

**Conseil économique et social**

Distr. générale
6 juillet 2018
Français
Original : anglais

Commission économique pour l'Europe

Conférence des statisticiens européens

**Groupe d'experts des recensements
de la population et des habitations****Vingtième session**

Genève, 26-28 septembre 2018

Point 2 de l'ordre du jour provisoire

**Méthodes et nouvelles sources de données,
notamment les mégadonnées**

**Approche adoptée concernant l'utilisation de nouvelles
sources de données pour appuyer le recensement
de 2021 en Angleterre et au Pays de Galles****Note du service de conception statistique du recensement de 2021
de l'Office for National Statistics (Royaume-Uni de Grande-Bretagne
et d'Irlande du Nord)****Résumé*

Pour le recensement de 2021, nous étudions la manière dont nous envisageons d'intégrer de bout en bout de nouvelles sources de données pour optimiser la qualité du recensement et mieux répondre aux besoins des utilisateurs. Les plans sont à divers niveaux d'élaboration, s'agissant de la préparation, de la collecte, du traitement et des produits. Le présent document expose notre approche et nos objectifs, les principes de conception et les critères d'intégration, ainsi qu'une actualisation de chaque aspect de la situation. Il présente également les résultats des tests, lorsqu'ils sont disponibles, y compris les utilisations qui ont été écartées.

* Note établie par Cal Ghee, Garnett Compton et Marie Haythornthwaite.



I. Introduction

1. Les données administratives et les données tirées d'enquêtes provenant des autorités locales et de l'administration centrale jouent un rôle essentiel dans la conception du recensement de 2011. Elles ont été largement utilisées dans le cadre de l'assurance de la qualité des estimations du recensement de 2011 et de la planification des opérations. L'Office for National Statistics (ONS) a l'intention d'utiliser une gamme plus large de données administratives, de données tirées d'enquêtes et d'autres données pour appuyer l'étude, la conception et la mise en œuvre du recensement de 2021 en Angleterre et au Pays de Galles. L'ONS veut aussi créer un plus large éventail de produits (pas seulement des statistiques de base) en intégrant les données de recensement et les données administratives conformément à la recommandation formulée par le Statisticien national en mars 2014 au sujet de l'avenir du recensement et des statistiques de la population (ONS, 2014).

2. Le document présente nos réflexions initiales concernant la manière dont ces données pourraient être utilisées pour la conception du recensement de 2021 en Angleterre et au Pays de Galles. Il est important de noter que ces réflexions ont pour le moment un caractère indicatif. La manière exacte dont nous utiliserons les données dépendra :

- a) Des données disponibles ;
- b) Du résultat de nos travaux de recherche sur la qualité des données disponibles ;
- c) De l'élaboration des méthodes visant à tirer parti de la puissance des données.

3. Le document ne contient pas d'observations à propos du fait que l'ONS pourrait utiliser plus largement des données administratives pour améliorer la qualité des statistiques démographiques, migratoires, économiques et sociales ou concernant la possibilité de passer à un recensement axé sur des données administratives (ONS, 2018) à l'avenir pour appuyer la stratégie « Meilleures statistiques, meilleures décisions » de l'ONS (UK Statistics Authority, 2014). Le Statisticien national fera en 2023 une recommandation sur l'avenir des recensements après 2021.

4. Le recensement comporte trois opérations distinctes : l'Écosse et l'Irlande du Nord établissent leur propre modèle selon les méthodes les plus appropriées. L'ONS, l'organisme National Records Scotland (NRS), la Northern Ireland Statistics Research Agency (NISRA) et le Gouvernement du Pays de Galles veillent à l'harmonisation des méthodes et des produits pour assurer la cohérence des statistiques au Royaume-Uni. Le présent document ne porte que sur les activités de l'ONS relatives au recensement en Angleterre et au Pays de Galles.

II. Objectifs du recensement de 2021

5. Les données administratives, les données tirées d'enquêtes et d'autres nouvelles sources de données peuvent aider à atteindre deux objectifs particuliers du recensement de 2021, à savoir :

- a) Produire des statistiques d'actualité et de bonne qualité pour répondre aux besoins des utilisateurs ;
- b) Créer des produits intégrés à partir des données de recensement, des données administratives et des données tirées d'enquêtes.

6. Ainsi, les données administratives, les données tirées d'enquêtes et d'autres nouvelles données pourraient servir à appuyer les opérations de planification et de préparation du recensement ainsi que les opérations de collecte et de traitement des données. Parmi ces activités figurent la validation des données et l'assurance de la qualité, ainsi que la création de produits et la communication d'informations sur la qualité. On trouvera de plus amples détails au chapitre IV.

7. Le recours à de nouvelles sources de données permet de mettre au point et d'appliquer des méthodes visant à créer de nouveaux produits améliorés dans le cadre du recensement de 2021 (par exemple corrélation entre les données administratives sur le revenu et les produits du recensement pour établir des statistiques aréolaires à plusieurs variables se rapportant au revenu). On trouvera d'autres détails au paragraphe 1 de la section C du chapitre ci-après.

III. Que sont les nouvelles données : données administratives, données tirées d'enquêtes et autres sources

8. L'ONS étudie un éventail de sources de données supplémentaires différentes pour atteindre les objectifs du recensement de 2021, notamment les suivantes.

A. Données administratives

9. Il s'agit des informations recueillies principalement pour des raisons administratives (autrement dit, elles ne sont pas initialement rassemblées à des fins statistiques 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, la conservation des informations, généralement en tant que produit dérivé de la prestation d'un service. Les données administratives souvent exploitées à des fins opérationnelles et leur utilisation en statistique est le plus souvent secondaire.

