

**Statistical Confidentiality  
and the Dissemination of Restricted-Access Integrated Census Microdata Extracts:  
The Case of Kenya, 1969-1999  
Robert McCaa and Agnes Odinga**

Research for this paper was funded in part by the  
National Science Foundation of the United States, grant SBR-9908380

**Abstract.** Kenya has one of the richest microdata collections in Africa and, indeed, the world. Beginning in 1969 the Central Bureau of Statistics has not only conducted censuses at regular decennial intervals but also conserved the complete microdata for the two most recent censuses. Before May 2002, when the IPUMS released five percent samples of the censuses of 1989 and 1999, this valuable trove was not readily available to scholars or public policy-makers. Access has been the principal obstacle, not only in the case of Kenya, but for many other countries around the world. An important step in providing broader, if still restricted access—and reaping the benefits to be gleaned from these valuable sources—is to ensure that the data are anonymized according to the requirements of the corresponding official statistical agency. The IPUMS International project, in cooperation with a group of National Statistical Agencies in Europe, the Americas, Asia, and Africa, is acquiring anonymized census samples of individuals and households from official statistical agencies, and, for others, constructing anonymized samples. This paper summarizes statistical confidentiality issues and, then as an example, discusses the requirements for release of five percent samples drawn from the 1989 and 1999 censuses of Kenya. The results are promising. Of the thirty-six person variables in the 1989 census microdata, five geographic variables are suppressed entirely (because they report finely detailed information on place of residence), and one social variable is suppressed: tribe/ethnicity. For the 1999 sample, “religion” is added to the list of suppressed variables. While suppression of these variables may disappoint purists who

demand total access to the original data, this approach strikes a balance between access and statistical confidentiality, sacrificing some degree of detail to safeguard statistical confidentiality to a maximum, yet still make it possible for researchers to gain access to the vast bulk of the Kenyan census microdata.

**Introduction.** Kenya has one of the richest collections of census microdata in the world, but also one of the least used. With five percent samples for the national censuses of 1979, 1989 and 1999 and a slightly smaller sample for 1969, the Central Bureau of Statistics of Kenya has produced an extraordinary statistical series with an unusually sophisticated set of variables (Table 1). The collection is all the more remarkable for its enormous size, its uniformity over time as well as its conformity with international standards. With records on more than four million individuals and households, the massive size of the Kenyan census samples has presented a substantial challenge to all but the best-endowed research institutions. Now however, the microcomputer revolution is overcoming the technical barriers to use these valuable data as well as other collections of census microdata around the globe.

Table 1

The Integrated Public Use Microdata Series International project assists researchers in unlocking the knowledge in census microdata. The list now exceeds 50 countries and encompasses more than half the world's populations. If the IPUMS International project is to succeed in lowering the barriers to knowledge from research based on high quality census microdata, the following three tasks must first be accomplished

1. Anonymize each census sample according to the requirements of the corresponding official statistical agency

2. Harmonize the samples according to a uniform design, census-by-census, variable-by-variable, code-by-code, and country-by-country
3. Disseminate harmonized microdata extracts to bona-fide researchers who agree to stringent usage and confidentiality restrictions. Extracts are custom-tailored with regard to countries, years, sub-populations, and variables according to the needs of each individual project. Data extracts are constructed by means of a series of selection menus at <http://www.ipums.org>.

Step two, harmonization, is the core of the project plan and the most intellectually challenging. First, though, the samples must be anonymized to safeguard statistical confidentiality. This paper addresses step one of the project, that is to explain the anonymization of the census microdata of Kenya, using the 1989 and 1999 samples as illustrations.

**Anonymizing census samples.** From the outset, it must be noted that notwithstanding the increasingly widespread access to census microdata there are no known cases of confidentiality violation. In the case of the United Kingdom, for example, Elliott and Dale observe that:

There has been no known attempt at identification with the 1991 SARs-nor in any other countries that disseminate samples of microdata (Elliott and Dale, 1999).

For the United States, the situation is also wholly positive:

In practice, such disclosure of confidential information is highly improbable. These microdata are samples, and none of them includes information on more than a tiny minority of the population. For this reason alone, any attempt to identify the characteristics of a particular individual, in say a five percent sample, would necessarily fail at least nineteen times out of twenty (McCaa and Ruggles, 2001).

Although there has never been even an allegation of confidentiality violation, statistical agencies remain vigilant to safeguard privacy, minimize the risk of disclosure,

protect the integrity and quality of statistical data, and at the same time, facilitate the use of an ever growing list of statistical data products, including microdata.

National statistical agencies have stringent regulations regarding access to census microdata, and Kenya is no exception. Indeed, of the 54 member-states of the International Monetary Fund's General Data Dissemination System, almost all are bound by law to respect the privacy of individuals and maintain statistical confidentiality of the information collected. Yet three of every four member-states make census microdata samples available to researchers either through third parties or upon direct application (see Table 2). In the case of the host country for this country, for example, the Australian Statistician declined to participate, not on the grounds of protecting statistical confidentiality (after all, Australian microdata, or CURFs, are readily available at most Australian universities), but “regardless of any proposed safeguards”<sup>1</sup>. For most countries, though, the issue is no longer a matter of "whether" census microdata can be anonymized and disseminated widely through the IPUMS-International system, but rather "how" the task should be accomplished. Before discussing our proposal for the Kenyan census microdata samples, it is fruitful to review some of the major developments in theory and practice in the field of statistical confidentiality protection over the past decade, particularly with regard to census microdata samples.

#### Table 2

Before detailing the measures used for minimizing disclosure risks in the 1989 and 1999 census sample, we begin by discussing the meaning of disclosure, and then the nature of disclosure risks.

---

<sup>1</sup> Email dated May 11, 2005, from Dr. Siu-Ming Tam, First Assistant Statistician, Information Management Division, Australian Bureau of Statistics in reply to query from Robert McCaa.

**Disclosure.** Disclosure refers to the possibility of, first, being able to identify individuals or entities in released statistical information and, second, revealing what the subject might consider to be “sensitive” information. Identification of an individual takes place when a one to one relationship between a record in released statistical information and a specific individual is established (Bethlehem, Keller and Pannekoek, 1990:38)<sup>2</sup>.

What are some of the ways in which disclosure can take place? In order for disclosure to occur an individual has to be within a sample of a population contained in the microdata. That individual also has to possess “unique” characteristics contained within the variables in the records. The information in the record consists of two disjoint parts: identifying and “sensitive” information (Bethlehem, 1990:39). Identifying information refers to those variables, called identifying variables or *key variables*, that allow one to identify a record—that is establish a one to one correspondence between the record and a specific individual. Well known *key variables* are name and address, but household composition, age, race, ethnicity, sex, place of residence, and occupation, or region of work can help identify individuals.

For disclosure to take place a snooper has to have *prior knowledge* or information about the individual.<sup>3</sup> If there is no prior information about a specific individual, identification and thus disclosure is impossible. *Prior knowledge* could be obtained from other databases, for instance those maintained by labor or employment departments, educational institutions, social security administration, registrars of births and deaths, the

---

<sup>2</sup> T. Dalenius (1977) “Privacy Transformation for Statistical Information Systems.” *Journal of Statistical Planning and Inference*, 1, 73-86, provides a slightly different definition of disclosure.

