



Conseil économique et social

Distr. générale
19 août 2009
Français
Original: anglais

Commission économique pour l'Europe

Conférence des statisticiens européens

Groupe d'experts des recensements de la population et des habitations

Douzième réunion

Genève, 28-30 octobre 2009

Point 6 de l'ordre du jour provisoire

Collecte de données en ligne

Utilisation de l'Internet pour la collecte de données de recensement au Royaume-Uni – Les leçons dégagées du recensement expérimental en Écosse

Note du General Register Office (Bureau général de l'état civil) pour l'Écosse (Royaume-Uni)

I. Introduction

1. Les statisticiens et les informaticiens ont collaboré étroitement, dans le cadre du processus de conception par itération des questionnaires de recensement en ligne, et ont constaté, d'entrée de jeu, que si un questionnaire de recensement en ligne comprend les mêmes questions que la version sur papier, il n'est pas toujours souhaitable ni possible que les deux types de questionnaire soient absolument identiques compte tenu des différences intrinsèques de format.
2. Le questionnaire en ligne rempli par les répondants lors du recensement expérimental en Écosse était conforme à ce principe et mettait en œuvre, au moment même où il était utilisé, des méthodes de validation basées sur des règles et d'autres caractéristiques intelligentes ayant pour but de tirer le meilleur parti envisageable de l'expérience des utilisateurs et des possibilités d'améliorer la qualité des données.
3. Le retour d'information obtenu des personnes qui ont rempli le questionnaire en ligne est positif. Les membres de ce groupe estiment aussi qu'il est plus rapide, plus simple et plus convivial de remplir le questionnaire en ligne que sur papier.
4. Les éléments d'information recueillis montrent également que les principes de conception énoncés ci-après, qui s'appliquent au questionnaire sur papier mais qui ne conviennent pas à la version en ligne, peuvent avoir un effet négatif sur l'expérience des utilisateurs et sur la qualité des réponses.

5. Les questionnaires et les questions devraient être mis au point en ayant à l'esprit, dès le départ, un cadre conceptuel pour les versions tant sur papier qu'en ligne en vue d'éviter les différences potentielles.

II. Contexte 1: Recensement au Royaume-Uni

6. La responsabilité du recensement de 2011 au Royaume-Uni incombe à trois bureaux de recensement:

a) L'Office for National Statistics (ONS) pour l'Angleterre et le pays de Galles (avec le concours, s'agissant du pays de Galles, du Gouvernement de l'Assemblée galloise);

b) Le General Register Office (Bureau général de l'état civil) pour l'Écosse (GROS);

c) La Northern Ireland Statistics and Research Agency (Agence de statistique et de recherche d'Irlande du Nord), pour l'Irlande du Nord.

7. Le recensement de 2011 est planifié de manière harmonisée, à l'aide d'un questionnaire de base commun, de méthodes communes de «traitement en aval» et d'une approche cohérente de la production destinée à répondre à la fois aux besoins d'Eurostat et aux besoins des utilisateurs britanniques en données relatives au Royaume-Uni dans son ensemble.

8. Même si la majeure partie du questionnaire de recensement sera commune à toutes les régions du Royaume-Uni, les questions incluses varieront en fonction des besoins différents des utilisateurs de chaque région. Une autre différence concerne les modalités de collecte des données, divers sous-traitants étant embauchés par l'ONS et la NISRA, d'une part, et par le GROS, d'autre part, en vue de la collecte de données, notamment la collecte via l'Internet qu'il est prévu d'introduire comme option dans l'ensemble du Royaume-Uni en 2011.

9. Une autre différence, associée à ces modalités contractuelles diverses, concerne les dates des recensements expérimentaux réalisés en 2009 en Écosse, de même qu'en Angleterre et au pays de Galles. En effet, ils étaient programmés en mars en Écosse et en octobre en Angleterre et au pays de Galles. Le présent document rend compte de la première collecte de données en ligne à des fins de recensement au Royaume-Uni qui a été effectuée en Écosse en mars 2009.

III. Contexte 2: Recensement expérimental via le Web en Écosse

10. En mars 2009, les personnes recensées dans l'ouest d'Édimbourg et dans les îles de Lewis et d'Harris ont été les premières, auxquelles l'occasion a été donnée au Royaume-Uni, de remplir des questionnaires de recensement en ligne au cours du recensement expérimental en Écosse.

11. Tous les ménages des zones où a eu lieu le recensement expérimental qui figuraient dans une base de données d'adresses avant le 29 mars 2009 ont reçu à leur adresse postale un questionnaire sur papier comportant un code unique à 12 chiffres leur donnant accès au questionnaire en ligne et des instructions, préimprimées sur la première page, concernant la manière de remplir le questionnaire en ligne.

12. Les experts de la conception de questionnaires et de l'ergonomie du Web ont unis leurs efforts pour mettre au point un questionnaire qui réponde aux besoins de la statistique et se sont conformés autant que possible au modèle sur papier sans pour autant le considérer

comme un carcan lorsqu'il y avait lieu de s'en éloigner. Le GROS a estimé que pour presque toutes les questions, l'ajout d'outils intelligents fondés sur des règles, grâce auxquels la version Web offrait davantage aux répondants que de simples champs de réponse, se justifiait.

13. L'insertion des formes même les plus élémentaires, d'intelligence artificielle dans le questionnaire en ligne le distingue de son équivalent sur papier. Par exemple, un dispositif de réponse automatique peut être intégré à la version en ligne lorsqu'un champ est laissé en blanc ou lorsqu'une réponse non valide est saisie, ce qui n'est pas possible dans le questionnaire sur papier.

14. Cette innovation peut être justifiée par le fait que comme le recensement expérimental l'a montré, les utilisateurs s'attendent à ce que les applications Web soient conviviales. Si ce n'est pas le cas, ils ne s'attarderont pas sur un site Internet. Il ne faut pas perdre de vue que si les organismes de statistique voient avant tout le questionnaire en ligne comme un instrument de recensement, les répondants considèrent en revanche qu'il s'agit d'une application Web comme une autre et, par conséquent, ils s'attendent à ce qu'elle fonctionne comme n'importe quelle application analogue. En d'autres termes, ils comptent avoir à l'écran une application intelligente qui les aide à s'acquitter de leur tâche. Dans le cas d'un questionnaire de recensement, cela signifie que l'application doit comporter les fonctions suivantes: validation des réponses, réutilisation des réponses, messages d'erreur et d'information, et acheminement automatique, le cas échéant.

IV. Éléments clefs de la conception

15. Le GROS a créé un questionnaire de recensement en ligne correspondant le plus possible au questionnaire sur papier en ce sens que l'ordre dans lequel les réponses possibles s'affichent est cohérent. Le questionnaire était pleinement conforme aux normes d'utilisabilité et d'accessibilité et mettait en œuvre une série de caractéristiques conceptuelles intelligentes telles qu'énumérées ci-après:

a) Éléments appropriés.

Des boutons radio ont été utilisés lorsqu'une seule réponse était attendue, et des cases à cocher ont été employées quand plusieurs options pouvaient être sélectionnées à la fois. Certaines questions sous forme de matrice ont été reformulées pour utiliser au mieux l'espace disponible à l'écran;

b) Questions et conseils formulés différemment.

Par exemple, des termes tels que «cocher» dans la version papier sont remplacés par «sélectionner» dans la version en ligne;

c) Messages de signalement d'erreur adapté.

Si une question reste sans réponse ou si une réponse non valide ou illogique est saisie, l'utilisateur est informé de l'erreur et, le cas échéant, de ce qu'il doit faire pour pouvoir entrer une réponse valide;

d) Validations «dures».

Cette procédure se met en route lorsqu'une question appelle une réponse valide sous un format déterminé; en cas de réponse non valide, un message d'erreur s'affiche et le répondant est informé qu'il doit vérifier la réponse. Cette procédure s'applique aussi à une combinaison non valide de réponses lorsque plusieurs réponses peuvent être sélectionnées. Le répondant ne peut pas continuer à remplir le questionnaire avant d'insérer une réponse valide;

e) Validations «douces».

Ce type de validation peut intervenir lorsqu'une réponse valide conforme à un format précis n'est pas saisie et qu'un message d'erreur s'affiche pour indiquer au répondant qu'il doit vérifier la réponse. Les validations «douces» permettent au répondant de passer à la question suivante même sans avoir entré une réponse valide ou sans avoir répondu à la question;

f) Prépopulation de champs.

Les réponses à des questions antérieures ou d'autres données sont utilisées pour alléger la tâche des répondants et les aider à remplir le questionnaire;

g) Acheminement automatique et messages.

Le questionnaire est individualisé automatiquement sur la base des réponses antérieures et des messages sont affichés pour informer les répondants que cette opération a été effectuée.

16. On trouvera ci-après quelques exemples pertinents.

17. Exemple de question 1: nombre de résidents habituels.

Validation dure. Une réponse numérique valide est requise pour pouvoir déterminer la longueur et la structure de la suite du questionnaire, par exemple une matrice relationnelle, le nombre d'ensembles de questions individuelles. Les recensés ne peuvent pas passer à la question suivante (noms des résidents habituels) sans avoir donné de réponse valide.

18. Exemple de question 2: noms des résidents habituels.

Validation dure. Une réponse en caractères alphabétiques doit être introduite dans chaque cas. Les informations saisies sont utilisées en vue d'établir une prépopulation aux fins des questions suivantes pour lesquelles un nom doit être indiqué et sont également utilisés comme référence pour toutes les questions par exemple individualisées: «Ce sont des questions se rapportant à la personne 2 – John Smith.». Le répondant ne peut pas continuer à remplir le questionnaire sans avoir entré une réponse valide.

19. Exemple de question 3: Date de naissance.

Validation dure. Cet élément est considéré comme une variable clef et il y a une très forte probabilité que les répondants connaissent la réponse. Un message d'erreur s'affiche si la réponse n'est pas donnée sous un format numérique valable jj/mm/aaaa. Les réponses illogiques ou invraisemblables telles que 14/14/1960 ou 1/1/1850 seront rejetées. Le répondant ne peut pas continuer à remplir le questionnaire sans avoir saisi une réponse valide.

20. Exemple de question 4: Adresse habituelle une année auparavant.

Validation douce. Un message d'erreur initial s'affiche en cas de réponse non valide, mais les répondants peuvent, s'ils le souhaitent, continuer à remplir le questionnaire même s'ils ont introduit une réponse non valide. La décision d'offrir cette possibilité repose sur le fait qu'il ne s'agit pas en l'occurrence d'une variable clef et qu'il est moins vraisemblable que les répondants connaissent la réponse. Ils risquent donc d'introduire des données fausses, voire renoncer complètement à remplir le questionnaire, si la réponse est considérée comme obligatoire pour passer à la question suivante. (Dans le cas de cette question, l'une des possibilités de réponse est l'adresse où le questionnaire a été envoyé. Dans le système de saisie en ligne, cette information fait partie des éléments utilisés pour établir une population de base car le code d'accès au questionnaire en ligne est lié à une seule adresse de la base de données des adresses.)

A. Exemple de mode de conception 1: utilisation d'outils intelligents dans le questionnaire pour assurer l'acheminement automatique

21. La grande majorité des réactions à l'utilisation expérimentale d'outils intelligents dans le questionnaire a été positive. Quant aux critiques formulées dans ce domaine, elles tendaient généralement à obtenir qu'un nombre encore plus grand d'outils intelligents soient intégrés dans le système pour alléger la tâche des réponses.

22. Il a été fait usage d'un système de cheminement automatique dans le questionnaire de manière à sauter les questions inutiles, par exemple celles sur le type de propriétaire n'ont été posées qu'aux personnes ayant déclaré être locataires en réponse à la question précédente. De toute évidence, les recensés s'attendaient à ce que cette caractéristique de conception permette également de sauter des questions qui, sur la base de la réponse donnée concernant la date de naissance, semblaient n'avoir aucune pertinence dans le cas des enfants. Cette attente valait notamment pour la question relative à la situation matrimoniale des enfants de moins de 16 ans, mais aussi pour d'autres questions. Par exemple, les deux observations suivantes ont été formulées à ce propos:

a) «Pourquoi poser une question concernant la situation matrimoniale d'un enfant? Si celui-ci a moins de 16 ans, on peut clairement considérer que la réponse est connue!»;

b) «Le questionnaire pourrait – semble-t-il, être rendu encore plus logique. Par exemple, pourquoi demander si un enfant de 2 ans est marié, s'il est scolarisé ou s'il se rend quotidiennement à son travail?».

23. Ces observations pourraient également s'appliquer à la version papier, mais pour la collecte en ligne, il paraît évident que puisque des éléments logiques sont intégrés dans d'autres parties du questionnaire, il devrait en être de même dans les cas précités. Le GROS est cependant conscient des risques ci-après: a) l'introduction possible d'un biais modal, et b) les incidences d'une réponse valide mais erronée qui pourrait aboutir à un mauvais aigüillage et, par conséquent, amener à poser un ensemble incomplet ou incorrect de questions. Après le recensement expérimental, cette approche sera reconsidérée.

B. Exemples de mode de conception 2: recréer le format des champs à compléter du questionnaire papier

24. À l'origine, il s'agissait de respecter strictement la structure du questionnaire sur papier pour présenter les réponses possibles, sauf si c'était complètement infaisable, par exemple pour certaines questions sous forme de matrices. Cette approche a également été suivie en ce qui concerne la taille des zones de réponse. Compte tenu des exigences et incidences qu'impliquaient l'obtention d'un ensemble de données cohérent avec les réponses aux questionnaires sur papier ainsi que la création d'images synthétiques du questionnaire à des fins d'archivage, il a été décidé que la taille des zones de réponse pour chaque question devrait être la même dans la version en ligne que dans la version sur papier, et que l'affichage devrait se faire de la même manière.

25. Pour plusieurs questions, il s'ensuit que trois zones de texte distinctes ont été prévues avec une limite de 17 caractères par zone. Sur papier, un espace est aussi délimité à la fin de la ligne qui précède la zone de texte située juste en-dessous. Mais dans la version en ligne, les choses ne sont pas limpides. En effet, si l'approche suivie correspondait exactement au questionnaire sur papier pour les raisons susmentionnées, il s'est avéré que c'était là une grave erreur de conception. Parmi les exemples de réactions, on peut citer les remarques ci-après:

a) «Pourquoi le champ du texte ne peut-il contenir que 17 caractères? Cette restriction est incroyablement gênante et prête à confusion lorsqu'il faut introduire des noms ou adresses d'une plus grande longueur, etc.»;

b) «Pourquoi avoir conservé cette stupide restriction des lignes à 17 caractères dans la version Web?»;

c) «Les zones de texte de trois lignes limitées à 17 caractères chacune prêtent à confusion. Faut-il taper des mots en allant à la ligne?».

26. La dernière remarque citée était effectivement correcte. Il fallait en fait taper les mots à l'intérieur de chacune des trois zones de texte, tout comme dans la version sur papier. Par conséquent, le champ de réponse à la question pouvait en réalité contenir 51 caractères, mais les utilisateurs ont généralement considéré que les trois zones de texte constituaient des champs distincts et ont donc réduit leur champ de réponse à un tiers de l'espace prévu dans la version papier. La qualité des données risque d'en pâtir et il faudra revoir la formulation de ces questions dans le questionnaire du recensement de 2011.

