

IPUMS-ASIA/PACIFIC: SYNOPSIS OF A PROPOSAL
FOR A RESTRICTED ACCESS WEB-SITE PROVIDING ANONYMIZED, INTEGRATED CENSUS MICRODATA
FOR SOCIAL SCIENCE AND POLICY RESEARCH FOR THE ASIA AND PACIFIC REGION

Robert McCaa, Steven Ruggles, Matt Sobek (University of Minnesota Population Center)
and Albert Esteve (Centre d'Estudis Demogràfics, Autonomous University of Barcelona)

21st Population Census Conference: Analysis of the 2000 Round of Censuses
Kyoto, Japan, 19-21 November 2003

Research for this paper was funded in part by the National Science Foundation of the United States,
grant SBR-9908380 'Integrated Public Use Microdata Series International'.

Abstract. This paper describes a proposal for the Asia and Pacific Region to develop a restricted access web-site providing anonymized, integrated census microdata for social science and policy research. With funding by the U.S. National Science Foundation and the National Institutes of Health we propose to build on the successful IPUMS-International protocols (www.ipums.org/international). This initiative is a global collaboratory in partnership with the Official Statistical Authority (OSA) in each member-state. Two region-wide efforts, one for Europe and the other for the Americas, are now underway with a total of 33 countries participating. In the Asia region, integrated census microdata for Vietnam first became available in May 2002 and China in 2003. Access to integrated microdata is restricted to bona fide university scientists with demonstrated research need who agree to abide by the conditions of use license. Custom-tailored extracts are delivered, at no charge via the Internet. At present more than forty official census agencies have formally ratified the IPUMS-International protocols. Statistical Agencies of the Asia and Pacific Area interested in joining a region-wide initiative are invited to contact Dr. Robert McCaa (rmccaa@umn.edu).

Introduction. Some forty countries, encompassing more than 2.5 billion people, have now formally joined the IPUMS-International project (Table 1). The rapid expansion from three countries in 1999 is due in part to the fact that there is growing recognition that anonymized census microdata samples constitute statistical data. As such, they do not violate even the most stringent laws on statistical confidentiality, privacy or human rights. In country-after-country, close scrutiny of statistical laws on census privacy reveals that the release of anonymized microdata samples, with names and detailed geographical identifiers suppressed, is not prohibited by law. In the rare instance where law are interpreted to the contrary, this is often based on a misreading of the statutes and a misunderstanding of the statistical nature of census microdata samples. The General Data Dissemination System (GDSS) of the International Monetary Fund is widely recognized as the gold standard with respect to issues of privacy and statistical confidentiality. As of 2001, census microdata samples were disseminated by 37 of the 52 member states of the GDSS (McCaa and Ruggles 2002). This change in legal interpretation as well as the recognition of the importance of making better use of the enormous amount of census data collected, coupled with both the recognition that qualified academic researchers have a need to access census data and the enormous advances in desktop computing power—all have led to a breakthrough in making these valuable resources available for scientific and policy research. At present, in addition to the 40 official statistical agency members, international partners of the IPUMS-International initiative include The UN Demographic Center for Latin America and the Caribbean (CELADE), the World Health Organization (Department of Health Service Provision, or OSD), and the Population Activities Unit of the UN-ECE.

Insert Table 1 near here

In May 2002, working in conjunction with official statistical agency (OSA) partners, the IPUMS-International project released the first preliminary group of harmonized census microdata samples for Colombia (1964-1993), France (1962-1990), Kenya (1989-1999), Mexico (1960-2000), the United States (1960-1990), and Vietnam (1989-1999), followed by China in 2003. A third group of harmonized samples for Brazil will be released in 2004. Over 60 million person records from more than 20 nationally representative samples consisting of more than 50 variables are now available from the IPUMS-International web-site (<http://www.ipums.org/international>).

