

III. STATISTICAL DEVELOPMENT - SELECTED AREAS

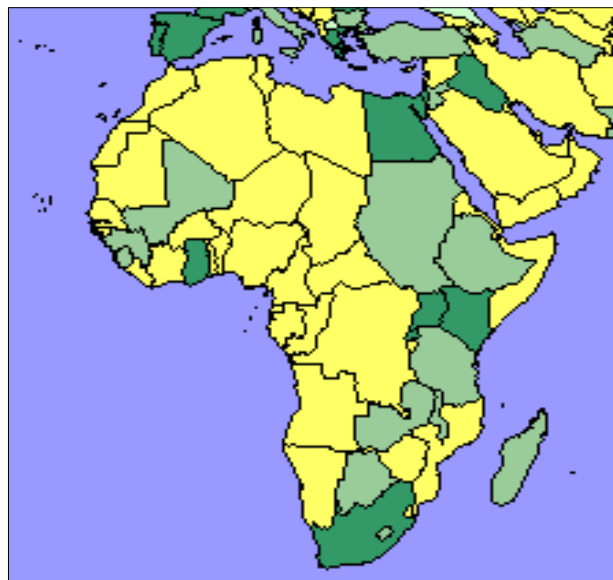
Integrated Microdata Samples for 111 Censuses Free to Accredited Researchers

Robert McCaa, University of Minnesota Population Center, USA

The Integrated Public Use Microdata Series (IPUMS) international is an integrated series of microdata samples from 1860 to present housed at the University of Minnesota Population Centre. It aims at collecting, preserving and distributing census data from around the world, harmonizing data and disseminating the data free of charge to accredited researchers. This article presents the current status of the participation of African countries in the IPUMS international project.

Thanks to the cooperation of the Central Agency for Public Mobilization and Statistics (CAPMAS) and Ghana Statistical Services, integrated samples for the 1996 census of Egypt and the 2000 census of Ghana are now available to accredited researchers from IPUMS International web-site: <https://international.ipums.org> Integrated census samples for Kenya, Rwanda, South Africa and Uganda are also available, along with those of 29 countries beyond Africa (see table and map).

Statistical offices of African countries that have not yet joined the IPUMS global initiative are cordially invited to contact Prof. Robert McCaa at the University of Minnesota (rmccaa@umn.edu). Additional details are available at www.hist.umn.edu/~rmccaa/ipums-global.



Light Green = Microdata entrusted
Dark Green = Samples launched

and dissemination. 18 African statistical agencies have endorsed the project memorandum of understanding and entrusted microdata to the Minnesota Population Center, including, in addition to the six countries listed above, Botswana, Ethiopia, Guinea (Conakry), Lesotho, Madagascar, Malawi, Mali, Mauritius, Sierra Leone, Sudan, Tanzania, and Zambia. World-wide, more than 80 statistical offices, encompassing almost 80% of the world's population, have joined the IPUMS initiative. Although, at present, African statistical institutes are relatively under-represented, the National Science Foundation (USA) advisory board strongly supports a concerted effort to attract widespread participation by African official statisticians and academic researchers.

Benefits are many and costs are negligible. The project pays the official statistical agency a license fee of US\$5,000 per census dataset of one million or more person records and then disseminates integrated samples to researchers without charge. Official statisticians have the opportunity to associate with statistical agencies in a global initiative at project sponsored workshops, such as the upcoming meeting in Cairo (Oct. 11-16) and, in 2009, the 57th International Statistical Institute Biennial Convention in Durban, South Africa. Participation in IPUMS promotes transparency and builds trust. Integrated microdata facilitate high-quality research and fact-based policy making. The project places few demands on its partners. The integration work is performed by the Minnesota Population Center according to national practices and international principles. Both microdata and documentation are disseminated with minimal risk and maximum benefit. The project pays the costs of recovering historical censuses and promotes the highest archival standards for long-term preservation. Once the integration of the samples is completed, typically there is a surge in use of the microdata by stakeholders and top-ranked researchers as well as university students. Increased usage of expensive to collect

