IPUMS-International: Making Confidentialized, Harmonized Census Microdata for 44 Countries Available Free-of-Charge to Academic and Policy Researchers World-Wide

Robert McCaa, Steven Ruggles, and Matt Sobek
University of Minnesota Population Center

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

Abstract. Thanks to cooperative undertakings with national statistical agencies world-wide and to sustained funding by the National Science Foundation and the National Institutes of Health of the United States, over the next five years the University of Minnesota Population Center is extending the IPUMS-International project to 44 countries, including at least 7 Commonwealth nations. Presently, 28 anonymized, integrated census microdata samples for 8 countries, totaling more than 122 million unit records, are available at no cost to accredited researchers on a restricted-access basis from https://www.ipums.org/international. This paper briefly summarizes the confidentiality protocols, harmonization methodology, and uses of the database over the first 33 months of operation. Commonwealth producers and users of census microdata are invited to contribute to and benefit from this global initiative.

Introduction. Census *microdata* provide information about individual persons and often families, households, and dwellings, usually in the form of one or more records per case, each consisting of a series of variables. Typical census microdata variables for person records include age, sex, marital status, family relationship, place of birth, educational attainment, employment status, etc. Microdata are exceedingly useful because they allow researchers to interrelate any desired set of population and housing characteristics (Dale, Fieldhouse and Holsworth, 2000). Remarkably, the United Nations Statistics Division has long remained silent with respect to the use of census microdata. For example, the Principles and Recommendations for the 2000 round of population and housing censuses (UNSD 1998) offers little advice regarding preservation of or access to microdata.

Nevertheless, the flexibility offered by microdata is essential for comparative research because aggregate tabulations produced by national statistical offices are usually not comparable across time or between countries. In the few countries where census microdata covering multiple census years have been easily available to researchers, these data are the most widely-used source for the study of large-scale economic and demographic transformations (McCaa and Ruggles, 2002).

See Appendix Table 1

The IPUMS-International project is a global consortium to confidentialize, harmonize and disseminate high-density census microdata samples (Ruggles et. al. 2003). Begun in 1999 with funding provided by the National Institutes of Health and the National Science Foundation of the United States, to date the initiative enjoys the endorsement of official statistical agencies in more than fifty countries, encompassing more than half the world’s population (Appendix Table 1). In May 2002, the first phase of integrated census microdata for Colombia (1964-1993), France (1962-1990), Kenya (1989-1999), Mexico (1960-2000), the United States (1960-2000), and Vietnam (1989-1999) were made available to researchers, followed by China (1982) in 2003 and Brazil (1960, 1970, 1980, 1991, 2000) in 2004.

Including the data for Brazil, the IPUMS-International website offers some 120 million person records consisting of more than 100 variables from 28 samples (Table 1). Over the next five years, thanks to sustained funding by the National Science Foundation and the National Institutes of Health, the database will expand to 44 countries with regional initiatives in Latin America (McCaa and Esteve 2005), Europe, Africa, Asia and the Pacific. We expect that researchers from Commonwealth countries will constitute a large group of users, once samples for the UK, Canada, South Africa, Pakistan, Malaysia and Fiji Islands are incorporated into the database. Negotiations are underway with the census authorities of Bangladesh, Ghana, India, New Zealand, Nigeria, Uganda, and a number of other Commonwealth statistical agencies.

Table 2. IPUMS-International Integrated Census Microdata Sample Characteristics
120 million person records
Source: https://www.ipums.org/international/sample_descriptions.html

Country census	Sample %	No. of Person records	Additional details
Brazil 1960	5.0	3,001,000	Long-form, cluster sample
1970	5.0	4,954,000	Same
1980	5.0	5,870,000	Same
1990	5.0	8,523,000	Same
2000	6.0	10,136,000	Same
China 1982	0.1	1,003,000	Every thousandth household
Colombia 1964	2.0	350,000	Every fiftieth person
1972	10.0	1,989,000	Every tenth household
1985	10.0	2,643,000	Long-form, cluster sample
1993	10.0	3,247,000	Every tenth household
France 1962	5.0	2,321,000	Every twentieth household
1968	5.0	2,488,000	Same
1975	5.0	2,629,000	Same
1982	5.0	2,714,000	Same
1990	4.2	2,361,000	Every twenty-fourth household
Kenya 1989	5.0	1,074,000	Every twentieth household
1999	5.0	1,410,000	Same
Mexico 1960	1.5	503,000	Every 67th individual
1970	1.0	483,000	Every hundredth household
1990	10.0	8,028,000	Every tenth household
2000	10.6	10,099,000	Long-form, cluster sample
USA 1960	1.0	1,800,000	Stratified, random sample
1970	1.0	2,030,000	Same
1980	5.0	11,337,000	Same
1990	5.0	12,500,000	Stratified, cluster sample
2000	5.0	14,082,000	Same
Vietnam 1989	5.0	2,627,000	Long-form, cluster sample
1999	3.0	2,368,000	Same