To begin an Asia/Pacific regional census harmonization project is the same as for the Americas and Europe. The first step is formal approval by the official statistical authority of each country wishing to participate. Table 2 lists statistical authorities of the region (with a minimum population of 400,000) as well as census microdata availability, as far as is presently known. The protocol of understanding, spelling out rights of the Official Statistical Agencies and the obligations of the Minnesota Population Center, is reproduced in Appendix A. Note that when the protocol is signed, informal agreement is reached on census microdata that will be made available to the project, likely sample density, a tentative calendar and division of tasks (drawing samples, data cleaning, designing harmonization, translating documentation into English or French, etc.). A budget is also agreed upon so that

the Minnesota Population Center may obtain the necessary funding. No formal contract is signed. Payments are made by means of a bill of sale, unless a formal contract is required. Where desired, consultants are contracted, individually, to work on the project for a month or two, as necessary. Official Statistical Agencies not presently associated with the project are invited to contact the International Project Coordinator, Dr. Robert McCaa at rmccaa@umn.edu.

Insert Table 2 near here

Two regional projects are already underway. In 2002, sixteen countries of North and South America (Table 3), joined together to undertake a region-wide integration of census microdata samples using the IPUMS-International protocols. In 2003, funding was requested for a project for seventeen European countries (Table 4). Notification of the outcome will be forthcoming in Spring 2004. Given that the Americas project was ranked in the top one percentile of NIH funded work, we expect a similar outcome with the European proposal. Other regional initiatives are being developed as a sufficient number of OSAs ratify the project protocols. Since only one new project may be funded per year, it is a bit of a race whether the next project will be in Asia, Africa or the Arab region. Over the longer term, if the IPUMS-International initiative is successful—and here each OSA will decide based on the experience with the current regional projects—it will continue through the 2010 census round, incorporating census microdata of member countries, as soon as they become available. Note that for the USA 2000 census, integrated microdata were made available from the <http://ipums.org/USA> web-site within two months of the day of release by the United States Census Bureau.

Insert Table 3 near here

Insert Table 4 near here

The ensuing sections of the paper discuss three issues that are of central concern to official statistical authorities considering membership in the initiative: confidentiality protections, principles of integration, and data quality enhancements. The paper concludes with a brief outline of a draft work plan for an Asia & Pacific census microdata integration project.

1. Confidentiality protections. IPUMS-International means Integrated Restricted-Access, Anonymized Microdata Samples. The IPUMS-International acronym carries “PUMS” embedded in its name, but in fact the data are available only as “Restricted-Access”, Anonymized Microdata Samples. Thus, “IRAAMS” 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 proposals, IPUMS-Latin America and IPUMS-Europe.

Nonetheless, it is important to understand that a comprehensive array of protections are in place to guarantee the privacy and statistical confidentiality of 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 OSA
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—indeed no funds are solicited—for a project without prior signed authorization from the corresponding OSA. 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 ten 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 OSAs suggest modifications. Any new provision is forwarded to current members of the consortium for their consideration and up-dating as desired.

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 authority of the Chamber of Commerce of Paris.

Second, due to confidentiality restrictions, researchers must apply to become approved to use the system (Appendix B). Typically, one-in-two 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, and 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 download data extracts that contain only the samples and variables of use to them. They import the extracts into their statistical software of choice to analyze in the convenience of their own institutional setting.

Table 5 lists projects approved for access by subject matter, university or research organization, funding agency, and human subjects protection boards, as reported at a EuroSTAT meeting last February. 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. While the vetting of applications is performed by the Principal Investigators of the IPUMS-International project, an international advisory board made up of distinguished statisticians and researchers is being constituted to review on a regular basis all aspects of the project to ensure compliance with the memoranda of understanding.

Insert Table 5 near here

Third are the technical measures taken to ensure statistical confidentiality. Where the OSA requests that the MPC apply anonymization procedures, we implement the following technical protections (based on Thorogood 1999):

1. adopt sample size according to official norms or conventions;
2. limit geographical detail to administrative units with a minimum number of inhabitants (as high as 100,000 for some countries and as low as 10,000 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 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 Integration. 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. Beginning in 1991, the IPUMS-USA project has worked to harmonize census data for the United States for the period since 1850, and IPUMS-International has capitalized on this experience.

