



Commission économique pour l'Europe

Conférence des statisticiens européens

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

Genève, 21-23 septembre 2022

Point 2 de l'ordre du jour provisoire

Enseignements à retenir du cycle des recensements de 2020**Sources administratives et sources de parodonnées de terrain
au service de l'assurance qualité et de la gestion des imprévus****Note de l'Office for National Statistics du Royaume-Uni****Résumé*

Avant le recensement de 2021 en Angleterre et au Pays de Galles, nous avons présenté nos propositions concernant l'utilisation des données administratives dans le cadre de notre plan de recensement, ainsi que les informations de gestion supplémentaires obtenues grâce à notre système de réponse en ligne et aux innovations technologiques en matière de suivi sur le terrain. Nous avons été en mesure d'exploiter les données recueillies de diverses manières, y compris à certaines fins que nous n'avions pas envisagées mais que la pandémie de COVID-19 a rendues nécessaires.

Le présent document rend compte de l'utilisation qui a été faite des différentes sources, notamment s'agissant de vérifier si la méthode d'estimation de la couverture était biaisée ou encore d'évaluer les stratégies de gestion des imprévus et de les mettre en place dans toute une série de domaines.

* Document établi par Cal Ghee.

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. Les sources administratives et d'autres sources ont joué un rôle important dans certaines étapes des recensements précédents. Elles ont, par exemple, été largement utilisées aux fins de la planification des opérations et de l'assurance qualité du recensement de 2011. Pour 2021, nous avons défini à l'avance les utilisations supplémentaires que nous recherchions entre le début et la fin de l'opération de recensement (UNECE, 2018).

2. Au Royaume-Uni, le recensement se compose de trois opérations distinctes. L'Écosse et l'Irlande du Nord ont établi leur propre plan de recensement, axé sur les méthodes et les données les plus adaptées à leur territoire. Le présent document ne porte que sur les activités de l'Office for National Statistics relatives au recensement en Angleterre et au Pays de Galles.

II. Cadre général

3. Au Royaume-Uni, le recensement repose principalement sur des données fournies par chaque personne et par chaque ménage suivant le principe de l'autodéclaration. Une lettre d'invitation contenant un code d'accès en ligne unique, ou un questionnaire papier où figurait ce même code, ainsi que des instructions permettant de communiquer ses informations en ligne, par courrier ou, si nécessaire, par téléphone en prévision du jour du recensement a été envoyée à chaque adresse. Tous les ménages n'y donnant pas suite immédiatement, plusieurs moyens de relance ont été déployés aux fins de la collecte, notamment :

- a) La diffusion régulière de messages dans les médias ;
- b) L'envoi de courriers de rappel ;
- c) La mise en liaison avec la population et l'organisation de manifestations ;
- d) La réalisation de visites de rappel par les agents de terrain.

4. Le tableau 1 ci-dessous résume les utilisations potentielles des données administratives que nous avons répertoriées avant le lancement du recensement. Voir aussi UNECE, 2018.

Tableau 1

Utilisations potentielles de nouvelles données dans le cadre des trois étapes de l'opération de recensement de 2021 (de bout en bout)

<i>Étape</i>	<i>Utilisation</i>	<i>Observations</i>
Préparation et collecte	Création de la base d'adresses à utiliser aux fins du recensement	Validation des adresses et ajout de nouvelles adresses par recoupement avec les sources
	Inventaire des types de réponse prévus destiné à améliorer l'efficacité de l'opération sur le terrain	Étude des caractéristiques de la population locale et de ses tendances en matière de non-réponse
Traitement et analyse	Actualisation des index et classifications	Les nouvelles catégories figurant dans les sources administratives et celles qui y ont été modifiées peuvent être ajoutées aux index et aux classifications à l'avance, ce qui fait augmenter le taux de codage automatique des variables.
	Nettoyage et vérification	Comparaison des réponses avec d'autres sources
	Ajustements liés aux biais de couverture	Voir les sections ci-après

<i>Étape</i>	<i>Utilisation</i>	<i>Observations</i>
	Recours à des paramètres fictifs aux fins de l'imputation de données	Les ménages dont l'adresse est manquante sont assignés à un emplacement géographique probable. L'idée est de déterminer les emplacements qui s'y prêtent à l'aide des données issues de l'opération de recensement et des données administratives.
	Assurance qualité/Validation des données	Diverses sources ont servi à valider les estimations finales du recensement concernant les personnes et les ménages ainsi que leurs caractéristiques.
	Ajustement des données collectées dans des logements collectifs (« communal establishments », « collective dwellings » ou « group quarters »)	Nous avons calculé le nombre de réponses manquantes dans les logements collectifs en recoupant les niveaux de réponse par âge et par sexe avec les sources de données administratives.
Produits	Remplacement des variables précédemment calculées ou création de nouvelles variables	Des données déjà disponibles dans une source administrative de haute qualité ont été utilisées en remplacement de la question portant sur le nombre de pièces. Des produits intégrés vont être créés concernant les revenus, la surface habitable et d'autres sujets.

5. Les étapes ci-dessus sont décrites plus en détail dans le plan de recensement publié sur notre site Internet (ONS, 2020). On trouvera de plus amples informations sur les ajustements liés aux biais de couverture et la sous-estimation du nombre d'étudiants dans les sections qui suivent.

III. Sources

6. Nous avons utilisé deux principaux types de données pour l'établissement du plan de recensement et la planification en cas d'imprévu : les données administratives et les parodontées.

A. Données administratives

7. Il s'agit des informations recueillies principalement pour des raisons administratives (autrement dit, elles ne sont pas initialement rassemblées à des fins de statistique ou de recherche). Ce type de données est collecté par les services publics et d'autres organisations à des fins telles que l'enregistrement, la réalisation d'opérations et la conservation des informations, généralement en tant que produit dérivé de la prestation d'un service. Les données administratives sont souvent exploitées à des fins opérationnelles et leur utilisation en statistique est le plus souvent secondaire.

8. En voici quelques exemples :

a) Impôts locaux : informations sur les adresses des lieux de résidence retenus pour le paiement de la contribution aux services locaux. Cette source est généralement considérée comme étant de bonne qualité car, d'une part, les collectivités locales en dépendent pour financer leurs services et la gèrent donc de près, et, d'autre part, les résidents qui partent préfèrent en toute logique cesser de payer dès leur déménagement. Cependant, ses définitions ne sont pas toujours cohérentes avec la définition d'un ménage privé occupant

une résidence principale utilisée aux fins du recensement. En outre, les règles varient d'une collectivité locale à l'autre, par exemple concernant l'application de dégrèvements ou d'exonérations pour les résidences secondaires ;

b) Affiliation au système de santé : données démographiques de base sur chaque personne affiliée au Service national de santé (NHS). Cette source présente un taux élevé de couverture de la population totale, avec un léger sous-dénombrement des personnes qui n'ont qu'une assurance privée et des migrants nouvellement accueillis. Les coordonnées mises à jour par les personnes qui déménagent mettent ainsi en évidence le sous-dénombrement et le surdénombrement au niveau local, en particulier dans les zones à forte rotation de population. Pour certains groupes (notamment les jeunes hommes, qui ont rarement besoin d'interagir avec les services de santé), les dossiers peuvent ne plus être à jour pendant plusieurs années ;

c) Services publics de distribution : informations sur la plupart des adresses actuellement approvisionnées en gaz ou en électricité. Cette source a permis d'obtenir des informations sur les adresses ou les zones à faible activité ;

d) Travail et retraite : données sur les interactions des personnes dans un large éventail de régimes de travail et de retraite ;

e) Enquête sur les résidences étudiantes et les internats : données administratives collectées directement auprès des responsables de l'hébergement des résidences étudiantes, sur le nombre d'étudiants dont le bail couvre la période de recensement ;

f) Registre des naissances : données collectées en vertu de la législation à des fins statistiques ;

g) Recensement des établissements scolaires : informations sur tous les élèves fréquentant des établissements publics en Angleterre et au Pays de Galles. La qualité des données est élevée car l'obtention de fonds est conditionnée par l'exactitude des renseignements que les établissements fournissent. Cette source ne prend toutefois pas en compte les élèves scolarisés dans le privé ou à domicile ;

h) Listes électorales : noms et adresses de tous les adultes inscrits sur les listes électorales au Royaume-Uni. Toute personne invitée à s'inscrire sur les listes électorales est tenue de le faire si elle remplit les critères correspondants. Toutefois, dans la pratique, le registre a tendance à être plus à jour au moment des élections importantes et peut ne pas l'être entre celles-ci.