Country and Census	Sample Density (%)	Person Records
Egypt 1996	10	5,902,243
Ghana 2000	10	1,894,133
Kenya 1989	5	1,074,098
Kenya 1999	5	1,407,547
Rwanda 1991	10	742,918
Rwanda 2002	10	843,392
South Africa 1996	10	3,621,164
South Africa 2001	10	3,725,655
Uganda 1991	10	1,548,460
Uganda 2002	10	2,497,449

Continent	Countries	Samples	Person records (millions)
Africa	6	10	23.1
Asia	8	17	52.9
Europe	10	31	37.2
Americas	11	53	149.8
Total	35	111	263.0

IPUMS projects consist of microdata recovery, integration of samples

III. STATISTICAL DEVELOPMENT - SELECTED AREAS

census data is also an important benefit. Finally, IPUMS solves the problem of how to disseminate microdata by minimizing the risks to the national statistical offices for maximum benefit.

Researchers, once accredited, use the IPUMS site to obtain custom-tailored samples by country, census year, sample density, and variables. With over 250 million person records and thousands of variables, the IPUMS database is too large to simply download. Instead the IPUMS extract engine is used to “place an order”, much like ordering music or videos over the Internet. Once the extract is ready for delivery, the researcher is sent an email to download the extract. The researcher is then free to analyze the dataset using his or her own hardware and software. Statistical instructions are provided in three flavours—SPSS, SAS, and STATA—and researchers may import the data into any statistical package that reads ASCII files.

Complete documentation is readily available from the website. The IPUMS dynamic metadata system makes it possible for researchers to compare the wording, in English, of any question for any combination of census years and countries. Instructions to enumerators, obtained verbatim from the official training manuals, may also be compared in the same way. Researchers may easily compare concepts, definitions, and instructions for one or one hundred censuses simply by making a few selections. In addition, the original source documentation for each census is available in the official language, so that if there is any question about how the IPUMS team may have interpreted a particular word, phrase, sentence or an entire text, there are no obstacles to readily checking the integration work.

With more than 2,000 registered users, satisfaction is exceedingly high both for researchers and participating national statistical offices. Support by the National Science Foundation and the National Institutes of Health of the USA is sustained and unstinting. An African regional initiative, modelled after successful projects for Latin America (2003-2012) and Europe (2004-2009), is currently being organized with the new African Centre for Statistics. National Statistical Offices that are not presently associated with the project are invited to email the project coordinator.



Best Practice in Mainstreaming Sectoral Concerns into the National Strategy for the Development of Statistics (NSDS) during Design Process

O.O. Ajayi, AfDB Consultant

Introduction and Rationale

The rationale for mainstreaming sector concerns into the NSDS is to achieve uniform and balanced strengthening of the National Statisti-

cal Systems (NSS). Statistical development has often been focused on the National Statistics office (NSOs) to the neglect of the other producers of data whereas there is need for holistic development covering all the sectors and sub-national domains. This ensures that comprehensive statistics delivered to users to meet national, sub-national, sectoral and international needs.

Objectives

To mainstream sectors means systematic involvement and integration of the sectors in the design of their own strategic plans as building blocks for the NSDS. This also ensures joint implementation of the various plans. One objective is to adopt a common best approach (a model) to design such an integrated strategy. Another objective is to improve the quality of data needed to monitor and evaluate the poverty reduction strategies (PRS) as well as progress towards the Millennium Development Goals (MDG'S).

Countries, some with centralized statistical systems and some with de-centralised systems, are at different stages in developing an integrated NSDS. These stages include:

- Stage I - where sectors have not been mainstreamed and is the existing Statistical Master Plan is effectively only for the NSO;
- Stage II - where sectors have been partially mainstreamed in a decentralized statistical system –here some sector concerns are addressed in the plan but only as a token;
- Stage III - where sectors have been partially mainstreamed without the sectors involvement in a centralized statistical system;
- Stage IV - when some sectors have been fully mainstreamed; and
- Stage V - where all the sectors have been fully mainstreamed.

Both Stages IV and V constitute good practice and IV is realistic and practical to achieve particularly where we have many Ministries, Departments and Agencies (MDAs) in a country. Stage V could eventually be accomplished over time by appropriate phasing of the involvement of the various sectors..

It is very important for countries to correctly assess the stages they have achieved so that they can plan the appropriate next steps for improvement with the objective of achieving a coordinated NSDS.