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.

For most variables, it is impossible to construct a single uniform classification without losing information. Some samples provide far more detail than others, so the lowest common denominator of all samples inevitably loses important information. Composite coding schemes offer a solution. 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 marital status, for example. In the IPUMS-International system, the first digit of this variable with four categories is comparable across all samples: single, married, widowed, and separated/divorced. The second digit delineates consensual unions from other forms of marriage (where possible) and distinguishes among the categories separated, divorced, and married with spouse absent. The final digit provides additional detail with the married and married-spouse-absent categories (such as polygamous marriages in Kenya).

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 6 illustrates the harmonization of codes for the variable "employment status".

Insert Table 6 near here

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. 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. Thanks to the advice of experienced consultants it is possible to readily identify problematic concepts and revise the harmonized codes accordingly. It is important to understand that no major decisions are made at the integration center without input by experts who work as paid consultants to the project. This decentralized approach allows multiple projects to go forward simultaneously, country-by-country, without duplication or wasted effort.

Geographic variables pose the greatest challenges. Within the cost constraints of the first-stage projects, full harmonization of the lowest level of geographic information available, even taking into account constraints imposed by statistical confidentiality measures, cannot be attempted. However an attempt is made to create a consistent definition of large metropolitan districts. Moreover, wherever feasible, maps are provided of administrative districts identified in the microdata and any other ancillary geographic information available.

3. Data Quality and Constructed Variables. In addition to providing harmonized codes for variables and accompanying documentation, the IPUMS-International project is carrying out a variety of other tasks to improve data quality. These tasks include the following:

- Clean data to eliminate duplicate records, inappropriately merged households, and other errors
- Develop internal consistency checks to maximize data integrity. This includes, for example, examining consistency between age and marital status, occupation, and school attendance; looking for persons with multiple spouses for countries in which this is not an accepted custom; and checking for agreement between household and individual characteristics.

- Implement allocation procedures to impute values for missing or inconsistent data items, using logical edits together with probabilistic "hot deck" methodology. A data quality flag identifies allocated data items.
- Create constructed variables to simplify data analysis, including family interrelationship variables.

Researchers tell us that the constructed family interrelationship variables constitute one of the most valuable enhancements of the integrated dataset. We use a system of logical rules to identify the record number within each household of the individual's mother, father, or spouse, if they were present in the household. These pointers allow users to automatically attach the characteristics of these kin or to construct measures of fertility and family composition. In addition, other constructed variables describe family and household characteristics at the individual and household level (such as family and subfamily membership, family and subfamily size, and number of own children).

4. A proposed initiative for the Asia/Pacific region.

4.a. Integration work plan. Typically only three years of effort are required to prepare a country's microdata for distribution, once endorsement of project protocols has been formally ratified. This shortness of time is due in part to the fact that the IPUMS International consortium is a partnership between the Minnesota Population Center, Official Statistical Agencies (OSAs), international statistical organizations, and researchers world-wide. The MPC obtains the funding for country-specific projects, coordinates the research effort, programs the anonymization and integration, and distributes the data. The integration work is a collective endeavor, which draws on the expertise of official census agencies and other experts. License fees are paid to the OSAs not only for dissemination rights, but also for the supply of ancillary materials (such as codebooks and technical publications) and technical support by the staff of these agencies. As needed, this pool of knowledgeable specialists is supplemented with the help of other experts. They answer questions on census enumeration procedures and post-enumeration data processing, the methodology employed to create existing samples, and specific integration problems (such as the details of economic, education, housing, and geographic variables for particular countries).

