



## Conseil économique et social

Distr. générale  
14 mars 2012  
Français  
Original: anglais

---

### Commission économique pour l'Europe

#### Conférence des Statisticiens européens

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

##### Quatorzième réunion

Genève, 24 et 25 mai 2012

Point 3 de l'ordre du jour provisoire

##### Accès aux microdonnées

### Plans pour l'accès aux microdonnées du recensement de 2011 du Royaume-Uni

#### Note de l'Office for National Statistics (ONS) du Royaume-Uni

##### *Résumé*

La présente note donne un aperçu de la conception des produits issus de microdonnées que le Royaume-Uni (RU) se propose de développer pour faire suite au recensement de 2011 en Angleterre et au pays de Galles, en Écosse et en Irlande du Nord, y compris une présentation générale des produits mis à disposition en 1991 et 2001.

L'incidence des changements apportés dans ce domaine à la législation du RU et aux méthodes de protection de la confidentialité des statistiques entre 2001 et 2011 est présentée, de même que les activités qui s'articulent autour de la consultation des utilisateurs et de la conception des spécifications de microdonnées.

Les plans élaborés à ce jour pour les produits de microdonnées du recensement de 2011 et les niveaux d'accès correspondants sont ensuite décrits.

## I. Introduction

1. Le présent document a été élaboré par Emma White avec le soutien de Denise McGregor, Keith Spicer et Joe Traynor de l'Office for National Statistics et Keith Dugmore de la société Demographic Decisions Ltd.
2. Le recensement, source d'information la plus complète sur la population du Royaume-Uni, se déroule tous les dix ans depuis 1801 (à l'exception de 1941, tandis qu'en 1966 il a été procédé au recensement d'un échantillon de 10 % de la population). L'éventail et le champ d'application des données de recensement sont vastes et détaillés, et les informations ainsi obtenues sont utilisées pour la planification et l'allocation des ressources, ainsi qu'à des fins universitaires, commerciales et autres. La prescription européenne selon laquelle les États membres doivent communiquer à Eurostat d'ici au 31 mars 2014 des ensembles de données d'information sur leur population apporte une nouvelle dimension à cette activité de collecte de données.
3. Le dernier jour du recensement en Angleterre, au pays de Galles, en Écosse et en Irlande du Nord était le 27 mars 2011. L'Office for National Statistics (ONS) est chargé de communiquer les résultats du recensement de l'Angleterre et du pays de Galles, tandis que le General Register Office (Bureau général de l'état civil) pour l'Écosse (GROS) et la Northern Ireland Statistics and Research Agency (Agence de statistique et de recherche d'Irlande du Nord), pour l'Irlande du Nord (NISRA) sont chargés de transmettre les résultats pour l'Écosse et l'Irlande du Nord respectivement. Une fois que les trois bureaux auront communiqué séparément leurs informations nationales à partir de juillet 2012, l'ONS coordonnera la publication des données du RU.
4. L'ONS est convaincu que l'objectif national d'un taux de réponse de 94 % pour le recensement de 2011 en Angleterre et au pays de Galles sera satisfait, de même que l'objectif d'un taux de réponse de plus de 80 % dans chaque administration locale, tandis que l'enquête sur la couverture du recensement a eu un taux de réponse d'environ 90 %. La réalisation de ces objectifs implique que l'ONS est en bonne voie de donner des estimations de la population de l'Angleterre et du pays de Galles avec une marge d'erreur de  $\pm 0,2$  % et des estimations de la population de toutes les administrations locales avec une marge d'erreur de  $\pm 3$  %. Les premières évaluations sont également prometteuses pour ce qui est de la réalisation de l'objectif d'un taux maximum de 5 % de non-réponse à 70 % des questions.
5. Une enquête indépendante réalisée auprès des consommateurs a conclu que les méthodes adoptées pour l'évaluation de la couverture, l'ajustement de la couverture et l'assurance qualité du recensement de 2011 en Angleterre et au pays de Galles donnaient la certitude que les estimations de population finales obtenues à partir du recensement seraient meilleures que celles obtenues par toute autre méthode et pourraient être utilisées pour l'allocation des ressources et la planification.
6. À ce jour, plus de 100 nouveaux tableaux ont été conçus pour le recensement de 2011 à la suite des consultations menées avec les utilisateurs, soit une augmentation de 25 % par rapport à 2001. Certains de ces tableaux présentent des données collectées pour la première fois en 2011, d'autres permettent de combler les lacunes recensées par les utilisateurs en ce qui concerne les données de 2001. Si la qualité des données le permet, des informations seront également communiquées sur d'autres éléments démographiques, tels que lieu de travail, journée de travail, étudiant hors contexte, étudiant, migrant à court terme, et majeure partie du temps. Ces plans élargissent la base des produits du recensement, offrent un meilleur accès aux données de recensement, satisfont aux besoins

d'un plus grand nombre d'utilisateurs et utilisent plus efficacement les données du recensement, tout en donnant une valeur ajoutée au travail de recensement.

