

ASSD V : "Technologie de l'information et des communications dans la diffusion de données : faire le lien entre producteurs et utilisateurs dans le cycle 2010 des recensements de population et de logement " (19-21 novembre 2009, Dakar, Sénégal)

Diffusion des microdonnées et des métadonnées intégrées du recensement : L'approche du projet IPUMS- international

Robert McCaa

Centre de population du Minnesota, Minneapolis, Mn États-Unis, rmccaa@umn.edu

Cette recherche est financée en partie par la National Science Foundation des États-Unis, subvention SES-0433654 'séries de microdonnées intégrées internationales (IPUMS-International cycle 10)'.

"Le projet IPUMS-International respecte sans aucun doute les quatre principes essentiels soulignés dans CES [Conférence des statisticiens européens] (2007). Il est cité dans le document CES(2007) comme exemple de bonne pratique. L'examen sur site du projet confirme son statut de bonne pratique pour les entrepôts de données. Il peut en effet fournir un exemple de meilleures pratiques pour un entrepôt de données statistiques internationales. "

— Dennis Trewin (2007), Président émérite de l'Institut International de Statistique
www.hist.umn.edu/~rmccaa/ipums-global/trewin_report_2007.pdf

1. Résumé.

1. Le site web du projet IPUMS-International du Centre de Population du Minnesota (MPC) diffuse les microdonnées anonymisées et intégrées et les métadonnées de 130 recensements de 44 pays. L'Afrique est présente avec seulement 13 recensements. Nous invitons cordialement les instituts statistiques africains à participer à l'initiative IPUMS pour rattraper ce retard. Au cours des cinq années à venir, la base de données internationale IPUMS devrait doubler de dimension ; toutefois, les recensements de l'Afrique risquent d'être sous-représentés parce que, d'une part, trop peu d'instituts statistiques y participent et d'autre part la fourniture des microdonnées au projet demande des délais assez longs. Plus de 90 instituts statistiques ont déjà approuvé l'initiative IPUMS, y compris tous ceux des pays d'Amérique avec plus de 1 million d'habitants ainsi que les instituts d'Europe occidentale.

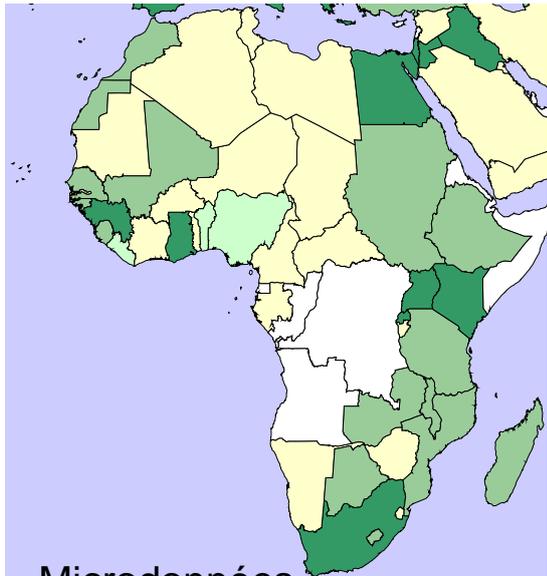
2. Le but de ce document est, d'abord, d'inviter les instituts statistiques africains à participer au projet IPUMS-international de microdonnées de recensement, ensuite de proposer de réduire le fossé entre producteurs et utilisateurs de microdonnées de recensement, en particulier, en préparant les microdonnées et les métadonnées de recensement pour permettre une utilisation efficace et régulière par le projet IPUMS-international et les chercheurs universitaires. Au cours de la décennie, plus de 250 ensembles de microdonnées de recensement et la documentation correspondante, sous des formes multiples, ont été confiées à MPC au nom du projet IPUMS. Cependant leur utilisation est facilitée et les erreurs minimisées lorsque les métadonnées et les microdonnées sont complètement documentées. Les instituts statistiques sont donc fortement encouragés à remplir un formulaire détaillé (voir l'annexe A) accompagnant chaque ensemble de micro et de méta-données de recensement.

3. Pour assurer une sécurité maximale, les microdonnées doivent être transmises dans des fichiers cryptés exécutables, le mot de passe étant emailé ou faxé de façon séparée au coordinateur du projet IPUMS-international. Les métadonnées peuvent être transmises sous forme d'images, mais aussi au format ASCII, CSp, PIM, NESSTAR, SPSS, STATA, SAS, tableur, ou fichiers de document, DDI (Data Document Initiative — NESSTAR est conforme au DDI), hypertexte ou d'autres normes émergentes. La documentation dans la langue officielle est essentielle et doit être aussi fournie en anglais si possible. Sinon, les textes seront traduits en anglais officieusement par les traducteurs - embauchés et payés par MPC - dans le format ASCII simple.

2. Introduction. IPUMS-International: “Les meilleures pratiques”.

4. Le qualificatif de "meilleures pratiques" de Dennis Trewin résume son évaluation des équipements, des politiques et des procédures d'IPUMS-international pour l'archivage, le traitement et la diffusion anonymisée des échantillons de données du recensement. M.Trewin, président de la task-force UNECE sur les bonnes pratiques pour la diffusion des microdonnées et la protection de la confidentialité, est largement reconnu comme faisant autorité dans ce domaine. Son évaluation positive du système de protection des données du projet Ipums-International doit rassurer les producteurs et utilisateurs de microdonnées. Les lecteurs peu familiers avec les protections des données et les mesures de confidentialité du projet IPUMS-International peuvent consulter notre document de la session UNECE/Eurostat sur la confidentialité de données statistiques publiée dans Monographies des Statistiques Officielles (www.unece.org/stats/documents/2005.11.confidentiality.htm voir wp.5 et McCaa et Esteve, 2006).

5. Via www.ipums.org/international, le site web d'IPUMS-International, 130 échantillons de données intégrées et anonymisées de recensements de population sont disponibles actuellement et gratuitement. La taille de la base de données devrait doubler au cours des cinq années à venir, principalement grâce au financement renouvelé jusqu'en 2014 de la National Science Foundation et des instituts nationaux de la santé (États-Unis) ainsi qu'avec l'aide généreuse des partenaires des offices statistiques nationaux. Plus de 3.000 chercheurs représentant 76 pays sont accrédités à l'accès aux microdonnées du site IPUMS-international, qui sont des microdonnées de recensement intégrées pour l'analyse comparative dans le temps et l'espace. Il faut noter que le projet Ipums-International ne diffuse que des microdonnées intégrées et anonymisées - et non les statistiques officielles ou les fichiers sources confiés au projet. Les utilisateurs qui recherchent des statistiques officielles doivent se diriger vers les sites web des offices statistiques nationaux.

Table 1. IPUMS-Afrique: Situation des données de recensement par pays (année en gras = données intégrées)				
 <p>Microdonnées</p> <p>■ Intégrées dans IPUMS</p> <p>■ En cours avec IPUMS</p>	Pays	2000	1990	1980
	Échantillon à la disposition des chercheurs			
	Égypte	2006	1996	1986
	*Ghana	2000		1984
	*Guinée, Conakry		1996	1983
	Kenya	1999	1989	1979
	*Rwanda	2002	1991	
	Afrique du Sud	2001,7	1996-1	1985-0
	*Ouganda	2002	1991	1980
	A lancer en 2010			
	*Mali		1998	1987
	*Sénégal	2002		1988
	*Tanzanie	2002		1988
	En cours			
	*Botswana	2001	1991	1981
	*Éthiopie	2007	1994	1984
	Guinée-Bissau	2009	1991	
	Lesotho	2006	1996	1986
	Liberia (négociation)	2008		1974
	*Madagascar		1993	
	*Malawi	2008	1997	1987
	*Maurice	2000	1990	1983
	Maroc	2004	1994	1982
	Mozambique	2007	1997	1980
	Nigeria (négociation)	2006	1991	
	*Sierra Leone	2004		1985
*Soudan	2008	1993	1983	
*Zambie	2000	1990	1980	

6. La base de données IPUMS-international comprend 44 pays, et pour le continent africain, les pays suivants : Égypte, Ghana, Guinée (Conakry), Kenya, Rwanda, Afrique du Sud et Ouganda (voir tableau 1). La base contient plus de 279 millions d'enregistrements, dont 77 millions de ménages. En 2010, il est prévu d'intégrer des échantillons de 3 pays africains - Mali, Sénégal, et Tanzanie-, 4 pays d'Europe -Asie Népal, Pakistan, Suisse et Thaïlande- et 3 pays d'Amérique -Cuba, Pérou, et Sainte Lucie-. Au cours des cinq années à venir, nous envisageons d'incorporer des échantillons de ménages et individuelles des recensements du cycle 2010 d'autres pays, y compris 13 pays d'Afrique (McCaa, Esteve, Ruggles et Sobek 2006). Néanmoins, nous ne devons pas nous satisfaire de ces 21 pays supplémentaires car de nombreux pays africains resteront encore absents de la base de données (tableau 1. pays africains présents et absents).

7. Pour l'Afrique, à peine 13 recensements africains (7 pays) sont intégrés et disponibles pour la diffusion actuellement représentant 9,1% de la base de données (tableau 2). Cela est plus que regrettable parce que, pour beaucoup de pays d'Afrique, à la différence de l'Europe, les microdonnées n'existent qu'à partir des années 1970. Le projet IPUMS est prêt à consacrer les ressources nécessaires pour renforcer sensiblement la représentation africaine. Ceci demande, tout d'abord, l'approbation du protocole d'accord-type par les agences statistiques concernées, comme vient de le faire le Haut Commissaire du Plan du Maroc(voir annexe B).

Table 2. IPUMS-Africa: Échantillons Intégrés (Septembre, 2009)

Pays	Année Recensement	Échantillon %	Ménages (N)	Personnes (N)
Égypte	1996	10	1,270,787	5,902,243
Ghana	2000	10	397,097	1,894,133
Guinée	1983	10	110,777	457,837
	1996	10	108,793	729,071
Kenya	1989	5	224,861	1,074,098
	1999	5	317,106	1,407,547
Rwanda	1991	10	153,041	742,918
	2002	10	191,719	843,392
Afrique du Sud	1996	10	993,801	3,621,164
	2001	10	991,543	3,725,655
	2007	2	345,170	1,047,657
Uganda	1991	10	339,166	1,548,460
	2002	10	529,271	2,497,449
Enregistrements par Continent				
Continent	Échantillon Intégré		Ménages	Personnes
Afrique	13		5,973,132	25,491,624
Amériques	56		44,018,013	151,649,996
Asie	26		12,642,375	56,803,289
Europe	35		14,753,767	45,193,375
Total	130		77,469,216	279,464,844

4. Descriptions succinctes du recensement et des données: formulaire "A".

8. La coopération des partenaires africains des agences statistiques nationales est essentielle pour la réussite du projet IPUMS-Afrique. Comme les statisticiens officiels sont surchargés par la réponse aux demandes croissantes de statistiques du gouvernement, des entreprises et du public, nous sommes prêts comme nous l'avons fait au cours de la décennie passée, à travailler sur toutes formes de métadonnées et de microdonnées, sans traitement ou

préparation spéciale. De plus, le processus d'intégration a été amélioré et les erreurs minimisées.

9. Le document "A" (voir annexe A) devra être utilisé pour décrire succinctement chaque recensement et ses métadonnées et microdonnées correspondantes. Il doit être rempli par un expert de recensement de l'agence statistique nationale. Un exemple de formulaires remplis pour 3 recensements de l'Afrique du Sud — 1996, 2001 et 2007 — figure en annexe C. D'autres exemples sont sur <https://international.ipums.org/international/samples.shtml>, où il suffit de cliquer le nom de pays

10. Le formulaire "A" est organisé en quatre catégories : la description du recensement, les caractéristiques des microdonnées, les unités identifiées dans les microdonnées et les définitions des unités.

1) Description du recensement. Les éléments suivants sont demandés

- i. Titre officiel du recensement,
- ii. Agence qui a conduit le recensement,
- iii. Population étudiée(indiquer si des populations particulières sont omises, comme les nomades, les étrangers, etc.),
- iv. de jure or de facto,
- v. jour du recensement
- vi. période de travail sur le terrain,
- vii. nombre and type de formulaires,
- viii. type(s) de travaux sur le terrain,
- ix. personnes interrogées et
- x. couverture.

2) Caractéristiques des microdonnées:

- i. source (généralement l'Institut National de la Statistique, le service des Archives national ou l'organisme de Recherche Universitaire),
- ii. méthode d'échantillonnage (de préférence, un dixième des ménages avec départ aléatoire),
- iii. unité témoin (ménage pour des ménages privés; personne pour des ménages collectifs ou groupes),
- iv. taux d'échantillon (10% à la fois pour les ménages privés et les groupes de façon séparée—voir ci-dessous),
- v. taille de l'échantillon (nombre d'enregistrements)
- vi. poids des échantillons, quand les protocoles IPUMS standards ne sont pas utilisés.

3) Unités des microdonnées (indiquer oui/non et ajouter tout commentaire désiré):

- i. Habitations
- ii. Habitations vacantes,
- iii. ménages,
- iv. individus,
- v. groupes
- vi. logement,
- vii. la plus petite unité géographique identifiée,
- viii. habitant régulier / habitant irrégulier/ populations spéciales
- ix. modules spéciaux (mortalité, émigration, agriculture, santé, handicap, etc.).

4) Définitions des unités :

- i. résidences,
- ii. ménages privés,
- iii. groupes

iv. populations régulière /irrégulière/ ou spéciales .

11. Des points supplémentaires peuvent être ajoutés au formulaire au besoin (par exemple, détails des modules mortalité, émigration, fertilité, agriculture, etc.). Le formulaire devra être soumis à MPC dans la langue officielle. Si le formulaire "A" pour votre pays figure déjà sur le site web Ipums-International (voir lien du Web "samples.shtml" ci-dessus), veuillez vérifier pour chacun des recensements afin de confirmer la qualité des informations et envoyer des suggestions, corrections ou commentaires à ipumsi@pop.umn.edu.

5. Besoins en métadonnées.

12. Les métadonnées sont nécessaires au système IPUMS-international. Une grande partie des métadonnées de base est indispensable au traitement et l'évaluation des microdonnées et à l'harmonisation de variables spécifiques. Des métadonnées complètes sont essentielles pour la réussite de l'intégration et une meilleure utilisation des données par les chercheurs (Statistiques Canada 2008 ; cf. aussi McCaa et Thomas 2009). Les métadonnées peuvent être transmises sur des fichiers images mais également au format ASCII, CPro, PIM, NESSTAR, SPSS, STATA, SAS, tableur, document, ou fichiers d'hypertexte et de préférence en plus d'une version. Quand les documents ne sont pas disponibles sous forme électronique, le projet les scannera et les placera sur le site web Ipums-International organisé par année, pays et recensement pour les rendre facilement accessibles. Des copies de ces documents scannés par MPC figureront également sur le CD/DVD fourni à l'agence statistique concernée ainsi qu'aux organisations de recherche nationales et internationales.

13. Nous avons trois objectifs en ce qui concerne les métadonnées.

14. **Tout d'abord**, les chercheurs doivent avoir un accès facile à la documentation originale du recensement dans la langue officielle : au minimum, les questionnaires, les instructions destinées aux agents recenseurs ou les manuels de formation, les dictionnaires de données et de codes. Des métadonnées supplémentaires concernant l'organisation, la préparation et des exemples concrets sur le recensement sont aussi précieuses et sont cataloguées et archivées avec tous les autres documents reçus. Il serait préférable de publier une copie des documents originaux ou en format PDF. Notre objectif est de fournir une collection archivée des fichiers au format PDF de haute qualité pour toutes les métadonnées concernant les données de recensement. Les documents suivants sont demandés aux agences statistiques nationales :

- 1) document de présentation du recensement.
- 2) instructions aux agents recenseur (parfois considéré comme manuel de formation).
- 3) "Dictionnaire des codes" ou "dictionnaires de données" pour chaque ensemble de données (définitions des structures du fichier, de l'emplacement des colonnes des variables et des étiquettes pour les codes, tels que les fichiers de dictionnaire de données du bureau de recensement américain "MIPS"), y compris le découpage administratif, les professions, etc.
- 4) Tableaux de correspondance indiquant l'équivalence des codages entre deux recensements ou davantage ou entre un recensement et une norme internationale (ISCO, CITE, etc.). Ils sont spécialement utiles pour suivre les modifications du découpage administratif et pour l'intégration de la profession, de la branche d'occupation et les variables de niveau scolaire.
- 5) Tableaux de base des résultats officiels publiés sur un site web, un livre ou un CD.
- 6) Rapports techniques et méthodologiques sur les opérations de recensement, les concepts, les nomenclatures, la comparabilité, la qualité, les enquêtes de post-énumération, etc.
- 7) Lorsque les microdonnées sont fournies en échantillons, le tirage d'échantillon devra être décrit en détail. Lorsque le standard de Ipums-International est utilisé avec un dixième des ménages à partir d'un point aléatoire, aucune documentation supplémentaire n'est exigée (voir les spécifications de microdonnées ci-dessous). Autrement, il serait utile de recevoir des estimations des erreurs d'échantillonnage pour une échelle des fréquences absolues ou

relatives (par exemple, par pourcentage d'échantillon = 2, 5, 10, 15, 20, 25, 30, 35, 40 et 50), et pour les variables clés telles que l'âge, le lien avec la personne de référence, l'enseignement et le statut d'emploi. Il faut souligner ici le cas de l'institut national des statistiques du Mozambique qui fournit la documentation la plus complète sur le tirage d'échantillon et les erreurs (Megill 2007).

8) Les fichiers correspondant au découpage administratif codé dans les microdonnées (correspondant au niveau européen des nomenclatures spatiales de NUTS1, NUTS2 et NUTS3) et approprié à la diffusion aux chercheurs. Si ces fichiers de limite ne sont pas fournis, nous construirons des fichiers officiels à partir des sources aisément disponibles.

15. **En second lieu**, nous construisons un système dynamique de métadonnées pour chaque variable, intégrée ou non-harmonisée. Il s'agit de faciliter la comparaison à la fois de la rédaction (en anglais) d'une question particulière et des instructions correspondantes aux agents recenseurs, pour toute combinaison de pays et de recensements.

16) **Troisièmement**, à partir de la documentation originale, nous rédigeons les métadonnées intégrées décrivant chaque variable de la manière suivante :

- 1) brèves définition et description de la variable sélectionnée,
- 2) disponibilité (liste de pays et d'années de recensement avec la variable),
- 3) comparabilité générale (nuances entre des définitions différentes),
- 4) univers (population à laquelle la question est abordée),
- 5) période de référence (par ex., pour l'activité économique, 7 jours, mois dernier, l'année),
- 6) variations dans les définitions des attributs spécifiques (par exemple, "employé"),
- 7) questions de comparabilité pour les recensements spécifiques organisés par pays.

Le chercheur trouvera ces pages simplement en cliquant le nom de la variable. Les pages sont construites sur demande par le système dynamique de métadonnées. Les questions de comparabilité ne s'appliquent que pour les recensements actuellement sélectionnés.

17. Nous préférons avoir des copies électroniques de la documentation source mais nous acceptons les publications sur papier ou des photocopies. Les fichiers électroniques peuvent être envoyés par e-mail ou par courrier sur CDs. Nous contacterons en cas de besoin des traducteurs professionnels en anglais qui effectueront des traductions officielles en format simple de texte. Pour éviter toute perte de document ou de CD et par économies, la collection entière de documents devrait être envoyée par courrier en un seul paquet, aux frais du projet.

18. Les normes émergentes pour les métadonnées structurées (dictionnaires de données, listes de code, définitions, formulaires, etc.) — comme la Data Documentation Initiative DDI (www.icpsr.umich.edu/DDI/codebook/) de NESSTAR ou le Toolkit de microdonnées (<http://www.surveynetwork.org/home/>) développé par le Réseau International d'enquêtes auprès des ménages et la Banque mondiale — facilitent le transfert de l'information vers le système de traitement IPUM-International. DDI est une technique de documentation basée sur le format XML (Extensible Markup Language). Il identifie les éléments trouvés dans le dictionnaire des codes accompagnant un fichier de données. Il permet l'identification des caractéristiques du fichier de données de recensement ou d'enquête, des caractéristiques d'échantillon, des définitions d'unités, de la méthodologie, des structures de fichier, de la structure et du contenu des variables, des questions et des relations aux variables, des listes de codes et des documents relatifs soit en ligne soit sur papier.

19. De nouvelles versions de DDI, disponibles depuis 2008, permettent d'offrir plus d'information sur le processus d'harmonisation complexe utilisé pour construire des variables intégrées. Bientôt, nous comptons offrir aux chercheurs accrédités qui demandent des

extractions de microdonnées, des dictionnaire des codes « customisés » construits à partir de la base métadonnées qui sous-tend l'interface et le système d'extraction du système Ipums.

6. Besoins en microdonnées.

20. Pour les microdonnées, nous avons deux objectifs principaux : tout d'abord, archiver de manière permanente les fichiers sources originaux des agences statistiques nationales, et ensuite, diffuser des extraits d'échantillon anonymisés intégrés et customisés de ménages aux chercheurs accrédités. Nous préférons que les offices statistiques nationaux nous confient des copies anonymisées (suppression des noms, adresses, et numéros d'identification) des fichiers sources complets (100% des données). Ainsi nous pouvons tirer des échantillons de façon cohérente et efficace sans effort supplémentaire des agences statistiques partenaires. En outre, les enregistrements imparfaits rencontrés pourront être facilement remplacés plutôt que corrigés par imputation. Tous les fichiers sources de microdonnées confiés au centre de population du Minnesota sont archivés dans un dépôt sécurisé (« la glacière »), les fichiers confiés ne sont pas reproduits ou cédés à quiconque, personne ou institution, en toute circonstance. Ainsi que le rapport Trewin le note, le centre de population du Minnesota cherche à maintenir une sécurité sans faille.

21. Des objectifs supplémentaires à l'étude sont :

1) Développer un tableur en ligne pour permettre des tabulations intégrées pour différents pays et années de recensement. De préférence, le tableur permettra d'exploiter 100% des microdonnées mais à des fins d'anonymisation, le niveau géographique le plus fin serait supprimé. Un prototype fonctionne déjà sur une douzaine de pays européens.

2) Échantillon sur-important, avec événements (mortalité maternelle) ou caractéristiques (handicaps) peu fréquents. Par exemple, à partir des 100% des microdonnées, nous proposons d'inclure tous les ménages ayant des décès maternels afin d'analyser et de mesurer les causes de phénomènes épidémiologiques rares (voir Garenne, McCaa, et Nacro 2008). Nous développons une méthode conviviale pour fournir des échantillons sans compromettre l'anonymisation des données. En outre, notre méthode assure que les chercheurs utilisent les méthodes appropriées d'extrapolation.

22. Quatre façons de confier les microdonnées ont été utilisées pendant la première décennie de partenariats Ipums-international (illustrées avec des exemples) :

1) Le centre de population du Minnesota se voit confier l'archivage des fichiers sources exhaustifs de microdonnées et la production des échantillons (38 pays).

2) Les échantillons sont produits entièrement par l'agence nationale selon les spécifications d'Ipums-International où 100% des microdonnées sont disponibles (25 pays).

- Office statistique fédéral(FSO) — l'Allemagne : Tout le travail est effectué par FSO, y compris les recensements de 1970 et 1987 de la République fédérale d'Allemagne et les recensements de 1971 de 1981 de la République démocratique allemande.

- Statistiques Pays-Bas (SPB). Années 1960 et 1971 et un échantillon basé sur registre pour 2001 — tout le travail a été effectué par SPB.

- Office statistique fédéral suisse (FSO): 1971, 1981, 1991 et 2001 - préparé par FSO.

3) Les échantillons confiés pour usage public ou restreint des chercheurs sont aussi confiés à Ipums-International avec ou sans paiement d'honoraires (12 pays) :

- Bureau national des statistiques, Chine (honoraires payés pour 1982 ; pas pour 1990)
- Organisation nationale des enquêtes statistiques, Inde (honoraires pour 5 échantillons)
- Statistiques Canada (aucun honoraire)
- Bureau des statistiques nationales, Royaume-Uni (aucun honoraire)
- Statistiques Afrique du Sud (aucun honoraire)

4) La tâche de produire des échantillons anonymisés est confiée à une institution ou à un expert individuel sous la surveillance de l'autorité statistique nationale (6 pays)

- INSEE- France recensements 1962, 1968, 1975, 1982, 1990 et 1999 - les données sont préparées par un chercheur travaillant au sein de l'INSEE dans le cadre du contrat avec le centre de population du Minnesota et sous contrôle de l'INSEE.
- INSSE - Roumanie : le travail est effectué par un institut de recherche universitaire pour les recensements de 2002, 1992, et 1977 dans le cadre du contrat avec MPC sous contrôle de l'institut roumain INSSE.

23. A partir des 4 modalités précédentes, chaque office statistique national détermine celle à utiliser. Le projet est toujours prêt à étudier d'autres dispositions. Indépendamment des modalités, le projet offre la somme de 5 000 dollars US pour les fichiers de 1 million ou plus d'enregistrements individuels pour couvrir les coûts de préparation et la documentation.

24. La "haute précision" est en général définie pour des taux d'échantillon de 10% ou plus (70 des 130 fichiers actuellement intégrés), suivis par le taux de 5% (28 fichiers). Des 32 échantillons à moins de 5%, treize sont des échantillons historiques comprenant toutes les données existantes. Lorsqu'il n'est pas possible de fournir 100% des données, il est préférable de constituer des échantillons aléatoires selon le protocole simple suivant :

- 1) Trier les fichiers selon les divisions administratives majeures et mineures au niveau le plus bas du recensement, unité d'habitation, ménage, famille et personne
- 2) A partir d'un point de départ aléatoire, prendre chaque $n^{\text{ème}}$ unité d'habitation privée (chaque dixième pour un échantillon de 10%).
- 3) Pour les ménages institutionnels — ou les grands ménages privés identifiables par leur taille — après un début aléatoire, tirer chaque $n^{\text{ème}}$ personne en utilisant le même taux de sondage que pour les unités d'habitation privées.

25. Les échantillons aléatoires systématiques se fondent sur le niveau géographique le plus bas. En assurant une répartition géographique représentative, ils sont équivalents à une stratification géographique extrêmement fine avec une pondération proportionnelle. Comme de nombreuses caractéristiques économiques et démographiques sont fortement corrélées avec le lieu géographique, cette précision est sensiblement plus grande qu'un échantillon aléatoire simple de ménages. Dans la mesure où les strates utilisées pour tirer un échantillon de haute précision sont associées aux variables d'intérêt (par exemple, orphelins, pauvreté, chômage, etc.), les estimations de ces variables auront de plus faibles risques d'erreurs que celles qui résultent d'un échantillon aléatoire simple (Davern, et Al, 2009).

26. L'un des principaux avantages à utiliser des microdonnées de recensement est dans leur repérage géographique, qui permet l'analyse sub-nationale sans obérer l'importance statistique. En raison des contraintes de confidentialité, le détail fin au niveau géographique doit être exclu même lorsque les microdonnées sont diffusées sur une base restreinte d'accès, comme dans le cas du projet IPUMS. En général seuls les deux premiers niveaux de détail géographique sont fournis, tel la province et la commune, l'état et le comté, le niveau NUTS1 et NUTS2, etc. En outre, une variable de taille de commune est préférable car elle facilite la cohérence de la définition de la résidence urbaine et rurale. La répartition des classes de population pour l'Allemagne et la France se présentent de la manière suivante :

<u>Germany (de préférence)</u>	<u>France</u>
1) 1 to 2,499 personnes	1 to 4,999
2) 2,500 to 9,999	5 to 9,999
3) 10,000 to 49,999	10 to 19,999

	20 to 49,999
4) 50,000 to 99,999	50 to 99,999
5) 100,000 to 499,999	100,000 to 1,999,999
6) 500,000 or 1,999,999	
7) 2,000,000 or more	2,000,000 or more

27. L'anonymisation peut être effectuée par l'agence statistique concerné ou, sur demande, par le centre de population du Minnesota. Les extraits de microdonnées sont fournis aux chercheurs accrédités sous contrôle juridique et administratif strict (McCaa et Esteve 2006 ; McCaa, Ruggles, et Al 2006). L'échantillonnage "fournit l'incertitude supplémentaire requise pour protéger les données.." Anderson et Fienberg (2001). Mais nous ne nous arrêtons pas là. Nous employons six niveaux de protections techniques. **Tout d'abord**, nous supprimons les codes d'identification du lieu de résidence, du lieu de travail ou du niveau d'instruction pour les unités géographiques inférieures à 20 000 personnes dans les recensements récents. (Certaines agences statistiques placent le seuil plus haut, tel le Royaume-Uni, où le seuil de diffusion est fixé à 65 000 personnes). **En second lieu**, pour les variables catégorielles, toute valeur d'une population de moins de 250 individus est supprimée (l'Allemagne applique un seuil de 2 500). Ces valeurs sont recodées comme "autres," "manquantes," ou dans le cas des codes composites, l'information est codée zéro. **Troisièmement**, pour les variables continues, telles que le revenu ou la surface du logement, les queues des distributions sont tronquées lorsque les effectifs deviennent faibles. **Quatrièmement**, certaines variables sensibles qui permettent l'identification des personnes, telles la date de naissance, sont supprimées. **Cinquièmement**, une petite fraction des ménages "est échangée" d'une unité géographique donnée et affectée à une autre unité pour contribuer à un degré supplémentaire d'incertitude. **Enfin**, on assigne aux ménages un nombre aléatoire unique pour un dernier tri.

7. Conclusion.

28. Les nouvelles technologies de l'information et des communications permettent mais aussi facilitent la diffusion de données de recensement et rapprochent les producteurs et les utilisateurs de données. Le projet IPUMS demande une quantité importante de métadonnées et microdonnées. Néanmoins leur préparation est facilitée et le retour sur l'investissement est substantiel. En confiant des données de recensement au projet IPUMS, les agences statistiques sont soulagés des risques, des tâches et responsabilités de diffuser des microdonnées aux chercheurs. De plus, les procédures standard d'IPUMS, employées dans une majorité des instituts statistiques du monde, assure la sécurité des données. L'agence statistique isolée qui diffuse des microdonnées, court des risques et engage des responsabilités substantielles, les coûts de traitements en ressources et en matériels sont significatifs pour un rendement relativement faible, vu le nombre restreint d'utilisateurs. Le projet IPUMS International permet de réaliser des économies d'échelle substantielles en offrant des niveaux élevés de sécurité et diffuse des méta- et des microdonnées intégrées, ce qui facilite la recherche scientifique. La tabulation interactive des variables intégrées permet d'envisager la croissance du nombre des utilisateurs et une utilisation plus importante des données de recensement sans coût supplémentaire pour l'office statistique national.

29. Les instituts statistiques participant à l'initiative Ipums-Afrique sont invités à confier des métadonnées et des microdonnées pour le cycle 2010 de recensement. Les instituts qui ne participent pas encore au projet sont invités à le faire.

Bibliographie

- Anderson, Margo and Stephen E. Fienberg. (2001). "U.S. census confidentiality: Perception and reality," International Statistical Institute Biennial Meeting (Seoul). (unpub.)
- CES (2007), "Managing statistical confidentiality and microdata access: Principles and guidelines on good practice", published by the Conference of European of Statisticians:
<http://www.unece.org/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf>
- Davern, Michael, Steven Ruggles, Tami Swenson, J. Trent Alexander and J. Michael Oakes. (2009) "Drawing statistical inferences from historical census data, 1850-1950," *Demography*, 46(3):589-603.
- Garenne, Michel; Robert McCaa and Kouroum Nacro. (2008) "*Maternal Mortality in South Africa in 2001: From Demographic Census to Epidemiological Investigation*," *Population Health Metrics*, 6:4(Aug) 1-13.
- McCaa, Robert and Albert Esteve. (2006). "IPUMS-Europe: Confidentiality measures for licensing and disseminating restricted access census microdata extracts to academic users," *Monographs of official statistics: Work session on statistical data confidentiality*. Luxembourg: Office for Official Publications of the European Communities, pp. 37-46.
- McCaa, Robert; Albert Esteve, Steven Ruggles and Matt Sobek. (2006) "Using integrated census microdata for evidence-based policy making: the IPUMS-International global initiative," *African Statistical Journal*, 2:83-100
- McCaa, Robert; Steven Ruggles, Michael Davern, Tami Swenson, and Krishna Mohan, Palipudi. (2006) "IPUMS-International High Precision Population Census Microdata Samples: Balancing the Privacy-Quality Tradeoff by Means of Restricted Access Extracts," *Privacy in Statistical Databases* (New York: Springer), 375-382.
- McCaa, Robert and Wendy Thomas. (2009) "IPUMS-International: lessons from 10 years of archiving and disseminating census microdata," 57th Session International Statistical Institute, Durban, South Africa (unpub.) www.hist.umn.edu/~rmccaa/ipums-global (scroll to "IPM 100 Microdata Session" and click "IPUMS")
- Megill, David. (2007). "Technical documentation for public use microdata samples files for the 1997 Mozambique census of population and housing," Instituto Nacional de Estatística, Maputo (unpub.). www.hist.umn.edu/~rmccaa/ipums-africa (click "electronic holdings," scroll to Mozambique, and click "sample design" for the 1997 census.
- Statistics Canada. (2008) "Metadata requirements for archiving structured data," Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS), Luxembourg.
- Trewin, Dennis (2007). *A review of IPUMS-International*. (unpub.).

Annexe A. Formulaire "A" Enregistrement des recensements et des échantillons

Instructions: Décrire brièvement le recensement/échantillon de données. Format libre	
Nom: _____ email: _____ date: _____	
Veuillez vérifier les données des recensements déjà intégrés de pays dans: https://international.ipums.org/international/samples.shtml	
Adresser les questions à Robert McCaa: rmccaa@umn.edu	
Caractéristiques de recensement (pays): _____	
Titre	
Agence responsable	
Population étudiée	
De jure or de facto	
Unité dénombrée	
Date du recensement	
Période du travail de terrain	
Formulaires du dénombrement	
Type de travaux sur le terrain	
Personnes interrogées	
Couverture	
Caractéristiques des échantillons de microdonnées	
Source des microdonnées	
Conception d'échantillon	
Unité d'échantillon	
Fraction d'échantillon	
Dimension de l'échantillon	
Poids d'échantillonnage (décrire)	
Unités identifiées ("oui" = unité identifiée ; introduire autrement " aucun")	
Unités d'habitation	
Unités vacantes	
Ménages	
Personnes	
Logements collectifs	
Habitants réguliers/irréguliers	
Populations spéciales	
plus petit niveau géographique	
module spécial (mortalité, etc.)	
Définitions des Unités	
Unités d'habitation	
Ménages privés	
Logements collectifs	
Populations sans logement	
Populations spéciales	
Métadonnées confiées (noms de fichier électroniques ou titres des copies sur papier)	
Fiches de recensement	
Manuels/Instructions des agents	
Dictionnaire de données	
Dictionnaire des codes (éducation, profession, branche, géographie,..)	
Tables de correspondance (éducation)	
Resultats officiels	
Rapports méthodologiques	
Rapport d'enquêtes	
Erreurs d'échantillonnage	
Découpage géographique (au besoin)	

Annexe B.

**Lettre d'accord entre
Integrated Public Use Microdata Series International et
le Haut-commissariat au Plan du Royaume du Maroc**

Objet: L'objet de cette lettre est de spécifier les termes et conditions régissant la diffusion par **Integrated Public Use Microdata Series International**, Université du Minnesota, des macro- et micro-données fournies par **le Haut-commissariat au Plan du Royaume du Maroc**.

1. **Propriété.** **Le Haut-commissariat au Plan du Royaume du Maroc** est le propriétaire et le détenteur des droits de propriété intellectuelle (incluant les droits de copie) des macro- et micro-données de son pays acquises auprès de lui par l'Université du Minnesota pour être distribuées par **Integrated Public Use Microdata Series International**.
2. **Usage.** Ces données sont destinées à l'usage exclusif de l'enseignement ainsi que de la recherche et édition scientifiques. Elles ne sauraient être utilisées à aucune autre fin, sauf accord écrit explicite du **Haut-commissariat au Plan** donné au préalable.
3. **Autorisation.** Pour consulter ou obtenir copie des micro-données intégrées du Maroc auprès d'**Integrated Public Use Microdata Series International**, l'utilisateur potentiel doit tout d'abord déposer un formulaire d'autorisation électronique l'identifiant comme chercheur principal, indiquant ses nom, adresse électronique et l'institution dont il relève. Obligation lui est faite d'exposer les fins poursuivies par son projet de recherche et de se conformer au règlement ci-inclus. Ce projet une fois approuvé, un mot de passe lui permettra d'obtenir des données auprès des serveurs et autres moyens électroniques de diffusion d'**Integrated Public Use Microdata Series International**, du **Haut-commissariat au Plan** ou d'autres distributeurs autorisés. L'approbation du projet entraîne pour l'utilisateur une licence d'acquérir les macro- ou micro-données intégrées du Maroc auprès d'**Integrated Public Use Microdata Series International** ou autres distributeurs autorisés. Aucun autre titre ou droit n'est concédé à l'utilisateur.
4. **Restrictions.** Il est interdit aux utilisateurs de données obtenues d'**Integrated Public Use Microdata Series International** ou autres distributeurs autorisés d'en faire quelque usage commercial que ce soit ou d'en tirer profit, à titre privé ou sous une autre forme.
5. **Confidentialité.** Les utilisateurs traiteront de façon rigoureusement confidentielle tout ce qui touche aux personnes et ménages. Toute tentative d'identification d'une personne, d'une famille, d'un ménage, d'un lieu de résidence, d'une organisation ou d'une entreprise à partir de ces micro-données est strictement proscrite. Il est également interdit de prétendre que telle ou telle personne ou entité a été identifiée à partir de ces données.
6. **Sécurité.** Les utilisateurs prendront les mesures de sécurité appropriées pour prévenir tout accès non autorisé aux micro-données recueillies auprès d'**Integrated Public Use Microdata Series International** ou de ses partenaires.

7. **Publication.** Est autorisée la publication des données et de l'analyse qui en est faite, après recherche à partir des macro- et micro-données concernant le Maroc, dans le cadre de communications scientifiques, d'articles de journaux de recherche ou autres parutions du même type. Il est exigé des auteurs qu'ils citent le **Haut-commissariat au Plan** et l'**Integrated Public Use Microdata Series International** comme sources des données concernant le Maroc. Ils doivent également mentionner que les résultats qu'ils ont obtenus et les opinions qu'ils émettent n'engagent que l'auteur/utilisateur.
8. **Violations.** Toute violation de la licence d'utilisation (§ 3) peut se conclure par un blâme professionnel, par la perte de l'emploi et/ou par des poursuites au civil. L'Université du Minnesota, des organisations scientifiques nationales et internationales et le **Haut-commissariat au Plan** veilleront à la bonne exécution des dispositions de cet accord.
9. **Partage des tâches.** L'**Integrated Public Use Microdata Series International** fournira au **Haut-commissariat au Plan** des copies électroniques de sa documentation et des données correspondant aux micro-données qu'il aura intégrées. Il transmettra également les rapports que lui auront adressés les utilisateurs des données.
10. **Juridiction.** Les désaccords qui puissent surgir seront traités par biais de la conciliation, la transaction et une attitude amicale. Si l'accord s'avérait impossible par ces moyens, un Tribunal d'Accord serait établi pour juger la situation d'accord avec la loi. Ce Tribunal serait composé par un arbitre sélectionné par le Tribunal International d'Arbitrage (TIA) . L'accord devra s'accorder aux principes généralement acceptés de la Loi Internationale.
11. **Ordre de priorité.** En cas de conflit entre un terme ou une condition de cette Lettre d'Accord et un terme ou une condition d'un contrat auquel cette Lettre est rattachée, le terme ou condition de cette Lettre d'Accord devra prévaloir.

Date: 04 mai 2009

Signature :

Le Haut-commissariat au Plan du Royaume du Maroc

Représenté par Monsieur Ahmed LAHLIMI ALAMI, Haut-commissaire au Plan

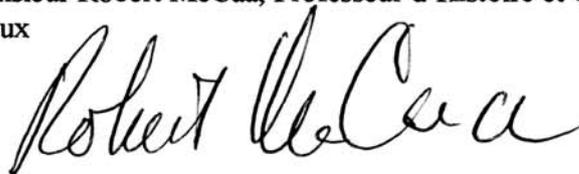


Date: 04 mai 2009

Signature :

Integrated Public Use Microdata Series International-Université du Minnesota

Représenté par Monsieur Robert McCaa, Professeur d'Histoire et Coordinateur des projets internationaux



Annexe C. Exemple de métadonnées internationales -Ipums Caractéristiques de recensements et d'échantillons : Afrique du Sud, 1996, 2001 et 2007			
https://international.ipums.org/international/sample_designs/sample_designs_za.shtml			
	1996	2001	2007
Caractéristiques du recensement			
Titre	Recensement 1996	Recensement 2001	Enquête communautaire 2007
Agence de recensement	Statistiques Afrique du Sud	Statistiques Afrique du Sud	Statistiques Afrique du Sud
Population recensées	Chaque personne présente en Afrique du Sud dans la nuit du 9 au 10 octobre 1996, doit être recensée	Les personnes présentes dans le pays dans la nuit du 9 au 10 octobre 2001. Les personnes appartenant à des ménages, dans les pensions, hôtels, hôpitaux et les lieux de vie communale et même les sans-abri, ont tous été visitées.	Les membres du ménage résidant dans l'unité d'habitation au moins quatre nuit par semaine et présentes au cours des quatre dernières semaines avant la date de recensement, plus les visiteurs qui passent la nuit avant l'interview.
De jure/de facto	De facto	De facto	De facto
unité dénombrée	Visites dans les zones de recensement	Ménages et personnes dans les zones (contient de 100 à 250 ménages).	Ménage
Date de début recensement	10 Octobre 1996	10 Octobre 2001	7 Février 2007
Période de collecte	Du 10 octobre au 30 octobre 1996 bien que, dans certains cas, la collecte a été menée jusqu'en décembre afin de s'assurer que toutes les personnes avaient été recensées.	—	Une période de trois semaines (7-28 février 2007) avec une période complémentaire d'une semaine pour les non-réponses (1-7 mars)
Questionnaires	Cinq questionnaires différents : 1) Un questionnaire ménages est rempli pour chaque ménage. Avec informations sur chaque personne du ménage, âge et sexe ainsi que sur le logement, par exemple accès à l'électricité et à l'eau du robinet. 2) Un questionnaire individuel, qui est rempli par les personnes vivant seuls, par exemple celles vivant dans les pensions ou les communautés. 3) Un questionnaire rapide pour les pensions. 4) Un questionnaire pour les institutions, prisons, hôtels de touriste ou maisons pour gens âgés. 5) Un questionnaire pour les sans-abri.	Trois questionnaires différents de recensement sont utilisés - ménages (formulaire A), personnes dans les institutions (formulaire B), les institutions (formulaire C). Le questionnaire A est également utilisé pour les pensions de travailleurs, les pensions pour étudiants, les hôtels et les résidences pour personnes âgées; les formulaires B et C sont utilisés pour hôtels de touriste et les sans-abri.	Un questionnaire unique « ménages » avec questions sur l'unité d'habitation, le ménage, et les personnes.
Type d'interviews Passation du questionnaire	On donne aux répondants le choix soit de la visite ou soit de remplir le questionnaire eux-mêmes. La grande majorité des personnes choisit d'être visitée. L'agent recenseur est un	Chaque agent recenseur doit remplir un ou plusieurs questionnaires pour chaque unité d'habitation visitée. Les ménages sont encouragés à être interviewés par un agent	Interviews directes