<sup>3</sup> Our discussion is based on the work of the following authors: G. Paas (1988), “Disclosure Risk and Disclosure Avoidance for Microdata” *Journal of Business and Economic Statistics*, 6, 487-500; G. Duncan and D. Lambert (1989), “The Risk of Disclosure for Microdata.” *Journal of Business*

postal service, ministry of health, etc. If the would-be intruder has access to some comprehensive list of the population or specific subgroups defined by a census variable, it would be possible to verify the identity of that person without the population list or other database. A snooper might also infer identity, particularly of a person in the public eye, such as a politician, actor or musician, who possesses unusual characteristics. In summary, in order to arrive at a match, an intruder who attempts to find information about an individual has to have access to prior information about the target individual whose identity and other key characteristics are known. In order to achieve disclosure, the intruder must link prior information for the target individual to the microdata records using the values of a set of *key variables* which are available both in the prior information and the microdata. A linkage is said to result in disclosure if each of the following two steps occur:

- a) Identification: whereby the snooper succeeds in linking an individual to microdata record and is able to verify with high probability that the link is correct.
- b) The snooper consequently obtains new information about this individual which was not available in the previous dataset (Skinner, Marsh, Openshaw and Wymer, 1994:33).

**Reflections on Disclosure Risks for Kenyan Census Microdata.** If disclosure can only take place when an intruder has *prior knowledge* or information about an individual with which a correct match is made using census microdata, thereby resulting in identification and subsequently disclosure, then other sources of information that both exist in Kenya and which a snooper might rely upon must be taken into account. We also

---

*and Economic Statistics*, 7, 207-217; G. Bethlehem, W. J. Keller and J. Pannekoek (1990), "Disclosure Control of Microdata," *Journal of the American Statistical Association*, 85, 38-45.

examine how accessible that information is to assess the likelihood of a snooper gaining prior information to make a match. Finally, we note ways of minimizing risks of identification in the Kenyan census microdata samples. Our analysis encompasses not only the pre-existing methods of disclosure control practiced by the Central Bureau of Statistics, but also those developed by the IPUMS International project.

A number of institutions and organizations in Kenya maintain data on different attributes of Kenyan subgroups and sub-populations. These organizations include the Registrar of Births and Deaths, Church Registries, the Registrar of Clubs and Societies, the Ministry of Labor, the Transportation Department, the Income Tax Authority, and the Ministry of Education, Health and Social Services. Unfortunately for the would-be intruder the databases of these organizations exist only in paper form. A few institutions such as the University of Nairobi and Kenyatta have computerized databases, but they are inaccessible to the “public” and even insiders (those who work within the institutions) have professional, legal and ethical obligations barring them from divulging private information to an outsider unless authorized and only then if that information is required for official purposes. This is not to say that there are no exceptional cases where information may be leaked by an ill-intentioned employee. It is however a very rare phenomenon.

There are a number of barriers that would limit a snooper’s ability to make a match. First and foremost, individual information filed and stored in paper form is inaccessible. Extracting records on individuals for the purpose of linking to a census database would constitute an extremely expensive process. Given the enormous resources required in terms of computing equipment and research time it is unlikely that anyone would engage in such an undertaking. Much more sensitive data are more easily, if also

illegally, obtained from other sources. Besides the technological barriers that limit intrusion into individuals' private information, records in paper form are subject to the 30 years rule while under the ministry or any government organization including the Kenya National Archives. Thirty years is a long time in a country, such as Kenya, where life expectancy is less than fifty. Then too, it would be folly to rely on such information for matching purposes since personal circumstances change with time.

Indeed, this is precisely the argument of an insightful study recently published in the *Journal of the Royal Statistical Society* (Dale and Elliott, 2001). Highly skilled researchers with unlimited resources working with the permission of the Office of National Statistics attempted to link an employment survey with the 1991 census microdata sample for the United Kingdom. The test demonstrated that the practical risks to identification are many orders of magnitude less than the theoretical risks (Dale and Elliott, 2001).

In the case of Kenya, far simpler ways of obtaining information exist, including word of mouth. Kenya, like many other African societies (with the exception of Islamic communities along the East Coast) until the early part of the 19<sup>th</sup> century relied almost exclusively on the transmission of information by word of mouth and lineage networks. Using lineage, friendship and community networks one can obtain far more information about an individual than is possible from paper records or census microdata. The risk of identification and subsequent disclosure may be somewhat greater for public individuals about whom more is known than for "ordinary" men and women. If an intruder intended to find out more about a public figure, for example a chief, a minister, church pastor or a renown healer—with some unique characteristics, then the possibility of making a match

would be heightened--unless measures are taken to further anonymize census microdata such as those proposed below.

**Disclosure Control in Kenya.** There are no known confidentiality violations of Kenyan census data, nor has there been a single allegation of a violation.<sup>4</sup> The Kenyan Central Bureau of Statistics and the Institute of Science and Technology through the office of the Vice President regulates all population research carried out in Kenya. This office only authorizes projects that are not prejudicial and guarantee anonymity and confidentiality of research subjects. In addition to obtaining a clearance, the researcher is required to sign a document stipulating that two copies of research findings will be deposited with the Kenyan government, which further protects the identity of research subjects.

The CBS has always taken great care to ensure that the statistical data are used for statistical purposes only. As a first step, and in conformity with standard practices of census agencies around the world, the Kenyan Central Bureau of Statistics never includes names or addresses in census data files. Computerizing such information would be prohibitively expensive and cause great delays in compiling even the simplest statistics on population. When conducting the census enumeration in the field, the KCBS assures respondents that:

the data requested from you and other persons by CBS officers will be used exclusively for the preparation of statistical publications. From these publications no identifiable information concerning separate persons can be derived by others, including other government agencies. As a result KCBS takes great care to ensure that the information provided by individuals can never be used for any other than statistical purposes.

---

<sup>4</sup> No such violations have occurred elsewhere see Marilyn McMillen, "Data Access: National Center for Education Statistics," paper presented for the US National Center for Education Statistics (2000).

As a member of the International Statistical Institute, the KCBS is obligated by the declaration on professional ethics to abide by the highest standards. The declaration states, in part:

Statisticians should take appropriate measures to prevent their data from being published or otherwise released in a form that would allow any subjects' identity to be disclosed or inferred (ISI, 1985).

Since Kenya relies on statistical information to make policies and to plan resource allocations, it is vital that respondents trust the KCBS with personal, even sensitive information, if accuracy is to be attained. Because of declining response rate in a number of countries, for example, in The Netherlands where the response rate in household surveys declined from 20% to 40% over the last decades and also in the United Kingdom,<sup>5</sup> statistical agencies are vigorously pursuing policies to promote public confidence.

There is a notion among some scholars that disclosure of certain "sensitive" information about an individual may result in the person being arrested for a crime, denied eligibility for welfare or subsidized medical care, charged with tax evasion, or lose a job or an election. The person could also face financial consequences such as being denied a mortgage or admission to college (Mackie 2002 cited in McCaa and Ruggles, 2001:8).

"Sensitive" information is culture, place and time specific as are the consequences. In Kenya, disclosure of one's "sensitive" information may not carry the consequences listed above since Kenya does not have a program similar to Medicaid or public welfare for its citizens. Even in situations where Kenyans are entitled to social security, the criteria for providing such services is not based on one's past earnings. Sensitive information for

---

<sup>5</sup> See Catherine Heeny, "Research on the Role of Privacy and Confidentiality in the Collection and Dissemination of Census and Survey Data," nd., <http://lesl.man.ac.uk/ccsr/rschproj/privacy>.

Kenyans include the following: ethnicity (even though this is public information), religious background, income and incapacitating illness.

