



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 d) 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 : Technologie****Élaboration de recommandations sur la technologie
appliquée aux recensements****Note de l'Équipe spéciale de la technologie appliquée aux recensements
de la Conférence des statisticiens européens****Résumé*

Le présent document contient le projet de chapitre consacré à la technologie 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 modifications apportées par rapport aux Recommandations pour les recensements de 2020. L'objectif principal est de recueillir les observations et les propositions des experts nationaux sur le projet de texte afin qu'il soit tenu compte des besoins des organismes nationaux de statistique et de l'évolution des technologies et des méthodes utilisées.

* L'Équipe spéciale de la technologie appliquée aux recensements se compose des membres suivants : Andres Vikat (CEE, Président), Régis Plamondon (Canada), John Conway (Irlande), Ana Binstok (Israël), Johan Åmberg (Norvège), Espen Andersen (Norvège), Galina Sheverdova (Fédération de Russie), Hannes Reuter (Eurostat) et Seiffe Tadesse (Division de statistique de l'ONU).

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. Dans les recommandations pour les recensements de 2030, l'Équipe spéciale de la technologie appliquée aux recensements a abordé les questions liées à la technologie, à l'exception de celles qui concernaient les systèmes d'information géographique (SIG), pour lesquelles une équipe spéciale distincte avait été créée.
2. On trouvera à la section II du présent document un résumé des modifications introduites par rapport aux Recommandations pour les recensements de 2020.
3. La section III présente le projet de chapitre sur la technologie destiné à figurer dans les recommandations de la CSE pour les recensements de la population et des habitations de 2030.
4. L'objectif principal est de recueillir les observations et les propositions des experts nationaux sur le projet de texte afin qu'il soit tenu compte des besoins des organismes nationaux de statistique et de l'évolution des technologies et des méthodes utilisées.

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

5. Le chapitre II des Recommandations de la CSE pour les recensements de 2020, consacré à la technologie, a servi de point de départ à la rédaction du texte de la section III du présent document¹. Pour rendre compte de l'évolution de la technologie et des pratiques depuis 2020, sept nouvelles parties ont été ajoutées dans les recommandations pour les recensements de 2030 :
 - a) Éléments qu'il est recommandé de mettre à l'épreuve (sect. III.B.3) ;
 - b) Élaboration du questionnaire électronique (III.C.4) ;
 - c) Technologies visant à faciliter le dénombrement des personnes handicapées et des personnes non connectées (III.C.5) ;
 - d) Plateformes interactives en ligne permettant de diffuser les résultats (III.D.3) ;
 - e) Stockage et archivage (III.D.4) ;
 - f) Informatique en nuage (III.D.5) ;
 - g) Technologies appliquées à l'utilisation des données administratives (III.E).
6. L'accent est mis sur les solutions technologiques dont l'utilisation s'est généralisée depuis les recensements de 2020, à savoir les outils électroniques de collecte de données, l'informatique en nuage et les solutions intégrées permettant d'automatiser toutes les étapes du recensement, depuis la planification de celui-ci jusqu'à la diffusion des résultats.
7. Les technologies liées aux systèmes d'information géographique (SIG) font l'objet d'un chapitre distinct, dont le projet est présenté dans un autre document, élaboré par une autre équipe spéciale.
8. Tout au long du chapitre, une place importante est accordée à la question de la sécurité des données, notamment en ce qui concerne la nécessité de mettre à l'épreuve les systèmes de recensement intégrés et leurs différentes composantes, la certification de cybersécurité, le cloisonnement des infrastructures essentielles, la possibilité de répondre par Internet, l'utilisation d'appareils portables et le stockage des données. Dans la nouvelle partie portant

¹ CEE, 2015. [Recommandations de la Conférence des statisticiens européens pour les recensements de la population et des habitations de 2020](#). Genève : Nations Unies.

sur les technologies appliquées à l'utilisation des données administratives, une section a été consacrée à la sécurité et à la confidentialité.

9. Les technologies éprouvées de collecte des données telles que les entretiens téléphoniques, la saisie au clavier et la lecture optique ne sont pas présentées en détail, car elles sont déjà bien connues et ont été largement décrites. Sur la base de l'expérience acquise lors des recensements de 2020, il n'est plus recommandé de recourir aux entretiens téléphoniques automatisés.

III. Projet de texte du chapitre consacré à la technologie appliquée aux recensements

A. Introduction

10. Depuis de nombreuses années, la technologie intervient à toutes les étapes des recensements de la population et des habitations. Le présent chapitre contient des considérations générales relatives à l'utilisation et à la mise à l'essai des technologies appliquées aux recensements ainsi que des recommandations sur les technologies pouvant faciliter la collecte et le traitement des données. Il y est également question des technologies utilisant des données provenant de sources administratives.

11. Les technologies déjà éprouvées de saisie des données telles que la saisie au clavier ne sont pas présentées en détail, car elles sont déjà bien connues et ont été largement décrites. L'accent est mis sur les solutions technologiques dont l'utilisation s'est généralisée depuis les recensements de 2020 : outils de collecte de données électroniques, informatique en nuage et solutions intégrées permettant d'automatiser toutes les étapes du recensement, depuis la planification de celui-ci jusqu'à la diffusion des résultats. Il est entendu que de nouvelles technologies telles que l'apprentissage automatique et l'intelligence artificielle pourraient encore se développer dans les années précédant le recensement de 2030.

B. Considérations générales

1. Facteurs à prendre en compte avant d'adopter une nouvelle technologie

12. Grâce à la technologie, il est *possible* d'améliorer la couverture des recensements et la qualité des données, de réduire les coûts et de diffuser les résultats plus rapidement. À court terme, cependant, l'adoption de nouvelles technologies peut augmenter les coûts. Les organismes de recensement doivent examiner la façon dont les nouvelles possibilités offertes par l'innovation technologique peuvent contribuer à améliorer la pertinence, l'exactitude et l'actualité des données de recensement². Par exemple, il serait possible, grâce à la réduction des coûts de traitement, d'étoffer le contenu du questionnaire de recensement ou d'accroître le taux d'échantillonnage pour les questions faisant l'objet d'un échantillon, ce qui permettrait d'améliorer la qualité des données recueillies concernant certains thèmes (petits groupes de population, par exemple) ou certaines zones géographiques (zones de petite taille, voire un quadrillage de 1 km par 1 km). Cela dit, un équilibre est à trouver entre tout accroissement du contenu et ses effets sur les personnes interrogées.

13. Pour que les politiques et la planification puissent reposer sur des données factuelles, les données de recensement doivent provenir d'*utilisateurs très variés*. Il faut donc que les systèmes de production répondent aux besoins d'un large éventail d'utilisateurs allant de

² On trouvera de plus amples informations sur les facteurs à prendre en compte avant d'adopter une nouvelle technologie aux fins d'un dénombrement direct à la section B.3 du document du Département des affaires économiques et sociales (DESA) de l'ONU (2019) intitulé « Guidelines on the use of electronic data collection technologies in population and housing censuses » (Directives sur l'utilisation des technologies de collecte électronique de données aux fins des recensements de la population et des habitations). New York : Nations Unies.

ceux qui cherchent à obtenir rapidement des chiffres clefs à ceux qui souhaitent effectuer des analyses approfondies.

2. Évaluation

14. *L'adoption de technologies ou de méthodes nouvelles ne doit être envisagée que lorsque leurs avantages sont bien compris et que leur évolution pourra être maîtrisée³.*

15. *Lorsqu'il est envisagé de recourir à une technologie qui n'a pas encore été éprouvée dans le cadre d'un recensement, il faut d'abord procéder à une évaluation minutieuse, en tenant compte du contexte national, du coût de la technologie concernée par rapport aux solutions traditionnelles, du travail nécessaire pour la développer et la mettre à l'essai, des incidences que son utilisation pourrait avoir sur l'organisation globale du recensement, des effets qu'elle pourrait avoir sur la qualité des résultats et de son incidence sur la population.*

16. Les organismes de recensement devraient procéder à une évaluation en amont du recensement pour *déterminer quels systèmes et processus sont adaptés* à leur situation. Les principaux aspects à prendre en compte sont :

a) Les coûts salariaux et les coûts des opérations manuelles par comparaison avec ceux de systèmes informatiques et de l'infrastructure associée ;

b) Les capacités et l'infrastructure technologiques dont disposent l'organisme national de statistique et le pays dans son ensemble ;

c) La capacité de l'organisme national de statistique à gérer le développement de systèmes élaborés et complexes ;

d) La disponibilité des fonds et du temps requis pour le développement et la mise à l'essai de la solution technologique.

17. *Les principes de la réflexion conceptuelle et de la science du comportement* doivent être pris en compte tout au long du recensement. La réflexion conceptuelle permettra de porter un regard critique sur les processus actuels et de s'adapter pour répondre aux besoins futurs des clients et des citoyens. Cela devrait permettre de réduire le nombre de questions restant sans réponse puisque les citoyens reçoivent des indications qui les aident à remplir le questionnaire.

18. Avant d'adopter des technologies ou des méthodes nouvelles, l'organisme national de statistique devrait être parfaitement conscient des *risques* qui y sont associés. La nouvelle technologie fonctionnera-t-elle ? Quelles seront les réactions des personnes interrogées ou du personnel chargé du recensement ? En raison des longs intervalles entre deux cycles de recensement (de cinq à dix ans), la possibilité de tirer directement les leçons de l'application d'approches nouvelles risque d'être limitée.

19. Lorsqu'ils envisagent d'innover dans leurs pratiques, les organismes nationaux de statistique seraient bien avisés de *s'inspirer de l'expérience de leurs homologues d'autres pays*. Une collaboration avec d'autres organismes pourrait également être envisagée dans le but d'éprouver conjointement de nouvelles approches ou de nouvelles technologies avant de les adopter.

20. Peu d'organismes nationaux de statistique disposent des capacités techniques nécessaires à la maîtrise de nouveaux logiciels souvent complexes et de l'infrastructure nécessaire pour utiliser nombre des technologies nouvelles ou émergentes. Il est donc probable que des pays se poseront la question de savoir si des éléments significatifs des solutions techniques applicables dans le cadre d'un recensement se prêteraient à une *externalisation*.

³ On trouvera de plus amples informations sur le processus décisionnel concernant le recours à une technologie donnée à la section B.4 du document du DESA (2019) intitulé « Guidelines on the use of electronic data collection technologies in population and housing censuses » (Directives sur l'utilisation des technologies de collecte électronique de données aux fins des recensements de la population et des habitations). New York : Nations Unies.

3. Mise à l'essai

21. L'organisme national de statistique doit appliquer une stratégie de *mise à l'essai* rigoureuse et en faire une priorité du recensement. Les essais doivent notamment porter sur l'exactitude fonctionnelle, les exigences non fonctionnelles, les interfaces, le traitement des erreurs, la performance, la fiabilité, l'utilisabilité et la sécurité. En plus d'être soumise à un cycle d'essais, la technologie adoptée doit être certifiée quant à la cybersécurité.

22. *Il est recommandé d'inclure les éléments ci-après dans le cycle d'essais :*

- a) Tests portant sur un domaine ou un contenu :
 - Tests qualitatifs ou cognitifs ;
 - Tests d'évaluation des nouveaux services en ligne ;
 - Tests quantitatifs relatifs au contenu du questionnaire ou tests de comportement ;
 - Tests A/B (split testing) ;
- b) Tests relatifs à l'informatique :
 - Tests unitaires ;
 - Tests des opérations sur le terrain ;
 - Tests d'interface ;
 - Tests de l'état de préparation des systèmes et tests de préparation opérationnelle ;
 - Tests de sauvegarde et de récupération des données ;
 - Tests simulant des attaques de logiciels rançonneurs ;
 - Tests de volume et de performance ;
 - Tests de charge ;
 - Tests de sécurité tels que les tests d'intrusion ;
 - Tests non fonctionnels (validation des performances, de la fiabilité, de l'utilisabilité et d'autres aspects non fonctionnels d'un logiciel) ;
- c) Tests fonctionnels :
 - Tests d'intégration (nombreuses itérations au cours du recensement) ;
 - Tests de bout en bout ;
 - Tests de validation.

23. Lorsque les répondants ont la possibilité de remplir le questionnaire sur différents supports (téléphone portable, tablette ou ordinateur de bureau), il est particulièrement important de vérifier que l'ensemble du système de recensement est cohérent. Chacun de ces supports doit être mis à l'essai du point de vue de l'accessibilité, de l'expérience utilisateur, de la disponibilité, de la performance et de la sécurité.

24. Dans le contexte du développement rapide de l'informatique et des changements importants qui en découlent pour de nombreux aspects des recensements menés après 2020, *il est particulièrement important, dans le cadre des recensements pilotes, de rendre obligatoire la mise à l'essai de l'ensemble de la solution technologique en vue des recensements de 2030.*

4. Gestion de projet

25. Les opérations de recensement requièrent une série de processus administratifs caractéristiques des *projets de grande ampleur*. Par exemple, la planification d'une opération aussi complexe qu'un recensement peut être facilitée par l'utilisation d'un logiciel approprié. Dans bon nombre de pays, il peut aussi être nécessaire d'avoir recours à différents systèmes et processus pour recruter et rémunérer un grand nombre d'agents recenseurs employés à titre

temporaire. L'organisme national de statistique devrait s'interroger sur la manière dont la technologie pourrait permettre de gérer ces opérations efficacement et à moindre coût. Celle-ci pourrait en effet contribuer à limiter le coût du recensement et à en améliorer la qualité générale en permettant d'affecter des ressources aux tâches essentielles de dénombrement, de traitement et de diffusion, plutôt qu'à des opérations administratives telles que la paie du personnel.

C. Collecte de données dans le cadre du dénombrement direct

26. Trois méthodes de collecte de données sont examinées ci-après : les questionnaires en ligne, les entretiens téléphoniques et les questionnaires sur des appareils portables tels que les ordinateurs portables, les tablettes et les smartphones⁴.

1. Questionnaires en ligne

27. *Pour le dénombrement direct, il est recommandé d'offrir en priorité la possibilité de répondre en ligne.*

28. Les questionnaires en ligne permettent de réduire le **coût** global de la collecte et d'obtenir des données de meilleure *qualité*. Les données collectées en ligne seront formatées et peuvent être corrigées et modifiées par les répondants eux-mêmes. À titre de comparaison, la saisie sur papier passe par de nombreuses étapes telles que la numérisation, la capture et, surtout, le codage par des opérateurs. Le fait que le système doive interpréter différentes écritures manuscrites ajoute à la complexité du processus et réduit la précision des données. Pour toutes ces raisons, il est clairement préférable de recourir aux questionnaires en ligne.

29. L'utilisation des questionnaires en ligne implique que les données soient collectées par *autodénombrement* plutôt que dans le cadre d'un entretien. La possibilité de répondre par Internet peut être intégrée à toutes les méthodes traditionnelles de distribution et de collecte des questionnaires (par exemple, distribution/ramassage, envoi et renvoi des questionnaires par la poste).

30. Il est essentiel de bien gérer les opérations de *contrôle* de la collecte, c'est-à-dire de faire en sorte que chaque ménage ou individu ne soit dénombré qu'une seule fois. Pour cela, il faut pouvoir attribuer à chaque ménage et à chaque individu au sein de ce ménage un code unique lié à une localisation géographique. Un problème supplémentaire se pose dans les pays où les questionnaires sont récupérés par les agents recenseurs (et non renvoyés par la poste), dans la mesure où il faut que ces agents soient avertis en temps utile pour qu'ils puissent mettre à jour leurs fiches de contrôle de la collecte et ne rendent donc pas visite aux ménages qui ont déjà renvoyé les questionnaires.

31. Pour déterminer dans quelle mesure les répondants pourraient avoir *recours* aux questionnaires en ligne, il faut évaluer la proportion de la population qui a accès à Internet depuis son domicile, la proportion qui utilise des connexions à haut débit et l'utilisation générale d'Internet à d'autres fins (par exemple, opérations bancaires, déclarations fiscales et achats en ligne).

32. En outre, il faut concevoir *des systèmes et des procédures* qui permettent de renvoyer les questionnaires par Internet, ce qui permet de réaliser des économies en diminuant la charge de travail des agents recenseurs ainsi qu'en réduisant les coûts liés à la saisie des données, à l'impression et à l'envoi par la poste.

33. Lors de la conception de l'infrastructure, il convient d'accorder une attention toute particulière à la sécurité. *Il faut mettre en place une infrastructure physiquement distincte pour recueillir les données de recensement et les traiter. Une fois collectés et saisis, les*

⁴ On trouvera de plus amples informations sur ces technologies, entre autres, dans le document du DESA (2019) intitulé « [Guidelines on the use of electronic data collection technologies in population and housing censuses](#) » (Directives sur l'utilisation des technologies de collecte électronique de données aux fins des recensements de la population et des habitations). New York : Nations Unies.

questionnaires individuels doivent être placés dans une infrastructure de traitement des données sécurisée, distincte de l'infrastructure de collecte.

