IPUMS-INTERNATIONAL: A RESTRICTED ACCESS WEB-SITE
PROVIDING ANONYMIZED, INTEGRATED CENSUS MICRODATA FOR SOCIAL SCIENCE AND POLICY RESEARCH
International Statistical Institute 54th Session (Berlin 2003):
Invited Paper Meeting 38: Microdata – managing the dilemma between access, privacy, and confidentiality

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

Abstract. Census microdata are an invaluable resource for social science and policy research. Until recently National Statistical Institutes (NSI) permitted little use of these data. This paper describes the IPUMS-International project (www.ipums.org/international), a global collaboratory of NSIs to anonymize, harmonize and provide access on a restricted basis to extracts of integrated census microdata samples. Access is limited to bona fide 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 forty official census agencies have formally ratified the IPUMS-International protocols: Argentina, Austria, Belarus, Brazil, Bulgaria, Chile, China, Colombia, Costa Rica, Czech Republic, Dominican Republic, Ecuador, El Salvador, France, Germany, Ghana, Greece, Guatemala, Honduras, Hungary, Israel, Kenya, Madagascar, Mexico, Netherlands, Nicaragua, Palestinian Authority, Panama, Paraguay, Peru, Portugal, Puerto Rico, Romania, Slovenia, Spain, Tajikistan, the United States, Turkmenistan, United Kingdom , Venezuela, and Vietnam. National Statistical Institutes interested in additional information about the initiative are invited to contact Dr. Robert McCaa (rmccaa@umn.edu).

Introduction. Census microdata are an invaluable resource for social science and policy research. Other sources—such as demographic and labor force surveys—often offer greater subject coverage and detail than do census data, but no alternate source offers comparable sample density, chronological depth, and geographic coverage. This paper describes the IPUMS-International project, a global consortium to anonymize, harmonize and distribute high-density census microdata of a large number of countries. Custom-tailored extracts are delivered, at no charge, to bona fide researchers via the Internet.

For much of the world, census microdata are either wholly unavailable or rarely released, and are therefore seldom used (McCaa and Ruggles 2002). In the United States and Canada, however, census microdata have been available to researchers for almost forty years and have become an indispensable component of social science infrastructure. For example, census microdata were the data source for nineteen of the fifty-one U.S. and Canadian articles that appeared in the 2000 and 2001 volumes of the journal *Demography*. Even though the United States has abundant high-quality survey data and the most recent census samples were over a decade old, U.S. census microdata were used three times as often as the next most popular data source. By contrast, during the same two years not a single article in *Demography* made use of census microdata from Africa, Asia, Europe or Latin America.

IPUMS-USA. The Integrated Public Use Microdata Series (IPUMS-USA) is partly responsible for the widespread use of census microdata by social scientists studying the United States. IPUMS-USA, developed by Steven Ruggles, Matthew Sobek, and others at the Minnesota Population Center, makes census microdata freely available to scholars in harmonized format with comprehensive documentation through a user-friendly data access system (Ruggles and Sobek 1997; http://www.ipums.org/usa). Since its preliminary release in 1995, the IPUMS has become one of the most widely used demographic resources in the world. Over 6,000 researchers have registered to use the IPUMS data extraction system. The user base continues to expand rapidly, with approximately 2,500 new registered users per year. We are now distributing about 140 gigabytes of data per month, or an average of 190 megabytes per hour, twenty-four hours a day. We have prepared approximately 60,000 custom extracts of IPUMS data since May 1996 and are now processing approximately 2,800 data extract requests per month. This massive data distribution is beginning to bear fruit. Although the IPUMS has been available for only eight years, our bibliography lists more than twenty-six books, seventy-one dissertations, 207 published research articles, and hundreds of working papers, conference presentations, and research reports.

IPUMS-International. In 1998 we proposed to extend the IPUMS paradigm to the censuses of Colombia. This pilot project, a collaboration with the Colombian National Statistical Office (DANE), was designed to demonstrate the feasibility of creating public use microdata for Latin America. Shortly after we proposed the Colombia project, the National Science Foundation of the USA announced a special program for "Enhancing Infrastructure for the Social and Behavioral Sciences" that offered one-time funding for major new data improvement initiatives. We proposed a large-scale international project with two major components. The first step was to identify and preserve surviving machine-readable census microdata from around the world for the period 1960 to 2000. The second step was to select seven countries with broad geographical distribution and to clean, harmonize, document, and disseminate microdata for those countries using the same principles and methods that underlie the original IPUMS-USA database.