7. Les consultations ont mis en évidence l'importance de communiquer des données détaillées et exactes concernant de petits groupes de la population. Un ensemble de produits conçus autour de ces groupes sera donc élaboré. En outre, comme des groupes ethniques et religieux n'ayant pas de case leur correspondant dans le questionnaire du recensement ont clairement indiqué qu'ils souhaitaient être reconnus de la même façon que ceux qui ont une case à cocher, la deuxième publication des tableaux de recensement (la première pour laquelle on dispose de données sur les groupes ethniques et religieux) inclura une estimation de tous les groupes ethniques et religieux concernant lesquels l'ONS a un code de classification, jusqu'au plus bas niveau de la classification géographique utilisée pour le recensement, la zone statistique (sous réserve des contraintes en matière de confidentialité des statistiques). De même que pour d'autres estimations de base de la population, il n'y a pas au Royaume-Uni de sources existantes équivalentes au recensement pour ce genre de données. Il en est de même des microdonnées.

## II. Microdonnées au Royaume-Uni

8. L'étude longitudinale de l'ONS rassemble des informations sur 1 % de la population de l'Angleterre et du pays de Galles. Cette étude offre des données sur les individus concernant plus de 500 000 personnes, y compris des données provenant de tous les recensements réalisés depuis 1971. Des microdonnées ont été élaborées auparavant en rendant anonymes les données de l'étude longitudinale moyennant la suppression des noms et adresses. Bien que l'échantillon à l'intérieur de cette série de données soit anonymisé, il s'agit toujours de données portant sur les individus et qui n'ont été ni déguisées ni agrégées. En conséquence, l'accès à ces données fait l'objet de mesures de protection.

9. Les premiers échantillons de microdonnées anonymisées (SAR) disponibles reposent sur le recensement de 1991. Deux produits ont été publiés – chacun pour la Grande-Bretagne et l'Irlande du Nord séparément. Des données détaillées figurent dans le tableau 1.

Tableau 1  
Échantillons anonymisés 1991

	<i>Échantillon</i>	<i>Niveau géographique</i>	<i>Licence</i>	<i>Disponibilité</i>	<i>Coût</i>
SAR individus	2 % des individus (>1,1 million de données)	288 zones de SAR comptant une population de >120 000	Licence utilisateur final (LUF). Sur demande adressée au Cathie Marsh Centre for Census and Survey Research (CCSR)	Utilisateurs des milieux universitaires: données téléchargeables sous le format SPSS, Stata ou TAB. Recherche en ligne possible également. CCSR  Utilisateurs autres que des milieux universitaires: CD	Sans frais (avant octobre 2011 une redevance était demandée aux utilisateurs autres que des milieux universitaires)
SAR ménages	1 % de ménages (216 000 données d'information, >500 000 personnes)	Régional	De même que pour les SAR individus	De même que pour les SAR individus	De même que pour les SAR individus

Source: Dugmore, Keith. A business microdata files from the 2011 Census. Rapport final, avril 2009. Établi par Demographic Decisions Ltd, Londres et l'ONS à la demande de l'Université de Manchester.

10. Comme décrit dans le rapport établi par Demographic Decisions Ltd, les échantillons anonymisés de 1991 ont favorisé une flambée de la recherche en science sociale axée sur les données, dans le cadre de laquelle des chercheurs de divers domaines se sont rapidement aperçus des avantages inhérents aux fichiers, étant donné surtout la facilité avec laquelle il est possible d'y accéder.

11. Toutefois, le recensement de 2001 a entraîné la création d'une gamme de produits plus complexe reflétant l'inquiétude de plus en plus grande concernant la divulgation de données personnelles. Cinq produits de différents ont été fournis dans le cadre de trois accords de licence différents, l'accès le plus restreint correspondant à un degré de détail des données plus important. Voir tableau 2.