One's ethnic background ("tribe") is a sensitive matter in Kenya because of the long history of ethnic struggles, later exacerbated by arbitrary colonial boundaries that separated families and combined people of different ethnic groups within administrative districts. Recently there have been antagonisms and struggles over land, distribution of resources, power sharing, etc. As a result disclosure of one's ethnic group may at times lead to discrimination, violence, and even death. In recent years, the Maasai and Gusii have been involved in an intensely fierce "tribal" struggle over land and cows. Those killed are members of minority ethnic groups. In these circumstances revealing ethnic identity through census microdata might contribute to violence. On the other hand, readily available information, such as mode of dress or language or a simple table from the published census, is more likely to be used for such purposes than census microdata!

The recent Gucha-Tansmara clash is not the only ethnically motivated clash Kenya has experienced. In the late 1990s, the Luo and Masaai also engaged in an ethnically motivated clash, but it was the conflict between the Gusii and the Luo which was most devastating, not only in terms of land and lives, but also in terms of personal relations. Inter-ethnic marriages, for example, were often condemned by both communities. Couples in such unions could no longer live in the Luo or the Gusii lands. There are many other ethnic conflicts that have not yet been resolved in Kenya. In all these instances it is clearly evident that one's ethnic community besides being "public", is also sensitive because minorities may be subjected to discrimination, violence and even loss of life. Hence statistical agencies especially in Africa strive to gain and maintain the cooperation of

respondents by assuring them that the information they provide will be held in strict confidence. The CBS decided to exclude information on “tribe” as well as “religion” from the microdatasets provided to the IPUMS project.

**IPUMS-International Disclosure Control Measures.** Holvast (Thessalonika, 1999) identifies three strategies for safeguarding statistical confidentiality of microdata: legal, organizational and technical. All must be used in combination to attain the highest possible level of statistical confidentiality and at the same time promote the highest levels of scientific usage of the data. While technical safeguards are likely to constitute the greatest intellectual challenge, it is important that these be designed within a framework of legal and organizational safeguards.

**Legal Safeguards.** IPUMS International has adopted legally enforceable measures to ensure user conformity with existing confidentiality regulations and guidelines. In order to comply with the international confidentiality standards, IPUMS International negotiates non-exclusive distribution licenses with National Statistical Agencies to disseminate integrated, anonymized microdata via the internet. The current version of the memorandum of agreement is listed in appendix A.

Potential users of the database must obtain permission from IPUMS International, sign a non-disclosure agreement and agree to abide by the stipulations governing the use of the data. In developing these procedures, IPUMS international has emulated successful guidelines used by other already established microdata distribution agencies, such as the United States Census Bureau, the Office of National Statistics and IPUMS–USA. IPUMS International, unlike its USA counterpart, requires users to sign a user license agreement before obtaining data. The online registration system requires users to provide biographical

information, institutional affiliation, contact information including e-mail address, academic background, field of study, research interests and a brief statement about the purpose for which the research data is intended. In addition to explicit acceptance of each clause in the user license agreement, IPUMS International has a disclaimer warning users that those who violate the terms of the agreement will be prosecuted for violation of privacy, their license may be revoked, the microdata in their possession may be recalled and IPUMS could file motions with professional organizations to censure such violators. Finally, official statistical agencies, signatory to the agreement, assist the project in enforcing the user license. The application form is in appendix B.

**Organizational Safeguards.** Organizational safeguards are key to attaining maximum microdata confidentiality protection. As we have explained under the legal safeguards, IPUMS International provides restricted access exclusively to bona-fide users who affirm to abide by the non-disclosure agreement. Data are stored on secure, password protected computers using industry standards to prevent unauthorized access. Data are transmitted using 128-bit encryption standards (SSL or Secure Sockets Layer) protocol, identical to that used by financial and other institutions where security and confidentiality are essential.

**Technical Safeguards.** Technical safeguards directly focus on issues of statistical confidentiality and making optimal use of microdata for scientific, social and policy analysis. The IPUMS International project seeks to design and implement technical safeguards that provide the highest level of statistical confidentiality and scientific usability. Four rules constitute the core of the process:

1. Suppress geographical details for administrative districts with fewer than 20,000 inhabitants, or whatever threshold the corresponding official statistical agency may establish..
2. Aggregate sensitive characteristics of individuals with other characteristics to exceed a minimum threshold.
3. Randomly distribute households within districts to disguise the order in which individuals were enumerated or the data processed.
4. Convert date variables such as birth to single years of age (at advanced ages this may require additional recoding)

For Rule 1, the suppression of geographical details, we adopt whatever threshold is required by each official statistical agency. There is a well-known trade-off between utility and risk. Some statistical agencies prefer greater utility while others favor reduced risk. In the case of Colombia and many other Latin American countries where there is high interest in studying smaller administrative districts, the threshold is set at 20,000 inhabitants. In the case of the 2000 census of the United States Census (USCB), the threshold is 100,000. The Office of National Statistics (United Kingdom) has embraced a 65,000 minimum for the 2001 SARS. In the case of Kenya, the smallest minor administrative district reported 57,960 inhabitants in 1989. CBS-Kenya agreed to permit the identification of all major and minor administrative districts.

Likewise for Rule 2, aggregation of sensitive characteristics, we endorse the USCB guideline, although neither the ONS nor the CBS-Kenya apply this rule. In the case of the United States, where the rule is applied, there is a debate about whether the population threshold should be an absolute or a percentage figure (10,000 or 0.004% as in the USCB microdata sample for 2000). In the case of Kenya, given that a 5% sample density, this translates into a threshold of 500 or 50, depending whether the rule is interpreted as absolute or relative. Occupational data in Kenyan censuses are now more than 15 years old. It was deemed unnecessary to suppress or aggregate any occupational codes. Rule 3,

randomly re-ordering households, is applied to the entire dataset. Rule 4 is not applicable because Kenyan censuses request age, not birthdate or date of marriage.

Table 3

Of the 38 person variables in the Kenyan census microdata sample for 1989, five geographic and one sensitive variables (ethnicity/tribe) are suppressed entirely. Twenty-eight require no treatment under the rules listed above. In the 1999 sample, one additional sensitive variable, religion, is suppressed.

**Sensitive variables.** Sensitive information is culture specific. While in the U.S., U.K., Canada and the Netherlands, for example, address and income may constitute unique identifiers, in Kenya this is not the case because a majority of the population uses institutional postal service. Under the institutional postal service system, a group of people, working or living within an area may use a particular box and often some have one or more postal service boxes. In so far as income is concerned, unless one is employed by the Civil Service, Kenya has a poor system of keeping track of how much money business men and women make. As a result determining an individual's accurate income is extremely difficult. Moreover the Kenyan censuses never request this information so there is no risk of disclosure by means of census microdata. Likewise, until the 1999 enumeration, information regarding religion was never requested.

"Tribe" (ethnicity or national origin) and Religion constitute the most sensitive information in the census. At the request of the Director General of the Central Bureau of Statistics of Kenya both these variables are currently suppressed in the datasets distributed by IPUMS-International.

**Conclusion.** Disclosure risks in the census samples of Kenya are minimal. On the one hand, Kenyan censuses contains only two sensitive variable, tribe/ethnicity/national origin and religion. On the other, thousands of people share characteristics with respect to most variables. Only two districts had fewer than 100,000 inhabitants according to the 1989 census. With respect to the tribe/ethnicity/national-origin variable there are nine categories with fewer than 500 individuals, but only four of these are of indigenous groups. The official decision to suppress these variables is unfortunate, but the project must comply with the wishes of the national statistical service. The hamronized Kenyan samples have now been available to researchers from the IPUMS-international website for some three years. Dozens of researchers have downloaded extracts including the Kenyan samples. Confidentiality has proven to be a non-issue because researchers are uninterest in attempting to breach confidentiality. On the contrary, they are its most ardent defenders.