These two international projects, collectively known as IPUMS-International, have been an unqualified success. Both projects are now in their fourth year and are well ahead of schedule. We have created a comprehensive inventory of known microdata, much of which is described in our award-winning book, *Handbook of International Historical Microdata* (Hall, McCaa, and Thorvaldsen 2000), and we have preserved microdata from over one hundred censuses. In May 2002, we released our 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. We plan to release a second group of harmonized samples for Brazil in 2004. Over 60 million person records consisting of more than 50 variables are now available from the international web-site (http://www.ipums.org/international).

Some forty countries, encompassing more than 2.5 billion people, have now formally joined the IPUMS-International project (Table 1). This is thanks in part to the fact that there is increasing recognition that anonymized census microdata samples constitute statistical data. As such, they do not violate national laws on statistical confidentiality and privacy. This change in legal interpretation, coupled with both the recognition that stakeholders have a right to access to census data and the enormous advances in desktop computing power, has led to a breakthrough in making these valuable resources available for scientific and policy research. 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 case where the law is 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 (GDDS) of the International Monetary Fund is widely recognized as the gold standard in this regard. As of 2001, census microdata samples were disseminated by 37 of the 52 member states of the GDDS (McCaa and Ruggles 2002).

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 UN/ECE Population Activities Unit (PAU-Geneva), and the World Health Organization (Department of Health Service Provision, or OSD). Funding is now available for a five year project to harmonize census microdata of 16 countries in Latin America, and a proposal for 14 European countries is under consideration by a scientific funding agency of the ECE. Other regional initiatives are being developed as a sufficient number of NSIs ratify the project protocols. National Statistical Institutes not presently associated with the enterprise are invited to contact the International Project Coordinator, Dr. Robert McCaa at rmccaa@umn.edu .

If this project is successful it will continue beyond the 2000 round of censuses, incorporating census microdata of member countries for the 2010 round of censuses, as soon as they become available. For example, the 2000 census microdata of the USA were made available from the ipums.org/USA web-site within two months of the day of release by the United States Census Bureau.

Insert Table 1 near here

Confidentiality protections. **The IPUMS-International differs from IPUMS-USA in one important respect: statistical confidentiality protections.** IPUMS-International means Integrated <u>Restricted-Access</u>, 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 proposal IPUMS-Latin America.

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 database. These protections involve three elements—legal, administrative and technical:

1. dissemination agreements between the University of Minnesota and each NSI
2. user licenses between the University of Minnesota and each researcher
3. technical data 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 NSI. 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 NSIs suggest modifications. Any new provisions are forwarded to current members of the consortium for their consideration and up-dating as necessary.

The Minnesota Population Center and its authorized partners are obliged to share the integrated data and documentation with the national 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 national 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 national 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 registered 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, 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. Table 2 lists projects approved for access by subject matter, university or research organization, funding agency, and human subjects protection boards, from May 2002 through January 2003. 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 2 near here

Third are the technical measures taken to ensure statistical confidentiality. In cases where the NSI requests that the MPC apply anonymization procedures, we implement the following technical protections (based on Thorogood 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 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 identifiable census microdata truly correspond to a targeted individual (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.

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, not all of which have been implemented in the first release of the data. 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 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).

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. 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.

For example, in the IPUMS-International system marital status variable, the first digit is comparable across all samples. 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 3 illustrates the harmonization of codes for the variable "employment status".

Insert Table 3 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 national consultants it is possible to readily identify problematic concepts and revise the harmonized codes accordingly. It is important to understand that no decisions are made at the central integration center without comprehensive input by national experts who work as paid consultants to the project. This decentralized approach allows multiple projects to proceed 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.

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, National Statistical Institutes (NSIs), 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 national census agencies and other experts. License fees are paid to the NSIs 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 complemented 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 Institute.
0. Obtain funding by Minnesota Population Center to license data, reimburse in-country expenses, develop the database, and maintain the extract engine.
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-continents.
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.

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 from the original enumerations.

Data Dissemination (Extracts). 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 NSI. 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 using the supplied integrated metadata consisting of variable names and labels.

New Regional Initiatives. In mid-2003, a Latin American initative, including 16 Latin American countries with populations totaling one-half billion people, was begun with funding by the National Institutes of Health (Table 4). A European-wide project with the participation of fourteen countries (Table 5) is under consideration for funding by the European Union under the 6th Framework Program for Research Infrastructures. Other regional initiatives are also being organized. Officials of statistical agencies interested in discussing membership in the initiative should contact the International Projects Coordinator, Dr. Robert McCaa (rmccaa@umn.edu).

Insert Table 4 (Latin America) near here

Insert Table 5 (Europe) near here

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.

Résumé. Les microdonnés des recensements sont une inestimable ressource statistique pour la recherche en sciences sociales et politiques. Jusqu'à présent, les Institutes Nationaux de Statistique (NSI) ont limité l'usage de ces donnés. Cet article décrit le projet IPUMS-International, un consortium international des NSIs qui a pour but d'assurer l'anonymat, d'harmoniser et de distribuer des microdonnés intégrés des recensements à des chercheurs de confiance qui ont respecté les conditions d'usage et d'autorisation. Des formulaires de demande personnalisés sont délivrés gratuitement par Internet.

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	
World Region	**Oficial Statistical Authority**
Africa	Ghana, Kenya
Americas	Argentina, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Venezuela, USA.
Asia	China, Tajikistan, Turkmenistan, Vietnam
Europe	Austria, Belarus, Bulgaria, Czech Republic, France, Germany, Greece, Hungary, Netherlands, Portugal, Romania, Slovenia, Spain, the United Kingdom
Middle East	Israel, Palestinian Authority

Table 2. Report on Approved Access to Restricted Microdata, IPUMS-International, May 2002 – January 2003	
Funding Agencies	**Approved Projects (key words only)**
Canadian Foundation for Innovation	Brain drain: sending and receiving countries
Council for the Development of Social Science Research in Africa	Calibration of birth registrations against census microdata for countries with strong border migrations.
Economic and Social Research Council, UK	Comparison of fertility patterns by migration status
National Science Foundation	Construction of life-tables for sub-national populations.
National Institutes of Health	Cross national studies of poverty and social issues
Norwegian University Development Aid Funding	Cross-national analysis of human health resources
Rockefeller Foundation	Cross-national analysis of wage structure/discrimination
Wellcome Trust	Cross-national comparison of the determinants of poverty
Over-sight Boards	Cross-national determinants of female labor force
CNIL: Commission Nationale Information et Liberte	Cross-national study of inequality
Comite National d'Ethique	Cross-national study of living standards and sanitation
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.*	Demographic and spatial dimensions of homicide rates in relation to demographic changes.
Inter-University Consortium for Political and Social Research	Demographic processes: fertility, mortality, migration
IRD scientific commission (Conseil Scientifique)	Demographic profiles of older populations
ISA and its research committees RC28 and RC33	Develop regional accounts systems
National Committees for Research Ethics in Norway	Development of cross national social interaction and stratification scales.
USA Federal Code title 13/title 26 /title 5	Disability and welfare expenditures
Vice-decanat a la recherche, Universite de Montreal, Documents pour l'ethique	Education stock estimates for evaluating the efficiency of health systems
Professional Associations	Educational gaps between minority and majority populations
American Economic Association	Effects of AIDS on school enrollments
American Public Health Association	Effects of economic growth on demand for skills and education and the returns to labor.
American Sociological Association	Effects of educational mismatches on wages and salaries
International Union for the Scientific Study of Population (IUSSP)	Effects of national poverty programs on child labor and school attendance
Latin American and Caribbean Studies Association	Effects of social networks on rural-ruban migration.
Population Association of America	Effects of urbanization on internal migration
Universities/Research Organizations	Emigration: the gender gap
Europe	Emission of green house gases: population and labor
Cardiff University	Evolution of non-agricultural employment in rural areas
Demographic Studies Center - University Auton. of Barcelona	Extent of death clustering by regions
Department of Statistics, University of Florence	Gender differences in educational attainment
INED Paris	Gender earnings differences by ruralurban areas
Institut d etudes politiques de Paris	Household structures of the elderly
Institut francais de recherche en Afrique (IFRA)	Human welfare, agriculture and the environment

Ministry of Economic Development and Trade of Russian Federation	Inequality of wages: instruction of advanced graduate students on the use of census microdata
Novosibirsk State Technical University	Immigration of specific nationalities
University College London	Impact of climate variation on poverty
Canada	Infrastructure and economic activities on public health
Department of Demography, University of Montreal	Labor supply and regional development
Queen's University	Living arrangements of the elderly around the world
Simon Fraser University	Marriage transitions in developing countries
Statistics Canada -Library and information centre	Marriage, child labor, and polygamy
University of Toronto	Material inequality
USA	Migrants by country of origin/destination & duration
Boston University	Migration from Mexico to the USA
Brown University	Occupational changes and reshaping of industrial policies
Columbia University	Period-cohort analysis of educational attainment in comparative perspective
Dept. of Econ., Massachusetts Instit. of Technology	Recalibration of survey data using census microdata
East-West Center	Regional clustering of infant and child mortality
Florida State University	Religion and nationalism
George Mason University	School and work in developing and developed countries.
Georgetown Public Policy Institute	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 3. 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, familiy 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.

Table 4. Latin America census microdata access project:
density (%) of source microdata by country and decade of census

Country	Millions	1960s	1970s	1980s	1990s	2000s
Argentina	37.0	**3**	**2**	**2**	**100**	**100**
Bolivia	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**

Note: "n.a." indicates a census was taken but microdata are not known to exist; "." Indicates no national census was taken in this decade.

Table 5. Europe: Microdata by Census Year (bold) and Country
"signed" IPUMS-International agreement as of 17 July 2003

Country	Millions	1960s	1970s	1980s	1990s	2000s
Albania	3.4	1960, 69	1979	1989		**2001**
Austria (signed)	8.1	1961	**1971**	**1981**	**1991**	**2001**
Belarus (signed)	10.0			**1989**	**1999**	…
Belgium	10.3	1961	1970	1981	**1991**	**2001**
Bosnia and Herzegovina	3.4				1991	**2001**
Bulgaria (signed)	8.1	1965	1975	1985	**1992**	**2001**
Croatia	4.7				1991	**2001**
Czech Republic (signed)	10.3	1961	1970	1980	**1991**	**2001**
Denmark	5.4	1960, 65	1970, 76	1981	1991	**2001**
Estonia	1.4			**1989**		**2000**
Finland	5.2	**1960**	**1970, 75**	**1980, 85**	**1995, 90**	**2000**
France (signed)	59.2	**1968, 62**	**1975**	**1982**	**1990**	**1999**
Germany (signed)	82.2	1961	**1970**	**1987**	micro	micro
Greece (signed)	10.9	1961	**1971**	**1981**	**1991**	**2001**
Hungary (signed)	10.0		**1970**	**1980**	**1990**	**2001**
Iceland	0.3	1960		**1980**		**2001**
Ireland	3.8	1961, 66	1971, 76	**1981, 86**	**1996, 91**	**2002**
Israel (signed)	6.4	**1967, 61**	**1972**	**1983**	**1995**	…
Italy	57.8	1961	1971	**1981**	**1991**	**2001**
Latvia	2.4			**1989**		**2000**
Liechtenstein	0.0	1960	1970	1980	1990	**...**
Lithuania	3.7			**1989**		**2001**
Luxembourg	0.4	1960, 66	1970	1981	1991	**2001**
FYR Macedonia	2.0				1994, 91	**2001**
Malta	0.4	1967		1985	1995	
Moldova, Republic	4.3			**1989**		**2003**
Netherlands	16.0	**1960**	**1971**		**1991**	**2003**
Norway	4.5	**1960**	**1970**	**1980**	**1990**	**2001**
Poland	38.6	1960	1970, **78**	**1988**		**2002**
Portugal (signed)	10.0	1960	1970	**1981**	**1991**	**2001**
Romania (signed)	22.4	1965	1977		**1992**	**2002**
Russia	144.4	1970	1979	**1989**	**1994**	**2002**
San Marino	0.0		1976			
Slovakia	5.4				1991	**2001**
Slovenia (signed)	2.0			**1981**	**1991**	**2002**
Spain (signed)	39.8	1960	1970	**1981**	**1991**	**2001**
Sweden	8.9	1960, 65	1970, 75	1980, 85	**1990**	...
Switzerland	7.2	1960	**1970**	**1980**	**1990**	**2000**
Turkey	66.3	1960	**1970, 75**	**1980, 85**	**1990**	micro
Ukraine	49.1			**1989**		**2001**
United Kingdom (signed)	60.0	**1961**	**1971**	**1981**	**1991**	**2001**
Yugoslavia	10.7	1961	1971	1981	1991	**2001**
Total extant microdatasets	799.4	8	13	29	25	33
Total sets in project	**349.0**	**3**	**6**	**10**	**14**	**14**

Appendix A Memorandum of Agreement.

Integrated Public Use Microdata Series International
and [National Statistical Agency of Country X]

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

1.   Ownership. The **[National 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 **[National 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 **[National 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 **[National 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 **[National 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 [National 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 **[National 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: _____

Rev. Aug. 1, 2003

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:

    Street Address 2:

    City, State/Province, Zip:

    Country:

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

E-mail address:

Field:	*Status:*
○ Demography	○ Faculty
○ Economics	○ Academic researcher
○ History	○ Support staff
○ Sociology	○ Student
○ Other academic	○ Non-Academic Researcher
○ Public Policy	

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 of countries):

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