C. Autres points à noter

a) Lors du recensement expérimental, le nombre total de questionnaires reçus en ligne représentait environ 11 % du taux de réponse total;

b) Chaque code d'accès au questionnaire en ligne a été relié à une seule adresse de la base de données et imprimé sur un seul questionnaire papier. Aucun code de substitution ou code supplémentaire n'a été fourni;

c) Les répondants ont pu se déconnecter puis revenir ultérieurement au questionnaire, toutes les réponses qu'ils avaient envoyées jusqu'au moment où ils se déconnectaient étant alors conservées et sauvegardées en toute sécurité;

d) Temps de réponse moyen (15 questions concernant les ménages et 38 questions individuelles):

Ménages comptant un membre: 15 mn 30 s

Ménages comptant deux membres: 26 mn 35 s

Ménages comptant trois membres: 31 mn 23 s

Ménages comptant quatre membres: 34 mn 26 s

Ménages comptant cinq membres: 39 mn 31 s

Ménages comptant six membres: 46 mn 23s;

e) Les versions complètes du questionnaire en ligne étaient disponibles à la fois en anglais et en gaélique écossais. Le questionnaire en ligne a offert les moyens logistiques de distribuer un questionnaire en gaélique aux ménages dans l'ensemble du pays (la version sur papier n'étant disponible qu'en anglais);

f) Le site est conforme aux normes d'accessibilité Web et a été conçu pour être entièrement compatible avec les outils technologiques;

g) Les clips vidéo des questions traduites en langue des signes britannique pouvaient être visualisés en cours d'opération ou téléchargés;

h) Le fait d'avoir placé le questionnaire dans le domaine générique public www.scotlandscensus.gov.uk a inspiré confiance aux utilisateurs;

- i) De nombreux utilisateurs estiment que l'option en ligne est sûre et apaise les inquiétudes concernant la sécurité du courrier;
- j) Les répondants ont pu suivre un lien les conduisant à un site où ils ont donné anonymement leurs impressions après avoir rempli le questionnaire. Plus de 25 % des répondants en ligne ont formulé des observations précises sur la conception, l'utilisation pratique et le contenu des questionnaires en ligne et ces commentaires ont été pris en compte dans notre évaluation;
- k) Les tests d'utilisabilité ont été répétés à maintes reprises avant la mise en ligne des questionnaires et une étude qualitative a été entreprise après l'achèvement du recensement expérimental.

V. Conclusions

27. En offrant aux répondants la possibilité de remplir un questionnaire de recensement, il faudrait répondre aux attentes du public qui souhaite le faire en ligne. Le respect des normes d'utilisabilité permet dans une large mesure d'assurer l'accessibilité de tous les utilisateurs du Web, mais au-delà de cela, l'ajout d'outils intelligents et un certain degré de technicité devraient faciliter la tâche des répondants et faire en sorte que l'expérience des utilisateurs soit positive.

28. Le caractère itératif de l'analyse et la conception, du développement, des essais et de l'évaluation est cruciale pour garantir le succès du questionnaire en ligne et les hypothèses de conception particulières qui s'appliquent à la version sur papier doivent être testées pour vérifier qu'elles s'appliquent aussi à la version Web.

29. Certes, il est prématuré pour le GROS de tirer des conclusions définitives au sujet de tout impact quel qu'il soit de l'approche suivie sur la qualité des données ou des taux de réponses. Néanmoins, il se dégage clairement de notre étude que l'expérience positive de nombreux répondants au questionnaire en ligne au cours du recensement expérimental de 2009 justifie que l'on préconise vivement l'emploi de ce mode de réponse pour le recensement de 2011 en Écosse.