## References

- Bethlehem, J.G., Keller, W.J., and Pannekoek, J. (1990), "Disclosure Control of Microdata," *Journal of the American Statistical Association*, 85, 38-45.
- Bryman, A. and Cramer, D. 1990, *Quantitative data analysis for social scientists*, London: Routledge.
- Carter, R., Boudreau, J.-R., and Briggs, M. 1991: "Analysis of the Risk of Disclosure for Census Microdata," *Statistics Canada Working Paper*. Ottawa: Statistics Canada.
- Cox, L.H. (1995), "Protecting Confidentiality in Business Surveys," in *Business Survey Methods*, Cox, B.G., Binder, D.A., Chinnappa, B.N., Christianson, A., College, M.J., Kott, P.S. (eds.). Willey: New York.
- Cox, L.H., McDonald, S and Nelson, D. (1986), "Confidentiality Issues at the United States Bureau of the Census," *Journal of Official Statistics*, 2, 135-160.
- Dale, A. (1998), "Confidentiality of Official Statistics: An Excuse for Privacy," in Dorling, D. and Simpson, S. (eds), *Statistics in Society*. London: Arnold, 29-37.
- Dale, A. and Elliott, M. (1999), *Proposals for 2001 SARs: An Assessment of Disclosure Risk*. Manchester: CCSR, Manchester University.
- Dale, Angela and Mark Elliott. 2001. "Proposals for 2001 SARs: An assessment of disclosure risk," *Journal of the Royal Statistical Society, Series A*, 164, part 3, pp.427-447.
- Dale, A. Marsh, C. (1993), *The 1991 Census User's Guide*. London: HMSO.
- Dalenius, T. (1977), "Towards a Methodology for Statistical Disclosure Control," *Statistisk Tidskrift*, 5, 429-444.
- De Waal, T. and Willenborg, L.C.R.J. (1996), "A View on Statistical Disclosure for Microdata," *Survey Methodology*, 22,1, 95-103.
- Duncan, G.T. and Pearson, R.W. (1991), "Enhancing Access to Microdata While Protecting Confidentiality: Prospects for the Future," *Statistical Science*, 6, 219-239.
- Duncan, G.T., and Lambert, D. (1986), "Disclosure-Limited Data Dissemination," *Journal of the American Statistical Association*, 81, 10-28.
- \_\_\_\_\_(1987), "The Risk of Disclosure for Microdata," in *Proceedings of the Third Annual Research Conference of the Bureau of the Census*. Baltimore, MD: U.S. Bureau of the Census, 263-278.

- Duncan, G.T., Jabine, T.B. and de Wolf, V.A. (1993), *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. Washington, D.C.: National Academy Press.
- Elliott, M. and Dale, A. (1999), "Scenarios of Attack: A Data Intruder's Perspective on Statistical Disclosure Risk," *Netherlands Official Statistics*, 14, 6-10.
- Elliott, M.J., Skinner, C.J. and Dale, A. (1998), "Special Uniques, Random Uniques and Sticky Populations: Some Counterintuitive Effects of Geographical Detail on Disclosure Risk," *Research in Official Statistics* 1(2), 53-68.
- Fienberg, S.E. and Makov, U.E. (1998), "Confidentiality, Uniqueness And Disclosure Limitation for Categorical Data," *Journal of Official Statistics*, 14,4, 385-397.
- Greenburg, B. and Voshell, L. (1990), "The Geographic Component of Disclosure Risk For Microdata," *SRD Research Report Census/SRD/RR-90/13*. Washington D.C: American Bureau of the Census.
- Heeny, Catherine. Unpublished. "The Role of Privacy and Confidentiality in the Collection and Dissemination of Census and Survey Data," available at: <http://lesl.man.ac.uk/ccsr/rschproj/privacy>
- Holvast, Jan. 1999. "Statistical Confidentiality at the European Level," Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, March.
- [International Monetary Fund. 2001. General Data Dissemination Bulletin Board: http://dsbb.imf.org/category/popctys.htm](http://dsbb.imf.org/category/popctys.htm)
- International Statistical Institute. 1985. Declaration of Principles on Statistical Confidentiality.
- Jabine, Thomas B. (1993a), "Statistical Disclosure Limitation Practices of United States Statistical Agencies," *Journal of Official Statistics*, 9(2): 427-454.
- \_\_\_\_\_ (1993b), "Procedures for Restricted Access Data," *Journal of Official Statistics*, 9(2) 537-589.
- Kelly Hall, Patrick, Robert McCaa, and Gunnar Thorvaldsen. 2000. *Handbook of International Historical Microdata for Population Research*. Minneapolis MN: Population Research Center.
- Kim, J. (1986), "A Method for Limiting Disclosure in Microdata Based on Random Noise And Transformation," in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 370-374.

- Mackie, Christopher. 2000. "Improving Confidentiality of and Access to Research Microdata: Summary of a Workshop." *Of Significance. . .Journal of the Association of Public Data Users*.
- Marsh, C. (1993), "Privacy Confidentiality and Anonymity in the 1991 Census," In Dale A, and Marsh, C. (eds.), *The 1991 Census User Guide*. London: HMSO.
- Marsh, C., Dale, A. and Skinner, C. (1994), "Safe Data versus Safe Setting: Access to Microdata from the British Census," *International Statistical Review* 62(1):35-53.
- Marsh, C. and Teague, A. (1992), "Samples of Anonymised Records from the 1991 Census," *Population Trends* 81, 37-39.
- McCaa, Robert and Ruggles, Steven, 2002. "The Census in global perspective and the coming microdata revolution," *Scandinavian Population Studies*, 17:7-30.
- McMillen, Marilyn. 2001. "Data Access: National Center for Education Statistics." *Of Significance. . .Journal of the Association of Public Data Users*.
- Paas, G. 1988. "Disclosure Risk and Disclosure Avoidance for Microdata," *Journal of Business and Economic Statistics*, 6, 487-500.
- Ruggles, Steven. 2000. "The Public Use Microdata Samples of the U.S. Census: Research Applications and Privacy Issues," A report of the Task Force on Census 2000, Minnesota Population Center and Inter-University Consortium for Political and Social Research Census 2000 Advisory Committee. Available at <http://www.ipums.org/~census2000>.
- Secretariat. 2001. "Report of the March 2001 Work Session on Statistical Data Confidentiality," Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Skopje. Available at: <http://www.unece.org/stats/documents/2001.03.confidentiality.htm>
- Skinner, C. J., Marsh, C. Openshaw, S. and Wymer, C. (1994), "Disclosure Control for Census Microdata," *Journal of Official Statistics* 10, 31-51.
- United Nations Department of Economic and Social Affairs Statistical Division.  
*International Standard Industrial Classification of All Economic Activities*.  
New York: 1990.

Table 1. Kenyan Census Microdata Samples

	1969	1979	1989	1999
Enumeration: de facto	yes	yes	yes	yes
ACAP Sample size (person records)	659,310	931,864	1,066,902	-
IPUMS Sample size (person records)	-	-	1,074,131	1,410,197
Sampling fraction	3%	5%	5%	5%
<u>Type of Variables</u>	<u>Number of Questions</u>			
Geographic Information	6	8	8	8
Housing Characteristics	0	0	8	10
Personal Characteristics	5	5	5	6
Economic Status, Employment	0	0	3	1
Education	1	2	3	3
Migration	1	2	2	3
Orphanhood	2	2	2	2
Fertility, Mortality	5	9	13	14

Sources: [www.acap.upenn.edu](http://www.acap.upenn.edu), [www.ipums.org/international](http://www.ipums.org/international)

Key: CBS = Central Bureau of Statistics of the Republic of Kenya

ACAP = African Census Analysis Project, University of Pennsylvania

IPUMS = Integrated Public Use Microdata Series, University of Minnesota

Note: Many years ago, a 10% sample was drawn for the 1969 census, but only a 3.3% sample is known to survive. For the 1979 census, the IPUMS project recovered 64% of the microdata and repatriated a copy to the Central Bureau of Statistics (see table 2 for details). The 1989 sample disseminated by IPUMS-International is distinct from that distributed by ACAP. .