9. D'autres sources ont été utilisées pour différentes parties du plan de recensement. Par exemple, les données sur le nombre de membres des forces armées ont été obtenues auprès du Ministère de la défense du Royaume-Uni et des Forces armées des États-Unis ; le Ministère de l'intérieur et le Ministère de la justice du Royaume-Uni ont fourni des données sur les personnes se trouvant dans les centres de rétention et les prisons. En ce qui concerne les établissements de soins, les données de santé susmentionnées ont été utilisées, ainsi que les données du système dédié de suivi de la capacité, qui ont été conservées à un haut niveau de qualité, notamment aux fins du suivi des taux de vaccination pendant la pandémie de COVID-19.

B. Paradoxxées

10. Il s'agit des données qui décrivent le processus de collecte, normalement dans le cadre d'une enquête. Ces données comprennent les informations de gestion recueillies lors de l'opération de recensement. En voici quelques exemples :

a) Moments de la journée où les réponses ont été données ;

b) Temps pris pour remplir le questionnaire ;

c) Nombre de tentatives faites pour remplir le questionnaire ;

d) Nombre de fois où des agents de terrain se sont rendus à une adresse ou ont appelé sans obtenir de réponse, jour de la semaine et moment de la journée, nombre de fois où les agents ont établi un contact et obtention d'une réponse à l'issue des démarches ;

e) Mode de communication (tel que téléphone, Internet, courrier électronique et visite).

11. Il existe donc des parodonnées pour chaque observation faite dans le cadre de l'enquête. Leurs attributs influent sur le coût, la gestion et les résultats d'une enquête, les évaluations des agents de terrain et les conclusions qui peuvent être tirées au sujet des non-répondants.

12. Pour la première fois, lors du recensement de 2021, nous avons eu accès à 100 % des informations de gestion de la collecte en ce qui concerne les soumissions de formulaires et les opérations sur le terrain.

IV. Utilisations des données autres que celles prévues dans le plan de recensement

13. Les sections suivantes résument certaines des fins auxquelles nous avons utilisé les sources administratives et les parodonnées en adaptant le plan de recensement. L'accent est mis sur les domaines dans lesquels il a fallu modifier le plan afin que les utilisateurs finaux disposent de données de recensement de haute qualité répondant à leurs besoins :

- a) Étudiants ;
- b) Estimation indépendante du nombre de logements occupés ;
- c) Comparaison nationale des principaux groupes d'âge ;
- d) Autres groupes d'âge ;
- e) Validation supplémentaire : logements à occupation multiple.

A. Étudiants

14. Même si, au Royaume-Uni, de nombreux étudiants résident hors du domicile familial, nous savons que bien souvent, leurs parents les inscrivent tout de même sur le formulaire de recensement. C'est pourquoi nous demandons expressément aux parents de remplir les champs correspondants mais également de préciser à quelle adresse leurs enfants résident pendant la période scolaire. Si l'adresse en question est différente de celle qui figure sur le formulaire que remplissent les parents, l'étudiant est exclu des questions suivantes. Si les parents confirment que l'adresse figurant sur leur formulaire est également l'adresse à laquelle l'étudiant réside pendant la période scolaire, ils doivent répondre à toutes les autres questions le concernant. Dans les faits, des données supplémentaires sont collectées, non pas aux fins principales relatives aux estimations finales du recensement, mais pour que les étudiants ne soient pris en compte qu'une fois, à l'adresse où ils résident pendant la majeure partie de l'année.

15. Sur le plan de la communication et de la promotion, nous avons insisté sur les instructions destinées aux étudiants, notamment concernant l'adresse qu'ils devaient indiquer comme étant celle occupée pendant la période scolaire s'ils étaient toujours titulaires d'un bail en résidence étudiante mais étaient temporairement retournés chez leurs parents en raison de la pandémie de COVID-19. Voir ONS, 2021 pour plus d'informations sur les mesures supplémentaires que nous avons prises concernant le recensement des étudiants au moment de la collecte.

16. Il est ressorti des données relatives aux taux de réponse que nous n'obtenions pas toutes les réponses que nous attendions de la part des étudiants. Cependant, les parents indiquaient, dans leur déclaration, que leur enfant résidait à une adresse différente pendant la période scolaire. Grâce à ces données, nous avons pu déterminer l'adresse à laquelle résidaient les étudiants concernés pendant la période scolaire (après avoir procédé à un recoupement pour répertorier ceux qui avaient déjà répondu à cette adresse).

17. Il s'agissait d'une mesure de circonstance, mise en place du fait de l'incidence de la pandémie. Malheureusement, en raison du processus législatif, il était trop tard pour modifier

le questionnaire de telle façon que les étudiants puissent répondre à toutes les questions et non plus seulement aux questions démographiques générales depuis l'adresse indiquée comme hors période scolaire, mais il a tout de même été intéressant de pouvoir exploiter davantage les informations de base provenant de données initialement destinées à éviter le surdénombrement.

18. Parallèlement, nous nous sommes aperçus qu'en raison du déménagement temporaire des étudiants pendant la pandémie, le dénombrement des résidences, tel qu'il avait été conçu, tiendrait difficilement compte de la façon dont il avait été décidé de définir les personnes qui auraient dû y être recensées. En effet, le Bureau de statistique sur l'enseignement supérieur (HESA), qui constitue la principale source de données relatives aux étudiants de l'enseignement supérieur, ne communiquait généralement les données requises pour la période que plus tard dans l'année.

19. Nous avons donc lancé une nouvelle enquête auprès des universités et des établissements privés, à qui nous avons demandé des informations provenant de leurs propres données administratives sur le nombre, ventilé par âge et par sexe, d'étudiants dont le bail couvrait le jour du recensement, même si ceux-ci n'occupaient pas le logement correspondant.

20. Grâce à un taux de réponse de 92 %, nous avons été en mesure d'utiliser les informations obtenues pour estimer la population réelle des résidences par un calcul du déficit d'occupation fondé sur les résultats de l'enquête et les données « reprises » susmentionnées. Afin d'estimer le déficit d'occupation des établissements qui n'avaient pas été en mesure de nous fournir les données demandées, nous avons utilisé d'autres informations disponibles lors de l'opération de collecte ainsi que le nombre de places d'hébergement de notre base de recensement initiale, et nous nous sommes appuyés sur les enseignements tirés des réponses obtenues en matière de sous-dénombrement.

B. Estimation parallèle du nombre de ménages

21. Au Royaume-Uni, les résultats publiés à l'issue du recensement brossent un tableau de la population qui prend en compte la non-réponse au moyen d'une méthode d'estimation à double système destinée à détecter et à corriger le sous-dénombrement et le surdénombrement. Cette méthode est décrite dans le plan de recensement (ONS, 2020), et on trouvera des explications simplifiées sur l'estimation à double système sous la référence ONS, 2011. L'estimation à double système repose sur deux sources de données : les données de recensement collectées et l'enquête sur la couverture du recensement, une enquête postcensitaire réalisée de manière indépendante et portant sur environ 1,5 % des ménages.

22. La méthode de l'estimation à double système s'appuie sur un certain nombre d'hypothèses concernant ces deux sources, notamment leur indépendance (hypothèse : la probabilité qu'une personne réponde à l'enquête sur la couverture du recensement n'est pas déterminée par le fait d'avoir ou non participé au recensement) et l'homogénéité des probabilités d'acquisition de données (hypothèse : les schémas de réponse de la population visée par l'estimation à double système sont similaires pour le recensement et l'enquête sur la couverture de celui-ci). Si ces hypothèses ne se vérifient pas, les estimations auront tendance à subir un biais négatif, c'est-à-dire qu'elles seront inférieures à ce qu'elles devraient être.

23. Il pourrait exister un biais si, par exemple, les ménages avaient tendance à refuser de participer au recensement et à l'enquête sur la couverture de celui-ci, ou si les ménages ayant participé au recensement étaient ensuite plus enclins à répondre à l'enquête. Nous avons cherché s'il existait des biais découlant de la non-vérification de ces hypothèses en comparant l'estimation à double système des ménages avec une estimation indépendante du nombre de logements occupés (Alternative Household Estimate (AHE)). L'AHE a été créée à partir d'informations tirées des réponses, d'observations de terrain et d'autres informations provenant des données collectées dans le cadre du recensement ainsi qu'à l'aide de diverses sources administratives disposant de données sur les ménages.

24. L'AHE repose sur un éventail de sources complémentaires fondées sur les adresses :

- a) Base d'adresses utilisée aux fins du recensement ;

- b) Intégralité des parodontées relatives à la gestion des réponses au questionnaire de recensement et aux visites sur le terrain ;
- c) Données relatives aux impôts locaux ;
- d) Affiliation au système de santé ;
- e) Recensement des établissements scolaires ;
- f) Services publics de distribution ;
- g) Recensement de 2011.

25. La création de l'AHE s'est appuyée sur un modèle basique exploitant la relation entre les réponses par lesquelles la présence ou l'absence de résidents a été confirmée et les observations des agents de terrain. Nous avons créé des indicateurs d'occupation ou de vacance à partir des données administratives (par exemple, il pouvait être déduit des données relatives aux impôts locaux que personne ne résidait à une adresse donnée lorsqu'un dégrèvement pour logement vide y était appliqué, ou il pouvait être supposé qu'un logement était occupé par ses résidents habituels lorsqu'un enfant était inscrit à l'adresse correspondante dans le cadre du recensement des établissements scolaires). Chaque adresse a été attribuée à un groupe, et dans chaque groupe, les taux d'occupation correspondant aux logements pour lesquelles une réponse avait été reçue ont été appliqués aux logements pour lesquels aucune réponse n'avait été reçue.

26. Nous avons également appliqué des règles déterministes découlant d'observations de terrain :

a) Si cinq visites sur le terrain au moins avaient été réalisées à une adresse, par au moins deux agents de terrain différents, et toutes avaient abouti à la conclusion que personne n'y résidait, le logement était défini comme vacant ;

b) Si l'agent de terrain avait obtenu ses informations sur l'adresse auprès d'une tierce partie de confiance (le chef de ménage lui-même, un voisin ou un gardien/concierge), le logement était défini en fonction de l'observation effectuée (occupé ou vacant).

27. À l'origine, nous souhaitions utiliser l'AHE pour déterminer s'il existait des biais entre les ménages, mais dans les faits, nous nous en sommes servis pour remédier à d'autres problèmes repérés dans les données :

a) Biais d'estimation : un biais d'estimation a été constaté pour une seule collectivité locale, concernant le nombre d'enfants scolarisés et le nombre de logements occupés. L'AHE a été appliquée dans la zone correspondante, conformément au scénario initialement prévu en pareil cas ;

b) Niveaux d'incertitude (variance) plus élevés dans les estimations : les résultats de l'enquête sur la couverture du recensement ont varié en fonction des zones concernées. Dans les zones où de moins bons résultats ont été obtenus, les estimations présentaient un niveau d'incertitude plus élevé, tel qu'indiqué par la variance estimée à partir des intervalles de confiance provisoires. Les observations aberrantes (nombre estimé de ménages inférieur à l'AHE) ont été ramenées à la valeur de l'AHE ;

c) Taux de réponse des personnes supérieur au taux de réponse des ménages : l'AHE a démontré que dans certaines zones où le nombre de ménages manquants était en apparence supérieur au nombre de personnes manquantes, les estimations des ménages étaient plus robustes. Dans ces zones, le taux de réponse des personnes a été aligné sur le taux de réponse des ménages.

C. Faible estimation du nombre de nouveau-nés et de nourrissons

28. L'estimation du nombre de nouveau-nés et de nourrissons (0 à 2 ans) a été jugée trop faible à l'issue d'un recoupement avec les données de santé et les données du registre des naissances. Nous avons estimé la population totale à chaque âge, à partir des données sur les naissances et les décès, et rapporté l'estimation de chaque collectivité locale à ce total pour

les âges de 0, 1 et 2 ans. Pour éviter une surestimation dans certaines zones, les données de santé ont été utilisées comme limite supérieure.

D. Faible estimation du nombre d'enfants de 3 à 15 ans

29. L'administration galloise gère une source de données englobant tous les enfants d'âge scolaire, qui repose sur le recensement des établissements scolaires publics et sur d'autres sources relatives aux enfants scolarisés dans le privé et à domicile. Cette source a permis de constater que l'estimation du nombre d'enfants de 3 à 15 ans était faible dans les collectivités locales galloises et a servi à opérer des ajustements. L'analyse des estimations obtenues pour les régions anglaises a révélé que le Nord-Est présentait une aberration similaire, de sorte que les facteurs d'ajustement calculés pour le Pays de Galles ont également été appliqués à chaque zone de cette région.

E. Évaluation de la couverture dans les ménages non liés

30. Il est particulièrement difficile de recenser la population résidant dans des logements à occupation multiple (HMO) car, de par leur nature, ceux-ci regroupent des personnes entre lesquelles il n'existe pas de lien. Le risque est que l'un des occupants soumette une réponse au questionnaire de recensement uniquement pour lui-même et que l'adresse en question ne fasse ensuite plus l'objet d'aucun suivi alors que les autres occupants n'ont pas été pris en compte.

31. Nous avons atténué ce risque autant que possible lors de la création de la base d'adresses en décomposant les logements à occupation multiple en adresses par unité ou par pièce dans la mesure du possible. Ainsi, même si une personne participait au recensement, tout occupant d'une autre unité ayant reçu un code distinct mais ne l'ayant pas utilisé recevrait une lettre de rappel et la visite d'agents de terrain jusqu'à l'envoi de sa réponse. Nous avons été mesure d'établir les adresses à l'échelle des unités pour environ 4 % des logements à occupation multiple.

32. Lorsque cela n'a pas été possible, nous avons évalué le niveau de réponse au sein des ménages en comparant les niveaux de réponse dans les ménages composés d'au moins trois adultes, pour lesquels trois noms de famille distincts au moins étaient répertoriés, avec les données des listes électorales, triées selon le même critère (trois adultes au moins et trois noms de famille distincts). Les niveaux de non-réponse obtenus ne semblaient pas différents de ceux des autres catégories.

33. Dans le cadre du processus par lequel les collectivités locales font remonter l'information, nous avons également reçu des données sur l'enregistrement des logements à occupation multiple au niveau local pour certaines zones, notamment des adresses précises et la capacité pour laquelle les logements correspondants étaient autorisés. Nous avons pu rattacher ces adresses aux données collectées pour vérifier une fois de plus que les niveaux de réponse qui y étaient associés (tant au niveau des ménages qu'au sein des ménages) concordait avec ceux des autres catégories de ménages.

V. Enseignements

34. Les principaux enseignements tirés de l'utilisation de bout en bout des données administratives dans le cadre du recensement de 2021 sont les suivants :

a) Préparation : il est important de disposer de sources de données utiles, bien conçues et reliées entre elles, suffisamment adaptables pour permettre de relever tout nouveau défi ;

b) Il ne faut pas sous-estimer le temps nécessaire à la définition des besoins et à l'acquisition des sources ;

c) L'ingénierie des données, qui comprend le géoréférencement, la mise en correspondance et le couplage, est un processus long et complexe ;

d) Il peut être difficile de déterminer quels éléments de données doivent être utilisés en priorité parmi la multitude de données disponibles.

35. Nous avons comparé les définitions et la couverture fournies par les sources administratives avec les définitions utilisées aux fins du recensement dans des situations normales. Nous avons ensuite dû évaluer l'incidence de la pandémie et ainsi déterminer si l'absence de mise à jour régulière avait des conséquences pour la source, si nous pouvions toujours y avoir accès compte tenu des conditions de travail en période de confinement, mais aussi si cette source restait représentative de la situation réelle des personnes.

36. L'importance d'un couplage fiable : par exemple, dans notre modèle pour l'AHE, il était de loin préférable de ne pas utiliser une source, plutôt que de ne pas coupler les bonnes informations et de prendre une mauvaise décision concernant les adresses.

37. Les principales mises en garde concernant la qualité des sources de données administratives ont trait aux aspects suivants :

a) **Actualité** : tous les groupes n'interagissent pas rapidement avec l'administration, de sorte que certaines informations ne sont plus à jour, ce qui entraîne un sous-dénombrement ou un surdénombrement selon les zones géographiques ;

b) **Cohérence** : il faut avoir conscience du degré de recoupement entre les sources et ce qui est mesuré dans le cadre du recensement, en particulier s'il évolue au fil du temps (par exemple en raison de la pandémie, ou de modifications de l'utilisation principale de la source administrative) ou varie d'une zone à l'autre ;

c) **Détermination des adresses** : la variabilité des pratiques de détermination des adresses entre les sources et la complexité de certains types de logement (notamment les appartements et les habitations réaménagées) rendent le géoréférencement de chaque adresse très difficile :

i) Par exemple, certaines sources fourniront des données déjà géoréférencées et pourront attribuer à un immeuble d'appartements un numéro de référence unique de propriété (UPRN) « parent », tandis qu'une autre source utilisera les UPRN « enfants » s'imbriquant dans le numéro parent ;

ii) D'autres sources encore nous ont fourni des adresses brutes que nous avons dû géoréférencer. Dans certains cas, le bloc d'adresse peut se présenter sous différentes formes (par exemple, « appartement 1, bloc A, rue Smith » peut être la même adresse que « premier étage, 93 rue Smith ») ;

d) Nos outils de géoréférencement se sont considérablement améliorés ces dernières années, mais nous travaillons toujours à l'élaboration de processus automatisés permettant de comprendre les questions très complexes posées par la détermination des adresses et de nettoyer celles-ci. À cet égard, le travail accompli dans le cadre des activités de collecte et de nettoyage liées aux recensements nous a énormément appris.

VI. Conclusion

38. L'utilisation des sources de données administratives et de parades dans le cadre du recensement de 2021 s'est dans une large mesure déroulée comme prévu : ces sources ont servi à préparer la base d'adresses et à rendre le suivi des non-réponses plus efficace qu'il ne l'avait été lors des recensements précédents. Nous avons utilisé les sources comme nous l'avions envisagé pour valider les estimations finales concernant les personnes et les ménages ainsi que leurs caractéristiques.

39. Nous avons même été en mesure d'utiliser les sources de manière encore plus poussée que nous ne l'avions imaginé dans le plan de recensement. Nous nous étions préparés à faire face à une série de scénarios probables : nous avons mis en place des données, des méthodes, des ressources et des principes de gouvernance (voir UK Statistics Authority, 2021), et nous avons mis ces scénarios à l'épreuve au moyen d'une série de simulations. En pratique, les situations auxquelles nous avons été confrontés étaient légèrement différentes de ce que nous avions prévu, mais notre approche demeurait valable.

40. En 2021, le taux de réponse a été plus élevé que lors des derniers recensements et n'a pas été aussi touché par la pandémie de COVID-19 que nous aurions pu le craindre avant le début de la collecte. Cependant, nous savions que tout le monde ne répondrait pas et qu'il faudrait analyser convenablement les schémas de non-réponse à la recherche d'éventuels biais et prendre des mesures correctives. Les sources de données administratives nous ont aidés dans cette tâche.

41. Pour information, les premiers résultats du recensement de 2021 en Angleterre et au Pays de Galles ont été publiés le 28 juin 2022 (voir ONS, 2022).

Bibliographie

- UNECE, 2018. Approach to using alternative data sources to support the 2021 Census in England and Wales. Document 15 de la réunion de 2018 du Groupe d'experts des recensements de la population et des habitations :
https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2018/Meeting-Geneva-Sept/ECE_CES_GE.41_2018_15-1811158E.pdf.
- UNECE, 2021. Progress of the England and Wales 2021 Census. Réunion de mai 2021 du Groupe d'experts des recensements de la population et des habitations (en ligne) :
https://unece.org/sites/default/files/2021-05/1%20Ghee%20UK%20ENG_2.pdf.
- ONS, 2020. Design for Census 2021 :
<https://www.ons.gov.uk/census/planningforcensus2021/censusdesign/designforcensus2021>.
- ONS, 2021. Students: Census 2021 :
<https://www.ons.gov.uk/census/planningforcensus2021/censusdesign/studentscensus2021>.
- ONS, 2011. Trout, Catfish and Roach: [ARCHIVED CONTENT] UK Government Web Archive – The National Archives.
- UK Statistics Authority, 2021. Methods for 2021 Census Playbook, EAP158 :
<https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2021/11/EAP158-Methods-for-Census-Playbook.pdf>.
- ONS, 2022. First results from Census 2021 in England and Wales – Office for National Statistics (<https://www.ons.gov.uk/>).
-