34. Un questionnaire traditionnel qui peut être téléchargé exige une infrastructure beaucoup moins lourde qu'un formulaire à remplir en ligne, mais aussi une meilleure maîtrise de la technologie de la part du répondant. Il ne sera pas nécessairement adapté aux différentes configurations informatiques possibles, et l'organisme national de statistique devrait être en mesure de faire face à chaque problème qui pourrait se poser. L'expérience récente a montré qu'il était beaucoup plus probable que les répondants préfèrent remplir le questionnaire en ligne. *C'est pourquoi il est recommandé d'utiliser un questionnaire en ligne.*

35. Pour que les répondants puissent remplir les questionnaires en ligne, il faut leur fournir un *moyen d'identification électronique*. Il existe différents moyens de communiquer l'identifiant et le mot de passe permettant d'accéder au questionnaire :

- a) Les faire figurer sur le questionnaire papier ou dans un courrier acheminé par la poste ;
- b) Les faire remettre par l'agent recenseur, directement à l'adresse postale du répondant ;
- c) Les envoyer par courrier électronique ;
- d) Les envoyer par SMS ;
- e) Utiliser les identifiants dont on se sert pour se connecter aux portails de services publics en ligne ou à d'autres services en ligne nécessitant un numéro d'identification personnel.

36. Le *questionnaire électronique*, contrairement au formulaire papier, permet de modifier la saisie, ce qui améliore la qualité des réponses. Les personnes qui remplissent un questionnaire électronique s'attendent à être guidées ; elles souhaitent au minimum que le questionnaire suive un ordre logique et qu'on ne leur pose pas de questions sans rapport avec leur situation. Pour que les données recueillies en ligne soient de qualité, il est important de prévoir des mécanismes permettant de détecter en temps réel les erreurs de réponse, que le répondant doit pouvoir modifier immédiatement. Si les réponses se contredisent, le répondant doit en être averti et avoir la possibilité de corriger une ou plusieurs réponses, de les supprimer ou de confirmer que le cas de figure existe bien dans la vie réelle, bien qu'il n'ait pas été prévu par les développeurs.

37. La possibilité d'utiliser Internet peut contribuer à améliorer la qualité du recensement en permettant à des groupes *difficiles à dénombrer* de répondre plus facilement. La plupart des pays signalent des difficultés de dénombrement, s'agissant par exemple des jeunes adultes et des personnes qui vivent dans des logements sécurisés, auxquels il est difficile d'avoir accès. En outre, il peut être plus facile pour certaines personnes handicapées de remplir un questionnaire en ligne qu'un formulaire papier. Il y a d'ailleurs plus de chances que ces catégories de personnes utilisent Internet à d'autres fins et, si cette option est disponible, il faut les encourager à participer au recensement de cette façon.

38. L'une des principales difficultés liées au questionnaire en ligne réside dans la disponibilité d'une *infrastructure* adéquate. Un recensement se déroule pendant une période relativement courte et vise l'ensemble de la population d'un pays, et il y a peu de chances que l'organisme national de statistique dispose déjà de l'infrastructure nécessaire pour faire face aux pics de demande qu'engendre une telle opération. Il est donc probable que cet élément au moins doive être externalisé. Il pourra alors être nécessaire de modifier les méthodes de collecte afin de moduler la demande. Par exemple, on peut restreindre le nombre d'utilisations simultanées du questionnaire en ligne en échelonnant dans le temps la remise des questionnaires ou des lettres d'invitation, ou en imposant aux personnes qui ne font pas partie des populations ciblées à une période donnée ou ne résident pas dans des zones déterminées de contacter l'organisme national de statistique avant d'utiliser ce questionnaire.

39. Il appartient par conséquent à l'organisme national de statistique d'évaluer de quelle façon il souhaite *promouvoir* l'utilisation des questionnaires en ligne. Il doit à cet effet prendre en compte la capacité de ses services à faire face à la charge attendue et assurer une coordination avec les autres méthodes de collecte adoptées. La stratégie en matière de

relations publiques devrait comporter des garanties quant à la sécurité et à la confidentialité des informations communiquées via Internet. Si la possibilité de répondre en ligne est offerte à tous, il est nécessaire, dans le cadre de cette stratégie, de satisfaire aux attentes de la population concernant la possibilité d'accéder au site aux moments où la demande est la plus forte. Des messages simples conseillant d'utiliser la possibilité de répondre par Internet pendant les heures creuses devraient être préparés et utilisés si nécessaire sur le site Web de l'organisme de recensement et par l'entremise de tout service de renseignements téléphoniques de l'organisme ou dans le cadre de la communication dans les médias.

40. On peut s'attendre à ce que, dans les recensements de 2030, les questionnaires en ligne soient davantage utilisés que pour ceux de 2020. Pendant la collecte des données, les organismes de recensement doivent suivre en permanence les taux de réponse du public et *s'efforcer d'augmenter* le nombre de réponses fournies en ligne si nécessaire.

2. Appareils portables

41. Puisque les ordinateurs portables, les tablettes et les smartphones sont de plus en plus élaborés et que les coûts de communication diminuent, ces appareils peuvent représenter une *solution économique* pour certains aspects de la collecte des données. Pour les agents recenseurs, ces appareils peuvent remplacer les cartes sur papier ainsi que les registres et les listes d'adresses et servir à collecter des données sur le terrain. Ils peuvent être utilisés dans le cadre de toute méthode de collecte, depuis la distribution et le ramassage jusqu'à la récupération des questionnaires remplis.

42. Les appareils portables présentent l'avantage de permettre la *gestion de l'information en temps réel et dans les deux sens*. Les gestionnaires du recensement peuvent être informés du progrès des opérations de collecte au fur et à mesure que les agents recenseurs livrent les formulaires de recensement et collectent ceux qui sont remplis. De même, les agents peuvent y trouver des versions actualisées des questionnaires reçus et des informations sur les ménages auprès desquels un suivi doit être assuré. En outre, les informations géospatiales nécessaires au recensement (par exemple, les adresses manquantes et les faits nouveaux) peuvent être communiquées afin de permettre une utilisation efficace des ressources. Les gestionnaires du recensement peuvent repérer en temps réel les secteurs où le dénombrement est en retard ou ne répond pas aux normes de qualité attendues, et prendre les mesures qui s'imposent.

43. L'utilisation d'appareils portables devrait permettre d'améliorer l'efficacité de la collecte de données. Lorsque l'on utilise ce type d'appareils, il faut tenir compte de plusieurs aspects techniques :

a) Du fait de la petite taille de l'écran, il peut être difficile pour les agents recenseurs d'enregistrer et de vérifier les réponses avec précision. Il y a également un risque de fragmentation des données ;

b) Les appareils compacts et légers dotés d'une capacité de stockage suffisamment grande sont ceux qui conviennent le mieux au travail des agents sur le terrain. La luminosité et le contraste de l'écran doivent être réglables afin que l'appareil puisse être utilisé aussi bien en pleine lumière que dans la pénombre ;

c) Pour la sécurité des données, celles-ci devraient être conservées dans les appareils le moins longtemps possible, de préférence pas plus de vingt-quatre heures ;

d) Il faut que les appareils puissent supporter une absence de connexion au réseau pendant un certain temps. L'autonomie de la batterie doit être évaluée en fonction de la charge de travail quotidienne du personnel de terrain. Il peut être utile de prévoir une batterie externe ;

e) Si des mises à jour des systèmes et des logiciels interviennent au stade de la collecte des données, il faut éviter que les données collectées se perdent ou qu'elles divergent par rapport aux données collectées après la mise à jour ;

f) La précision du GPS (par exemple, dans des zones urbaines densément peuplées) et la réception du signal mobile (par exemple, dans les zones montagneuses ou

forestières) peuvent être dégradées dans certaines régions du pays. Le degré de connectivité mobile doit être évalué, surtout si l'appareil utilisé est tributaire d'un système en ligne.

44. Toute solution faisant appel à des appareils portables doit être dûment *testée* avant d'être appliquée dans le cadre d'un recensement, aussi bien seule qu'en interaction avec les autres méthodes de recensement utilisées.

45. Pour utiliser des appareils portables, il faut aussi tenir compte de tout un ensemble de considérations liées à la *sécurité*.

a) Le risque de vol ou de perte est par exemple beaucoup plus grand avec ces appareils qu'avec des formulaires papier. Le téléchargement systématique des données vers un serveur devrait toutefois réduire au minimum la nécessité d'un nouveau dénombrement en cas de perte ;

b) Il faut prendre des mesures pour protéger le caractère confidentiel des données soit lorsqu'elles sont encore dans l'appareil si celui-ci venait à être perdu, soit au cours de leur transmission. Les données stockées sur ces appareils doivent être cryptées et leur accès doit être protégé (par exemple, par un mot de passe ou par des empreintes digitales) ;

c) Il faut également sécuriser la transmission des données en les cryptant et en utilisant des circuits sécurisés de bout en bout ;

d) Les appareils portables devraient être protégés par un logiciel de sécurité compatible avec les autres applications. Les logiciels de sécurité et les mots de passe compliquent toutefois l'utilisation de ce type d'appareils. Le coût de ces mesures de sécurité viendra s'ajouter aux dépenses d'appui ;

e) La formation et l'assistance technique destinées au personnel chargé du dénombrement revêtent une grande importance. On ne devrait pas attendre des personnes recrutées à cette fin qu'elles possèdent des compétences techniques. Ces facteurs deviennent de plus en plus complexes et difficiles à gérer à mesure que le nombre d'agents recenseurs et les distances physiques augmentent. Dans les pays de grande taille, il se peut que les agents recenseurs puissent se contenter d'une formation et d'une assistance technique fournies à distance par Internet ou par téléphone.

46. *Des outils de formation aux appareils portables doivent être téléchargés sur les appareils pour faciliter leur utilisation par les agents recenseurs pendant leur formation et leurs activités sur le terrain. Ces outils doivent couvrir tous les éléments du travail de l'agent, être interactifs et faciles d'utilisation, et contenir des exemples parlants de la conduite à adopter dans toutes les situations possibles d'utilisation de l'appareil.*

47. Les organismes de recensement doivent réfléchir à l'usage qui sera fait de ces nombreux appareils un fois le recensement terminé. Il est difficile de les stocker en vue du prochain recensement, car ils peuvent devenir obsolètes ou inutilisables s'ils n'ont pas été rechargés pendant une période allant de cinq à dix ans. Les organismes de recensement peuvent transférer certains de ces appareils (par exemple, à des organismes publics) et conserver les autres.

3. Téléphone

48. Par le passé, il a été envisagé, pour faire des économies, de procéder à des entretiens téléphoniques automatisés en utilisant un questionnaire abrégé pour recueillir uniquement des données essentielles d'ordre démographique. Cependant, aucun pays n'a eu recours à cette méthode lors du recensement de 2020. *Il n'est pas recommandé de recourir aux entretiens téléphoniques automatisés.*

49. Il est possible de recourir à la méthode des entretiens téléphoniques assistés par ordinateur (CATI) pour recueillir des données sur la base du questionnaire et/ou vérifier les données recueillies dans un long questionnaire en les complétant au besoin. Les pays doivent toutefois savoir que ces systèmes deviennent difficiles d'utilisation lorsque le nombre et la complexité des questions ou le nombre de personnes composant le ménage sont élevés.

4. Conception du questionnaire électronique

50. La conception du questionnaire électronique est un élément très important de la solution technologique lorsque les réponses sont fournies en ligne ou collectées à l'aide d'appareils portables.

51. Lors de l'élaboration du questionnaire électronique, il faut tenir compte de ce qui suit :

a) Le questionnaire doit contenir un ensemble complet de questions – aussi bien des questions ouvertes ou que des questions fermées – qui s'enchaînent dans un ordre logique ;

b) S'agissant des questions fermées, le questionnaire doit donner la possibilité de choisir une ou plusieurs réponses prédéfinies, et le répondant doit pouvoir saisir une réponse personnalisée en sélectionnant l'option « Autre » ;

c) Si possible, l'intégralité de la question et de la réponse doit tenir sur l'écran de l'appareil sans que le répondant n'ait besoin de faire défiler les éléments à l'écran ou de passer à l'écran suivant, car il pourrait passer à côté de la partie masquée de la question ou ne pas voir des choix de réponse ;

d) Il faut qu'une aide soit disponible, par exemple un indice textuel ou un lien vers l'élément approprié des métadonnées ou du support de formation ;

e) Chaque répondant et les membres du ménage auquel il appartient doivent pouvoir facilement naviguer d'une question à l'autre et d'une section du questionnaire à l'autre (s'agissant, par exemple, des conditions de logement et de la situation du ménage et du répondant concerné) ;

f) Des contrôles intégrés doivent permettre de valider les données saisies compte tenu des informations connues sur cette personne et d'autres membres de ce ménage ;

g) Une barre de progression doit s'afficher dans le questionnaire, de même que certaines informations quantitatives générales, par exemple le nombre de personnes composant le ménage et le pourcentage de progression des réponses pour chacune d'entre elles.

52. Lorsque le même questionnaire électronique est utilisé à la fois par les répondants et les agents recenseurs, l'interface graphique peut être différente, car les répondants n'ont aucune connaissance dans le domaine, alors que les agents recenseurs ont reçu une formation préalable et connaissent bien la terminologie employée et les métadonnées intégrées dans le questionnaire.

53. *Le questionnaire en ligne doit également contenir des explications de base à l'intention du répondant, par exemple :*

a) Une description de la structure générale du questionnaire et de l'ordre dans lequel il doit être rempli ;

b) Une estimation du temps dont aura besoin un répondant ou un ménage pour répondre aux questions ;

c) Une description des moyens d'accéder à l'aide en ligne et de réagir aux messages d'erreur ;

d) Une description de la manière dont le répondant peut au besoin corriger, supprimer ou ajouter des informations ;

e) L'indication de la réussite ou de l'échec du processus de réponse au questionnaire et les démarches suivantes à faire par le répondant (par exemple, pour obtenir une confirmation de sa participation au recensement) ;

f) Un moyen de contacter l'organisme national de statistique, tel que le numéro de téléphone ou l'adresse électronique de l'organisme ou du service d'assistance afin que le répondant puisse évaluer la qualité des services en ligne ou poser des questions auxquelles il n'a pas trouvé de réponse en remplissant le questionnaire ;

g) La possibilité d'obtenir une traduction du questionnaire dans les langues les plus courantes du pays ;

h) Il peut également être utile de fournir des réponses, en termes simples, aux questions fréquemment posées et de fournir un lien cliquable vers la page du site Web de l'organisme national de statistique où sont décrits les principes juridiques, méthodologiques et organisationnels du recensement.

5. Technologies visant à faciliter le dénombrement des personnes handicapées et des personnes non connectées

54. Lorsqu'ils adoptent de nouvelles technologies, les organismes de statistique doivent garder à l'esprit qu'il faut que le recensement *couvre l'entièreté de la population*, indépendamment des moyens techniques dont disposent les répondants et de leurs connaissances informatiques.

55. La technologie peut contribuer de *deux principales manières* au dénombrement des personnes handicapées et des personnes non connectées, à savoir : a) elle peut permettre aux organismes de statistique d'atteindre les répondants qui n'ont pas l'accès Internet nécessaire ; b) elle peut permettre au plus grand nombre de personnes possible de saisir leurs réponses par voie électronique.

56. *Personnes non connectées.* S'ils veulent atteindre le plus grand nombre de personnes possible, les organismes de statistique doivent tenir compte des zones dans lesquelles l'accès à Internet est limité. Dans ces cas, il est malheureusement nécessaire de prévoir un questionnaire papier. D'autres options sont néanmoins envisageables : par exemple, des agents disposant d'un accès à Internet par satellite pourraient recueillir les informations grâce à des appareils portables ou les répondants pourraient joindre un opérateur par téléphone qui les aiderait à remplir leur questionnaire.

57. *Réponse par Internet et accessibilité.* Il convient de tenir compte des difficultés que peuvent rencontrer les personnes handicapées lorsqu'elles répondent au questionnaire par Internet. La marche à suivre sur Internet doit respecter les normes définies dans les Règles pour l'accessibilité des contenus Web (WCAG) (version 2.2). Si les organismes de statistique doivent se mettre en conformité avec ces normes d'un point de vue technique, il est aussi important qu'ils tiennent compte de l'accessibilité dès l'élaboration de leurs contenus. Ils devraient ainsi prendre conseil auprès d'un organisme spécialisé dans ces problématiques et effectuer des tests d'expérience utilisateur au moment de concevoir de nouvelles fonctionnalités susceptibles d'avoir une incidence sur l'accessibilité, afin de veiller à les rendre pleinement accessibles. Les paragraphes qui suivent comportent des exemples des mesures qu'ils pourraient prendre.

58. *Texte invisible et masques de saisie introduisant automatiquement certains caractères.* Les masques de saisie permettent de faciliter la saisie de données dans un format prédéfini, en générant automatiquement des séparateurs (par exemple, barres obliques dans la date). Ils permettent également d'ajouter des instructions lisibles uniquement par les lecteurs d'écran, afin que les utilisateurs malvoyants puissent prendre connaissance des données à saisir et des éléments renseignés automatiquement par des commandes JavaScript.

59. *Contraste de couleurs nécessaire à la lisibilité du contenu textuel.* Pour que le contenu textuel soit lisible, il convient de mener une analyse destinée à vérifier que les couleurs utilisées (code RVB et code hexadécimal) présentent un rapport de contraste conforme à celui prescrit dans les WCAG 2.0 (niveau AA). Ce rapport doit être d'au moins 4,5:1 pour le texte normal et d'au moins 3:1 pour le texte agrandi (on entend par « texte agrandi » un texte de 14 points ou plus s'il est en gras (généralement 18,66 px) et de 18 points au moins s'il n'est pas en gras (généralement 24 px)).

60. *Utilisation d'étiquettes, des taquets de tabulation et d'attributs « aria-described by » destinés à représenter la hiérarchie entre les champs.* Il peut arriver, dans une enquête, que la présentation des questions soit complexe ; dans de tels cas, en se conformant aux normes WCAG concernant l'ordre séquentiel logique, on pourrait prévoir la génération automatique de taquets de tabulation (rectangles bleus) suivant les informations sur la page dans le sens de lecture, c'est-à-dire de gauche à droite, de haut en bas. En se conformant aux normes

WCAG concernant les étiquettes ou instructions, on pourrait associer à tous les champs nécessitant une saisie utilisateur (par exemple, bouton radio et cases à cocher⁵) des étiquettes lues par les lecteurs d'écran fonctionnant en mode formulaire. Enfin, en se conformant aux normes WCAG concernant l'information et les relations, on pourrait ajouter des attributs « *aria-described by* » à des zones de saisie, afin que le lecteur d'écran lise de manière séquentielle d'abord la case à cocher, puis le texte associé. L'utilisateur obtient ainsi des informations sur le contexte, qui l'aident à décider s'il doit cocher ou non la case.

6. Logiciel de gestion des recensements

61. À un niveau élémentaire, les opérations de collecte multimodales exigent la *fourniture d'informations en temps utile* aux agents recenseurs, pour qu'ils ne rendent pas visite à des ménages qui ont déjà remis un formulaire de recensement. Il s'agit d'une question d'efficacité mais aussi de relations publiques. Les nouvelles technologies permettent d'améliorer la gestion des opérations sur le terrain et donc la qualité du recensement lui-même.

62. Si la transmission en temps utile d'informations aux agents recenseurs est le point essentiel, un même système devrait aussi assurer la *circulation des informations dans les deux sens*, pratiquement en temps réel, *entre les gestionnaires du recensement et les agents*. Le suivi du travail de ces derniers permet davantage d'intervenir en temps opportun lorsque le travail de collecte des données prend du retard ou lorsque les données recueillies présentent des problèmes de qualité.

63. Les organismes nationaux de statistique peuvent devoir recourir à des entreprises extérieures pour certaines activités essentielles. Qu'ils soient internes ou externes, les systèmes utilisés doivent répondre à des *normes de cybersécurité* reconnues aux niveaux international (par exemple, famille de normes ISO 27000), régional (par exemple, NIS2⁶ dans l'Union européenne) ou national. Cet aspect est d'autant plus important qu'il existe des solutions logicielles diverses (guichet unique, licence propriétaire, développement sur mesure, par exemple) pour plusieurs opérations menées en interne ou en externe (collecte, traitement, systèmes d'information géographique, imputation, diffusion).

64. Un système intégré de communication sur le terrain peut utiliser une infrastructure informatique déjà existante. Il est par exemple possible de prévoir un accès direct à l'infrastructure nationale de géocodage, qui contient des données relatives aux adresses individuelles. *Les informations géospatiales (adresses, bâtiments, cadastre) devraient être assorties d'identifiants pérennes.*

65. *Toute infrastructure de gestion du recensement devrait être composée des éléments suivants :*

- a) Un registre des adresses des habitations, assorties de coordonnées géospatiales et réparties par district de recensement ;
- b) Un registre des agents recenseurs comprenant leurs coordonnées et permettant d'associer à chacun d'entre eux un district de recensement donné (issu du registre des adresses des habitations) et les adresses correspondantes (dans le cas où cette méthode de collecte de données est utilisée) ;
- c) Un registre des appareils utilisés pour la collecte des données (par exemple, tablettes numériques ou smartphones) dans lequel figure leur numéro de série unique et permettant d'attribuer chaque appareil à un agent recenseur (issu du registre des agents recenseurs) (dans le cas où cette méthode de collecte de données est utilisée) ;
- d) Un centre de stockage des données de recensement dans lequel sont collectées, traitées et rassemblées tous les données relatives aux adresses des répondants (provenant du registre des adresses des habitations), obtenues grâce aux différentes méthodes de collecte

⁵ World Wide Web Consortium (W3C), 2023. [Web content accessibility guidelines \(WCAG\) 2.2](#) (Règles pour l'accessibilité des contenus Web, version 2.2). Recommandation du W3C datée du 5 octobre 2023.

⁶ Directive (UE) 2022/2555 du Parlement européen et du Conseil du 14 décembre 2022 [concernant des mesures destinées à assurer un niveau élevé commun de cybersécurité dans l'ensemble de l'Union.](#)

utilisées au cours du recensement (réponses en ligne, collecte par les agents recenseurs, sources administratives). La méthode de collecte doit être renseignée ;

e) Un logiciel permettant aux agents recenseurs, aux superviseurs et à l'équipe de gestion du recensement de communiquer rapidement.

66. Toutes les composantes de cette infrastructure devraient *être connectées entre elles et gérées depuis un point central* à l'aide de logiciels et d'outils technologiques spécialement mis au point pour les besoins du recensement. *Il est recommandé de procéder à l'initialisation, à la configuration et au contrôle des performances des appareils depuis ce point central avant de confier ces derniers au personnel de terrain.*

67. La gestion centralisée des appareils portables destinés à la collecte de données de recensement implique *l'automatisation des fonctionnalités ci-après* (dans le cas où cette méthode de collecte de données est utilisée) :

a) Installation, sur les appareils, des éléments suivants : logiciel de collecte de données utilisé par les agents recenseurs (questionnaire électronique), liste des adresses des habitations et cartes du district de recensement, métadonnées et éléments de classification utilisés dans le questionnaire électronique, outils permettant de contrôler le fonctionnement des appareils ou de suivre le travail des agents recenseurs, supports de formation à destination des agents recenseurs ou des utilisateurs de l'appareil ;

b) Établissement d'un lien unique entre appareil, agent recenseur et district de recensement dans les registres correspondants, et mise à jour des registres et du lien en cas de remplacement de l'agent recenseur ou de l'appareil ;

c) Gestion et suivi en ligne des opérations effectuées depuis tous les appareils, après leur initialisation par le personnel de terrain ;

d) Suppression, sur les appareils, de toutes les informations et de tous les logiciels utilisés après le transfert des données collectées vers le centre de stockage à la fin du recensement et préparation des appareils en vue de leur conservation à long terme ou de leur utilisation à d'autres fins (par exemple, transfert à un autre organisme).

68. *La gestion et le suivi en ligne des opérations effectuées depuis les appareils* couvrent :

a) L'obtention d'informations sur l'heure à laquelle l'appareil a été allumé et éteint ;

b) L'obtention des coordonnées géographiques pour chaque adresse où un questionnaire électronique doit être rempli et de statistiques concernant la saisie (par exemple, temps de saisie, nombre de réponses données, erreurs et corrections apportées) ;

c) La transmission, aux appareils, d'informations concernant l'URL des questionnaires en ligne et de données de suivi concernant les répondants, qui permettent aux agents recenseurs de vérifier que le questionnaire a bien été rempli ou d'apporter des corrections aux données à la demande du répondant, si nécessaire ;

d) L'installation de mises à jour logicielles à distance en cas d'urgence. Cette installation n'est recommandée que lorsque des erreurs logicielles majeures sont détectées en cours de recensement, puisqu'il est impératif de garantir la cohérence des données recueillies et que l'utilisation de différentes versions logicielles rend la tâche plus ardue ;

e) Le verrouillage de l'appareil en cas de perte ou de vol, de manière à empêcher son utilisation illégale ou la fuite d'informations ;

f) L'établissement d'un moyen de communication à distance entre l'agent recenseur et le bureau central de recensement.

69. *Les outils de gestion du stockage des données employés par le bureau central de recensement* intègrent les fonctionnalités suivantes :

a) Téléchargement des formulaires remplis en ligne par les répondants (dans le cas où cette méthode de collecte de données est utilisée) ;

b) Vérification de la conformité des formulaires aux fins de leur traitement (le répondant a bien renseigné les informations nécessaires à son identification) ;

c) Établissement d'un lien entre les formulaires remplis et les adresses des habitations ;

d) Création d'une confirmation de participation invitant le répondant à donner son avis (par exemple, code QR envoyé sur son compte personnel, par courrier électronique ou par un autre moyen, en fonction de la méthode qu'il a utilisée pour retourner le formulaire complété), ensuite transmise à l'agent recenseur (dans le cas où cette méthode de collecte de données est utilisée) ;

e) Depuis l'appareil utilisé par l'agent recenseur, depuis des sources administratives et depuis d'autres moyens de collecte des données utilisés, téléchargement des questionnaires électroniques complétés et établissement d'un lien avec les adresses des habitations ;

f) Fusion des versions des formulaires complétés concernant la même adresse d'habitation, obtenues par tous les moyens de collecte de données utilisés (choix de la version de référence, enrichie des données provenant d'autres versions, et suppression des doublons, le cas échéant) ;

g) Élaboration de statistiques sur l'avancement du recensement et suivi des travaux afin de repérer les districts de recensement où des problèmes se posent et où une intervention est nécessaire en cours de recensement.

70. Les *sciences comportementales* peuvent en outre apporter un éclairage sur les motivations, les sentiments et les expériences des répondants et ainsi contribuer à augmenter le taux de réponse. Par exemple, elles permettent de reconnaître le moment opportun pour procéder au recensement, de mieux communiquer avec les répondants et de rendre les enquêtes en ligne ou sur papier plus faciles à comprendre.

71. Les relations avec les principaux *partenaires technologiques* constituent un autre facteur clé de réduction des risques. Une gouvernance solide, assurée principalement par l'équipe interne de gestion, est essentielle si l'on veut que les systèmes de recensement soient conçus et mis en place correctement, et qu'ils donnent de bons résultats.

72. Il existe désormais des solutions technologiques capables de combiner *plusieurs fonctionnalités de gestion sur le terrain*. Ainsi, les progiciels de gestion de la relation client peuvent permettre de gérer les demandes sur le terrain, de concevoir un site Internet, d'améliorer la communication grâce à un dialogueur (*chatbot*) et de mettre en place un service d'assistance aux utilisateurs, regroupant les informations recherchées.

73. Les organismes nationaux de statistique pourraient envisager de recourir à des *technologies d'accès à distance* afin de permettre aux membres du personnel chargé du traitement des données de recensement d'aménager leurs modalités de travail sans pour autant mettre à mal la sécurité et la confidentialité des données.

D. Traitement des données

74. Ces dernières années, les *technologies de traitement ont fait de grands progrès*, qui ont réduit le coût du traitement pour les organismes de recensement et amélioré la qualité des données.

75. En se fondant sur les conclusions tirées d'une analyse des pratiques des pays lors de la série de recensements de 2020, on peut supposer que la plupart des pays ayant recours au dénombrement direct utiliseront l'option Internet lors de leurs prochains recensements. On peut également supposer qu'ils auront un recours limité aux formulaires papier et aux technologies de reconnaissance optique de caractères et des marques.

76. Il sera nécessaire de planifier dans le détail *l'intégration entre les principaux systèmes de recensement* afin de faire en sorte que le parcours des données soit fluide de bout en bout, de leur collecte à leur traitement. Il sera en outre essentiel d'instaurer une gouvernance solide auprès des différents partenaires susceptibles d'être concernés, si l'on veut réussir l'intégration entre les systèmes de terrain, de traitement et de diffusion.

1. Traitement des questionnaires papier

77. Il peut être nécessaire de recourir à des questionnaires papier pour recueillir des données, soit parce que les répondants préfèrent ce mode de réponse, soit parce qu'ils n'ont pas accès à Internet. *Il est recommandé, aux fins du traitement de ce type de questionnaires, d'utiliser des procédés automatisés, tels que la reconnaissance intelligente des caractères.*

78. La reconnaissance optique de marques peut être une solution économique lorsque le formulaire de recensement ne contient que des *cases à cocher*. D'autres moyens de saisie des données ou de codage assisté par ordinateur sont nécessaires pour la saisie de réponses écrites. La reconnaissance optique de marques a toutefois été supplantée dans une large mesure par la reconnaissance intelligente de caractères sous ses différentes formes.

79. L'option la plus économique consiste probablement à *combiner l'imagerie numérique, la reconnaissance intelligente de caractères, la réparation de données et le codage automatique*. Elle est brièvement décrite ci-après :

a) Les formulaires de recensement sont traités par des numériseurs qui produisent une image. Le logiciel de reconnaissance est utilisé pour identifier les réponses données à l'aide de cases à cocher et traduire en caractères d'imprimerie les réponses manuscrites. On définit des niveaux de confiance pour déterminer quelles réponses sont d'une qualité acceptable et quelles sont celles qui nécessitent une réparation ou une validation ;

b) La réparation automatique vise à réduire la nécessité d'une intervention humaine et comprend généralement l'utilisation de tableaux de recherche dans un dictionnaire et d'une édition contextuelle. Les dictionnaires varient en fonction des questions. Ainsi, le dictionnaire concernant la question relative au pays de naissance ne contient que des noms de pays. Le travail de préparation à l'élaboration de dictionnaires de termes en langues naturelles augmentera considérablement l'efficacité du codage ;

c) Il est possible de procéder à une réparation manuelle des images non reconnues. Cette méthode ne vaut que pour les questions à propos desquelles les données réparées pourront ensuite très probablement faire l'objet d'un codage automatique ;

d) Le codage automatique utilise des algorithmes informatisés pour établir une concordance entre les réponses saisies et les index. Les réponses pour lesquelles il n'est pas possible d'établir une concordance font l'objet d'un codage assisté par ordinateur. Dans ce cas, il est recommandé d'utiliser un algorithme d'apprentissage automatique, qui pourrait remplacer une intervention humaine et permettre d'obtenir des données de bonne, voire meilleure qualité tout en réduisant considérablement les coûts. Les données issues de recensements antérieurs peuvent être utilisées pour entraîner l'algorithme, de même que des données du cycle d'essais en cours, en particulier si de nouvelles variables doivent être codées. Pour limiter encore les coûts et améliorer la qualité, les réponses qui ne sont pas codées peuvent être analysées en quête de réponses courantes. Ces réponses peuvent soit être ajoutées aux index de codage et soumises une nouvelle fois par codage automatique, soit faire l'objet d'un autre type de codage en bloc.

80. D'autres considérations sur le recours à l'imagerie numérique, à la reconnaissance intelligente de caractères, à la réparation des données, au codage automatique ainsi qu'à la reconnaissance optique des marques et des caractères sont présentées dans les *Recommandations de la Conférence des statisticiens européens pour les recensements de 2020*⁷. On peut s'attendre à ce que l'intelligence artificielle générative ouvre de nouvelles possibilités et remplace les clavistes. Cependant, il faut investir dans l'élaboration de modèles appropriés si l'on veut maintenir une qualité élevée de données.

2. Produits

81. Les produits du recensement sont communément des tableaux, des statistiques, des illustrations et des cartes accompagnés des métadonnées correspondantes. (Voir [renvoi au chapitre concerné]).

⁷ CEE, 2015. [Recommandations de la Conférence des statisticiens européens pour les recensements de la population et des habitations de 2020](#). Genève : Nations Unies.

82. La *diffusion en ligne* permet en outre la conception de produits répondant mieux aux besoins de différents types d'utilisateurs, la diffusion avantageuse d'une gamme beaucoup plus large de données de recensement et l'élargissement des possibilités d'utilisation de ces données. Ces possibilités pourraient être encore élargies grâce à la mise à disposition d'interfaces de programmation d'applications (API).

83. La fonctionnalité et le contenu des données peuvent viser à répondre aux *besoins d'utilisateurs de différents niveaux de compétence*. Cette fonctionnalité doit être fluide de bout en bout, quel que soit le niveau de difficulté de la question, les utilisateurs étant guidés par la nature de cette dernière ou l'analyse qu'ils souhaitent entreprendre au moyen des différents produits.

84. L'un des principaux objectifs du recensement est de produire des informations concernant *de petites zones géographiques ou de petits groupes de population* (sur le plan tant social qu'économique). La diffusion par Internet se prête aux deux types d'utilisation. S'agissant des petites zones, il est possible de recourir aux systèmes d'information géographique pour définir celles qui sont pertinentes lors de la recherche de données et pour cartographier les résultats de la recherche. Il existe divers progiciels qui permettent d'extraire les informations sur les populations étudiées à partir de grandes matrices prédéfinies.

85. Le système de diffusion par Internet doit donner aux utilisateurs une certaine *souplesse* et leur permettre d'exporter les résultats dans divers progiciels courants d'analyse statistique, de présentation en tableaux ou de cartographie.

3. Plateformes interactives en ligne aux fins de la diffusion

86. La quantité de données spatiales, temporelles et sémantiques recueillies augmente (en raison, par exemple, de la grille de référence de 1 km² utilisée par l'Union européenne), de même que le recours à des *moyens de visualisation*, destinés à aider les particuliers, les chefs d'entreprise, les universitaires et les responsables à tous les niveaux à comprendre les principales informations tirées de ces données. Ces outils⁸ peuvent être utilisés pour communiquer rapidement une information, présenter plus facilement une grande quantité de données, faire apparaître les structures et les liens entre les données et suivre l'évolution de variables dans le temps.

87. *Il est recommandé de diffuser les données sur des plateformes interactives en ligne, grâce auxquelles différents groupes d'utilisateurs peuvent directement générer des tableaux et des supports de visualisation (par exemple, graphiques, cartes) en fonction de leurs besoins.*

88. Il convient de veiller à appliquer les principes FAIR de gestion et d'intendance des données scientifiques (*FAIR Guiding principles for scientific data management and stewardship*)⁹. Ces principes mettent l'accent sur la capacité des systèmes informatiques à trouver des données, à y accéder, à les échanger et à les réutiliser sans intervention humaine ou avec une intervention humaine minimale (acronyme FAIR, de l'anglais *Find, Access, Interoperate, Reuse*). Les solutions analytiques proposées sur les plateformes peuvent demeurer limitées, car l'analyse n'est pas l'objectif principal des recensements.

89. Les utilisateurs de données de recensement présentent des niveaux différents de culture statistique et de compétences informatiques. *Il est recommandé de tenir compte des groupes d'utilisateurs suivants :*

- a) Les utilisateurs moins avertis, qui souhaitent principalement visualiser les données ou les points saillants du recensement ;
- b) Les utilisateurs intermédiaires, qui personnalisent les supports de visualisation ou sélectionnent des données particulières ;
- c) Les utilisateurs expérimentés sur le plan technique, qui téléchargent les données pour mener leur propre analyse et utilisent des API.

⁸ Par exemple, [Eurostat Census Hub](#).

⁹ [GO FAIR Initiative](#).

90. Les *normes internationales*, telles que la norme SDMX (Données statistiques et échange de métadonnées), doivent servir en priorité au formatage des données de sortie et des métadonnées mises à la disposition des utilisateurs.

91. Quels que soient les moyens d'accès ou de diffusion, la *protection de la confidentialité statistique* des données du recensement apparaît essentielle et des procédures de contrôle de la divulgation des statistiques doivent être mises en place.

4. Stockage et archivage

92. Les technologies de stockage des données ont progressé au même rythme que les autres évolutions techniques. Auparavant, le stockage avait lieu sur site, à l'aide de supports traditionnels à bande parfois complétés par des disques. Toutefois, au vu de la croissance exponentielle du volume de données, il a fallu adapter les capacités de stockage. De nombreuses entités envisagent désormais de passer à un *stockage hybride*, optant pour un stockage sur site et dans le *cloud* adapté à leurs besoins.

93. Les organismes de statistique devront prendre certains facteurs en compte dans le *choix de leur stratégie de stockage*, à savoir :

- a) L'adoption de solutions de stockage hors site comporte des risques ;
- b) De nouvelles compétences seront nécessaires pour mettre au point des solutions avancées de stockage et pour les mettre à jour, dans le cas d'une approche hybride par exemple. Les personnes disposant de ces compétences peuvent être difficiles à trouver et coûteuses à retenir ;
- c) Il leur sera essentiel de comprendre la manière dont leurs données sont classifiées, car cela facilitera la prise de décisions visant à renforcer leurs capacités de conception du stockage ;
- d) Lorsqu'ils envisagent de nouvelles solutions de stockage, il leur faut prendre des mesures de cybersécurité visant à identifier et détecter les risques et menaces, à garantir une protection contre ces derniers, à réagir en cas d'attaque ou d'incident et à assurer le retour à la normale. Leur stratégie de cyberdéfense devra prévoir des mesures de protection du stockage dans un environnement hybride ;
- e) Il existe de nombreuses méthodes de stockage en nuage, notamment des sites permettant l'archivage à long terme de données qui ne sont pas fréquemment utilisées (en anglais, *glacier storage*). Là encore, les organismes qui ont toujours opté pour des sauvegardes complètes incrémentielles ou différentielles de l'ensemble de leurs données devront mieux comprendre celles-ci s'ils veulent tirer parti de ces nouvelles méthodes ;
- f) S'ils confient le stockage à des prestataires externes, les organismes devront veiller à définir des accords de niveaux de services adaptés, qui répondent à leurs attentes en matière de prestation de services ;
- g) Le stockage et l'archivage doivent respecter les législations et règles en vigueur. Par exemple, dans l'Union européenne, toute donnée scientifique manuscrite doit être conservée pendant dix ans¹⁰ ;
- h) Les organismes doivent mettre en place et tester des procédures de sauvegarde et de récupération, qui leur permettent de prendre toutes les mesures nécessaires pour redémarrer les systèmes et les services, afin de vérifier le fonctionnement de ces derniers et l'intégrité des données ;
- i) Il est recommandé pour les organismes de mettre en place et de tester un plan de continuité des opérations axé sur la prévention et sur la récupération des systèmes et des données correspondantes en cas de crise telle qu'une catastrophe naturelle ou une cyberattaque.

¹⁰ Décision (UE) 2021/2121 de la Commission du 6 juillet 2020 [concernant la gestion des documents d'activité et les archives](#).

5. Informatique en nuage (*cloud*)

94. L'informatique en nuage progresse rapidement et devrait devenir *l'infrastructure la plus répandue dans les entreprises* du monde entier. Les stratégies d'adoption de cette pratique sont variées, certains optant pour une migration intégrale vers le nuage (*cloud-only*), d'autres donnant la priorité à l'adoption de technologies d'informatique en nuage (*cloud-first*) et d'autres encore, plus récemment, optant pour une combinaison réfléchie d'infrastructures cloud et sur site (*cloud-smart*)¹¹.

95. Les administrations publiques du monde entier sont de plus en plus nombreuses à se tourner vers l'informatique en nuage, reconnaissant qu'elle peut véritablement renforcer leur efficacité opérationnelle, leur capacité à transposer des activités à plus grande échelle et leurs prestations de services. En intégrant ces technologies, elles peuvent mieux utiliser les ressources dont elles disposent, rendre les données plus accessibles et encourager l'innovation dans différents domaines.

96. L'informatique en nuage a évolué pour inclure de multiples plateformes et modèles de services. On compte, parmi ces modèles, l'infrastructure en tant que service, la plateforme en tant que service et le logiciel en tant que service, et, parmi ces plateformes, le nuage public, le nuage privé et le nuage communautaire. L'adoption et le déploiement de chacun de ces modèles comportent leur lot d'avantages et de difficultés pour les organismes de statistique.

97. Les organismes de statistique doivent *tenir compte de nombreux aspects* avant de passer à l'informatique en nuage. Par exemple, les prestataires mondiaux de services d'informatique en nuage (dits « *hyperscalers* ») ont recours à des contrats généraux qui peuvent ne pas respecter la législation et les règles locales. Les organismes de statistique doivent donc réfléchir à la question de la compétence juridictionnelle en cas de différend. Il leur faudra passer des accords particuliers, convenus au niveau national, ce qui contribuera à lever d'importants obstacles à l'adoption du *cloud*. La souveraineté numérique est également un élément important qui doit faire l'objet d'accords.

98. En 2024, la CEE a publié des *lignes directrices destinées à encourager l'adoption de l'informatique en nuage*, dans lesquelles elle examine un grand nombre des principaux thèmes, avantages et défis concernant cette technologie¹². Ces lignes directrices peuvent aider les organismes de statistique à mieux comprendre les enjeux qui concernent leurs activités.

99. Une *approche graduelle*, par étapes, pourrait être bénéfique à de nombreux organismes de statistique souhaitant adopter le *cloud*. S'ils commencent par des applications et systèmes pour lesquels les enjeux sont moins cruciaux et qu'ils tirent les enseignements voulus de cette expérience, ils pourront être plus confiants dans leur décision de passer à l'informatique en nuage. Leur transition sera d'autant plus simple s'ils veillent à évaluer et mesurer les risques en permanence.

100. L'infrastructure en nuage présente d'importants avantages *dans le cas où les activités de recensement s'inscrivent dans un cycle de cinq à dix ans*. Du fait de son caractère modulable, elle peut facilement s'adapter aux besoins. Les programmes de recensement pourraient ainsi être moins coûteux, mais les organismes de recensement doivent néanmoins rester prudents et tenir compte des implications plus larges. Ils pourraient par exemple réutiliser à d'autres fins l'infrastructure de stockage sur site établie pour le recensement, ce qui permettrait de réduire les coûts du programme, mais il leur faut veiller à bien prendre en compte tous les facteurs en jeu. Dans le cas où le recensement s'inscrit dans un cycle plus court (recensement annuel, par exemple), une autre approche serait peut-être préférable.

101. Des *avantages* peuvent aussi être tirés de l'investissement dans les nouvelles technologies d'informatique en nuage, telles que les applications conçues et déployées dans le *cloud* (*cloud-native*) et les services de plateforme en ligne. Le recours à une architecture multi-nuages présente également un intérêt, car elle permet d'exploiter les atouts des différents prestataires de services d'informatique en nuage. Il serait ainsi possible qu'une

¹¹ CEE, 2024. [Cloud for official statistics](#) (L'informatique en nuage au service de la statistique officielle). Genève : Nations Unies.

¹² Ibid.

infrastructure soit consacrée au logiciel de gestion du recensement et une autre à l'hébergement de la base de données.

102. Le recours à une infrastructure en nuage offre une grande souplesse, mais il doit être *rigoureusement encadré*, sous peine de payer pour des installations inutilisées. À défaut, le *cloud* peut se révéler l'option la plus coûteuse. Il convient de réduire la capacité des serveurs en nuage lorsqu'ils ne sont pas utilisés, puis de l'augmenter au gré des besoins.

103. La migration vers l'informatique en nuage doit être bien évaluée et bien planifiée. Il ne serait pas recommandé de l'entreprendre simplement pour changer de centre de données. Dans certains pays, il a été prouvé qu'une telle migration était plus coûteuse que le maintien sur site. Les pays devraient s'efforcer d'adopter une approche fondée sur le principe *cloud-smart*, c'est-à-dire de passer à l'informatique en nuage en faisant migrer les applications pertinentes vers le *cloud* ou en tirant parti des possibilités offertes par les applications *cloud-native* (conteneurisation, service de plateforme en ligne, intégration et déploiement continus).

104. *Les avantages du recours à l'informatique en nuage aux fins du recensement peuvent être considérables, mais ce n'est pas toujours la meilleure approche. Il est recommandé de procéder à une évaluation approfondie avant d'entamer la transition.*

E. Technologies appliquées à l'utilisation des données administratives

1. Champ d'application

105. Les technologies appliquées à l'utilisation des données administratives sont différentes de celles appliquées à la collecte de données sur le terrain. Grâce à l'évolution technique des outils informatiques, il est de plus en plus facile d'utiliser les registres administratifs pour les recensements de la population et des habitations. Compte tenu du développement de technologies de pointe en la matière et de l'engagement pris par les organismes de statistique d'appliquer des solutions innovantes dans le cadre des recensements du cycle de 2030, il faudra nécessairement créer une infrastructure matériels/logiciels ou moderniser l'infrastructure existante aux fins de la collecte, du stockage et du croisement des données issues des sources administratives et du stockage des métadonnées sur les processus et les produits.

106. De la qualité des données sources dépend en grande partie celle des données produites (produits). Il est donc essentiel d'*améliorer la qualité des données provenant de sources administratives*, c'est-à-dire de les adapter aux besoins de la statistique. Des outils informatiques de pointe pourraient ici se révéler très utiles et jouer un rôle clé dans l'amélioration de l'efficacité et de la rentabilité de ces opérations. Pour évaluer la qualité des sources administratives utilisées dans les recensements, il convient de consulter les directives sur la question¹³ publiées par la CEE en 2021.

107. Au stade de la préparation d'un recensement, en particulier lors de la conception, il importe de préciser les *conditions techniques* d'utilisation des données issues de registres administratifs et dont peut dépendre le besoin de moderniser ou non l'infrastructure existante, *s'agissant* :

- a) De la collecte de données ;
- b) Du stockage des données ;
- c) Du couplage des données ;
- d) Du stockage des métadonnées relatives aux processus et aux produits.

108. L'application de plusieurs techniques de collecte des données issues de registres administratifs et d'autres sources pour utilisation dans les recensements de la population et des habitations nécessitera *une organisation et une gestion* plus globales et introduira

¹³ CEE, 2021. [Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses](#) (Directives pour l'évaluation de la qualité des sources administratives utilisées dans les recensements). Genève : Nations Unies.

davantage de complexité dans les systèmes. Les nouvelles technologies offrent des possibilités d'amélioration dans ce domaine également. Le processus de collecte des données issues de registres administratifs devrait comporter l'élaboration d'une stratégie faisant appel à différents modes de collecte.

2. Sécurité et confidentialité des données

109. Il est essentiel de tenir compte de l'*importance croissante* accordée à la sécurité, la confidentialité et la protection des données dans la société, comme en témoignent les cadres juridiques et statistiques de nombreux pays. À l'échelle de l'Union européenne, les États membres doivent en outre se conformer au règlement général sur la protection des données (RGPD)¹⁴. Ainsi, les organismes effectuant un recensement de la population et des habitations fondé sur les registres doivent respecter des exigences accrues concernant la collecte et le traitement des données. Pour répondre à ces exigences, il leur est recommandé d'appliquer le principe du besoin d'en connaître et de privilégier le traitement de données anonymisées.

110. La collecte de données à partir des registres administratifs suppose l'existence d'une infrastructure informatique sécurisée. En effet, la protection des données constitue un enjeu majeur de cette collecte. Quelle que soit la technique employée, la stratégie de collecte doit garantir la *sécurité de l'information*. Cette nécessité doit être prise en compte dès la conception à la fois du processus d'obtention et de collecte des données issues de registres et de l'infrastructure relative aux matériels et logiciels. Les aspects techniques liés au codage de la transmission des données doivent être étudiés en détail, en même temps que l'utilisation de voies de transmission sécurisées.

111. Des *mesures techniques et organisationnelles appropriées* doivent être prises de manière à protéger les données contre la destruction accidentelle ou illicite, la perte accidentelle (y compris des sauvegardes), l'altération, ou la divulgation ou l'accès non autorisés.

112. La sécurité doit être assurée à *plusieurs niveaux*. Les propriétaires de registres, les organismes nationaux de statistique collectant des données et les autres partenaires doivent établir des voies de transmission sécurisées et délivrer des certificats d'accès aux responsables de la collecte.

113. En organisant les données en fonction de leur « état » (données sources, données préparées, données statistiques et données de sortie), l'organisme national de statistique peut *gérer efficacement les autorisations* accordées à différents acteurs et équipes en son sein. L'accès aux données doit se faire sur la base du besoin d'en connaître ; elles doivent être associées à des identifiants anonymisés, et des mesures de contrôle relatives à la divulgation doivent être appliquées à tous les niveaux d'agrégation.

3. Couplage, transfert et stockage des données

114. Les nouvelles technologies sont indéniablement utiles pour croiser les fichiers et les données. Après avoir répertorié les sources administratives à utiliser pour mener le recensement, il serait nécessaire de les cartographier et de créer des API permettant d'assurer un *flux automatique de données* vers le bureau central.

115. Les sources administratives sont différentes selon les domaines, tant au sein d'un pays qu'entre les pays. Il est conseillé, dans la mesure du possible, d'utiliser des API uniformisées. Indépendamment de l'interface utilisée ou des différences rencontrées, il est très important *de disposer de métadonnées de qualité et de bien comprendre les données* pour pouvoir les utiliser aux fins du recensement.

116. De nombreux pays sont en train de passer au stockage sur le *cloud* et à l'informatique en nuage. Contrairement aux solutions traditionnelles sur site, où les données sont stockées, structurées et mises à jour de manière thématique, cette évolution offre des possibilités

¹⁴ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 [relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données](#).

nouvelles, notamment des environnements de données plus flexibles dont les variables peuvent être contrôlées, dits « *lac de données* ». Ces lacs de données requerront des quantités importantes de métadonnées, surtout pour la précision du classement chronologique, mais de telles solutions deviennent possibles grâce à la capacité de traitement accrue de l'informatique en nuage. La centralisation des données administratives dans des lacs de données faciliterait le fonctionnement opérationnel et pourrait fournir une infrastructure solide et fiable pour le stockage, la création de métadonnées et la synchronisation des différentes sources de données.

4. Amélioration des données administratives

117. De nombreuses techniques permettent de convertir automatiquement les données administratives en données statistiques. Une fois qu'une procédure de *nettoyage automatique des données* a été mise en place, elle permet d'éliminer les erreurs dans les données sources issues de registres administratifs et de modifier ces données efficacement.

118. Des *modèles d'apprentissage automatique* pourraient être utilisés, par exemple, pour déterminer l'« adresse de recensement » de la personne ou le type de ménage privé. Pour ce faire, il faudrait développer un réseau de neurones profond entraîné par un système d'intelligence artificielle.

119. L'intelligence artificielle (IA) pourrait être utilisée pour déterminer dans quelle branche économique et quelle catégorie d'emploi l'activité professionnelle du répondant peut être classée. Une équipe informatique composée d'ingénieurs de données et de spécialistes de l'IA pourrait être constituée pour développer des modèles de programmation neurolinguistique et de réseaux de neurones profonds. Ces modèles de classification neurolinguistique doivent constamment être mis à jour et leurs résultats soumis à des analyses de qualité. La classification à l'aide de l'IA peut se révéler plus efficace, donner de bons résultats et faire gagner du temps (90 % des cas couverts). En raison des exigences de sécurité et de confidentialité à respecter, il n'est pas toujours possible de recourir à ces données et modèles dans le nuage. On sait aussi qu'il peut être difficile d'utiliser des modèles de classification développés en anglais dans d'autres langues.

IV. Conclusion

120. Le projet de recommandations sur la technologie appliquée aux recensements de la population et des habitations de 2030 est présenté pour observations et examen.