<b>Table 2. Statistical Confidentiality and Census Microdata Dissemination Practices</b>			
<b>Repositories of anonymized census microdata samples for scientific research</b>			
<b>Acronym</b>	<b>Institution and Dissemination Policy</b>		
<b>ACAP</b>	African Census Analysis Project, Philadelphia USA. Permission of ACAP director.		
<b>CELADE</b>	Centro Latino Americano de Demografía, Santiago Chile. Application to National Statistical Agency.		
<b>ECE/PAU</b>	ECE Population Affairs Unit, Geneva Switzerland. Written application to PAU.		
<b>EWC</b>	East-West Center, Honolulu USA. Restricted to institution use only.		
<b>ICPSR</b>	Inter-University Consortium for Political and Social Research, Ann Arbor USA. Member university.		
<b>IPUMS</b>	Integrated Public Use Microdata Series International, Minneapolis USA. Electronic application.		
<b>CMCCSR</b>	Cathie Marsh Center for Census and Survey Research, Manchester UK. Written application to CMCCSR.		
<b>Synthesis of Confidentiality Provisions, 52 member-states:</b>			
<b>Country</b>	<b>Law</b>	<b>International Monetary Fund's General Data Dissemination System</b>	<b>Samples</b>
<b>Argentina</b>	1968	Individual reports and/or data may not be communicated to third parties or used or disseminated in such a way as to make it possible to <u>identify the reporting person or entity</u> .	CELADE, IPUMS
<b>Australia</b>	1905	The Census Act protects the confidentiality of persons and organisations by requiring that information not be published in a manner likely to enable the identification of a particular person or organisation. Notwithstanding this, the CSA provides for the Minister to make determinations providing for the release of certain classes of information which would not otherwise be permitted to be released under the Act; except that <u>personal or domestic information may not be disclosed under the provisions of a determination in a manner that is likely to enable the identification of a person</u> .	Australian National University
<b>Austria</b>	2000	Strict provisions on statistical confidentiality are contained in the Federal Statistics Act. The field on protection of personal data is covered by the Data Protection Act.	IPUMS
<b>Bangladesh</b>		There are no regulations enforcing confidentiality of reporting, but <u>strict confidentiality is maintained in practice</u> .	
<b>Belgium</b>	1994	According to the rules of the Official Statistics Act..., the <u>confidentiality of individual responses is protected</u> .	ECE/PAU
<b>Brazil</b>	1999	Decree 74.084 of May 20, 1974... and Decree 3.272 of December 3, 1999...provide <u>assurances of confidentiality of individual responses</u> so that the data can be used only for statistical purposes.	CELADE IPUMS

<b>Canada</b>	1985	[Under the Statistics Act of 1985,] Statistics Canada cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity.</u>	ECE/PAU
<b>Chile</b>	1970	Law No. 17-374 and its Regulations .... All individuals and legal entities are required to provide any information requested by the INE, which in turn is required to <u>maintain strict confidentiality and is prohibited from explicitly referring directly or indirectly in its publications to individuals or legal entities.</u>	CELADE, IPUMS
<b>Colombia</b>	1960	Article 75 of Decree 1633 of 1960...establishes the <u>principles of confidentiality and discretion; thereby forbidding communication of data by name or individually.</u>	CELADE IPUMS
<b>Croatia</b>	1994	Under Law N.N. 52/94, the CBS cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity.</u>	
<b>Czech Republic</b>	2000	The State Statistical Service Act No. 89/1995 Coll. Which came into force on June 15, 1995 and was amended by Act No. 220/2000 Coll. And Act No. 411/2000 Coll. ... <u>Protection of individual data represents an important section of this Act.</u>	ECE/PAU, IPUMS
<b>Denmark</b>		According to the "Public Authorities' Registers Act", <u>data attributable to identifiable individuals (or enterprises) shall not be passed on.</u>	
<b>Ecuador</b>	1976	The Official Registry Law No. 82 establishes the principles of confidentiality and discretion, thereby <u>forbidding disclosure of information for any individual person or private entity.</u>	CELADE, IPUMS
<b>El Salvador</b>	1955	...data compiled by the DIGESTYC are <u>confidential and may be used solely for statistical purposes.</u>	CELADE, IPUMS
<b>Estonia</b>	1997	The SOE may transmit or disseminate collected data only in a form which <u>precludes the possibility of direct or indirect identification of the respondents.</u>	ECE/PAU
<b>Finland</b>	1994	Under the terms of Act 62/1994, Statistics Finland cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity.</u>	ECE/PAU,
<b>France</b>	1978	INSEE cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity..</u>	IPUMS
<b>Germany</b>	1987	[no specific statement on confidentiality.] (Collection and current updating of population data are regulated by the Law on the Statistics of Population Movement and Adjustment of the Population State dated March 14, 1980 in conjunction with the Law on Statistics for Federal Purposes of 1987.)	German Research Institute, IPUMS

<b>Hong Kong</b>	1993	The 1978 Ordinance updated in 1993 stipulates that: ... (2) Only aggregate information will be published such that information relating to any particular individual or undertaking will be kept <u>strictly confidential and will not be divulged to other parties.</u>	EWC
<b>Hungary</b>	1993	The 1993 Law on Statistics of Hungary (XLVI/1993) and the 1992 Law on Protection of Personal Data and the Disclosure of Data of Public Interest (Law LXIII/1992) ... (4) All statistics collected and published by the HCSO are governed by the confidentiality provisions which specify that the HCSO cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity.</u>	ECE/PAU IPUMS
<b>Iceland</b>	2000	<u>Individual data are kept strictly confidential and care is taken that the data released cannot be traced directly or indirectly to an individual entity.</u> Researchers may be given access to information on individuals with the permission of the Data Protection Authority under strict rules and conditions.	
<b>India</b>	1948	Data relating to individuals have to be kept confidential.	
<b>Indonesia</b>	1997	The BPS (Law 16, 1997) cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual or entity.</u>	EWC, IPUMS
<b>Ireland</b>	1983	The Statistics Act of 1993 ... sets stringent confidentiality standards: the information collected may be used only for statistical purposes, and <u>no information that could be related to an identifiable person or undertaking may be released.</u>	IPUMS
<b>Israel</b>	1978	The Law on Statistics (1972 as amended in lawbook 908, 1978): ... (3) Stipulates that the CBS cannot publish, or otherwise make available to any individual or organization, <u>statistics that would enable the identification of data for any individual person or entity.</u>	ECE/PAU, IPUMS
<b>Italy</b>	1989	The Law on the National Statistical System (Legislative Decree n. 322, September 6, 1989) which is consistent with the U.N. Fundamental Principles of Official Statistics ... establishes: ... Strict confidentiality rules for data included in the National Statistical Program, approved yearly by Decree of the President of the Council of Ministers (D.P.C.M.) ( <u>Dissemination occurs only in an aggregate form and in a manner by which it is not possible to identify data for any individual person or entity.</u> )	ECE/PAU
<b>Japan</b>	1999	Law to Establish the Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT) of July 16, 1999, and the Cabinet Order on the Organization of the MPHPT. ... - [no specific confidentiality statement on GDDS web-site.]	