1. Confidentiality protections. IPUMS means Integrated Restricted-Access, Confidentialized Microdata Samples. The IPUMS-International acronym carries "PUMS" embedded in its name, but in fact the data are available only as "Restricted-Access", Confidentialized Microdata Samples. Thus, "IRACMS" would be the more literal acronym, and indeed when the IPUMS was internationalized in 1998, the Principal Investigators discussed replacing "PUMS" with a more accurate moniker. We also discussed inserting "scientific" in place of "public". However, a decade-long unbroken string of successes in obtaining monetary resources from the National Science Foundation and the National Institutes of Health dissuaded us then from adopting a more politically-correct name, as it does now with the sister projects, IPUMS-Latin America and IPUMS-Europe.

A comprehensive array of protections are instituted to confidentialize census microdata samples incorporated into the IPUMS-International database. These protections involve three elements:

1. legal: dissemination agreements between the University of Minnesota and each National Statistical Authority.
2. administrative: licenses specifying conditions and restrictions of use between the University of Minnesota and each researcher
3. technical: protection measures to prevent the identification of individuals, families or other entities in the data.

While much of the published literature on statistical confidentiality ignores the legal and administrative environment (and in doing so exaggerates the risk of improper use), we remain firmly persuaded that the strongest system of protections must take into account all three types of guarantees (Thorogood 1999).

First, with regard to legal mechanisms, IPUMS-International projects are undertaken only in countries where a memorandum of understanding signed by the official statistical agency authorizes a project. No work is begun for a project without prior signed authorization from the corresponding NSA. The IPUMS-International memorandum of understanding is entirely general in nature, yet it provides a legal framework for the project to proceed (please see Appendix A). Its clauses spell out: 1) rights of ownership, 2) rights of use, 3) conditions of access, 4) restrictions of use, 5) the protection of confidentiality, 6) security of data, 7) citation of publications, 8) the enforcement of violations, 9) sharing of integrated data, 10) and arbitration procedures for resolving disagreements. There are no secret clauses or special considerations. All members of the consortium are treated equally. Nonetheless, the protocols are revised, indeed expanded, as NSAs suggest modifications.

See Appendix Table 2

The Minnesota Population Center and its authorized partners are obliged to share the integrated data and documentation with the official statistical agencies and to police compliance by users. The signed agreements are highly general and uniform across countries. Details specific to each country such as fees and sample densities are negotiated separately with each official agency and do not form part of the agreement. Under a carefully worded legal arrangement, the Regents of the University of Minnesota are responsible for enforcing the terms of these accords. Any disputes with official statistical agencies that cannot be resolved through amicable negotiations are subject to arbitration under the International Court of Arbitration.

Second, confidentiality restrictions require researchers to apply individually to become approved to use the system (Appendix B). Typically, one-in-three applications are denied. Administrative measures limit access to the extract system to researchers, who:

1. sign an electronic non-disclosure license;

2. endorse prohibitions against a) attempting to identify individuals or the making of any claim to that effect and b) redistributing data to third parties;
3. agree to use the data solely for non-commercial ends and to provide copies of publications to ensure compliance;
4. place themselves under the authority of employers, institutional review boards, professional associations, or other enforcement agencies to deal with any alleged violation of the license;
5. demonstrate a need to use some portion of the database, according to a project description which must be submitted with the electronic application for access;
6. and, finally, demonstrate sufficient research competence and infrastructural support required to use the data properly.

Once registered, users are permitted to create data extracts that contain only the samples and variables of interest to them. It is noteworthy that approximately one-half of applications are denied access because of a failure to adequately satisfy one or another of the specified conditions. It is gratifying to report that no user has yet appealed a denial of access.

Third are the technical measures taken to ensure statistical confidentiality. Where the NSA requests that the MPC apply anonymization procedures, we implement the following technical protections (based on Thorogood 1999; see also Eurostat Secretariat 2001 and Holvast 1999):

1. adopt sample size according to national norms or conventions;
2. limit geographical detail to administrative units with a minimum number of inhabitants (as high as 1,000,000 for some countries and as low as 20,000 for others);
3. top and bottom code unique categories of sensitive variables;
4. round, group, or band age as necessary;
5. suppress date of birth (only age is reported);
6. suppress detailed place of birth (<10/100,000 population);
7. suppress detailed place of residence, work, study, and migration (<10/100,000 population);
8. systematically "swap" (recode) place of enumeration for a fraction of households;
9. randomly order households within administrative units;
10. and, conduct a sensitivity analysis once these measures are imposed to determine what additional measures may be required.

We continue to evaluate emerging methods and technologies for disclosure protection (McCaa and Ruggles 2002). At present we have decided against automatic data protection methods such as *µ-Argus* (Hundepool et al, 1998). In practice, disclosure of confidential information is highly improbable, requiring an enormous investment of resources to obtain rather trivial details invariably with a high degree of uncertainty about whether any one record truly corresponds to a targeted individual or entity (Dale and Elliot 2001). Indeed, over the past forty years of disseminating census microdata in the United States, Canada, and elsewhere there is not a single *allegation* of misuse or breach of statistical confidentiality. The IPUMS-International procedures are designed to extend this perfect record.

2. Microdata Harmonization. Harmonizing census data is not a new idea. First proposed in 1872 at the International Statistics Congress held in St. Petersburg, not much progress was made until the last half of the twentieth century. One of the signal achievements of the United Nations Statistics Division has been in the international harmonization of census concepts from the enumeration form to the publication of final tables. While incomplete, the effort has enjoyed widespread support by statistical agencies around the globe.

The IPUMS-International projects adopt uniform coding schemes, nomenclatures and classifications, based where possible on the United Nations Statistics Division's *Principles and Recommendations for Population and Housing Censuses* (1998) and other international standards such as:

- UNESCO (1997) *The International Standard Classification of Education (ISCED 1997).*
- International Labor Office (1990) *International Standard Classification of Occupations (ISCO-88).*
- United Nations Statistics Division (1990) *International Standard Industrial Classification of All Economic Activities* (ISIC-88).
- United Nations Economic Commission for Europe (1999). *Recommendations for the 2000 Censuses of Population and Housing in the ECE Region* (Statistical Standards and Studies No. 49)

International census samples employ differing numeric classification systems and reconciliation of these codes is a major effort. Variables must be easy to use for comparisons across time and space. This requires that we provide the lowest common denominator of detail that is fully comparable. On the other hand, we must retain all meaningful detail in each sample, even when it is unique to a single dataset.

Composite coding, using multiple digits, offer an elegant solution to this problem. Similar to those used by the International Labor Organization for occupations and industries, we apply composite coding to each variable to retain all original detail, and at the same time provide comparable codes across countries and censuses. The first one or two digits of the code provide information available across all samples. The next one or two digits provide additional information available in a broad subset of samples. Finally, trailing digits provide detail only rarely available.

Consider economic activity, for example (see table 2). The original codes in the census microdata are translated into a composite harmonized four-digit coding scheme. The range of concepts and coding schemes in this table hints at the complexities involved in developing a comprehensive system for a single variable. In the IPUMS-International system, the first digit of this variable with three categories is comparable across all samples: employed, unemployed, inactive. The second digit delineates different types of activity or inactivity. The final digits provides additional detail that may be available in only one country or even one census. The researcher then may recode to suit the particular research issues and methods of analysis.

As more experience is gained by incorporating more countries and censuses, the table will surely be modified, but the basic structure of the composite coding scheme will remain (McCaa, Esteve, Gutierrez and Vasquez 2003). The basic goal of our harmonization efforts is to simplify use of the data while losing no meaningful information. The IPUMS harmonization strategy has proven flexible enough to accommodate the integration of data across broad spans of time (the United States for 1850-2000) and space (China, Colombia, France, Kenya, Mexico, the United States, and Vietnam).

Table 2. Harmonization Table for Employment Status

Harmonized Codes and Labels **Source Data Codes (selected samples)**

IPUMS-International		Co	Co	Fr	Fr	Kn	Mx	Mx	US	Vn	Vn
Code	Label	1964	1993	1962	1975	1999	1970	2000	1960	1989	1999
0000	N/A	*,5	B	*	B	BB	0	BB	0	B	B,1
	ACTIVE (In Labor Force)										
1000	EMPLOYED, not specified	1								1	
1100	At work		4	1	1	1	1	10	10		
1101	At work, and 'student'							14			
1102	At work, and 'housework'							15			
1103	At work, and 'seeking work'							13			
1104	At work, and 'retired'							16			
1105	At work, and 'no work'							18			
1106	At work, public emergency								11		
1107	At work, family holding, not specified										
1108	At work, family holding, not agricultural					3					
1109	At work, family holding, agricultural					4					
1110	Working and studying (France)										
1200	Have job, not at work last week		3			2		20	12		
1300	Armed forces								13		
1301	Armed forces, at work								14		
1302	Armed forces, not work last week								15		
1303	Military trainee (France)			8	6						
2000	UNEMPLOYED, not specified	2			3	5	2	30	20		
2001	Unemployed (Vietnam)									4	5
2002	Worked less than 6 months, permanent job									2	
2003	Worked less than 6 months, temporary job									6	
2100	Unemployed, experienced worker		1						21		
2101	Seeking work, worked less than 3 months			2							
2102	Seeking work, worked 3 to 6 months			3							
2103	Seeking work, worked 6 to 12 months			4							
2104	Seeking work, worked more than 1 year			5							
2105	Seeking work, experience unspecified			6							
2200	Unemployed, new worker		2	7					22		
3000	INACTIVE (Not in Labor Force)								30		
3100	Housework	3	6			10	3	50	31	6	2
3200	Unable to work/disabled	7	7			9		70	32	7	4
3300	In school	4	5	9	5	7		40	33	5	3
3400	Retirees and living on rent	8						60			
3401	Living on rent payments										
3402	Retirees/pensioners		8		4	8					
3500	Elderly	6									
3600	No work available/discouraged					6					
3700	Inactive, other reasons	9	0	0	0	11	4	80	34		6
9000	UNKNOWN/MISSING		9			0	9	99			9

Note: In the source data columns: a comma indicates more than one code was coded to the respective IPUMS-International value; an asterisk means programming logic was used; B indicates a blank in the source data.

3. Dissemination ("Extracts"). Researchers must first be approved before any data may be acquired (please see Appendix Table 3 for an image of the electronic application form). Moreover users are never permitted access to data containing the original codes provided by the

National Statistical Institutes. Instead, only integrated data are provided, and these are only in the form of extracts, custom tailored to each researcher's needs. What this means is that there is no distribution of entire datasets by means of compact discs. Since each dataset is custom tailored "collecting" or "boot-legging" datasets is not only illegal, but effectively curtailed.

See Appendix Table 3

To request an extract, the researcher must first sign in by entering the registered password, then a series of selections are made by means of point-and-click menus. The researcher selects the country or countries, census years, samples, and variables as well as the form of metadata required for the statistics package to be used (SAS, SPSS, or STATA are supported). The IPUMS-International extract engine also makes it possible to select sub-populations, such as, say, females aged 15-19 in the workforce.

One of the most valuable enhancements of the database is the "SUBSAMPLE" feature. With SUBSAMPLE, the research may request any of 100 sub-samples each of which is nationally representative and preserves any stratification of the larger sample from which it was drawn. This tool may be used to test procedures, economize resources, where the research does not require large samples, or estimate variances through the replicate method.

Once the selections are complete, there is an opportunity to review or revise before final submission of the request. Then, once submitted, the extract engine registers the request and places it in a data processing queue. When the extract is ready (usually in a matter of minutes), the researcher is notified by email that the data should be retrieved within 72 hours. A link is provided to a password-protected page for downloading the specific extract. SSL (Secure Sockets Layer) protocol is fully functional on the IPUMS-International website. Data are encrypted during transmission using a 128-bit encryption standard, matching the level used today by the banking and other industries where security and confidentiality is essential. The researcher must log-in and enter the password to obtain the extract. The researcher then downloads the file, decompresses it and proceeds with the analysis using the supplied integrated metadata consisting of variable names and labels. The metadata are in ASCII format so that a researcher may readily adapt them for use by any statistical software.

Users and Uses. The IPUMS-International project offers bona fide researchers custom-tailored extracts at no charge via the Internet. During the first 33 months of operation, 766 applications were received, of which 39% were denied. The principal reason for rejecting an application is that the proposed research (as described by the applicant in the registration request—see Appendix Table 3) does not seem to require access to the available microdata. In some cases, researchers request microdata for countries which are not presently integrated in the database. In others the proposed analysis requires information, such as certain environmental or economic variables, that is not present in the data. Then too, because of the anonymization methodology, fine-grained geographic identifiers are suppressed so requests requiring information about localities, villages, or even towns, must also be rejected. In each case, the reason for rejection is communicated to the researcher, so that a revised application may be re-submitted, if desired.

The following statistics are derived from applications for access of the first 469 approved users of the IPUMS-International database. Please note that incomplete applications are not included in this tally nor is any supplemental information which may have been requested from applicants in weighing a decision on whether to grant access or not. Approval is based solely on criteria of scientific feasibility (that census microdata are essential for the proposed research), including credentials of the researcher.

Who uses the data and what do they use it for? The succinct answer is university professors, students and policy researchers use the data to investigate economic, demographic and social issues in comparative perspective.

In a very brief period, IPUMS-International has become an indispensable component of social science infrastructure. Hundreds of projects by scholars in more than thirty-four countries are already underway. The United States accounts for the largest number of applicants (72%), followed by Canada (4%) (see Table 3). Switzerland, thanks to the presence of a large number of international organizations, ranks third (3%). Every continent is represented. Over 5% or researchers are working in Europe. Commonwealth users, at less than 4% of the total, are under-represented at present, but this is because only two samples are from a Commonwealth country.

The application does not inquire as to country of origin, citizenship or identity. Nevertheless, it is apparent from names and project descriptions, that a considerable fraction of researchers at US and Canadian universities are nationals using the IPUMS-International database to study their country of origin, including not only Kenya, Mexico, Colombia, and Vietnam but also France.

Countries of research interest. Research interest is limited to countries in the database at the time of the application. Researchers often express interest in countries for which no data are available, such as India, but these are not included in the following table. Brazil is tallied only if the country was specifically mentioned in the project description.

The most noteworthy point here is that 72% of approved projects indicated comparative research, involving more than one country. Percentages indicate the proportion of approved applicants expressing an intention to study a specific country, excluding applications before August 2002, when this information first began to be requested by means of a check box.

Use of the database to study Mexico and the United States stands out, particularly by researchers interested in studying Mexicans, regardless of whether they reside in their country of birth or in the United States. It is gratifying that for both France and Colombia, many researchers are using the IPUMS-International database rather than the national sources of census microdata. Thanks to the easy availability of the data and documentation, preliminary preparations are reduced from a matter of weeks or even months, to a day or two. Original source documentation, including census forms, enumerator instructions and the like, are available from the project website.

Table 3. Country of residence and Countries of research interest (since August 2002)

Country of residence	%		Country/ies of interest	%
USA	72		Brazil (since Sept. 2004)	4
Canada	4		China (since May 2003)	11
Switzerland	3		Colombia	13
Brazil, Colombia, Kenya (total)	8		France	12
France, Italy, Mexico, Spain, UK, Vietnam (total)	6		Kenya	12
China (includes Hong Kong, etc.)	1		Mexico	20
Australia, Germany	1		USA (excludes IPUMS-USA)	17
19 other countries (total)	5		Vietnam	11

Institutional affiliation and position (Table 4). Almost 90% of users are university based, and almost half the users are students. It is gratifying to see that researchers at national policy institutes are using the IPUMS-International harmonized microdata in preference to the privileged access that they often enjoy to data from there own country. National statistical agencies are registering with the idea of evaluating the site rather than doing research.

Table 4. User Profile: Institutional affiliation and Position

Institutional affiliation	**%**		**Position**	**%**
University	88		Student	48
Regional/International organization	8		Researcher	26
National policy institute	2		Professor	21
National statistical agency	2		Other	6

"Field" (academic discipline) is a "radio dial," which means that applicants must select from among the options available. One-third of users are economists, followed by demographers at one-fourth. "Outcome" is inferred from the project description, which means that many users do not state what they expect to produce (57%). Most of the usage is for teaching (16%), followed by papers (10%), dissertations (9%) and finally books (2%). A common, but somewhat surprising application, is to complement survey data (DHS, employment, special one-of-a-kind surveys, etc.), to estimate population weights for the surveys.

Table 5. Academic discipline and Expected outcome

Academic discipline	**%**		**Expected outcome**	**%**
Economics	37		Teaching, B.A./M.A. thesis	16
Demography	26		Paper, article, policy report	10
Sociology	13		PhD dissertation	9
Public policy	6		Book	2
History	5		Enhance DHS/other survey	6
Other	13		Other, Not mentioned	57

Research topics. Applicants are required to submit a description of the proposed research. I have classified these, somewhat arbitrarily, into 26 categories (Table 6). They demonstrate the wide range of research uses for which census microdata may be used.

Research topics include the living arrangements of the aged, female labor-force participation and educational attainment, regional inequality differentials, patterns of age hypergamy, international migration, relationship between divorce and family composition, between disease factors and education, and between marriage and socio-economic conditions. Most of these studies incorporate both cross-national and cross-temporal comparisons. For example, a National Academy of Sciences panel on "Transitions to Adulthood in Developing Countries" is using the data from Colombia, Kenya, Mexico, and Vietnam to analyze changing outcomes such as schooling, work, fertility, and marriage as a function of age, gender, and household characteristics. A scattering of studies propose to analyze various needs at the level of minor administrative districts for various institutions or professions, such as schools, teachers, clinics, health professionals, etc. While one might expect that these studies would be better served by access to 100% microdata, the many high-density harmonized samples available from the IPUMS website make the results of such studies suggestive if not conclusive.

Table 6. 446 Research topics classified in 26 categories ordered by frequency

Migration	64	Marriage	12
Schooling	57	Aging	12
Gender	30	Equality/inequality	12
Data management/development	26	Mortality	12
Teaching	37	Development	10
Health	21	Statistics	9
Fertility	21	Sampling	9
Methods	17	Demography	7
Wages	17	Brain drain/gain	6
Urbanization	15	Religion	4
Family	15	Population projection	3
Children	13	Disability	3
Poverty	12	Vital statistics evaluation	2

The following abridged and edited project description is a good example of a policy study which couples economic data from an official source with census microdata over four decades:

> analyze the impact of public investment in [Country N] on a number of social and economic indicators over the last 40 years at the [major administrative district, MADs] level. There is evidence that despite high periods of overall growth in [Country N] very little economic convergence across [MADs] has occurred. This phenomenon has raised questions about the lack of ability (or willingness) of the central government to reduce disparities using national resources. This study tries to estimate the impact of different kinds of national investment and the role they have played over four decades of development in [Country N].

A quite different example is provided by a researcher who applied to the IPUMS-International database to design a new system for delivering harmonized census microdata. As far as we are aware the researcher has not yet implemented the system.

Conclusion. Now that the construction of anonymized microdata data samples is becoming an increasingly widespread practice, integration of census microdata is an obvious next step to enhance use. With the emergence of global standards for harmonizing census data and the massive power of ordinary desktop computers, the major challenge that remains is the actual construction of integrated census microdata samples. Thanks to the cooperation of some 50 official census agencies worldwide and with the financial support of the National Science Foundation and the National Institutes of Health, the IPUMS-International project is committed to integrating microdata for 150 censuses by 2010. If the IPUMS-International project is truly successful it will continue beyond the 2000 round of censuses, incorporating samples of participating countries for the 2010 censuses shortly after they become available. The number of users and uses may increase proportionately as well.

References.

Dale, A., Fieldhouse, E. and Holdsworth, C. (2000) *Analyzing census microdata*. Arnold: London.

Esteve, Albert and Matthew Sobek. (2003). Challenges and Methods of Census Harmonization. *Historical Methods* 36: 66-79.

Eurostat Secretariat. (2001) *Report of the March 2001 work session on statistical data confidentiality*. Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Skopje. March.

Holvast, J. (1999) 'Statistical confidentiality at the European level.' Paper presented at: Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, March.

Hundepool, A., L. Willenborg, A. Wessels, L. van Gemerden, S. Tiourine and C. Hurkens. (1998) *μ-Argus User's Manual*. Statistics Netherlands: Voorburg.

International Labor Office (1990) *International Standard Classification of Occupations (ISCO-88).* Geneva.

McCaa, R. and A. Esteve. (2005). La integración de los microdatos censales de América Latina: el proyecto IPUMS, *Estudios Demográficos y Urbanos* 20:1(58) 37-70.

McCaa, R. and Ruggles, S. (2002). The Census in Global Perspective and the Coming Microdata Revolution. In Vol. 13, *Nordic Demography: Trends and Differentials, Scandinavian Population Studies*, edited by J. Carling. Oslo: Unipub/Nordic Demographic Society, pp. 7-30.

Ruggles, Steven, Miriam King, Deborah Levison, Robert McCaa, and Matthew Sobek. (2003). "IPUMS-International: An Overview". *Historical Methods*, 36: 60-65.

Thorogood, D. (1999). 'Statistical Confidentiality at the European Level.' Paper presented at: Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, March.

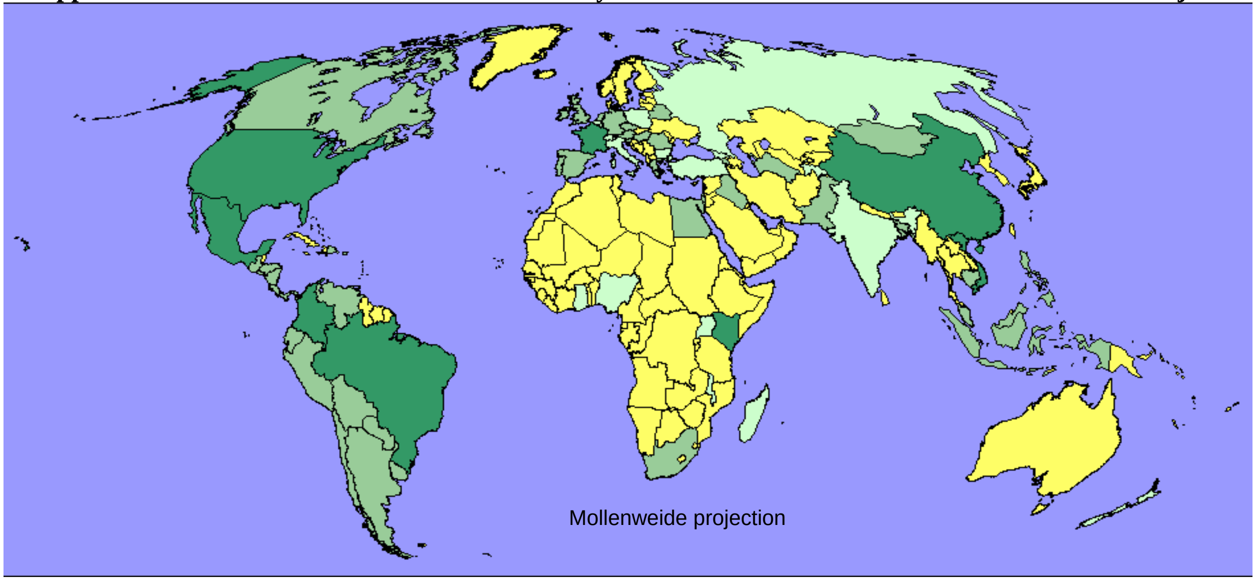
United Nations Economic Commission for Europe and Statistical Office of the European Communities. (1999). *Recommendations for the 2000 Censuses of Population and Housing in the ECE Region.* Statistical Standards and Studies, No. 49. New York and Geneva.

United Nations Educational, Scientific and Cultural Organization (1997) *The International Standard Classification of Education (ISCED 1997),* Paris.

United Nations Statistics Division (1990) *International Standard Industrial Classification of All Economic Activities* (ISIC-88). Department of Economic and Social Affairs, New York.

United Nations Statistics Division. (1998). *Principles and recommendations for population and housing censuses*. Department of Economic and Social Affairs, New York.

Appendix Table 1. IPUMS-International Country Partners and Census Microdata Entrusted to Project



Key: dark green = disseminating; medium green = data received; light green = negotiating
received = microdata for **bolded** census years entrusted to the Minnesota Population Center
Pending = agreement in principle, but official endorsement of MoU is pending
Year = census conducted; **Bold year** = microdata survive; m = microcensus

Status	**Statistical Agency of**	**2000s**	**1990s**	**1980s**	**1970s**	**1960s**
Phase I, 1999-2004 (8 countries)						
Received	Brazil	**2001**	**1991**	**1980**	**1970**	**1960**
	China (only '82 'til now)	2000	1990	**1982**		1964
Received	Colombia		**1993**	**1985**	**1973**	**1964**
Received	France	**1999**	**1990**	**1982**	**1975**	**1968, 62**
Received	Kenya	**1999**	**1989**	**1979**	**1969**	
Received	Mexico ('80 in recovery)	**2000**	**1990**	1980	**1970**	**1960**
Received	United States	**2000**	**1990**	**1980**	**1970**	**1960**
Received	Vietnam		**1999**	**1989**	1979	
Phase II, 2004-9						
Asia and the Pacific (13 countries)						
Received	Armenia	**2001**		1989	1979	1970
	Bangladesh	**2001**	**1991**	**1981**	**1974**	1961
Received	Cambodia		**1998**			1962
Received	Fiji Islands		**1996**	**1986**	1976	**1966**
Soon	Indonesia	**2000**	**1990**	**1980**	**1971**	1961
Received	Iraq		**1997**	1987	1977	1967
Received	Israel		**1995**	**1983**	**1972**	**1961**, 67
Received	Malaysia	**2000**	**1991**	**1980**	**1970**	1960
Received	Mongolia	**2000**		1989	1979	1970
Received	Pakistan		**1998**	**1981**	**1973**	1961
Received	Palestinian Authority		**1997**			
Received	Philippines	**2000**	**1990**	**1980**	**1970**	**1960**
Received	Turkmenistan		**1995**	1989	1979	1970

Europe, 2004-8 (18 countries)						
Received	Austria	**2001**	**1991**	**1981**	**1971**	1961
Received	Belarus		**1999**	1989	1979	1970
	Bulgaria	**2001**	**1992**	**1985**	1975	1965
Soon	Czech Republic	**2001**	**1991**	1980	1970?	1961
Soon	Germany (Ro and DR)	**2001m**	**1991m**	**1987, 81**	**1970, 71**	1961
Received	Greece	**2001**	**1991**	**1981**	**1971?**	1961
Received	Hungary	**2001**	**1990**	**1980**	**1970**	
	Ireland	**2001**	**1991**	1981	1971	1961
Pending	Italy	**2001**	**1991**	**1981**	1971	1961
Received	Netherlands	**2001m**			**1971**	**1960**
Pending	Poland	**2001**		**1988**	**1978**, **70**	1960
Soon	Portugal	**2001**	**1991**	**1981**	1970	1960
Received	Romania	**2001**	**1992**		**1977?**	1965
Pending	Russia (-1989 USSR)	**2002**	**1994m**	**1989**	1979	1970
Soon	Slovenia	**2001**	**1991**	1981		
Received	Spain	**2001**	**1991**	**1981**	1970	1960
Pending	Turkey	**2000m**	**1990**	**1980**, 85	**1970, 75**	1960, 65
Received	United Kingdom	2001	**1991**	1981	1971	1961
North America and the Caribbean, 2003-7 (8 countries)						
Received	Canada	2001	**1991,** 96	**1981,** 86	**1971,** 76	1961, 66
Received	Costa Rica	**2000**		**1984**	**1973**	**1963**
Received	Dominican Republic	**2003**	**1993**	**1981**	**1970**	1960
Received	El Salvador		**1992**		**1971**	1961
Received	Guatemala	**2003**	**1994**	**1981**	**1973**	**1964**
Received	Honduras	2000		**1988**	**1974**	**1961**
Received	Nicaragua	2005	**1995**		**1971**	1963
Received	Panama	**2000**	**1990**	**1980**	**1970**	**1960**
Received	Puerto Rico	**2000**	**1990**	**1980**	**1970**	1960
South America, 2003-7 (8 countries)						
Soon	Argentina	**2001**	**1991**	**1980**	**1970**	1960
Received	Bolivia	**2001**	**1992**		**1976**	
Received	Chile	**2002**	**1992**	**1982**	**1970**	**1960**
Received	Ecuador	**2001**	**1990**	**1982**	**1974**	**1962**
Received	Paraguay	**2002**	**1992**	**1982**	**1972**	**1962**
Received	Peru		**1993**	**1981**	1972	1961
Received	Uruguay		**1996**	**1985**	**1975**	**1963**
Received	Venezuela	**2001**	**1990**	**1981**	**1971**	**1961**
Africa, 2006 (3 countries)						
Received	Egypt		**1996**	**1986**, 81	1976	1964
	Gambia	**2003**	**1993**	**1983**	**1973**	
Received	South Africa	**2001**	**1996,** 91	1985, 80	1970	1960
Datasets per Census Round (n)		**44**	**53**	**42**	**40**	**19**

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.

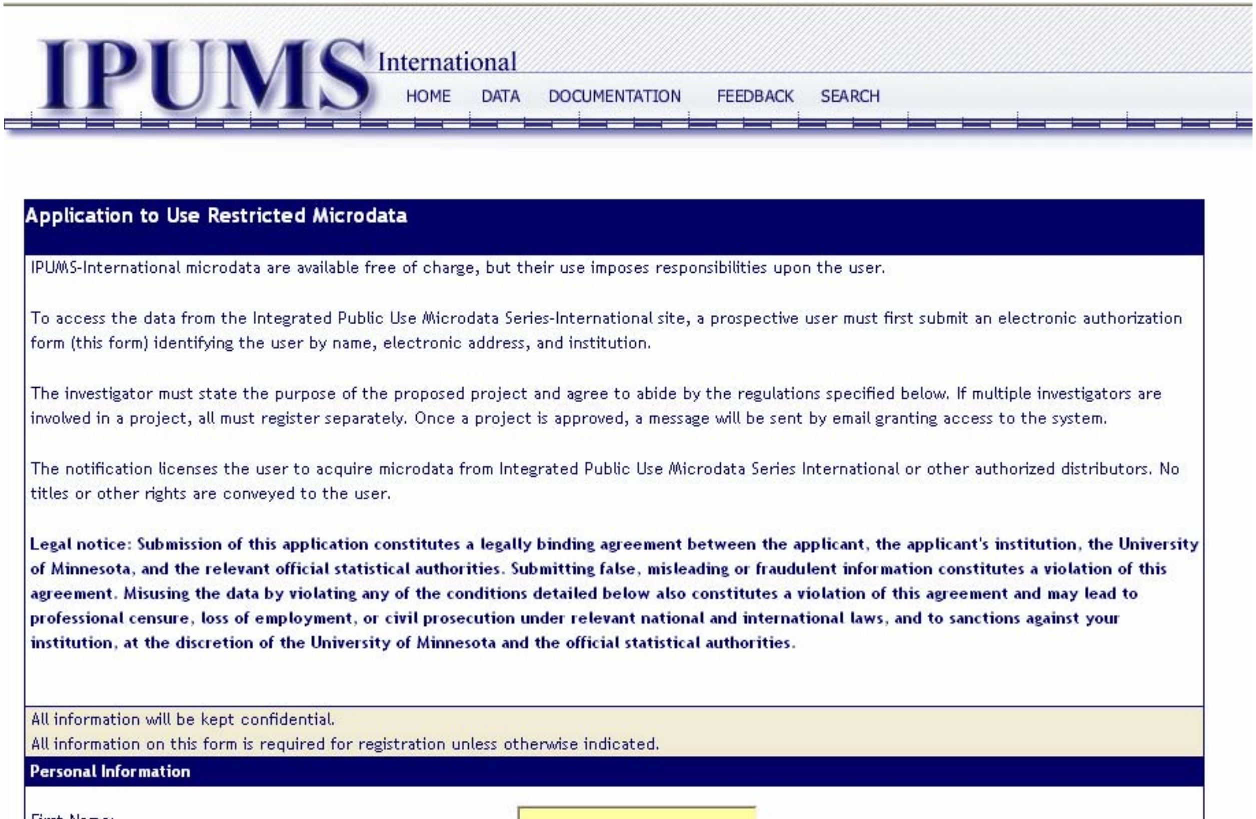
8. Violations. Violation of the user license may lead to professional censure, loss of employment, and/or civil prosecution. The University of Minnesota, national and international scientific organizations, and the [Official Statistical Agency of X] will assist in the enforcement of provisions of this 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.

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. Misusing the data by violating any of the conditions detailed below also constitutes a violation of this agreement and may lead to professional censure, loss of employment, or 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