The work proceeds in nine stages, upon completion of two preliminary steps, as follows:

- 1. Formally ratify the IPUMS-International project protocols between the University of Minnesota and the Official Statistical Agency (Appendix A).
0. Obtain funding by Minnesota Population Center to license data, reimburse in-country expenses, develop the database, and maintain the extract engine (a copy of the NIH proposal for the IPUMS-Europe project submitted October 1, 2003 is available upon request).
1. Acquire census documentation (enumeration forms, enumerator instructions, codebooks, record layouts, etc.) and microdata.
2. Clean raw data files (e.g., identify and correct data format problems; carry out internal consistency checks; identify coverage problems through comparison with published statistics).
3. Draw high-density samples from 100 percent internal census files, where available.
4. Impose confidentiality protections (e.g., top-codes, geographic swapping, category blurring, and randomization of household sequence within geographic units).
5. Recode variables into the IPUMS-International harmonized coding system to permit analysis across countries and time periods; develop and apply new harmonized coding designs optimized for regions or sub-continent.
6. Allocate missing and inconsistent data values through probabilistic and logical editing procedures.
7. Create a set of consistent constructed variables describing household composition, family interrelationships and socioeconomic status.
8. Develop harmonized English-language documentation (e.g., census enumeration procedures and instructions; post-enumeration processing; sample designs; variable-level documentation on census questions, universe definitions, variable category availability, and frequency distributions; definitions of households, dwellings, group quarters and other enumeration units; and comparability issues across census years and countries).
9. Convert all documentation to the Data Documentation Initiative (DDI) international metadata standard.

4.b. Documentation. The bulk of the web site documents the available samples and variables. Of particular note are the variable comparability discussions. These are designed to indicate where there are notable issues for interpreting a variable's codes for purposes of temporal and spatial comparison. In addition to these discussions, the web site contains the original census questionnaires and instructions so users can examine the full text of the original documentation developed by the OSA for conducting census operations.

4.c. Data Dissemination (Extracts). The final stage is to make the data available as "extracts" to accredited users through the IPUMS-International restricted access website. Researchers must first be approved, as explained above, before any data may be acquired. Moreover, once approved, only "integrated

extracts" are disseminated. Researchers are never provided complete copies of any sample nor are they given access to data containing the original codes developed by the OSA. No CDs are distributed by the project, unless requested by an OSA partner. Instead, researchers obtain custom extracts by means of a series of selection screens. After signing-in and entering the corresponding password, the researcher selects the country or countries, census years, samples, and variables required as well as the statistical analysis package desired (SAS, SPSS, or STATA). The extract engine also makes it possible to select sub-populations, such as females aged 15-19 in the workforce. Once the selections are complete, there is an opportunity to review or revise all selections before submission. Then, the extract engine places the request in a 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 in the message for downloading the specific extract. The extract is password protected and registered. The researcher may then download the file, decompress it and proceed with the analysis. The supplied integrated metadata consisting of variable names and labels consist of the instructions required for analysis using SAS, SPSS, or STATA.

4.d. Commercialization. Neither the IPUMS-International project nor the Minnesota Population Center sells any products. However our OSA partners do. We are developing an interface to assist our OSA partners in informing our users about products that may be of interest, such as digitized boundary files, household surveys, publications, documentation, and even census microdata. Because the project distributes only integrated census microdata, there will be users who will need to use the original data. Then too, under the agreement negotiated with each OSA partner, the project cannot supply microdata to commercial users. When they apply for integrated data, we direct them to the OSA, where they may purchase the data.

Conclusion. Now that the construction of anonymized microdata data samples is becoming an increasingly widespread practice, harmonization of census microdata is an obvious next step to enhancing use. With the emergence of global standards of statistical confidentiality and the massive power of ordinary desktop computers, the major challenge that remains is the actual construction of integrated, anonymized of census microdata samples.

References.

- Dale, A. and Elliot, M. (2001) 'Proposals for 2001 SARS: An assessment of disclosure risk.' *Journal of the Royal Statistical Society, Series A*, 164, part 3, pp.427-447.
- 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 Netherland: Voorburg.
- Kelly Hall, P., McCaa, R. and Thorvaldsen, G., eds (2000) *Handbook of international historical microdata for population research*. Minnesota Population Center: Minneapolis. (Updated microdata inventory available at www.IPUMS.org/international/iiinventory2.html.)
- McCaa, Robert, and Steven Ruggles. 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, S. (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: www.IPUMS.org/~census2000.)
- Ruggles, Steven, and Matthew Sobek, et. al. 1997. *Integrated Public Use Microdata Series: Version 2.0*. Minneapolis: Historical Census Projects, University of Minnesota.
- 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 Statistics Division. (1998). *Principles and recommendations for population and housing censuses*. Department of Economic and Social Affairs, New York.

United Nations Economic Commission for Europe and Statistical Office of the European Communities. (1998). *Recommendations for the 2000 Censuses of Population and Housing in the ECE Region*. Statistical Standards and Studies, No. 49. New York and Geneva.

Table 1. IPUMS-International consortium members, November 1, 2003	
World Region	Official Statistical Authority
Africa	Ghana, Kenya, Madagascar, Malawi, Uganda
Americas	Argentina, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Venezuela, USA (pending: Bolivia).
Asia	China, Tajikistan, Turkmenistan, Vietnam
Europe	Austria, Belarus, Bulgaria, Czech Republic, France, Germany, Greece, Hungary, Netherlands, Portugal, Romania, Slovenia, Spain, the United Kingdom (pending: Belgium, Poland, Russia).
Middle East	Israel, Palestinian Authority

Table 2. Asia/Pacific censuses and microdata availability by place and year

Note: Census year in bold indicates microdata are available from NSO or other source
Please email corrections, clarifications and revisions to: rmccaa@umn.edu. Thank you!

<i>Place</i>	<i>Pop (000)</i>	<i>2000s</i>	<i>1990s</i>	<i>1980s</i>	<i>1970s</i>	<i>1960s</i>	Pre-census 2000 microdata availability
IPUMS-International							
China	1264.5	2000	1990	1982		1964	EWC
Vietnam	78.6		1999	1989	1979		MPC
IPUMS-Asia/Pacific							
Afghanistan	26.6				1979		
Australia	19.2	2001	1991/6	1981/6	1971/6	1961/6	ANU
Bangladesh	128.1	2001	1991	1981	1974	1961	
Cambodia	12.1		1998			1962	
Fiji	0.8		1996	1986	1976	1966	EWC
Hong Kong	7.0	2001	1991/6	1981/6	1971/6	1961/6	
India	1002.1	2001	1991	1981	1971	1961	
Indonesia	212.2	2000	1990	1980	1971	1961	EWC, UW, ANU
Iran	67.4		1996	1986	1976	1966	
Japan	126.8	2000	1990/5	1980/5	1970/5	1960/5	
Korea, RO	47.2	2000	1990/5	1980/5	1970/5	1960/6	EWC
Korea, DPR	21.6		1993				
Lao PDR	5.2		1995	1985			
Macao	0.4		1991	1981	1970	1960	
Malaysia	23.2	2001	1991	1980	1970	1960	EWC, UW, ANU
Mongolia	2.4	2000		1989	1979		
Myanmar	48.8		1994	1983	1973		
Nepal	23.9	2001	1991	1981	1971	1961	
New Zealand	3.8	2001	1991/6	1981/6	1971/6	1961/6	
Pakistan	150.6		1998	1981	1973	1961	EWC
Papau New Guinea	4.8	2000	1990	1980	1971	1961	
Philippines	80.2	2000	1990	1980	1970	1960	EWC
Microcensus			1995				
Singapore	4.0	2000	1990	1980	1970		
Sri Lanka	19.1	2001		1981	1971	1960	EWC
Taiwan and Fukien Islands	22.2	2000	1990	1980	1970	1966	
Thailand	620	2000	1990	1980	1970	1960	EWC, ANU

**Table 3. Latin America census microdata access project:
density (%) of source microdata by country and decade of census (began July 2003)**

Country	Millions	1960s	1970s	1980s	1990s	2000s
Argentina	37.0	3	2	2	100	100
Bolivia (not signed)	8.3	.	100	.	100	100
Brazil	170.1	25	25	25	12	10
Chile	15.2	1	5	100	100	100
Colombia	40.0	2	100	100	100	100
Costa Rica	3.6	6	100	100	.	100
Cuba (not signed)	11.1	.	n.a.	n.a.	.	100
Dominican Republic	8.4	7	7	8	n.a.	100
Ecuador	12.6	3	17	100	100	100
El Salvador	6.3	1	5	.	100	100
Guatemala	12.7	5	5	5	100	100
Honduras	6.1	1	10	100	.	100
México	99.6	1.5	1	n.a.	100	100
Nicaragua	5.1	n.a.	10	.	100	.
Panama	2.8	5	20	100	100	100
Paraguay	5.5	5	10	100	100	100
Peru	27.1	n.a.	n.a.	n.a.	100	100
Puerto Rico	3.9	10	3	7	6	6
Uruguay (not signed)	3.3	5	100	100	100	.
Venezuela	24.2	2	22	100	30	100
Total extant datasets	502.9	16	18	14	16	18
Total datasets in project	488.5	15	17	13	15	17

**Table 4. IPUMS-Europe: Available Censuses and Expected Sample Sizes (in 000s), by Country
(Funding proposal submitted to the National Institutes of Health, October 1, 2003)**

	Sample Density (%)	Census	N	Census	N	Census	N	Census	N	Census	N	Census	N
Austria	5	2001	(405)	1991	(390)	1981	(380)	1971	(375)	1961	(360)	.	.
Belarus	5	1999	(520)	1989	(510)
Belgium	10	2001	(1,030)	1991	(1,000)	1981	(990)	1971	(970)
Bulgaria	5	2001	(395)	1992	(425)
Czech Rep.	5	2001	(515)	1991	(515)
France	5	1999	(3,005)	1990	(2,361)	1982	(2,714)	1975	(2,629)	1968	(2,488)	1962	(2,321)
Germany	.5	2000	(414)	1991	(400)	1987	(306)	1982	(308)	1970	(305)	1962	(285)
Greece	10	2001	1090)	1991	(1,020)	1981	(970)	1971	(880)
Hungary	5	2001	(505)	1990	(520)	1980	(535)	1970	(515)
Netherlands	1	2001	(160)	1991	(150)	1981	(141)	1971	(130)	1960	(115)	.	.
Poland	5	2002	(1,930)	1995	(1,940)	1988	(1,900)	1984	(1,850)	1978	(1,745)	1974	(1,680)
Portugal	5	2001	(500)	1991	(495)	1981	(490)
Romania	10	2002	(2,240)	1992	(2,280)
Russia	5	2002	(7,200)	1989	(7,400)
Slovenia	10	2001	(200)	1991	(200)	1981	(180)
Spain	5	2001	(2,040)	1991	(1,940)	1981	(1,875)
UK	1	2001	(600)	1991	(574)
												Total person records =	71,101

Notes:

Microcensuses: Germany 1982, 1991, 2000; Netherlands 1981, 1991, 2001; Poland 1974, 1984, 1995.

Our present agreement calls for a .5% sample for France 1999. We hope to expand that sample in the course of the project.

A sample for 1960 Poland is included in the total case count.

Final agreements for Belgium, Poland and Russia are pending, and some of the earliest censuses may not be recoverable.

Table 5. Report on Approved Access to Restricted Microdata, IPUMS-International,**Funding Agencies**

Canadian Foundation for Innovation
 Council for the Development of Social Science Research in
 Africa
 Economic and Social Research Council, UK
 National Science Foundation
 National Institutes of Health
 Norwegian University Development Aid Funding
 Rockefeller Foundation
 Wellcome Trust

Over-sight Boards

CNIL: Commission Nationale Information et Liberte
 Comite National d'Ethique
 Institutional Review Board (IRB) on research involving human
 subjects. *Note: Every university or research group funded
 by the National Institutes of Health must establish an IRB
 or equivalent.*
 Inter-University Consortium for Political and Social Research
 IRD scientific commission (Conseil Scientifique)
 ISA and its research committees RC28 and RC33
 National Committees for Research Ethics in Norway
 USA Federal Code title 13/title 26 /title 5
 Vice-decanat a la recherche, Universite de Montreal,
 Documents pour l'ethique

Professional Associations

American Economic Association
 American Public Health Association

 American Sociological Association
 International Union for the Scientific Study of Population
 (IUSSP)
 Latin American and Caribbean Studies Association
 Population Association of America
Universities/Research Organizations
Europe
 Cardiff University
 Demographic Studies Center - University Auton. of Barcelona
 Department of Statistics, University of Florence
 INED Paris
 Institut d etudes politiques de Paris
 Institut francais de recherche en Afrique (IFRA)
 Ministry of Economic Development and Trade of Russian
 Federation
 Novosibirsk State Technical University
 University College London

Canada

Department of Demography, University of Montreal
 Queen's University
 Simon Fraser University
 Statistics Canada -Library and information centre
 University of Toronto

USA

Boston University
 Brown University
 Columbia University
 Dept. of Econ., Massachusetts Instit. of Technology
 East-West Center
 Florida State University
 George Mason University
 Georgetown Public Policy Institute

Approved Projects (key words only)

Brain drain: sending and receiving countries
 Calibration of birth registrations against census microdata for countries
 with strong border migrations.
 Comparison of fertility patterns by migration status
 Construction of life-tables for sub-national populations.
 Cross national studies of poverty and social issues
 Cross-national analysis of human health resources
 Cross-national analysis of wage structure/discrimination
 Cross-national comparison of the determinants of poverty
 Cross-national determinants of female labor force
 Cross-national study of inequality
 Cross-national study of living standards and sanitation
 Demographic and spatial dimensions of homicide rates in relation to
 demographic changes.

 Demographic processes: fertility, mortality, migration
 Demographic profiles of older populations
 Develop regional accounts systems
 Development of cross national social interaction and stratification scales.
 Disability and welfare expenditures
 Education stock estimates for evaluating the efficiency of health systems

 Educational gaps between minority and majority populations

 Effects of AIDS on school enrollments
 Effects of economic growth on demand for skills and education and the
 returns to labor.
 Effects of educational mismatches on wages and salaries
 Effects of national poverty programs on child labor and school attendance

 Effects of social networks on rural-ruban migration.
 Effects of urbanization on internal migration
 Emigration: the gender gap
 Emission of green house gases: population and labor
 Evolution of non-agricultural employment in rural areas
 Extent of death clustering by regions
 Gender differences in educational attainment
 Gender earnings differences by ruralurban areas
 Household structures of the elderly
 Human welfare, agriculture and the environment
 Inequality of wages: instruction of advanced graduate students on the use of
 census microdata
 Immigration of specific nationalities
 Impact of climate variation on poverty
 Infrastructure and economic activities on public health
 Labor supply and regional development
 Living arrangements of the elderly around the world
 Marriage transitions in developing countries
 Marriage, child labor, and polygamy
 Material inequality
 Migrants by country of origin/destination & duration
 Migration from Mexico to the USA
 Occupational changes and reshaping of industrial policies
 Period-cohort analysis of educational attainment in comparative perspective
 Recalibration of survey data using census microdata
 Regional clustering of infant and child mortality
 Religion and nationalism
 School and work in developing and developed countries.
 Social determinants of marital fertility