<b>Korea</b>	1993	The Statistics Act of 1993 ... sets stringent confidentiality standards: the information collected may be used only for statistical purposes, and <u>no information that could be related to an identifiable person or undertaking may be released.</u>	EWC
<b>Latvia</b>	1997	The Law on State Statistics adopted on November 6, 1997 ... provides that the CSB cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity.</u>	ECE/PAU
<b>Lithuania</b>	1999	Under the Law on Statistics (1999, No. VIII-1511) ... Statistics Lithuania cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity.</u>	ECE/PAU
<b>Malaysia</b>	1989	Under the terms of the Statistics Act, 1965 (Revised 1989), DOSM: (2) Cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity.</u>	EWC, IPUMS
<b>Mexico</b>		All data provided by individuals or obtained from administrative or civil registers are treated with <u>strict confidentiality and discretion, and in no case may they be communicated by name or individually (Article 38).</u>	CELADE IPUMS
<b>Netherlands</b>	1996	<u>"Data gathered on the basis of this law will not be disclosed in such a form that returns and information about an individual person, company, or institutions can be deduced, unless the individual, the head of the company, or the governing board of the institution have no objection to such disclosure."</u>	IPUMS
<b>Norway</b>	1989	Statistics Norway is <u>prohibited to publish or disclose data from which information about individual persons or firms can be derived.</u> (Researchers may be given access to such information under strict rules and conditions. Guidelines provided by the Norwegian Data Inspectorate form the framework for internal management of data security.)	ECE/PAU
<b>Peru</b>	1990	INEI's Organization and Functions Law (Legislative Decree No. 604) of May 3, 1990 ... establishes the technical autonomy of INEI, details the norms concerning compilation of the data, and stipulates that information provided to the Peruvian statistical system is <u>confidential and cannot be disclosed individually</u> , even under an administrative or judicial order, and requires that the organization publish the data on population.	IPUMS

<b>Philippines</b>	1987	The ... Commonwealth Act No. 591 (August 19, 1940), Executive Order No. 121 (January 30, 1987), and Batas Pambansa Blg. 72 (June 11, 1980). ... Section 4 provides that data furnished to NSO will be kept strictly confidential and shall not be used as evidence in court for purposes of taxation, regulation or investigation; nor shall such data or information be <u>divulged to any person except in the form of summaries or statistical tables in which no reference to an individual, corporation, association, partnership, institution or business enterprise shall appear.</u>	EWC, IPUMS
<b>Poland</b>	1995	Under the Law on Official Statistics, which was passed on 29 June 1995 (Dz. U. Nr. 88) ... the CSO cannot publish, or otherwise make available to any individual or organization, statistics that would allow the <u>identification of data of any individual person or entity.</u>	ECE/PAU
<b>Portugal</b>	1989	The National Law on Statistics (Law 6/1989 of April 15, 1989), ... establishes the principle of the technical independence of the INE, as well as the principle of confidentiality under which <u>no individual information about people can be disseminated.</u>	
<b>Singapore</b>	1991	The Statistics Act, Revised Edition, 1991 ... specifies that the disseminating agencies cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity without prior consent.</u>	
<b>Slovak Republic</b>	1992	All statistical information collected, processed and released by SO SR is regulated by the Law on State Statistics (Law of SNC No. 322/92 Digest, in wording of latter regulations). This Law: ... - Specifies that individual responses to statistical surveys <u>cannot be used for other than statistical purposes without the permission of the legal or physical person in question.</u>	
<b>Slovenia</b>	1995	The Law on National Statistics ... (UrL RS No. 45/95) ... Emphasizes the importance of data confidentiality and stipulates that the Statistical Office cannot publish, or otherwise make available to any organization or individual, statistics that would <u>enable the identification of data for any individual person or entity.</u>	IPUMS
<b>South Africa</b>	1999	The Statistics Act, 1999 (Act No. 66 of 1999) ... - Stipulates that Stats SA cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity.</u>	ACAP, IPUMS
<b>Spain</b>	1996	Statistical Law No. 12/1989 ... and Law No. 13/1996: ... INE cannot publish, or make otherwise available, individual data or statistics that would <u>enable the identification of data for any individual person or entity.</u> (Article 13)	ECE/PAU IPUMS

<b>Sri Lanka</b>	1981	The DCS produces and disseminates data under the Statistical Ordinance and Census Ordinance (1981) ... Confidentiality of reporters is guaranteed under the 1981 Ordinance which states " <u>...no publication ... shall disclose or facilitate the identification of any particulars as being particulars relating to any individual person</u> ".	EWC
<b>Sweden</b>	1992	Data protection is ensured by prescriptions in the Data Act of 1973 (1973:289) and the Secrecy Act of 1980 (1980:100).	ECE/PAU
<b>Switzerland</b>	1992	The Federal Law on Data Protection (06/19/92) specifies that the Swiss Federal Statistical Office cannot publish, or otherwise make available to any individual or organization, statistics that would <u>enable the identification of data for any individual person or entity</u> .	ECE/PAU
<b>Thailand</b>		[No statement on confidentiality provided.]	EWC
<b>Turkey</b>	1989	The 1962 Statistical Law, as well as the 1984 Decree 219 and 1989 Decree 357: ... Data may be collected only for statistical purposes and confidentiality is assured. ... (3) The <u>confidentiality of individual responses is guaranteed</u> .	ECE/PAU
<b>Uganda</b>	1998	The Uganda Bureau of Statistics Act, 1998 ... Article 19 <u>ensures confidentiality of reported data</u> and Article 29 provides for substantial penalties to employees of the Bureau who violate the confidentiality provisions.	ACAP
<b>United Kingdom</b>		The Registrar General is required to compile and publish statistics on the number and condition of the population (1920 Census Act). Births and deaths from the National Registration System are subject to specific <u>statutory confidentiality constraints</u> , in addition to the general confidentiality policy of the ONS.	ECE/PAU CMCCSR IPUMS
<b>United States</b>	1954	"No individual-level input data are released." [Title 13 United States Code Section 9 prohibits "any publication whereby the data furnished by any particular establishment or individual under this title can be identified".]	IPUMS, ECE/PAU
<b>Venezuela</b>	1999	Law on National Statistics and Censuses of November 27, 1944 ... Article 10: "The Ministry of Development may officially order aggregate or average data, or statistical series, but in no way and under no pretext may it order or authorize the <u>disclosure of individual data or the dispatch of single copies...</u> related to a given individual or legal entity or to a given family or group of families."	CELADE, IPUMS
<b>Sources:</b>		Confidentiality provisions: <a href="http://dsbb.imf.org/category/popctys.htm">International Monetary Fund GDDS bulletin board (http://dsbb.imf.org/category/popctys.htm)</a> Microdata availability: Kelly Hall, McCaa and Thorvaldsen (eds.), <u>Handbook of International Historical Microdata for Population Research</u> 2000:388-395 (updated: <a href="http://www.ipums.org/international/iiinventory2.html">http://www.ipums.org/international/iiinventory2.html</a> )	

Table 3 Anonymization of Kenyan Census Microdata: Unique Characteristics Threshold  
(50,000 for geographic variables; 1,000 for other variables)

Type	Procedure	Variable Name
Key	Suppressed	Division, Location, Sublocation, Enumeration area, Tribe/Ethnicity , Religion
	Aggregated (none required)	50,000 minimum: Province, District of Residence, Birth and Past Residence
	None	Sex, Marital Status, Relationship to Head
Transitory (information is considered too changeable to be used to identify individuals from microdata).		
	None	Age, Urban/Rural Residence, Literacy, Educational Status, Educational Level, Labor Activity, Occupation, Children Everborn/Alive/Dead, Last Birth Year, Mortality variables

