4472525> ou <http://dx.doi.org/10.2139/ssrn.4472525>.

²⁶ Voir par exemple Asghar H. J. et Kaafar D. (2020), *Averaging Attacks on Bounded Noise-Based Disclosure Control Algorithms*, *Proceedings on Privacy Enhancing Technologies*, 2020, 358 à 378.

C. Diffusion géographique

18. Selon l'Équipe spéciale, l'évolution la plus notable depuis la dernière mise à jour des recommandations de la Conférence des statisticiens européens est liée à un fort regain d'intérêt pour les données géoréférencées et l'analyse. Il s'agit notamment d'une tendance générale à proposer des données de recensement géoréférencées par mailles, mais aussi à offrir des outils de visualisation géographique plus puissants et interactifs, tels que les logiciels de cartographie. Les nouveaux défis à relever dans le domaine du contrôle de la divulgation des données statistiques ont également été pris en considération (voir le paragraphe 13).

19. Une plus grande disponibilité des données géoréférencées et des données à plus petite échelle, ainsi que des outils d'analyse correspondants, exige généralement aussi une adaptation des métadonnées fournies. Par exemple, toute méthode d'estimation visant à produire des résultats de recensement au plus petit niveau géographique (par exemple, la maille) s'appuierait probablement sur une documentation distincte. D'ailleurs, les pays peuvent joindre des informations auxiliaires aux résultats de recensement qu'ils publient afin de rendre possibles des analyses géospatiales plus puissantes (par exemple, des données sur les réseaux routiers ou les limites de zones géographiques fixes ou personnalisées).

20. Par conséquent, l'Équipe spéciale a proposé diverses mises à jour du questionnaire en vue d'améliorer dans ce sens le retour d'informations. Notamment, les questions relatives aux petites zones géographiques ont été actualisées de manière à couvrir explicitement les mailles, s'agissant notamment de la résolution du maillage et des unités statistiques (personnes, logements, ménages, etc.) disponibles sur une maille, et certaines questions ont été ajoutées pour prendre en compte les méthodes d'estimation géographique et les informations auxiliaires publiées avec les données relatives aux petites zones ou aux mailles.

21. L'Équipe spéciale s'est en outre demandé si les questions relatives à la définition et à l'utilisation des « localités » à des fins de délimitation urbaine ou rurale devaient être adaptées ou complétées afin de tenir également compte des méthodes modernes de délimitation des villes, des zones urbaines et des zones rurales. Plus précisément, lors de sa cinquante et unième session, en 2020, la Commission de statistique de l'ONU avait approuvé une nouvelle méthodologie et un nouveau manuel ayant pour objet de délimiter les villes, les agglomérations et les zones rurales en s'appuyant essentiellement sur une analyse combinée des petites zones administratives et des mailles de population afin d'en déduire des grappes de population. Cependant, l'Équipe spéciale a décidé de transmettre ces considérations à l'Équipe spéciale 11, en charge des informations géospatiales et des statistiques sur de petites zones servant aux recensements, plutôt que de proposer elle-même des mises à jour du questionnaire à cet égard.

IV. Conclusion et prochaines étapes

22. L'Équipe spéciale a examiné le questionnaire utilisé pour l'enquête de 2013 en fonction des thèmes qui entrent dans son champ d'étude et a proposé au Groupe directeur un certain nombre de mises à jour et de nouvelles questions dans trois domaines : la diffusion générale, les métadonnées et l'archivage ; le contrôle de la divulgation des données statistiques ; la diffusion géographique. À la connaissance de l'auteur au moment d'achever le présent rapport, une grande majorité des propositions ont été intégrées dans le questionnaire final de l'enquête prévue en 2023.

23. Dès qu'ils seront disponibles, l'Équipe spéciale analysera les résultats de l'enquête de 2023 et reprendra son travail d'examen des recommandations les plus récentes de la Conférence des statisticiens européens.