The suggestions for next steps relating to each of the stages are correspondingly:

- Stage I- conduct an integrated NSDS using the guide to planning a coordinated NSS.
- Stage II – review the NSDS document and mainstream sectoral strategies using the same guide
- Stage III – review and update the NSDS documents with sectors using the guide.
- Stage IV – phase in more sectors in planned stages also using the guide
- Stage V - begin to monitor and evaluate the NSDS implementation.

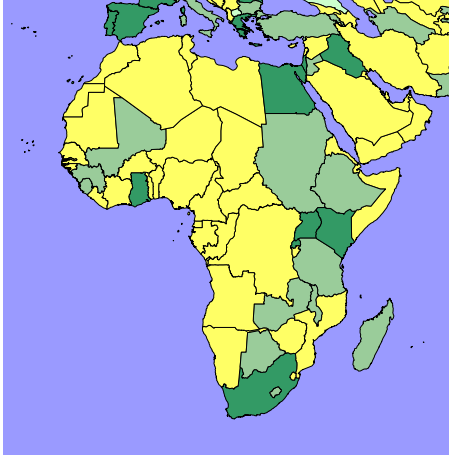
Article pour le Bulletin du Centre Africain pour la Statistique, parution de juin 2008 :
 Échantillons harmonisés de microdonnées pour 111 Recensements librement accessibles aux
 chercheurs régulièrement autorisés à <https://international.ipums.org/international>
 Robert McCaa, Professeur d'Histoire, Centre de Population du Minnesota

Grâce à la coopération de l'Agence Centrale pour la Mobilisation Publique et la Statistique (CAPMAS) et des Services de Statistique du Ghana, des échantillons harmonisés du recensement 1996 de l'Egypte et du recensement 2000 du Ghana sont maintenant à la disposition des chercheurs régulièrement autorisés sur le site web d'IPUMS-International :

<https://international.ipums.org/international>

Les échantillons harmonisés de recensement pour le Kenya, le Rwanda, l'Afrique du Sud et l'Ouganda sont également disponibles, avec ceux de 29 autres pays en dehors de l'Afrique (voir le tableau et la carte ci-après).

Les Offices de statistique des pays africains qui n'ont pas encore joint l'initiative mondiale d'IPUMS sont cordialement invités à contacter le Professeur Robert McCaa de l'Université du Minnesota (rmccaa@umn.edu). Des détails complémentaires sont disponibles à l'adresse suivante : www.hist.umn.edu/~rmccaa/ipums-global.

Échantillons harmonisés lancés par pays/continent :				
https://international.ipums.org/international				
Pays et Recensement (Afrique seulement)	Taille de l'échantillon (%)	Données individuelles		
Egypte 1996	10	5.902.243	 <p>Vert Clair = Microdonnées confiées Vert Foncé = Echantillons lancés Jaune = ne participant pas, pour le moment</p>	
Ghana 2000	10	1.894.133		
Kenya 1989	5	1.074.098		
Kenya 1999	5	1.407.547		
Rwanda 1991	10	742.918		
Rwanda 2002	10	843.392		
Afrique du Sud 1996	10	3.621.164		
Afrique du Sud 2001	10	3.725.655		
Ouganda 1991	10	1.548.460		
Ouganda 2002	10	2.497.449		
Continent	Pays	Echantillons		Données individuelles (millions)
Afrique	6	10		23,1
Asie	8	17	52,9	
Europe	10	31	37,2	
Amériques	11	53	149,8	
Total	35	111	263,0	

Les projets d'IPUMS consistent en la récupération de microdonnées, l'harmonisation d'échantillons et la diffusion. 18 Agences africaines de statistique ont adhéré au protocole d'accord de projet et confié des microdonnées au Centre de Population du Minnesota, y compris, en plus des six pays énumérés ci-dessus, le Botswana, l'Éthiopie, la Guinée (Conakry), le Lesotho, Madagascar, le Malawi, le Mali, les Îles Maurice, la Sierra Leone, le Soudan, la Tanzanie, et la Zambie. Dans le monde entier, plus de 80 offices de statistique, couvrant presque 80% de la population du monde, ont joint l'initiative d'IPUMS. Bien que, actuellement, les instituts africains de statistique sont relativement sous-représentés, le comité consultatif de la Fondation Nationale pour la Science (États-Unis) soutient fortement un effort concerté dans le but d'attirer la participation plus large des statisticiens officiels africains et des chercheurs universitaires.

Les avantages sont nombreux et les coûts sont négligeables. Le projet paie à l'agence officielle de statistique une redevance de 5.000 US\$ par échantillon de données de recensement d'un million d'individus ou plus et dissémine les échantillons harmonisés aux chercheurs sans frais. Les statisticiens officiels ont ainsi l'occasion de s'associer aux agences de statistique à travers une initiative mondiale au cours d'ateliers parrainés par le projet, tels que la prochaine réunion au Caire (11-16 octobre) et, en 2009, la 57^{ème} convention bisannuelle de l'Institut International de Statistique à Durban, Afrique du Sud. La participation à IPUMS favorise la transparence et établit la confiance. Les microdonnées harmonisées facilitent la recherche de haute qualité et la prise de décision politique basée sur les faits. Le projet sollicite très peu ses partenaires. Le travail d'harmonisation est effectué par le Centre de Population du Minnesota selon des pratiques nationales et des principes internationaux. L'ensemble des microdonnées et de la documentation sont disséminées avec le minimum de risques et le maximum d'avantages. Le projet supporte les coûts de récupération des recensements historiques et fait la promotion de niveaux d'archivage les plus élevés en vue de la conservation à long terme des données. Une fois l'harmonisation des échantillons accomplie, en général il y a une forte croissance dans l'utilisation des microdonnées par les partenaires et des chercheurs de haut niveau aussi bien que par des étudiants d'université. Une utilisation accrue des données de recensement très coûteux en matière de collecte est également un avantage important. En conclusion, IPUMS résout le problème de la façon dont il faut disséminer des microdonnées en réduisant au minimum les risques aux offices nationaux de statistique pour le maximum d'avantages.

Les chercheurs, une fois accrédités, exploitent le site d'IPUMS pour obtenir les échantillons sur mesure par pays, année de recensement, taille de l'échantillon, et variables. Avec plus de 250 millions de données individuelles et des milliers de variables, la base de données d'IPUMS est trop grande pour être téléchargée simplement. En conséquence, le moteur d'extraction d'IPUMS est doté d'un utilitaire " lancer une commande ", tout comme on peut commander de la musique, ou de la vidéo sur l'Internet. Une fois que l'extrait est prêt pour la livraison, le chercheur reçoit un courriel qui l'autorise à effectuer le téléchargement. Le chercheur est alors libre pour analyser l'ensemble de données en utilisant ses propres matériels et logiciels. Des instructions statistiques sont fournies dans trois logiciels — SPSS, SAS, et STATA — et les chercheurs peuvent importer les données dans n'importe quel autre logiciel statistique qui lit des fichiers ASCII.

La documentation complète est aisément fournie par le site web. Le système dynamique de métadonnées d'IPUMS permet aux chercheurs de comparer l'écriture des mots, en Anglais, de n'importe quelle question pour n'importe quelle combinaison des années de recensement et des

pays. Les instructions aux agents recenseurs, obtenus in extenso à partir des manuels officiels de formation, peuvent également être comparées de la même manière. Les chercheurs peuvent facilement comparer des concepts, des définitions, et des instructions pour un ou cent recensements tout simplement en faisant quelques sélections. En outre, la source originale de la documentation pour chaque recensement est disponible dans la langue officielle (du pays concerné), de telle sorte qu'en cas de préoccupation au sujet de la façon dont l'équipe d'IPUMS a pu avoir interprété un mot particulier, une expression, une phrase ou un texte entier, il n'y a aucun obstacle à vérifier aisément le travail d'harmonisation.

Avec plus de 2.000 utilisateurs enregistrés, la satisfaction est excessivement forte pour l'ensemble des chercheurs et des offices nationaux de statistique participants. L'appui apporté par la Fondation Nationale pour la Science et les Instituts Nationaux de la Santé des Etats-Unis est soutenu. Une initiative régionale africaine, conçue sur le modèle des projets réussis pour l'Amérique Latine (2003-2012) et l'Europe (2004-2009), est actuellement en cours d'organisation. Les Offices Nationaux de Statistique qui ne sont pas actuellement associés au projet sont invités à adresser un courriel au coordonnateur du projet.