Appendix Table 1. Kenyan Census Microdata Sample Characteristics  
and: Variable Availability

	1969	1979	1989	1999
Enumeration: de facto	X	X	X	X
Sample size	659,310	931,864	1,074,131	1,410,197.
Sampling fraction	3%	5%	5%	5%
<b>Geographic Information</b>				
Province	X	X	X	X
District	X	X	X	X
Division	X	X	X	X
Location	X	X	X	X
Sub-Location	X	X	X	X
E. A. Number	X	X	X	X
Household number	.	X	X	X
E.A. Type Urban/Rural	.	X	X	X
<b>Housing Characteristics</b>				
Number of dwelling units	.	.	.	X
Number of habitable rooms	.	.	.	X
Tenure status	.	.	X	X
Dominant construction material: Roof	.	.	X	X
Dominant construction material: Wall	.	.	X	X
Dominant construction material: Floor	.	.	X	X
Main source of water	.	.	X	X
Main type of human waste disposal	.	.	X	X
Main cooking fuel	.	.	X	X
Main type of lighting	.	.	X	X
<b>Personal Characteristics</b>				
Relationship to head	X	X	X	X
Sex	X	X	X	X
Age	X	X	X	X
Marital status	X	X	X	X
Tribe/Nationality	X	X	X	X
Religion	.	.	.	X
<b>Economic Status, Employment</b>				
Occupation	.	.	X	.
Economically Active	.	.	X	X
Position in workforce	.	.	X	.
<b>Education</b>				
Literacy	.	.	X	.
School Attendance	.	X	X	X
Level of Education	.	.	.	X
Education attained	X	X	X	X
<b>Migration</b>				
Birthplace	X	X	X	X
Previous residence	.	X	X	X

Duration of previous residence	.	.	.	X
Orphanhood				
Orphanhood of father	X	X	X	X
Orphanhood of mother	X	X	X	X
Live Births				
Born alive: boys	.	.	.	X
Born alive: girls	.	.	.	X
Home alive: boys		X	X	X
Home alive: girls		X	X	X
Home alive: total	X	.	.	.
Live elsewhere: boys		X	X	X
Live elsewhere: girls		X	X	X
Live elsewhere: total	X	.	.	.
Died: boys		X	X	X
Died: girls		X	X	X
Died: total	X	.	.	.
Last live birth				
Month of birth	X	X	X	X
Year of Birth	X	X	X	X
Sex	.	X	X	X
Multiple birth	.	.	X	X
Alive/dead	.	.	X	X
Dead multiple	.	.	.	X
Month of death	.	.	X	.
Year of death	.	.	X	.

---

## **MEMORANDUM OF UNDERSTANDING (MOU) BETWEEN INTEGRATED PUBLIC USE MICRODATA SERIES INTERNATIONAL (IPUMS), MINNESOTA UNIVERSITY AND CENTRAL BUREAU OF STATISTICS (CBS)**

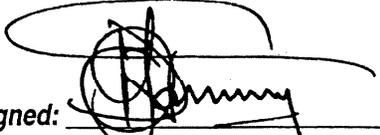
---

**Purpose:** The purpose of this letter of understanding is to specify the terms and conditions under which integrated metadata and microdata provided by **Central Bureau of Statistics of Kenya** shall be disseminated by **Integrated Public Use Microdata Series International** of the University of Minnesota.

1. **Ownership:** The **Central Bureau of Statistics** of Kenya is the owner and licensee of the intellectual property rights (including copyright) in metadata and microdata supplied to the University of Minnesota to be distributed by **Integrated Public Use Microdata Series International**.
2. **Use:** These data are provided for the exclusive purposes of teaching, academic research and publishing, and may not be used for any other purposes without the explicit written approval, in advance, of the **Central Bureau of Statistics**.
3. **Authorization:** To access or obtain copies of integrated microdata of Kenya from **Integrated Public Use Microdata Series International**, a prospective user must first submit an electronic authorization form identifying the user, that is, principal investigator by name, electronic address, and institution. The principal investigator must state the purpose of the proposed project and agree to abide by the regulations contained herein. Once a project is approved, a password will be issued and data may be acquired from servers or other electronic dissemination media maintained by **Integrated Public Use Microdata Series International**, the **Central Bureau of Statistics**, or other authorized distributors. Once approved, the user is licensed to acquire integrated metadata and microdata of Kenya from **Integrated Public Use Microdata Series International** or other authorized distributors. No titles or other rights are conveyed to the user.

4. **Restriction:** Users are prohibited from using Kenyan data acquired from the **Integrated Public Use Microdata Series International** or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise.
5. **Confidentiality:** Users will maintain the absolute confidentiality of persons and households as provided for under the Statistics Act Cap.112 of the Laws of Kenya. Any attempt to ascertain the identity of persons or households from the microdata is strictly prohibited. Alleging that a person or household has been identified in these data is also prohibited.
6. **Security:** Users will implement security measures to prevent unauthorized access to Kenyan microdata acquired from **Integrated Public Use Microdata Series International** or its partners.
7. **Publication:** The publishing of data and analysis resulting from research using metadata or microdata of Kenya is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite the **Central Bureau of Statistics and Integrated Public Use Microdata Series International** as the sources of Kenyan data, and to indicate that the results and views expressed are those of the author/user.
8. **Sharing:** **Integrated Public Use Microdata Series International** will provide electronic copies to the **Central Bureau of Statistics** of documentation and data related to its integrated microdata as well as timely reports of authorized users.
9. **Violations:** Violation of this agreement may lead to professional censure and/or civil prosecution.

Signed:   
Edward F. Wink, Assoc. V.P.  
University of Minnesota  
Sponsored Projects Admin.  
200 Oak Street SE, Suite 450  
Minneapolis, MN 55455-2070

Signed:   
Director of Statistics  
P.O. Box 30266  
NAIROBI, Kenya

## Letter of Understanding

### **Integrated Public Use Microdata Series International** and **[Official Statistical Agency of X]**

**Purpose.** The purpose of this letter is to specify the terms and conditions under which metadata and microdata produced by the **[Official Statistical Agency of X]** shall be distributed by **Integrated Public Use Microdata Series International** of the University of Minnesota.

1. **Ownership.** The **[Official Statistical Agency of X]** is the owner and licensee of the intellectual property rights (including copyright) in the metadata and microdata of **[X]** acquired by the University of Minnesota to be distributed by **Integrated Public Use Microdata Series International**. This agreement explicitly authorizes release to the University of microdata of **[X]** that may be in the possession of third parties. The University is obligated to provide to the **[Official Statistical Agency of X]** timely notice of any such acquisitions and, upon request and without cost, provide copies of same.
2. **Use.** These data are for the exclusive purposes of teaching, scientific research and publishing, and may not be used for any other purposes without the explicit written approval, in advance, of the **[Official Statistical Agency of X]**.
3. **Authorization.** To access or obtain copies of integrated microdata of **[X]** from **Integrated Public Use Microdata Series International**, a prospective user must first submit an electronic authorization form identifying the user (i.e., principal investigator) by name, electronic address, and institution. The principal investigator must state the purpose of the proposed project and agree to abide by the regulations contained herein. Once a project is approved, a password will be issued and data may be acquired from servers or other electronic dissemination media maintained by **Integrated Public Use Microdata Series International**, the **[Official Statistical Agency of X]**, or other authorized distributors. Once approved, the user is licensed to acquire integrated metadata and microdata of **[X]** from **Integrated Public Use Microdata Series International** or other authorized distributors. No titles or other rights are conveyed to the user.
4. **Restriction.** Users are prohibited from using data acquired from the **Integrated Public Use Microdata Series International** or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise.
5. **Confidentiality.** Users will maintain the absolute confidentiality of persons and households. Any attempt to ascertain the identity of a person, family, household, dwelling, organization, business or other entity from the microdata is strictly prohibited. Alleging that a person or any other entity has been identified in these data is also prohibited.