Harvard University	Substitution of wooden housing materials and effects on forest and environment
Illinois Wesleyan University	Teach advanced graduate students how to use census microdata for the study of public health issues
International Program Center-U.S. Census Bureau	Teach advanced graduate students to use census microdata to analyze labor markets
Johns Hopkins Bloomberg School of Public Health	Teach advanced graduate students to use census microdata to study aging and household structures
Johns Hopkins Population Center	The marriage squeeze and marriage rates: comparisons
Marshall University	Transitions from adolescence to adulthood: education, work, marriage, child-rearing
Northwestern University	Transitions to adulthood: life course trajectories by gender and household characteristics.
Office of Population Research - Princeton University	Trends in educational attainment; impact of work force.
ORC Macro International	Well being of the elderly
Population Research Institute Penn State University	Why the brain drain is more severe in some countries.
Population Studies Center University of Michigan	Women in the labor market
San Diego State University	Other World Regions
Stanford University	African Population and Health Research Center
Tufts University	Centro de Investigacion y Docencia Economicas.
Tulane University School of Public Health	Hong Kong University of Science and Technology
United States Bureau of the Census	National University of Singapore
University at Albany, SUNY	The University of Nairobi
University of California Riverside	The World Bank
University of California, Berkeley	Universidad Externado de Colombia
University of Chicago	Universidad Pedagogica Experimental Libertador
University of Illinois at Chicago	World Agro-Forestry Centre
University of Maryland	World Health Organization
University of Minnesota	
University of North Carolina School of Public Health	
University of North Carolina at Chapel Hill	
University of Pennsylvania	
University of Pittsburgh	
University of Southern California	
University of the Pacific	
University of Wisconsin--Demography and Ecology	
Yale University	

Table 6. 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.

Appendix A Memorandum of Agreement.
**Integrated Public Use Microdata Series International
 and the [Official Statistical Agency of Jurisdiction 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 census 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]**. A copy of both the original census microdata and integrated samples will be deposited with the World Health Organization, Geneva Switzerland for the exclusive research needs of that institution.
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, Official 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 (1) arbitrator, which shall be elected by lot from the list of Arbitrators of the Chamber of Commerce of Paris. This agreement shall be governed by, and construed in accordance with, generally accepted principles of International Law.

Date: _____

Signed: _____

Regents of the University of Minnesota

By: Kevin J. McKoskey, Sponsored Projects Administration

Date: _____

Signed: _____

Appendix B: Application to Use Restricted Microdata

IPUMS-International Data Extraction System 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.

*All information will be kept confidential.
All information on this form is required for registration.*

Personal Information

First Name: Last Name:

Employer/Institutional Affiliation (Note: change requires re-application):

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

Institutional Review/Data Safety Board, Office for Human Research Protections, or Scientific Conduct Committee. Indicate name at your institution, or state "None":

Address:

Street Address 1:	<input type="text"/>
Street Address 2:	<input type="text"/>
City, State/Province, Zip:	<input type="text"/>
Country:	<input type="text"/>

Phone Number: (include country and area codes) Fax Number: (optional)

E-mail address:

<p>Field:</p> <p><input type="radio"/> Demography</p> <p><input type="radio"/> Economics</p> <p><input type="radio"/> History</p> <p><input type="radio"/> Sociology</p> <p><input type="radio"/> Other academic</p> <p><input type="radio"/> Public Policy</p>	<p>Status:</p> <p><input type="radio"/> Faculty</p> <p><input type="radio"/> Academic researcher</p> <p><input type="radio"/> Support staff</p> <p><input type="radio"/> Student</p> <p><input type="radio"/> Non-Academic Researcher</p>
---	---

**Usage License
for Integrated Public Use Microdata Series International
(IPUMS-International) and its partners**

Please check all 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.

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.

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 is prohibited. Alleging that a person or household has been identified in these data is also prohibited.

The microdata must always be safely secured.
Users will implement security measures to prevent unauthorized access to microdata acquired from Integrated Public Use Microdata Series International, its partners or authorized distributors.

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 asked to provide the IPUMS-International staff with a full citation for any publications resulting from their work with these data.

Any 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 the relevant 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.

Description of Project Proposal:

Please provide a clear description of the proposed use of the data (25 words minimum). This description will be used to evaluate your application.

Data to be extracted (Enter names):

Contingent upon acceptance of the application, your User Name will be set to the following email address: (Please make sure it's correct; change at the top of this form.)

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

Confirm Password:

Submit Registration