Tableau 2

**Échantillons de données anonymisées du recensement 2001**

	<i>Échantillon</i>	<i>Niveau géographique</i>	<i>Licence</i>	<i>Disponibilité</i>	<i>Coût</i>
SAR individus	3 % (1,84 m d'entrées)	Fichier RU (Government Office Regions (régions administratives) (GOR), centre de Londres et banlieue de Londres)	LUF par l'intermédiaire du CCSR	Utilisateurs des milieux universitaires: données téléchargeables sous le format SPSS, Stata ou TAB. Recherche en ligne possible également. CCSR	Sans frais (avant octobre 2011 une redevance était demandée aux utilisateurs autres que des milieux universitaires)
				Utilisateurs autres que des milieux universitaires: CD	
Microdonnées portant sur de petites zones	5 % (2,96 m d'entrées sur les individus)	Fichier RU (administrations locales ou circonscriptions d'Irlande du Nord)	De même que pour les SAR individus	De même que pour les SAR individus	De même que pour les SAR individus
Licence spéciale SAR ménages	1 % (>200 000 entrées, >500 000 personnes)	Angleterre et pays de Galles uniquement (sans indication géographique)	Licence spéciale et chercheurs agréés	Archives de données du RU et ONS	De même que pour les SAR individus
Échantillon de microdonnées, sur les individus, à accès protégé (CAMS) individus	3 % (même échantillon que pour les échantillons anonymisés des individus)	Fichier du RU (administrations locales ou circonscriptions d'Irlande du Nord)	Chercheurs agréés	Laboratoire de microdonnées virtuelles de l'ONS	De même que pour les SAR individus
CAMS ménages	1 %	De même que pour les CAMS individus	De même que pour les CAMS individus	De même que pour les CAMS individus	De même que pour les SAR individus

12. Le programme chercheurs agréés a été introduit en 2007. Une demande formelle doit être présentée et les produits du laboratoire de microdonnées virtuelles (VML) de l'ONS sont évalués par des spécialistes de l'ONS pour peser le risque de divulgation, avant d'être envoyés au chercheur qui en a fait la demande (généralement dans un délai de sept jours). L'accès physique au VML passe uniquement par l'un des trois bureaux de l'ONS en Angleterre et au pays de Galles. De nombreux utilisateurs des données du recensement ont été déçus par les dispositifs d'accès mis en place pour les microdonnées de 2001, constatant que dans certains cas l'accès à des données équivalentes était devenu plus difficile qu'en 1991. Néanmoins, les produits ont été largement utilisés au niveau de la recherche et de l'analyse, aussi bien indépendamment qu'en liaison avec d'autres ensembles de données.

### III. Évolution depuis 2001

13. Plusieurs changements sont intervenus entre 2001 et 2011, affectant la façon dont l'ONS pourrait aborder l'élaboration de microdonnées:

a) La loi de 2007 sur la statistique et le service d'enregistrement (Statistics and Services Registration Act 2007) donne à l'ONS le pouvoir juridique de communiquer aux chercheurs agréés des données, et définit les critères selon lesquels ces données peuvent être communiquées;

b) La puissance des ordinateurs augmente tant à l'intérieur de l'ONS qu'à l'extérieur. Il s'ensuit que les données peuvent être communiquées de différentes façons, tandis que les utilisateurs demandent plus souvent un accès plus rapide et plus facile aux statistiques en général et aux microdonnées en particulier, tout en essayant d'obtenir des statistiques plus détaillées de l'ONS;

c) Sur la base du programme du Gouvernement axé sur la transparence et l'ouverture, il est prévu que les ministères doivent communiquer gratuitement un plus grand nombre d'informations;

d) La protection des données personnelles et de la confidentialité des données est un sujet de préoccupation de plus en plus grand au niveau du Gouvernement comme au niveau du public. Le nombre de prescriptions visant à protéger les données a nettement augmenté depuis 2001;

e) L'introduction de la licence «information publique librement réutilisable» permet aux utilisateurs de copier, publier, distribuer et transmettre les données d'information (les utilisateurs doivent indiquer la source de ces informations); de les adapter et de les exploiter, commercialement par exemple, en les associant à d'autres données d'information ou en les incorporant dans leurs propres produits ou logiciels. Néanmoins, les utilisateurs doivent continuer de respecter la loi de 1998 sur la protection des données et ne pas déformer les données.

### IV. Méthodes de protection de la confidentialité des statistiques pour le recensement de 2011

14. À la lumière des observations faites concernant la méthode utilisée en 2001, à savoir l'inversion des entrées de manière aléatoire, associée à des ajustements mineurs des cellules, une inversion *ciblée* des ménages sur la base des entrées à risque sera appliquée aux données du recensement de 2011 dans tout le Royaume-Uni. Les directeurs généraux de l'état civil ont convenu que les valeurs peu élevées (0, 1 et 2) pouvaient être présentées dans des tableaux de recensement diffusés auprès du public à condition:

a) Qu'ait été systématiquement créée une incertitude suffisante quant à la question de savoir si la petite valeur inscrite dans la cellule correspond à une valeur réelle; et

b) Qu'une telle incertitude ne nuise pas gravement aux données.

15. Une «incertitude suffisante» est évaluée par deux facteurs; le pourcentage de divulgation d'attributs réels protégés par l'inversion et l'imputation et le pourcentage de divulgation d'attributs apparents qui ont été créés. Le niveau d'inversion des entrées dans une zone est déterminé par le niveau de non-réponse et d'imputation en ce sens que l'inversion sera d'autant plus faible que le nombre de données imputées aura été grand.

16. Les méthodes convenues de protection de la confidentialité des statistiques auront des effets plus importants sur les petites valeurs de cellules mais cet effet se dissipera à de plus hauts niveaux géographiques. Comme il n'y aura pas d'ajustement mineur des cellules, la cohérence entre les tableaux ne sera pas parfaite. Le tableau 3 montre les principales différences entre les méthodes suivies par l'ONS pour les recensements de 2001 et de 2011 respectivement, et leurs effets.

Tableau 3

**Méthodes de protection de la confidentialité des statistiques:  
différences entre 2001 et 2011**

	2001	2011
Inversion des entrées	Aléatoire	Cible les entrées à risque
Après élaboration du tableau	Ajustements mineurs des cellules en Angleterre, au pays de Galles et en Irlande du Nord	Aucune mesure
Additivité	Pas dans tous les tableaux	Oui
Cohérence	Non	Oui
Harmonisation des méthodes de protection de la confidentialité des statistiques	Non	Oui – mais les paramètres peuvent différer selon les pays
Effets sur les petites valeurs	Effets importants	Quelques effets

*Source:* ONS, SDC (protection de la confidentialité des statistiques).

## V. Considérations en matière de protection de la confidentialité des statistiques pour les microdonnées de 2011

17. Les plans pour les microdonnées de 2011 viseront à assurer le meilleur équilibre possible entre risque et utilité, compte tenu des éléments suivants:

- Équilibre entre recodage et perturbation;
- Décision quant à l'opportunité d'avoir des échantillons mutuellement exclusifs;
- Nouvelles variables et catégories; et
- Contrats de licence et modalités d'accès.

## VI. Consultations

18. Les consultations sur les principaux produits statistiques du recensement de 2011 se sont déroulées en deux étapes distinctes sur les périodes allant du 14 décembre 2009 au 26 mars 2010, et du 7 février 2011 au 28 avril 2011.

19. Ces étapes prévoyaient un retour d'information sur la base de documents à remplir et renvoyer dans le cadre d'un processus de consultation formelle, renforcé par des consultations publiques au niveau national, ainsi que des contacts directs avec les principaux utilisateurs et principaux groupes d'utilisateurs.

20. Ces activités ont été suivies d'ateliers ciblés sur des produits de spécialistes, y compris des microdonnées. Un atelier sur les microdonnées s'est tenu le 11 novembre 2011, à la suite duquel des notes, matériaux et questions et réponses obtenus ce jour-là ont été mis à la disposition des utilisateurs, qui ont été invités à présenter de nouvelles observations d'ici au 19 décembre 2011. Les réactions aux documents et au processus de consultation ont été rassemblées, et un expert extérieur spécialisé dans les microdonnées a été chargé en janvier 2012 de passer en revue les observations communiquées et de présenter un ensemble détaillé de spécifications de microdonnées pour le recensement de 2011. À la date de rédaction du présent document, ces travaux sont en cours. L'ensemble de spécifications de microdonnées essentielles et souhaitables et le rapport final devraient être prêts le 30 mars 2012. Ces travaux ont été préparés par le Groupe de travail sur la stratégie en matière de microdonnées du recensement de 2011 (CMSWG), qui a été mis sur pied en mai 2007, et qui se compose de représentants de l'ONS, du Bureau général de l'état civil pour l'Écosse et de l'Agence de statistiques et de recherche d'Irlande du Nord ainsi que du CCSR de l'Université de Manchester.

## **VII. 2011: Une gamme de produits plus simple qu'en 2001, d'accès plus facile**

21. Trois niveaux de microdonnées sont prévus pour le recensement de 2011.

### **A. Fichier à la disposition du public**

22. C'est là une innovation pour l'ONS. La publication de ce fichier anonymisé portant sur les individus est conforme au programme pour la transparence du Gouvernement du Royaume-Uni et pourra être téléchargé à partir du site Web de l'ONS. Il sera moins détaillé que le fichier SAR individus mis à disposition en 2001 sur la base d'une licence pour les utilisateurs, à moins qu'il ne soit publié avec plus de détails mais sans aucune indication géographique. Ces données seraient faciles d'accès et pourraient être utilisées comme des instruments d'enseignement et d'apprentissage ou pour l'étude approfondie des relations entre variables du recensement. Ce fichier pourrait être utilisé de façon limitée pour vérifier la syntaxe avant de se rendre sur le laboratoire de microdonnées virtuelles (VML). Aucun fichier sur les ménages ne serait mis à la disposition du public.

### **B. Fichier(s) protégé(s)**

23. Ce fichier, lui aussi sur les individus, serait mis à la disposition des utilisateurs par informatique. Il devrait être aussi détaillé que le fichier de 2001 de microdonnées sur les petites zones (SAM) et que les fichiers SAR individus, peut-être dans des conditions analogues à celles fixées pour la licence d'utilisation finale de 2001. Ce fichier pourrait être considéré comme un retour à la spécification de 1991 relative au fichier SAR individus, mais avec un échantillon plus important. Un fichier sur les ménages est également envisagé. Des travaux sont en cours avec les services juridiques de l'ONS en vue de définir ce que l'on entend par «protégé» et d'envisager des modalités de licence appropriées.

### **C. Fichier du laboratoire de microdonnées virtuelles (VML)**

24. Ces données anonymisées mais identifiables seront gardées dans un environnement sûr et seront de façon générale analogues à l'échantillon de microdonnées d'accès contrôlé de 2001 (CAMS), tant au niveau des ménages qu'au niveau des individus; elles

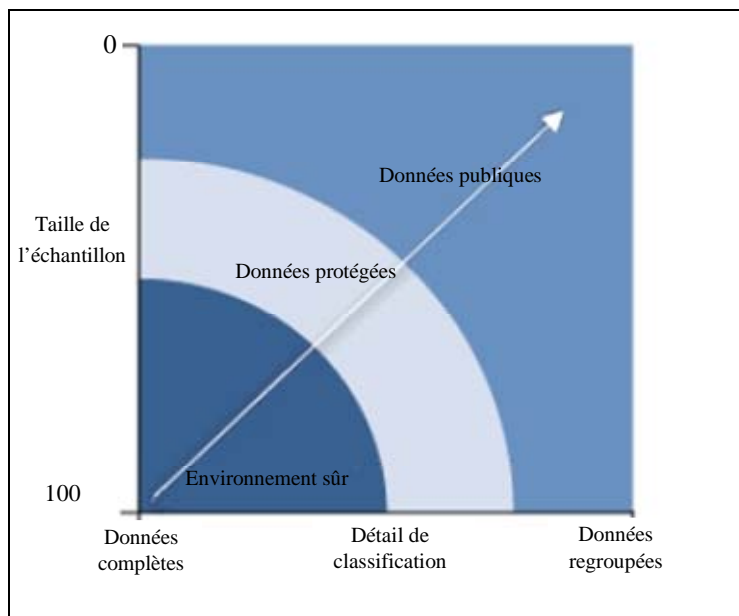
correspondront probablement à des échantillons dont la taille se situe entre 3 et 5 % de la population.

La figure 1 ci-dessous récapitule les niveaux d'accès pour 2011.

Le tableau 4 donne une description des produits de microdonnées proposés pour 2011.

Figure 1

**Diagramme des niveaux d'accès proposés: taille de l'échantillon (en pourcentage) par rapport au niveau de détail**



Source: Data Visualisation, ONS.

Tableau 4

**Produits de microdonnées proposés pour 2011**

Fichier	Détail	Niveau géographique	Licence	Disponibilité
SAR à la disposition du public	3 %	Régions administratives et moins de détails ou aucune indication géographique et plus de détails	OGL (open general licence)	Téléchargement
SAR individus	5 %	Administrations locales – même degré de détails que les SAR individus de 2001	Données protégées	Accord pour utilisation spéciale Téléchargement
SAR ménages	1 %	Régions administratives ou aucune indication géographique	De même que pour les SAR individus	De même que pour les SAR individus
Échantillon de microdonnées à accès contrôlé (CAMS) portant sur les individus	Plus détaillé qu'en 2001? Possibilité d'accroître la taille des échantillons	Au moins aussi détaillée qu'en 2001	Chercheur agréé	Laboratoire de microdonnées virtuelles de l'ONS (environnement sûr)



<i>Fichier</i>	<i>Détail</i>	<i>Niveau géographique</i>	<i>Licence</i>	<i>Disponibilité</i>
Échantillon de microdonnées à accès contrôlé portant sur les ménages	De même que pour les CAMS individus	De même que pour les CAMS individus	De même que pour les CAMS individus	De même que pour les CAMS individus

Source: ONS.

## VIII. Calendrier de publication

25. Selon les plans actuels, les fichiers de microdonnées du recensement de 2011 devraient pouvoir être disponibles en 2013. D'après les commentaires des utilisateurs, l'ordre de préférence pour la publication est comme suit: d'abord les SAR individus et ménages (protégés), puis les CAMS (VML) individus et ménages, et enfin les SAR à la disposition du public. Les utilisateurs ont également demandé à rester informés de l'évolution de ces produits et du calendrier de leur publication.

## IX. Affichage dynamique en ligne

26. La brochure de l'ONS sur le recensement de 2011 a été conçue comme un document dynamique en ligne qui sera publié pour la première fois en mars 2012. Les utilisateurs peuvent s'inscrire pour recevoir par courriel les mises à jour au fur et à mesure qu'elles sortiront. Les fichiers de microdonnées constituent un produit du recensement essentiel et sont traités comme tels dans la brochure.

## X. Conseils d'utilisation, formation et soutien en ce qui concerne les produits

27. Un plan a été engagé pour assurer qu'un personnel correctement formé donne des conseils, une formation et un soutien pour les produits de microdonnées du recensement de 2011. L'ONS, en liaison avec le Groupe de travail sur la stratégie en matière de microdonnées du recensement de 2011 (CMSWG) et l'Economic and Social Research Council (ESRC), examinera la meilleure façon de faire avancer ce plan, étant entendu que l'ESRC, qui a auparavant financé les services de recensement y compris le CCSR de l'Université de Manchester, lancera un nouveau service de collecte de données au niveau national lorsque le financement du cycle de recensement en cours aura pris fin en juillet 2012.

## XI. Conclusion

28. L'ONS est conscient de l'importance des SAR pour les utilisateurs – taille importante des échantillons, lesquels offrent la possibilité de procéder à des analyses exploratoires sur la base d'ensemble de données riches et adaptables, de tester les produits avant de demander l'élaboration de tableaux, et d'effectuer des analyses statistiques multivariées, le tout sur la base d'entrées qui sont comparables aux produits standard du recensement, puisque la source est la même.

29. Le contexte est tel qu'il est maintenant possible au Royaume-Uni de procéder à de puissantes analyses chronologiques sur la base d'entrées de microdonnées, puisque l'on

dispose des fichiers de recensement pour les périodes 1991 et 2001, tandis que les fichiers pour 2011 seront disponibles en 2013. En outre, la production de fichiers de microdonnées pour les années de recensement antérieures à 1991 est à l'examen.

30. L'ensemble de produits du recensement de 2011 a été conçu et élaboré sur la base de principes axés sur l'utilisateur. Ces principes s'appliquent également aux microdonnées, en tenant compte de la plus grande complexité de la protection de la confidentialité, des contrats de licence et de l'accès vu le degré de détail de l'ensemble des données. L'ONS n'en est pas moins convaincu que des fichiers de microdonnées du Royaume-Uni adaptés et dûment accessibles devraient pouvoir être disponibles en 2013.

## Références

Dugmore, Keith. A business case for microdata files from the 2011 Census. Rapport final, avril 2009. Étude commandée par l'Université de Manchester et réalisée par Demographic Decisions Ltd, Londres.

---