B. Données tirées d'enquêtes

10. Ces données sont issues d'enquêtes statistiques, y compris les recensements précédents, et d'enquêtes telles que l'enquête sur la population active (ONS, 2015) et l'enquête annuelle sur la population (ONS, 2012a).

C. Autres sources

11. Dans le contexte du présent document, il s'agit d'une expression collective qui désigne toutes les données non utilisées traditionnellement au sein de l'ONS. Ce sont les données suivantes :

a) *Mégadonnées* : l'ONS définit les mégadonnées comme étant de vastes ensembles de données souvent non structurés qui sont disponibles, éventuellement obtenus en temps réel, et qui sont difficiles à traiter efficacement à l'aide des méthodes et des techniques classiques. Le volume et la diversité des données disponibles augmentent rapidement. De nombreuses données sont disponibles dans de nombreux formats, notamment les suivants : audio, vidéo, journaux d'exploitation, opérations d'achat, capteurs et sites de réseaux sociaux. Certaines de ces données sont librement accessibles sur le Web, alors que d'autres sont détenues par le secteur privé.

b) *Informations de gestion* : il s'agit des informations agrégées compilées et utilisées lors d'activités normales pour éclairer les prestations opérationnelles, l'élaboration de politiques ou la gestion du rendement organisationnel. Elles sont généralement fondées sur des données administratives mais peuvent aussi être un produit des données tirées d'enquêtes. Les expressions « données administratives » et « informations de gestion » sont parfois interchangeables.

c) *Métadonnées* : ce sont des données qui décrivent ou définissent d'autres données. Il s'agit de tout ce que les utilisateurs doivent savoir pour faire un usage adéquat et correct des données réelles, s'agissant de la lecture, du traitement, de l'interprétation, de l'analyse et de la présentation des informations. Les métadonnées comprennent donc des descriptions de fichiers, des listes de codage, des informations de traitement, des modèles d'échantillons, des rapports de terrain, des motivations conceptuelles, etc.

d) *Paradonnées* : il s'agit des données qui décrivent le processus de collecte, normalement dans le cadre d'une enquête. L'objet des paradonnées est par exemple l'un des suivants :

- i) Moments de la journée où les réponses ont été données ;
- ii) Temps pris pour remplir le questionnaire ;
- iii) Nombre de tentatives faites pour remplir le questionnaire ;
- iv) 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 question de savoir si une réponse a été reçue ;
- v) Mode de communication (tel que téléphone, Internet, courriel et visite).

Par conséquent, il existe des paradonnées pour chaque observation faite dans l'enquête. Ces attributs influent sur le coût et la gestion d'une enquête, les conclusions d'une enquête, les évaluations des agents de terrain et les inférences que quelqu'un pourrait faire au sujet des non-répondants.

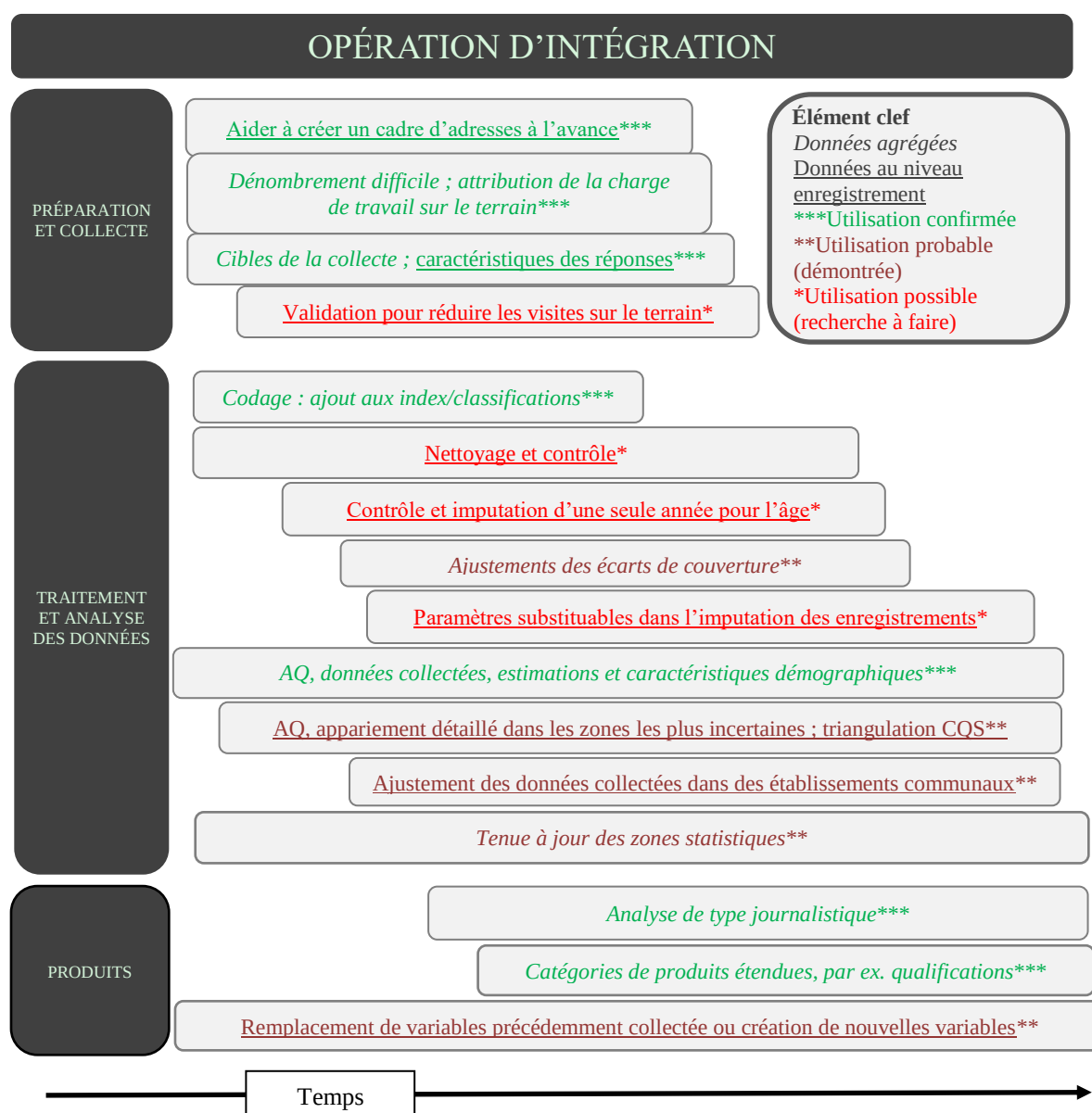
12. Les chapitres ci-après décrivent certains domaines que nous étudions dans notre modèle de 2021, mais ne portent pas sur toutes les utilisations confirmées ou potentielles.

IV. Comment les nouvelles données peuvent aider à atteindre les objectifs du recensement de 2021

13. Le déroulement d'un recensement peut se résumer à trois principales phases comme l'indique la figure I. Les sections ci-après donnent plus de précisions sur certaines des utilisations potentielles des nouvelles données. Nous sommes encore dans la phase de recherche du programme de recensement, de sorte que les possibilités mentionnées ne figureront pas toutes dans le modèle final de 2021.

Figure I

Utilisations potentielles de nouvelles données dans les trois phases de bout en bout du recensement de 2021



A. Préparation et collecte

14. Pour optimiser les réponses à un recensement touchant plus de 25 millions de ménages et assurer une utilisation efficace des ressources, certains faits et chiffres de base sont indispensables à la planification et aux tests. Les questions de type « Combien... » « Quelle quantité... » et « Où... est le plus efficace pour... » requièrent des estimations de travail raisonnables, afin de pouvoir mettre en place la bonne quantité de ressources pour le recensement.

15. Nous étudions le potentiel des nouvelles sources de données pour ce qui est de nous aider à répondre à ces questions. Ces sources peuvent être précieuses pour la planification et la collecte de données de quatre principales façons.

1. Établissement et validation de la liste d'adresses

16. Il est essentiel de disposer d'une liste d'adresses fiable et de bonne qualité pour appuyer les opérations de recensement, optimiser la couverture et s'assurer que les informations adéquates soient envoyées tant aux ménages qu'aux établissements communaux.

17. L'élément central de ce registre est la National Address Gazetteer (AddressBase). Cette base d'adresses est gérée par GeoPlace pour le compte du gouvernement et comprend des renseignements sur les adresses fournis par les administrations locales, la Poste et le Service cartographique. Au niveau des adresses, les renseignements issus des nouvelles sources de données serviront à évaluer et, dans la mesure du possible, à améliorer la qualité du registre des adresses. Par exemple :

a) Données du registre foncier sur les « prix payés » pour les propriétés résidentielles (pour comparer les types de propriété à ceux de la base d'adresses et pour d'autres usages possibles, par exemple l'occupation des nouvelles constructions) ;

b) Données sur le nombre de compteurs électriques par code postal issues du Département de l'énergie et des changements climatiques (DECC) (utilisation possible, par exemple pour comparer le nombre de compteurs résidentiels dans une zone aux adresses résidentielles de la base d'adresses et mettre en évidence toute anomalie) ;

c) Données provenant de sites Web consacrés aux habitations (comme Zoopla) pour recenser les petites zones où sont concentrées les propriétés appartenant à des retraités, les résidences secondaires et les résidences fermées.

18. De plus, les nouvelles sources de données serviront à évaluer la qualité des informations concernant les établissements communaux. Y sont comprises les données de la Commission de la qualité des soins, qui permettent d'évaluer les renseignements concernant les établissements de soins.

2. Échantillonnage – aux fins des tests et de l'enquête sur la couverture du recensement

19. De nouvelles sources de données ont été utilisées pour échantillonner les zones géographiques afin de tester les processus de recensement et sous-tendront la conception des échantillons pour l'enquête sur la couverture du recensement de 2021.

20. Pour certains essais, nous souhaitons tester des résultats particuliers fondés sur des populations spécifiques. Il s'agit notamment de tests sur la question de l'ethnicité dans les zones à forte concentration de groupes ethniques particuliers ou sur les procédures appliquées sur le terrain, notamment les types d'établissements communaux.

21. Ainsi, un échantillon représentatif des zones d'administration locale a été choisi pour les tests de 2017 fondés sur les éléments suivants :

- a) Taux de réponses au recensement de 2011 ;
- b) Données du recensement de 2011 concernant la langue, l'ethnicité, le nombre d'étudiants et le nombre de personnes âgées de 65 ans et plus ;
- c) Prise en compte, en tant qu'indicateur indirect du taux de réponses en ligne prévu, des connexions à large bande par code postal et de la puissance du signal de l'opérateur Ofcom, classée comme étant élevée, moyenne et faible ;
- d) Données de l'enquête annuelle sur la population ayant trait à l'identité sexuelle.

22. Plus d'informations sur le choix des zones de test ainsi qu'un résumé des résultats sont publiés dans le document (ONS, 2017a).

23. L'évaluation de la couverture consiste à mener une enquête indépendante à grande échelle appelée enquête sur la couverture du recensement peu de temps après le recensement. Cette enquête sera analogue à celle réalisée en 2011 (ONS, 2012b), sa conception reposant sur de nouvelles sources de données. Ainsi, la conception de l'enquête sur la couverture du recensement, notamment la stratification et la taille de l'échantillon, dépendra des taux de réponses probables au recensement et à l'enquête sur la couverture du recensement.

3. Établissement des priorités et ciblage : estimation des volumes pour la planification

24. Les recensements précédents avaient montré que certains groupes de population étaient moins susceptibles de remplir les questionnaires de recensement pour plusieurs raisons. De par leur nature, ces groupes sont particuliers et il faut leur prêter une attention spéciale pour les encourager à participer au recensement. De tout temps, ces groupes « difficiles à dénombrer » sont, par exemple, les hommes jeunes, les étudiants, certains groupes ethniques et les personnes très âgées. Dans un recensement mené essentiellement en ligne, les groupes qui respectent traditionnellement les règles, tels que les personnes âgées, mais qui pourraient manquer de compétences dans le domaine numérique, pourraient devenir plus difficiles à dénombrer. Partant, de nouveaux groupes englobant les exclus du numérique seront envisagés.

25. L'ONS aura recours à de nouvelles sources de données pour aider à recenser et à établir des priorités pour ces groupes de population en créant un « index des personnes difficiles à dénombrer » et d'autres outils et en fournissant des renseignements actualisés sur leurs caractéristiques. C'est ainsi que des procédures spéciales peuvent être mises au point pour veiller à ce que ces groupes soient encouragés et aidés à prendre part au recensement. Pour cela, il faudra prendre contact avec eux à un stade précoce afin de comprendre leurs besoins et étudier les moyens adéquats pour obtenir des niveaux de réponse plus élevés, faire appel à des conseillers communautaires, offrir un soutien approprié en matière de langue, ainsi qu'attribuer et hiérarchiser les charges de travail sur le terrain dans les zones où l'on prévoit que le taux de réponses initial sera le plus faible.

26. En l'espèce, un exemple critique est l'établissement des priorités pour le travail de terrain à partir des caractéristiques d'une zone et du type de réponse prévu. Cela aidera à réaliser nos objectifs en termes de qualité, à savoir un taux de réponses élevé, mais aussi à réduire au minimum la variation du taux de réponses d'une zone à l'autre. Par exemple :

- a) Les taux de réponses spontanées observées dans d'autres enquêtes et les recensements précédents peuvent nous aider à évaluer le nombre de personnes qui nécessiteront une certaine forme de suivi, leur type et leur localisation ;
- b) Les données des compteurs communicants peuvent identifier les communautés susceptibles de comporter des niveaux de résidences vacantes ou secondaires plus élevés que la moyenne en raison d'une faible utilisation de l'électricité ou de l'eau ;

c) Les données provenant de sites Web consacrés aux habitations comme Zoopla peuvent permettre de détecter les petites zones susceptibles de contenir de nombreuses adresses difficiles d'accès (caravanes utilisées uniquement pour les vacances, propriétés appartenant à des retraités et résidences fermées).

27. Ces renseignements permettraient d'établir des priorités pour les ressources sur le terrain, par exemple de les attribuer aux ménages plus difficiles à contacter ou devant être encore plus encouragés à remplir leur questionnaire.

28. Un autre exemple concerne les renseignements d'autres sources que nous pouvons utiliser pour identifier et cibler les populations exclues du numérique afin de leur permettre de répondre sans leur imposer de charge indue. On peut ainsi leur montrer où envoyer les questionnaires sur support papier pour un premier contact (sous forme de rappel) ou leur dire de remettre les questionnaires lors d'une visite des recenseurs. Les sources peuvent aussi indiquer les populations qui répondraient en ligne, mais avec une aide supplémentaire. Les données sur le recours à d'autres services publics en ligne et d'autres sources de données sur l'inclusion numérique peuvent nous aider à estimer les lieux probables où des questionnaires sur support papier sont demandés et le volume probable de ces demandes. Seront aussi utiles les conclusions de nos tests de 2017 (ONS, 2017) concernant les types de population susceptibles d'avoir besoin d'une plus grande aide dans le domaine numérique.

29. Nous étudions également les résultats des tests de 2017 pour voir s'il est possible d'identifier des groupes de population susceptibles de ne pas répondre spontanément, mais qui répondent à des rappels envoyés par la poste. De nouvelles sources actualisées serviraient alors à déterminer où ces groupes seraient probablement en 2021.

4. Établissement des priorités et ciblage : nouveau ciblage des ressources pendant la collecte directe

a) Validation des résultats obtenus sur le terrain

30. Une fois l'opération sur le terrain commencée, d'autres sources peuvent aider à valider certains résultats obtenus par les agents de recensement sur le terrain. Par exemple, si un agent tente une visite mais indique que le bâtiment à l'adresse indiquée :

- a) N'existe pas ;
- b) Est destiné à un usage différent (hébergement de vacances, établissement communal ou propriété commerciale) ;
- c) A été démoli ou est abandonné ;
- d) N'est pas accessible ;
- e) Est en construction ;
- f) Est vacant (inoccupé ou correspondant à une deuxième adresse).

31. De nouvelles données actualisées peuvent servir à valider ces résultats obtenus sur le terrain : par exemple, en vérifiant que le bâtiment situé à une adresse est abandonné depuis la création de la liste d'adresses ou par contre en vérifiant que des impôts locaux sont payés pour une adresse, de sorte qu'il est improbable que le bâtiment soit vacant. Cela permettra d'utiliser plus efficacement des ressources limitées sur le terrain en les réorientant au cours de l'opération vers les zones ayant le plus besoin de visites ou d'être plus certain qu'une réponse est attendue d'une adresse, ce qui aide à minimiser la variabilité des taux de réponses entre les zones.

b) *Réaction aux schémas des réponses pendant l'opération de collecte*

32. Comme dans les recensements précédents, nous appliquerons diverses stratégies pour contacter les ménages qui ne répondent pas et les encourager à participer au recensement. De nouvelles sources de données peuvent être exploitées avant le recensement pour nous aider à comprendre et à modéliser les volumes et les schémas des non-réponses pendant la période de collecte. Elles serviront également à prévoir l'incidence de différentes méthodes opérationnelles ainsi que le calendrier le plus efficace pour les interventions. Nous nous employons à déceler ces données et à évaluer la possibilité de prévoir les schémas des non-réponses.

33. Au cours de la collecte directe, nous contrôlerons les niveaux de réponses par rapport aux profils prévus et mettrons en évidence les zones ou groupes de population risquant le plus de ne pas fournir des réponses suffisantes. Nous évaluerons aussi les interventions adéquates (par exemple davantage de visites sur le terrain, rappels, interventions auprès de certaines communautés ou activités de communications de plus vaste envergure).

B. Traitement et analyse des données

1. Saisie, nettoyage, codage, contrôle et imputation

34. Grâce à la disponibilité accrue des données administratives et autres, nous pouvons améliorer la qualité et l'actualité du traitement des informations issues du recensement de 2021. Nous étudierons de quelle manière ces données supplémentaires peuvent nous aider dans les diverses phases de saisie, de codage, de nettoyage et d'imputation des données. Toutefois, nous devons peser la question de savoir si l'amélioration de la qualité des données obtenue grâce au recours à ces données est contrebalancée par une augmentation des délais de traitement.

35. Les sources en question sont déjà utilisées pour actualiser les index et les classifications, les réponses aux enquêtes en cours permettant d'ajouter de nouveaux groupes.

36. Les nouvelles sources de données pourraient être utilisées dans la phase de contrôle et d'imputation. Comme avec tout questionnaire rempli par les répondants eux-mêmes, ces derniers font parfois des erreurs lorsqu'ils enregistrent leurs réponses. Il en résulte des données manquantes ou des réponses non valables car elles sont incompatibles avec d'autres réponses au questionnaire. Comme pour les recensements précédents, nous élaborerons une stratégie de contrôle et d'imputation pour corriger les incohérences et estimer les données manquantes, tout en préservant la corrélation entre les caractéristiques du recensement. Les nouvelles sources de données pourraient aussi fournir des informations auxiliaires qui aideraient, conjointement avec les données de recensement, à établir notre modèle d'imputation. Ainsi, nous étudions s'il est possible de faire appel à des sources administratives pour améliorer l'imputation de l'âge dans le recensement de 2021. Les preuves établies concernant le concept indiquent à ce jour que cette approche pourrait être prometteuse, mais de nouvelles recherches sont nécessaires pour s'assurer qu'elle réponde à tous nos critères de qualité.

2. Évaluation et ajustement de la couverture

37. Aucun effort n'est ménagé pour s'assurer que chacun soit dénombré lors du recensement. Cependant, aucun recensement n'est parfait et certaines personnes en sont absentes. Ce sous-dénombrement ne se produit pas uniformément dans toutes les zones géographiques ou dans l'ensemble de la population (sous-groupes spécifiques par exemple). Nous évaluerons donc la couverture du recensement et ajusterons les données pour les personnes et les ménages manquants à l'aide des renseignements issus du recensement et de l'enquête sur la couverture du recensement.

38. Une fois déterminé le nombre de personnes et de ménages manquants, nous pouvons ajuster les données à l'aide de méthodes statistiques. L'approche adoptée pour 2021 reposera sur le cadre et les méthodes efficaces appliqués lors du recensement de 2011 et peut entraîner une plus grande exploitation des données administratives et autres. Par

exemple, de nouvelles sources de données pourraient servir à améliorer la faisabilité du processus d'imputation des enregistrements des personnes et des ménages manquants créés suite à l'évaluation de la couverture, grâce à la fourniture d'informations concernant les types de ménages ayant été, à un moment donné, enregistrés comme résidant à des adresses où il n'y a pas eu de réponse et concernant la taille de ces ménages. Ce processus fait fond sur la méthode appliquée en 2011, selon laquelle nous avons utilisé les renseignements enregistrés par le personnel de terrain au sujet du type d'hébergement probable et du nombre de résidents probable aux adresses où il n'y a pas eu de réponse.

39. Une telle approche dépend de la qualité et de la disponibilité des données administratives adéquates et du résultat de nos travaux, le but étant de comprendre les conséquences sur les délais de traitement et l'amélioration de la qualité.

40. En 2011, nous avons eu recours aux données administratives de grands établissements communaux pour estimer la couverture insuffisante au niveau des réponses. Nous prévoyons de suivre une méthode analogue en 2021. Nous ferons également appel à de nouvelles sources pour détecter et corriger les distorsions dans le processus d'évaluation et d'ajustement de la couverture, nous fondant de nouveau sur la méthode de 2011.

3. Assurance de la qualité

41. Les résultats du recensement de 2021 feront l'objet d'un rigoureux processus d'assurance de la qualité. Comme pour 2011, ce processus consistera à mener une série de vérifications pour comparer les données de recensement à un éventail d'autres sources de données. Nous prévoyons recourir à une gamme plus large de sources en 2021.

42. Les nouvelles sources de données seront utiles car elles fourniront des données agrégées permettant de vérifier le nombre total de personnes et de ménages. Elles pourront aussi aider à vérifier les comparaisons démographiques pour s'assurer, par exemple, que les caractéristiques ayant trait à l'âge et au sexe soient conformes aux prévisions. Une plus grande disponibilité des nouvelles sources nous aidera à valider la faisabilité des estimations du recensement concernant un plus grand nombre de sujets et les résultats seront utilisés dans des documents d'analyse de style journalistique diffusés avec les produits. En outre, une assurance qualité dynamique des nouvelles estimations du recensement apportera des avantages, s'agissant d'un meilleur ciblage du suivi pendant la collecte directe ainsi que de la détection et de la correction rapides des anomalies dans les données.

43. Ainsi, pour le recensement de 2011 (ONS, 2011), les principales vérifications des caractéristiques en termes d'âge et de sexe faisaient fond sur les sources de comparaison suivantes :

- a) Enregistrement des naissances ;
- b) Estimations démographiques effectuées en milieu d'année par l'ONS ;
- c) Registre des patients du Service national de santé ;
- d) Recensement des établissements scolaires du Département de l'éducation et du Gouvernement de l'Assemblée galloise ;
- e) Données de l'Administration fiscale et douanière sur les allocations familiales ;
- f) Données du Département du travail et des retraites sur les demandeurs d'une pension de retraite ;
- g) Données du Service d'information des clients du Département du travail et des retraites/de l'Administration fiscale et douanière concernant les interactions avec les systèmes de prestations et d'imposition.

4. Mise à jour des zones géographiques

44. Un domaine dans lequel les délais de traitement pourraient être raccourcis est celui de la création de zones géographiques statistiques (zones statistiques et zones de lieux de travail).

45. Au lieu d'évaluer si les frontières des zones statistiques se situent au-dessus ou au-dessous du seuil après ajustement en fonction du sous-dénombrement, nous étudierons si des sources de données complémentaires pourraient servir à prévoir s'il faut modifier les zones géographiques. Nous pourrions alors les ajuster en conséquence avant les résultats du recensement de 2021.

46. Les changements prévus seraient ensuite comparés aux résultats finals du recensement et tout ce qui ne correspond pas aux seuils requis sera ajusté en conséquence.

C. Résultats

1. Amélioration des résultats du recensement

47. Après la collecte et le traitement des données, cette phase consiste à analyser, à produire et à diffuser les résultats.

48. Pour 2021, nous avons l'intention d'intégrer les données de recensement à des sources de données autres que les enquêtes pour pouvoir remplacer ou compléter les informations collectées et créer de nouveaux produits améliorés. Ainsi, des données administratives pourraient être intégrées aux données de recensement pour fournir les caractéristiques de la population jusqu'à présent non incluses dans le recensement. Un exemple pourrait être l'intégration d'informations du Département du travail et des retraites et de l'Administration fiscale et douanière aux données de recensement afin de produire des statistiques sur les revenus des ménages et/ou des personnes, recoupées avec d'autres variables telles que l'ethnicité, sur la base de nos recherches dans ce domaine (ONS, 2017b).

49. Nous nous employons actuellement à comprendre la qualité statistique des données du Bureau d'évaluation pour déterminer comment nous pourrions utiliser ces informations dans les opérations de recensement et les produits. Ces travaux sont principalement axés sur le potentiel des données susmentionnées, s'agissant du remplacement ou de l'amélioration de la question « nombre de pièces » dans le recensement de 2021. Nous évaluerons également le potentiel des données du Bureau d'évaluation lorsqu'il s'agit de produire de nouveaux résultats contenant des informations non collectées précédemment, y compris la superficie totale de la propriété et une ventilation plus détaillée du type de propriété. Outre les exemples ci-dessus, nous planifions aussi un programme de recherche visant à déterminer d'autres produits améliorés à partir des sources de données administratives et de sources autres que les enquêtes qui sont à notre disposition.

50. Pour créer des produits améliorés, nous devons mettre au point et appliquer des méthodes visant à rassembler ces données d'une façon à produire des statistiques de la qualité requise. Des recherches sont déjà en cours concernant les méthodes de contrôle et d'imputation. Des méthodes de nettoyage et d'appariement seront également nécessaires, ainsi qu'une meilleure compréhension de la manière dont nous évaluons la qualité des sources et en rendons compte. Nous devons aussi déterminer si ces processus conviennent à tous les usages ou s'ils doivent être spécifiques à l'usage voulu.

51. Enfin, faisant fond sur les travaux issus du recensement de 2011, nous continuerons à élaborer des méthodes pour recourir à de nouvelles données tirées d'enquêtes ou provenant d'autres sources pour produire des catégories de produits non disponibles directement à partir des catégories de réponses issues des données collectées (nous nous assurerons par exemple que les catégories de qualifications soient conformes aux prescriptions d'Eurostat en matière de produits).

2. Informations sur la qualité

52. Comme en 2011, nous informerons les utilisateurs des méthodes élaborées pour produire les statistiques dans le cadre du recensement de 2021 ainsi que de la qualité des données. Nous définissons la qualité d'un produit statistique comme étant son adéquation avec le but poursuivi en ce qui concerne ses aspects, à savoir la pertinence, l'exactitude, l'actualité, l'accessibilité, l'interprétabilité et la comparabilité, comme le définissent les Recommandations de la CEE pour les recensements de la population et des habitations de (CEE, 2015).

53. La disponibilité prévue d'un plus large éventail de données pour comparaison avec les estimations du recensement nous donne la possibilité de fournir plus d'informations sur la qualité des estimations du recensement de 2021.

V. Possibilités offertes par l'utilisation de ces données dans le recensement de 2021 et défis connexes

54. Nous sommes convaincus que les nouvelles sources de données peuvent nous aider à produire, dans le cadre d'un recensement, des statistiques de bonne qualité qui répondent aux besoins des utilisateurs et cette conviction présente plusieurs aspects. L'utilisation accrue des ensembles de données, existants ou nouveaux, offre de nombreuses possibilités, notamment :

- a) Un emploi plus rentable des ressources de recensement grâce aux renseignements provenant d'autres sources de données pour cibler ces ressources plus efficacement ;
- b) Une amélioration de l'exactitude des statistiques obtenues ;
- c) Une meilleure compréhension des nouvelles sources de données et de la façon dont elles peuvent être plus largement utilisées dans la production de statistiques ;
- d) Une amélioration de la qualité des données administratives clefs grâce à la collaboration avec les fournisseurs de données pour améliorer la qualité des données ;
- e) La production de statistiques plus d'actualité qu'en 2011 ;
- f) La capacité de créer de nouveaux produits améliorés pour répondre à la demande des utilisateurs s'agissant de statistiques à plusieurs variables concernant des sujets non demandés dans le recensement, par exemple des données sur le revenu ;
- g) Une plus grande réactivité face à l'évolution des besoins des utilisateurs grâce à une amélioration des données par de nouvelles sources ;
- h) Une réduction de la charge imposée aux répondants lorsque d'autres sources de données sont considérées comme assez bonnes pour remplacer les questions posées lors du recensement, par exemple concernant le nombre de pièces (ONS, 2017c) ;
- i) La mise au point de nouvelles méthodes statistiques qui peuvent être adoptées à plus grande échelle pour aider à améliorer des produits statistiques.

55. Le recours aux nouvelles sources de données décrites dans ce document présente aussi quelques défis. Nous étudierons leur incidence sur les progrès réalisés en matière de recherche. Ces défis sont les suivants :

- a) Pertinence des nouvelles sources de données pour les concepts statistiques mesurés en termes de disponibilité des variables et de leur définition ;
- b) Compréhension des processus et des procédures liés à la collecte, à la compilation, au traitement et à la validation de nouvelles données souvent complexes et compréhension des conséquences éventuelles de ces processus sur la qualité des données ;
- c) Accès aux sources de données (en particulier les microdonnées) sous une forme appropriée et dans le délai imparti ;

d) Modification des sources ou de leur accès, qui pourrait influencer sur nos méthodes ou nos produits ;

e) Élaboration et vérification de méthodes suffisamment solides concernant l'utilisation de données administratives pour fournir des statistiques utiles, outre les données du recensement de 2021 ;

f) Équilibre entre une plus grande utilisation de nouvelles données pour améliorer l'exactitude et le désir de diffuser des statistiques plus rapidement qu'en 2011 ;

g) Préoccupations supplémentaires concernant le respect de la vie privée découlant d'un recours plus important aux données personnelles ;

h) Circularité éventuelle de l'utilisation des mêmes données dans la collecte et le traitement, mais aussi dans l'assurance de la qualité ;

i) Nécessité d'une étude comparative des résultats du recensement administratif de 2021 par rapport aux estimations du recensement, compte tenu de l'utilisation déjà faite des mêmes sources.

56. Considérée dans son ensemble, notre stratégie visant à inclure davantage de nouvelles données dans la conception du recensement de 2021 se résume comme suit en termes de vision et d'objectifs :

57. *Vision* : intégrer de nouvelles données pour offrir la meilleure qualité pour le recensement de 2021 et laisser un enseignement positif ;

58. *Objectifs* : améliorer et valoriser :

a) La collecte – rentabilité et optimisation des taux de réponses ;

b) Le traitement – amélioration de la qualité et de la confiance grâce à l'assurance de la qualité ;

c) Les produits – amélioration de l'actualité et meilleure réponse aux besoins des utilisateurs ;

d) La phase postérieure au recensement – action visant à continuer à ajouter de la valeur aux produits grâce aux enseignements acquis.

59. Nous énonçons des principes de conception fondés sur un équilibre entre les coûts et les gains, en nous fondant sur les principes de qualité, de valeur et de confiance du Code de bonnes pratiques (UK Statistics Authority, 2018) et en gardant à l'esprit le fait que ce n'est pas parce que nous pouvons faire quelque chose que nous devrions le faire.

VI. Conclusions et prochaines étapes

60. Nous sommes encore dans la phase de recherche afin de déterminer si certaines des utilisations potentielles précitées répondront à nos critères pour être incluses dans la conception du recensement de 2021. Cette phase de conception s'achèvera par une « répétition » à grande échelle des systèmes et des services nécessaires pour réaliser les opérations de collecte, de traitement, d'analyse et de diffusion. À cet effet, nous rassemblerons et utiliserons autant de nouvelles sources de données que possible. Les renseignements et réactions obtenus nous permettront de finaliser la conception du recensement de 2021.

61. Dans le cadre de notre stratégie et de notre approche, nous avons dressé une liste de critères d'intégration pour faciliter la prise de décisions (voir la figure II), en nous inspirant du Code de bonnes pratiques relatif aux statistiques des autorités statistiques du Royaume-Uni (UK Statistics Authority, 2018) et des recommandations de la CEE sur la gestion de la qualité (CEE, 2015). Cela contribuera aux travaux de l'Équipe spéciale de la CEE chargée d'évaluer la qualité des sources administratives utilisées pour les recensements (CEE, 2018).

Figure II

Utilisations potentielles de nouvelles données dans les trois étapes des opérations de bout en bout du recensement de 2021

CRITÈRES D'INTÉGRATION

- Équilibre** – Équilibre entre l'amélioration de la qualité, de la confiance et de la valeur et les risques pour l'exactitude, l'actualité et l'interprétabilité.
- Détail** – Amélioration de la qualité pour un sous-ensemble de la population ou les principales variables – Qualité par rapport à l'effort.
- Qualité** – Pertinence – Est-il encore répondu aux besoins des utilisateurs ? Ce critère comprend la granularité/un faible niveau géographique.
Exactitude – Amélioration ? Dans quelle mesure ajoutons-nous plus d'incertitude ?
Problèmes avec la collecte et les nouvelles données susceptibles d'être regroupées dans les mêmes zones/mêmes groupes de personnes – ajoutons-nous des risques pour l'ensemble sans réellement apporter une amélioration suffisante là où cela est nécessaire ?
Actualité – L'ajout de plus de processus présente-t-il un risque pour notre calendrier ?
Accessibilité – L'utilisation accrue de données disponibles dans le domaine public présente-t-elle un risque pour ce que nous pouvons publier ?
Interprétabilité – Pouvons-nous expliquer clairement l'utilisation des données ?
En particulier concernant la qualité des sources utilisées. Circularité de l'utilisation.
Cohérence – Les nouvelles données visent-elles les mêmes définitions ?
Harmonisation au Royaume-Uni, cohérence dans le temps.
- Valeur** – Intégration adéquate des données. Utilisation de programmes de transformation d'entreprises et contribution à ces programmes ; amélioration de l'efficacité de la collecte et du traitement.
- Confiance** – Considérations éthiques, réduction de la charge imposée aux répondants, instauration de la confiance parmi les utilisateurs.

62. Après la collecte d'informations, cette phase de recherche consistera à analyser les données ainsi qu'à produire et à diffuser les résultats du recensement, améliorés par les données administratives.

Bibliographie

ONS 2011-2011 Census – Methodology for Quality Assuring the Census Population Estimates
https://www.ons.gov.uk/file?uri=/census/2011census/howourcensusworks/howwetookthe2011census/howweprocessedtheinformation/dataqualityassurance/qualityassuranceoverview_tc_m77-224750.pdf

ONS 2012a – ONS Annual Population Survey
<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/annualpopulationsurveyapsqmi>

ONS 2012b – 2011 Census Coverage Survey
<https://www.ons.gov.uk/census/2011census/howourcensusworks/howdidwedoin2011>

ONS 2014 – National Statistician’s recommendation on the future of population statistics in England and Wales

<https://www.ons.gov.uk/census/censustransformationprogramme/beyond2011censustransformationprogramme/thecensusandfutureprovisionofpopulationstatisticsinenglandandwalesrecommendationfromthenationalstatisticianandchiefexecutiveoftheukstatisticsauthorityandthegovernmentsresponse>

ONS 2015 – ONS Labour Market Survey

<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurvey/lfsqmi>

ONS 2017a – 2017 Test

<https://www.ons.gov.uk/census/censustransformationprogramme/progressanddevelopment/the2017test>

ONS 2017b – Research Outputs: income from Pay As You Earn (PAYE) and benefits for tax year ending 2016

<https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/populationcharacteristics/researchoutputsincomefrompayasyouearnpayeandbenefitsfortaxyearending2016>

ONS 2017c – Estimating the number of rooms and bedrooms in the 2021 census in England and Wales: An alternative approach using Valuation Office Agency data

<https://www.ons.gov.uk/census/censustransformationprogramme/progressanddevelopment/questiondevelopment/estimatingthenumberofroomsandbedroomsinthe2021censusinenglandandwalesanalternativeapproachusingvaluationofficeagencydata>

ONS 2018 – Administrative Data Census Project

<https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject>

UK Statistics Authority 2014 – Better Statistics, Better Decisions

<https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject>

UK Statistics Authority 2018 – Code of Practice for Statistics

<https://www.statisticsauthority.gov.uk/code-of-practice/>

CEE 2015 – Recommandations de la Conférence des statisticiens européens pour les recensements de la population et des habitations de 2020

<http://www.unece.org/index.php?id=41398>

CEE 2018 – Équipe spéciale de la CEE chargée d’évaluer la qualité des sources administratives utilisées pour les recensements

https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/bur/2018/February/06Add.1-TF_on_quality_of_admin_data_for_censuses__ToR__apr.pdf