	<p>membre temporaire de Stats SA nommé pour collecter des informations sur les personnes qui vivent dans les ménages dans le logement privé, par exemple, un appartement dans un bloc d'appartements, une cabane ou une unité d'habitation traditionnelle.</p> <p>Un agent recenseur spécial est nommé pour recenser des personnes dans les institutions tel que pensions, prisons, hôtels et les hôpitaux. Les agents recenseurs spéciaux collectent également des informations sur les sans-abri</p>	<p>recenseur si possible. Alternativement, un répondant pouvait remplir le questionnaire pour une collecte plus tard, si les circonstances le permettent. Les agents recenseurs fournissent les traductions des questions dans les dix autres langues officielles.</p>	
répondant			Le chef de famille ou du ménage, ou le membre du ménage responsable le plus âgé si le chef de ménage n'est pas présent.
Sous-compte	10,7%	Varie par province : 14,07% à 22,51% pour les personnes ; 15,55% à 26,21% pour les ménages.	Les institutions et certains ménages dénombrés et classés dans les secteurs des loisirs ou les institutions.
Couverture			Habitations privées et résidences privées de vacances ou saisonnières; pensions de travailleurs et couvent, monastère, retraite religieuse, mais sans autres lieux de vie collectifs
Caractéristiques de l'échantillon de microdonnées			
Source des microdonnées	Statistiques Afrique du Sud	Statistiques Afrique du Sud	Statistiques Afrique du Sud
Conception de l'échantillon	<p>Échantillon systématique de 10% des ménages du recensement stratifié comme spécifié ci-dessous. L'échantillon des personnes de 10% est obtenu en incluant toutes les personnes dans ces ménages plus les personnes tirées dans un échantillon systématique séparé de 10% à partir des personnes vivant dans les institutions et les pensions spéciales</p> <p>NOTE : 19 secteurs dans la province orientale du Cap ne sont pas organisés en ménages, en raison d'une erreur dans le fichier de données original. 1,3% de l'échantillon est touché.</p>	Échantillon stratifié systématique tiré par l'institut de statistiques d'Afrique du Sud.	<p>Échantillonnage aléatoire systématique à deux niveaux et stratifié. Stade 1 : Le cadre d'échantillonnage contient 79.466 secteurs de collecte (unités primaires) stratifiés par municipalité. L'échantillonnage aléatoire systématique est utilisé sur les secteurs de collecte au sein des municipalités. Dans les municipalités avec moins de 30 secteurs, tous les secteurs sont automatiquement sélectionnés. Dans les municipalités avec plus de 30 + secteurs, une proportion de 19% des secteurs sont retenus. Si le nombre de secteurs de dénombrement d'une municipalité est inférieur à 30, l'échantillon municipal est augmenté jusqu'à 30. Stade 2 : Une proportion fixe de 10% des unités d'habitation dans un secteur de collecte est choisie. S'il y a moins de 10 unités</p>

			d'habitation dans un secteur, la sélection est augmentée à 10 unités d'habitation. Tous les ménages dans les unités d'habitation sélectionnées sont couverts. Aucun remplacement des refus de réponse des unités d'habitation ou des absences. Taux de réponse 93,9%.
Unité témoin	Ménages et personnes	Ménages	Région de recensement, séjour
Taux sondage	10%	10%	2.2%
personnes enquêtées	3,621,164	3,725,655	1,047,657
Pondérations du sondage	Calculées par l'agence de recensement et doivent être utilisées pour la plupart des types d'analyses. La pondération est le coefficient de correction pour le sous-dénombrement (pour les ménages ou les personnes selon le cas); il est multiplié par 10 pour extrapoler l'échantillon de 10% de la population.	Calculés par l'agence de recensement et doivent être utilisées pour la plupart des types d'analyses.	Calculées par l'agence de recensement et doivent être utilisées pour la plupart des types d'analyses.
Unités identifiées			
Unités d'habitation	Non	Non	Non
Unités vacantes	Non	Non	Non
Ménages	Oui	Oui	Oui
personnes	Oui	Oui	Oui
Collectifs	Oui	Oui	Oui
+ petit échelon géographique	Secteurs judiciaires avec 20.000 + de population en 2001	Secteurs judiciaires et municipalités avec 20.000 + de population en 2001	Municipalités avec 20.000 + de population
définitions des Unités			
Unités d'habitation	Une unité d'habitation occupée est un local habité par un ou plusieurs ménages la nuit du recensement. Une unité d'habitation occupée peut être un logement, une salle, un appartement, une cabane, une hutte, une tente, une caravane, une péniche aménagée en habitation, un magasin, une école, etc.	Toute structure destinée ou utilisée pour l'habitation humaine.	Une structure ou une partie d'une structure ou d'un groupe de structures occupés ou censés être occupés par un ou plusieurs ménages
Ménages Privés	Un ménage est composé de personne, ou d'un groupe de personnes, qui occupent une unité d'habitation commune (ou une partie) pour au moins quatre jours par semaine et qui se partagent ensemble des produits alimentaires ou essentiels. Les personnes qui occupent la même	Un groupe de personnes qui vivent ensemble, et se fournissent conjointement avec les produits alimentaires pour vivre, ou une personne seule qui vit seulement.	Un ménage est un groupe de personnes qui vivent ensemble et se fournissent conjointement avec les produits alimentaires pour la vie, ou une personne seule qui vit seulement.

	<p>unité d'habitation, mais qui ne partagent pas des produits alimentaires ou essentiels, sont recensées comme des ménages distincts. Par exemple, les personnes qui partagent une unité d'habitation, mais qui achètent des produits alimentaires et mangent séparément, sont comptées comme les ménages distincts.</p>		
Logements collectifs	<p>Une unité d'habitation spéciale n'est pas occupée par un ménage. Comme une prison, un hôtel, une pension, une maison pour personnes âgées, etc. En outre, il y a des « hostels » : une forme collective de logement construit spécifiquement pendant l'ère d'apartheid pour les employés, l'usine, les employés municipaux ou autres.</p>	<p>Lieux de vie où certains équipements sont partagés par des groupes de personnes ou de ménages. Ils incluent des pensions, des hôtels et des institutions.</p>	<p>Les Lieux de vie collectifs sont (1) des domiciles distincts et indépendants destinés à l'habitation par des grands groupes de personnes ou de plusieurs ménages. Ces lieux ont des équipements communs, équipements de cuisson et d'ablution, salons ou dortoirs qui sont partagés par les occupants</p>
Institution			<p>Un type particulier de lieu de vie collective, pour les personnes avec une caractéristique commune qui vivent sous un régime commun. Les exemples sont : hôpital/clinique, centre de soins, institution de protection d'enfance, orphelinat, etc.</p>