6. Security. Users will implement security measures to prevent unauthorized access to microdata acquired from **Integrated Public Use Microdata Series International** or its partners.
7. Publication. The publishing of data and analysis resulting from research using metadata or microdata of [X] is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite **[Official Statistical Agency of X] and Integrated Public Use Microdata Series International** as the sources of the data of [X], and to indicate that the results and views expressed are those of the author/user.
8. Violations. Violation of the user license may lead to professional censure, loss of employment, and/or civil prosecution. The University of Minnesota, national and international scientific organizations, and the [Official Statistical Agency of X] will assist in the enforcement of provisions of this accord.
9. Sharing. **Integrated Public Use Microdata Series International** will provide electronic copies to the **[Official Statistical Agency of X]** of documentation and data related to its integrated microdata as well as timely reports of authorized users.
10. Jurisdiction. Disagreements which may arise shall be settled by means of conciliation, transaction and friendly composition. Should a settlement by these means prove impossible, a Tribunal of Settlement shall be convened which will rule upon the matter under law. This Tribunal shall be composed of an arbitrator, which shall be selected by the ICC International Court of Arbitration. This agreement shall be governed by, and construed in accordance with, generally accepted principles of International Law.
11. Order of Precedence. In the event of a conflict between a term or condition of this Letter of Understanding and a term or condition of any Contract, to which this Letter of Understanding is attached, the term or condition in this Letter of Understanding shall prevail.

Date: \_\_\_\_\_

Signed: \_\_\_\_\_

**Regents of the University of Minnesota**

By: Kevin J. McKoskey, Sponsored Projects Administration

Date: \_\_\_\_\_

Signed: \_\_\_\_\_

Rev. Jan. 27, 2005

## Application to Use Restricted Microdata

IPUMS-International microdata are available free of charge, but their use imposes responsibilities upon the user.

To access the data from the Integrated Public Use Microdata Series-International site, a prospective user must first submit an electronic authorization form (this form) identifying the user by name, electronic address, and institution.

The investigator must state the purpose of the proposed project and agree to abide by the regulations specified below. If multiple investigators are involved in a project, all must register separately. Once a project is approved, a message will be sent by email granting access to the system.

The notification licenses the user to acquire microdata from Integrated Public Use Microdata Series International or other authorized distributors. No titles or other rights are conveyed to the user.

Errors discovered in the microdata must be reported to: [ipumsi@pop.umn.edu](mailto:ipumsi@pop.umn.edu)

**Legal Notice:** Submission of this application constitutes a legally binding agreement between the applicant, the applicant's institution the University of Minnesota, and the relevant official statistical authorities. Submitting false, misleading or fraudulent information constitutes a violation of this agreement. Mis-using the data by violating any of the conditions detailed below also constitutes a violation of this agreement. Violation of this agreement may lead to professional censure, loss of employment, civil prosecution under relevant national and international laws, and to sanctions against your institution, at the discretion of the University of Minnesota and the official statistical authorities.

All information will be kept confidential.

All information on this form is required for registration unless otherwise indicated.

### Personal Information

Given Name:

Family Name:

Employer/Institutional affiliation:

Name, Position and identifying e-mail address and webpage, if any:

Supervisor/Department Head/Class Instructor Name:  
email address:

Address (institutional; if none, enter personal):

Street Address 1:

Street Address 2:

City, State/Province, Zip:

Country:

Phone Number(s): (include country and city/area codes)

Fax Number: (optional)

E-mail address to be used for this project:

Funding research, other than employer, if any:  
Indicate name of granting institution, grant #, title, and year(s) of award, or state "None":

Does your institution have a Institutional Review Board (IRB), Office for Human Subject Protections, Professional Conduct or similar committee? If so, please indicate name:

- No [Learn more](#)  
 Yes

Field:

- Demography
- Economics
- History
- Sociology
- Statistics
- Public Policy
- Other academic

Status:

- Student
- Teacher
- Academic Researcher
- Support Staff
- Non-Academic Researcher
- Consultant
- Other

Anticipated Results:

- Class Exercise/Paper
- Thesis
- Article
- Policy Report
- Doctoral Dissertation
- Book
- Other

## Usage License

Integrated Public Use Microdata Series International (IPUMS-International) and its partners

Please check each of the following boxes to indicate that you have read about the limitations of the IPUMS-International data and you agree to abide by the conditions of use. The purpose of this license is to specify the terms and conditions under which integrated microdata samples distributed by Integrated Public Use Microdata Series International of the University of Minnesota may be used. The license is valid for one year and may be renewed.

Data must not be redistributed without authorization.

All data extracted from the IPUMS-International database are intended solely for the use of the licensee. Under IPUMS-International agreements with collaborating agencies, redistribution of the data to third parties is prohibited. Each member of a research team, group or class using the data must apply for access and be licensed individually. A change of status or affiliation requires the submission of a revised application if the microdata are to remain in the custody of the licensee. If responsibility for the microdata is to be passed to another individual, that person must apply for access.

The microdata are intended only for scholarly research and educational purposes.

These microdata are provided for the exclusive purposes of teaching and scholarly research, and may not be used for any other purposes without explicit written approval.

Commercial use and redistribution of the microdata is strictly prohibited.

Users are prohibited from using microdata acquired from the Integrated Public Use Microdata Series International or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise.

Use of the microdata must follow strict rules of confidentiality.

Users will maintain the confidentiality of persons and households. Any attempt to ascertain the identity of persons or households from the microdata, or to link individual records between microdatasets, is prohibited. Alleging that a person or household has been identified in these data is also prohibited. Statistical results that might reveal the identity of persons or entities may not be reported or published in any form.

The microdata must always be safely secured.

Users will implement security measures, such as the use of passwords, to prevent unauthorized access to microdata acquired from Integrated Public Use Microdata Series International, its partners or authorized distributors. Upon the completion of this research, data and results may be retained only if they can be securely guarded. If security cannot be guaranteed, the microdata and results must be destroyed.

Scholarly publications are permitted, and must be cited appropriately.

The publishing of research results based on IPUMS-International microdata is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite Integrated Public Use Microdata Series-International as the source of the microdata, and to indicate that the results and views expressed are those of the author. Users are requested to provide the IPUMS-International staff with a full citation for any publications resulting from their work with these data.

Violation of this license agreement will result in disciplinary action, including possible loss of employment.

Violation of this agreement will lead to a revocation of this license, recall of all microdata acquired, a motion of censure to professional organization(s) and civil prosecution under the relevant national or international statutes, at the discretion of the Regents of the University of Minnesota and the national statistical agencies. Sanctions likewise may be taken against the institution with which the violator is affiliated.

User agrees to notify [ipumsi@pop.umn.edu](mailto:ipumsi@pop.umn.edu) of errors in the data or violation(s) of this agreement by individuals or institutions.

Description of Project Proposal:

Please provide a clear, substantial description of your research project or educational use for the data. This description will be used to evaluate your application.



Which country samples do you intend to use in your research?

Note: Approval will be restricted to the countries selected.

Brazil

China

Colombia

France

Kenya

Mexico

United States

Vietnam

The addition of data for what country(s) or region(s) would you consider most valuable for your future research (optional)?

---

Please enter your Preferred Password:  
(at least 7 characters, using at least one alphabetic and one numeric character)

Confirm Password: