

IPUMS-International: Changing Patterns Of Use And Analysis Of National Population Census Microdata And Metadata, 1999-2009

www.ipums.org/international

Robert McCaa, Steven Ruggles, and Matt Sobek

University of Minnesota Population Center

The IPUMS-International projects receive major funding from the National Science Foundation of the United States, grants sbr-9908380 and ses-0433654.

I am conducting research on ■■■ emigration. I need data on ■■■ emigrants, such as income, education, age, and gender to compare earnings of emigrants and the earnings they would have gained had they not emigrated. In this way I will test the hypothesis that there is negative selection of emigrants, which means that those with below-average skills and education are more likely to migrate.

--project description, application #2422, www.ipums.org/international

Introduction. Census *microdata* consist of records of persons, households and dwellings as recorded in quinquennial, or more commonly decennial, enumerations of national populations. Census *metadata* are the documentation which describe each census: forms, manuals, data dictionaries, codebooks, technical reports, etc. Beginning with the 1960 round of population censuses, statistical offices began to disseminate anonymized samples of census microdata, but access was often limited to researchers resident in the country. Over the decades the number of statistical offices offering census samples—and their availability world-wide—increased, so that with the 2000 round, almost all statistical offices plan to disseminate such samples. What has not changed is the metadata. These are typically limited to a data dictionary to describe the record structure as well as codes and labels of the microdata. As far as I am aware, no statistical agency has attempted to re-write or restructure metadata for historical census microdata to attain standards of the 21st century.

In 1999, the IPUMS-International initiative began a global effort to preserve and integrate census microdata and metadata cross-nationally and chronologically. The

IPUMS project traces its beginnings to the 19th ANCSDAAP conference held in Beijing in April 2000. The Central Statistics Office of Viet Nam and the National Bureau of Statistics of China were founding partners of the IPUMS collaboratory, joining in 2001 and 2002, respectively. Over the past ten years eighty-three National Statistical Offices (NSO) have endorsed the IPUMS project Memorandum of Understanding with the University of Minnesota Population Center (MPC). Seventy-eight—encompassing four-fifths of the world's population—have entrusted microdata to the MPC for a total of 241 censuses, verified and documented. Integrated microdata and metadata is available on a restricted access, no-cost basis for 43 countries (130 samples, 77.5 million households, and 279,464,935 person records). 2,484 researchers representing 61 countries are accredited for access to the data.

The Asia and Pacific region is represented in the IPUMS project by 12 countries, 22 samples and 55.8 million person records (see Tables 1 and 2 for details). The three countries with the largest contributions are China (21.9 million person records in 2 samples), Philippines (20.3 in 3) and Viet Nam (5.0 in 2). 10% household samples are the norm, accounting for 69 of 130 in the IPUMS database. For Asia and Pacific region, integrated 10% samples are available for Armenia, Cambodia, Jordan, Kyrgyz Republic, Mongolia, Palestine, and The Philippines. In addition, the MPC is authorized to integrate 10% samples for Bangladesh, Indonesia, Nepal, and Pakistan into the IPUMS-International database, once the tasks of translating documentation and preparing anonymized, integrated microdata samples are completed.

Over the next ten years, using the microdata and metadata entrusted to the MPC, as well as more enroute, the IPUMS-International project will more than double—thanks to

the cooperation of national statistical offices, funding of the National Science Foundation and the National Institutes of Health (USA), and the avid interest of researchers and policy makers world-wide. This paper reports on accomplishments to date with respect to the integration and dissemination of internationally integrated microdata and metadata, including a brief description of users and uses of the IPUMS database.

Statistical offices that are not yet participating in IPUMS are respectfully invited to consider doing so.

A. Why census microdata? Census *microdata* provide information about individual persons, families, households, and dwellings, usually in the form of one or more records per case, each consisting of a series of variables. Typical census microdata for persons include age, sex, marital status, family relationship, place of birth, educational attainment, employment status, etc. Microdata are exceedingly useful because they allow researchers to interrelate any desired set of population and housing characteristics (Dale, Fieldhouse and Holsworth, 2000). The flexibility offered by microdata is essential for comparative research because aggregate tabulations are often not comparable across time or between countries. In the few countries where census microdata covering multiple census years have been easily available to researchers, these data are the most widely-used source for the study of large-scale economic and demographic transformations (McCaa and Ruggles, 2002).

B. Why IPUMS? IPUMS adds value to census microdata in five important ways: data recovery, microdata integration, metadata integration, confidentiality protections and dissemination. IPUMS provides significant benefits both to researchers and to NSOs—so much so that the National Science Foundation and the National Institutes of Health USA

are providing sustained, major funding to the IPUMS initiatives now extended to a period of fifteen years. The likelihood of continued renewals is exceedingly high because peer reviewers typically rank IPUMS proposals for funding at the top. With the completion of the IPUMS-Europe project (2004-2009), a new five year effort is currently under review, bringing together some 30 National Statistics Offices of Europe and Asia (IPUMS-EurAsia) into a combined effort. Consider the five ways that IPUMS adds value to census microdata.

1. Microdata and documentation recovery, archiving and preservation. Until recently few national statistical offices assiduously archived census microdata and documentation. Processing of a census slowly winds down, as there is the need for revised or additional tables declines. Over the past decade, the transition from main-frame to mini-computers meant that many statistical offices no longer possess functioning hardware to migrate microdata and documentation to the new computing environment. Data recovery now is critical because as the old data tapes and cartridges deteriorate with age, the likelihood of successful recovery declines. Scarce human and material resources means that preservation is a low priority for many statistical agencies.

The IPUMS-International project has funded data recovery initiatives in Bangladesh, Fiji Islands, Indonesia, Israel, Mongolia, Nepal, Pakistan, Philippines, and Viet Nam. In some cases, the project has repatriated microdata that no longer existed within the national statistical office. In others, we have assisted in assembling complete documentation and shipping old data tapes to a commercial data recovery company. The most costly recovery is that of Bangladesh, where more than 7,000 tapes are being recovered with equipment and expertise provided by the IPUMS project. Meanwhile the

microdata of the 1981 census is almost entirely recovered, while work continues on. Where the original enumeration forms survive, but the microdata have been lost, we assist with the re-keying, cleaning, and rehabilitation of nationally representative samples. In Europe, our most ambitious project is with the Federal Statistical Office of Germany, where the microdata and documentation have been recovered, verified and archived for the 1971 and 1981 censuses of the German Democratic Republic. By 2012 it is expected that an entire integrated series of 8 nationally representative samples will be made available to researchers for the first time.

2. *Microdata Integration.* The principal benefit of IPUMS to researchers and NSOs alike is integration—integration of both microdata and metadata. For decades, many NSOs have provided census samples for academic and policy research, but few have ever integrated even two samples, much less a complete series. At best, the more modern statistical offices construct a census sample and a data dictionary for researchers, as soon as the final data cleaning is complete. Five or ten years later, with the ensuing census, the process is repeated. Few statistical offices re-examine earlier samples or documentation to harmonize successive datasets or draft documentation to facilitate comparative analysis of two or more censuses.

The project does not disseminate census files entrusted by NSOs. Instead high-precision census samples are anonymized (McCaa et. al. 2006) and integrated, variable-by-variable, using a composite coding system (Esteve and Sobek, 2003). Samples are integrated both cross-nationally and chronologically. Integrated metadata are constructed not only by means of meticulous study of comprehensive original source documentation but also after extensive analysis of the microdata. Thousands of hours are devoted at the

Minnesota Population Center to analyze, discuss, debate, draft, test and re-test until the microdata integration is validated for dissemination to researchers. The entire process is repeated each year as 15-20 new samples are added to the database.

To obtain an outside evaluation of the integration process performed in Minnesota, the National Statistical Office of Argentina (INDEC) was contracted to conduct an exhaustive analysis of the integration of samples of the Argentine censuses of 1970, 1980, 1991 and 2001. The INDEC team compared the frequencies for each variable and code against the original microdata and metadata entrusted to the MPC. From the tens of thousands of words and codes of metadata, barely a handful of minor errors, misinterpretations or misunderstandings were discovered. This important outside evaluation—accomplished on-site in INDEC’s Buenos Aires offices without the presence of IPUMS personnel—attests to the trustworthiness of IPUMS integrations. What INDEC did can be done from the IPUMS website by any statistical agency working in the convenience of their own offices. The IPUMS database provides tools for the expert user to cross-check every integration decision made by the IPUMS team so that little doubt remains about the significance, quality, or transparency of the work.

DANE-Colombia, the first statistical agency to participate in the IPUMS collaboratory, is using the five integrated Colombian census samples (12.3 million person records) to construct a nationally integrated dataset with metadata in the Spanish language. With IPUMS assistance, DANE is simplifying internationally integrated datasets to a national system. It is more efficient to simplify an international integration. If national integration is performed first, it may unwittingly sacrifice important details

that are considered trivial at the national level but that are exceedingly important for international integration.

As an example, consider the marital statuses of “single” and “never-married”. Few censuses distinguish between these categories. A census typically has one category or the other—rarely both. In recent decades, as the United Nations Statistics Division Principles and Recommendations have evolved, the more precise category (“never-married”) has become more common. The IPUMS system retains all significant distinctions in the original microdata. National integrators will more readily overlook these and other seemingly trivial distinctions. Successful international integration must document these distinctions so that all researchers may readily be informed of these and thousands of other details.

3. Metadata integration. Metadata integration is essential if microdata integration is to succeed. By integrating metadata, researchers are relieved of the task of studying every census for changes or deviations in concepts and definitions. Instead, experts carefully consider all the documents and analyze the microdata to be able to write new, comprehensive documentation that spells out common practices and discusses significant differences and discrepancies. Researchers then can study the integrated metadata confident that most attention can be directed to the issues of greatest salience for the research issues at hand.

The IPUMS eXtensible Language Markup tool facilitates navigation of both source and integrated metadata in any way desired, by means of a few clicks. For example, to compare the wording of the employment status variable, select the countries and census years desired, then click “employment status”, and “enumeration text”. This allows the

researcher to compare the precise wording, in English, of the question on the form as well as the instructions to the enumerators for each selected census.

a. Censuses and samples. IPUMS metadata offer detailed descriptions of each census in the database, listing for each the title, year, universe, de jure/de facto, enumeration unit, census day, forms, field work period and type, respondent and estimates of undercount, if any. Images of census enumeration forms and instructions manuals are available in the official language and English translation. Each sample is described with regard to source, sample design, sampling unit, sample fraction, number of person records, sample weights, dwelling or housing units, vacant dwellings, households, group quarters and special populations.

b. Variable descriptions, source texts, and codes. IPUMS metadata define each integrated variable and describe basic characteristics: censuses for which the variable is available, universe to which the variable or question applies, codes available, source (enumeration) text, and the non-harmonized variables from which the integrated equivalents are derived. Clicking a hypertext word (“universe”) or phrase (“enumeration text”) provides a window with additional information. A general comparability discussion is provided for every variable, with country or census specific discussions focusing on specific departures from usual practices. The purpose of these discussions is to highlight important contrasts. Click “enumeration text” to read the question as it appeared in each selected census as well as the associated instructions to field workers. To view full original sources, click “image” to view image or “text” to view “text”.

Coding of the microdata is readily examined by clicking “codes” (see figure #). “X” indicates that a code exists in the microdata. Click “case count view” to view the number

of records in the full dataset with the corresponding code. Click “detailed codes” to view the complete, multi-digit composite codes available for the selected variable. This table is handy so that researchers may plan recodes for any selected variable according to specific analytical needs.

4. *Statistical confidentiality.* In 2006, the Conference of European Statisticians (CES) published guidelines for Managing Statistical Confidentiality and Microdata Access. An IPUMS-Europe case study, using the specific case of France, is appended to the CES report as an example of good practice. The case study describes how IPUMS manages access to microdata, explains why it is a good practice, identifies the target audience, explains confidentiality measures, specifies the rules and procedures regarding user access, summarizes supporting legislation, and lists strengths and weaknesses as well as bibliographical references. While the IPUMS case study is European in scope, the details are identical for the International project. For the complete study, see: <http://www.unece.org/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf> In this on-line version, the IPUMS case study begins on page 98.

In cases where fully anonymized samples are entrusted to the project, no further statistical confidentiality measures are imposed. In many cases, full information contained in the census is provided to the project, including detail sufficient to pose a theoretical risk of re-identification. To minimize risks, statistical confidentiality edits are performed by the IPUMS project. The lowest level of geography to be released is identified (e.g., for European countries, typically NUTS3) and suppress all finer geographic variables. Other variables identified as sensitive are also suppressed. Any technical variables that could be used to identify records within the original data are also

suppressed. Very small population categories for specific variables are recoded into larger groups (for example, grouping rare occupations with more common pursuits), and top- or bottom-coding some variables (for example, income). Finally, the sequence of dwellings within the smallest geographic unit identified in the data is randomized, so geography cannot be inferred. An undisclosed fraction of cases is randomly swapped across geographic districts to add uncertainty about the origin of any particular record. Finally, a new serial number is generated to reflect the final ordering of the file.

5. Microdata access. Researchers must first be approved before access to any microdata is permitted. Moreover users are never permitted access to the original source files provided by the NSOs. Instead, data are provided in the form of extracts, custom tailored to each researcher's needs. What this means is that there is no distribution of entire datasets by means of compact discs. Since each dataset is custom tailored "collecting" or "boot-legging" datasets is not only illegal, but effectively curtailed.

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

Once the selections are complete, there is an opportunity to review or revise before final submission of the request. Then, once submitted, the extract engine registers the

request and places it in a data processing queue. When the extract is ready (usually in a matter of minutes), the researcher is notified by email that the data should be retrieved within 72 hours. A link is provided to a password-protected page for downloading the specific extract. Soon an SSL (Secure Sockets Layer) protocol will be implemented at the Minnesota Population Center. After SSL is in place, the data will be encrypted during transmission using a 128-bit encryption standard, matching the level used today by the banking and other industries where security and confidentiality is essential. The researcher may then securely download the file, decompress it and proceed with the analysis using the supplied integrated metadata consisting of variable names and labels. The metadata are in ASCII format so that a researcher may readily use them with any statistical software.

Dennis Trewin, the former Australian Statistician and the chair of the CES report on Statistical Confidentiality and Microdata Access, visited the Minnesota Population Center, Oct. 30-Nov. 4, 2007, to conduct an on-site assessment of microdata security and access management. His detailed report concludes: “Without question IPUMS International meets the four Core Principles outlined in CES [Conference of European Statisticians] (2007). It is cited in CES (2007) as a Case Study of good practice. This review confirms its status as good practice for Data Repositories. Indeed it is likely to provide the best practice for a Data Repository for international statistical data. ... The security of the computing environment used by IPUMS-International is first class and appears to be of the standard of the best statistical offices.”

C. Whither IPUMS-EurAsia?

1. *IPUMS users.* With these value-added and others not discussed here it is not surprising that IPUMS-International is the favored microdata repository for both National Statistical Offices and researchers alike. The number of accredited researchers has reached 2,482, representing 69 countries and entities. 90% of users are university-based researchers. The user list includes representatives of many national statistical offices and international statistical agencies such as the World Health Organization, the International Labour Office, and the World Bank. Economists account for 45.7% of IPUMS users, followed by demographers (19%), sociologists (10%), public policy analysts (7%), statisticians (2.7%) and historians (2.4%).

As the epigraph indicates, microdata are essential for cross-national and cross-temporal comparisons. For example, a National Academy of Sciences volume, *Growing Up Global: Transitions to Adulthood in Developing Countries*, used IPUMS-International data from multiple countries to analyze changing outcomes such as attending school, working, childbearing, and marrying as a function of age, gender, and household characteristics. Other important strands of research have sought to understand the global returns to education; assess trends in Intergenerational coresidence in developing countries; and explore the relationship between fertility transition and the family life of children.

Regrettably only sixteen percent of IPUMS users reside in the Asia and Pacific region, with China topping the regional rankings (n=37), followed by Japan (25), Australia (24), Singapore (10), India (8), Philippines (7), and Hong Kong SAR (6). Three researchers have registered from each of the following--Republic of Korea, Malaysia, New Zealand, Thailand, and Vietnam. It is remarkable that of the seven top

user countries in the Asia-Pacific region, five have no microdata in the IPUMS database: Japan, Australia, Singapore, India and Hong Kong SAR. A reasonable question might be: What additional insights might these researchers gain, if integrated census samples of their own countries were also available in the IPUMS-database? It is hoped that, with the addition of samples for India this year and samples for more Asian countries in future years, the IPUMS database will attract more researchers from the Asian and Pacific region.

2. Looking ahead. 2009-2014. A new value-added, a product of the IPUMS-EurAsia project, will be the construction of variables necessary to exploit variance estimation software. This work will capitalize on IPUMS methods for imputing pseudo-strata for systematic random samples (Davern, et. al., forthcoming). The new variables will allow users to make use of Taylor-series linearization procedures now available in the leading statistical packages to simplify variance estimation for complex samples. IPUMS-EurAsia will also develop user-friendly documentation and recommendations for variance estimation using each statistical package, with examples of typical analyses. Because statistical tests are at the core of quantitative research in the social sciences, reliable variance estimation is vital; otherwise, the cumulative process of scientific research rests on a foundation of unstable inferences. Tools developed by the IPUMS-EurAsia project to help researchers perform valid statistical tests based on accurate variance estimates will therefore represent a significant contribution to the global scientific community.

Looking ahead, the biggest challenge for the IPUMS collaboratory will be the integration of 2010 round census microdata samples: to zealously protect statistical

confidentiality, attain the highest standards of integration, and manage access to extracts of samples by researchers around the globe in a timely way and at no cost—all accomplished with a minimum of delay.

Conclusion. Now that the construction of anonymized microdata data samples is becoming an increasingly widespread practice in Asia and the Pacific Region, integration of census microdata is an important next step to enhance use. With the emergence of global standards for harmonizing census data and the massive power of ordinary desktop computers, the major challenge that remains is the actual construction of integrated census microdata samples. Thanks to the cooperation of more than 80 official census agencies worldwide and with the financial support of the National Science Foundation and the National Institutes of Health, the IPUMS-International project is committed to integrating microdata for 200 censuses by 2015. The numbers of users and uses are likely to increase by an order of magnitude as IPUMS becomes the most widely used demographic database in the world.

References.

- Conference of European Statisticians. (2007). *Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines on Good Practice*. <http://www.unece.org/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf> Geneva: United Nations Economic Commission for Europe.
- Dale, A., Fieldhouse, E. and Holdsworth, C. (2000). *Analyzing census microdata*. Arnold: London.
- Davern, M.E., S. Ruggles, T. Swenson, and J.M. Oakes. Forthcoming. *Drawing Statistical Inferences from Historical Census Data*. Demography.
- Esteve, Albert and Matthew Sobek. (2003). Challenges and Methods of Census Harmonization. *Historical Methods* 36: 66-79.
- 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.
- McCaa, Robert, Steven Ruggles, Michael Davern, Tami Swenson, Krishna Mohan Palipudi. (2006). "IPUMS-International High Precision

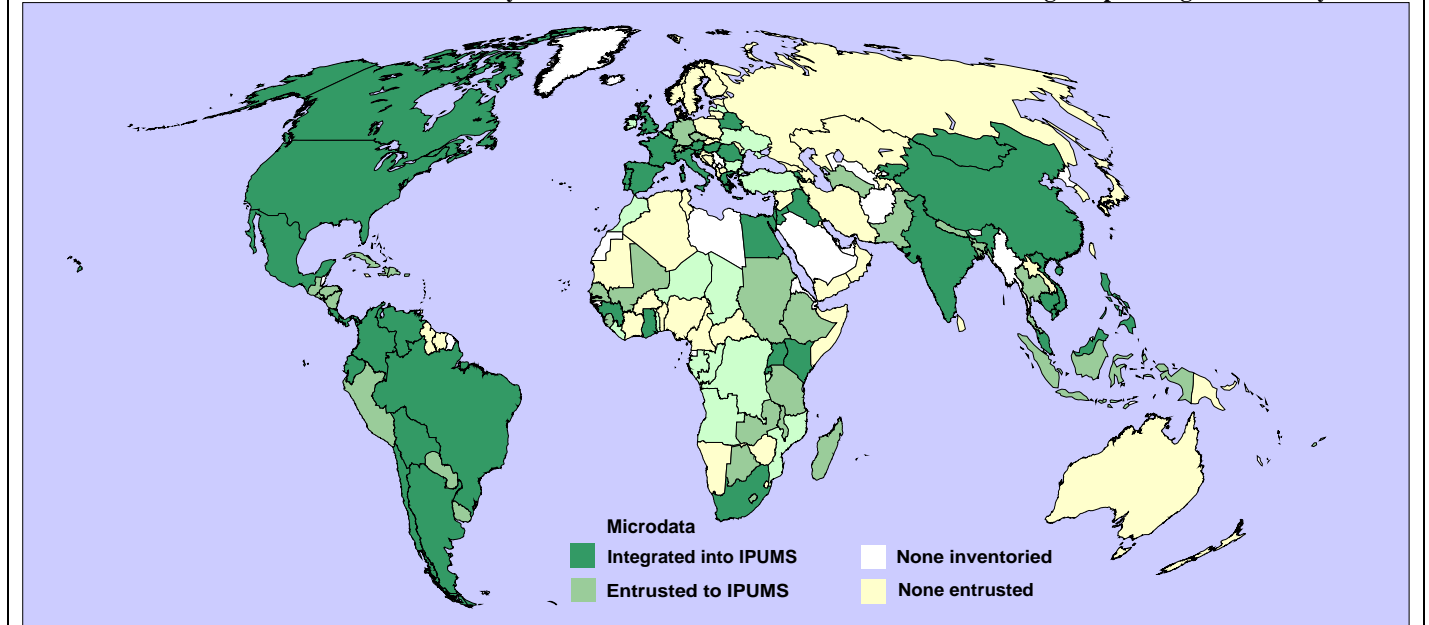
Population Census Microdata Samples: Balancing the Privacy-Quality Tradeoff by Means of Restricted Access Extracts,” Privacy in Statistical Databases. Berlin: Springer, pp. 375-382.

Ruggles, Steven, Miriam King, Deborah Levison, Robert McCaa, and Matthew Sobek.

(2003). “IPUMS-International: An Overview”. *Historical Methods*, 36: 60-65.

Trewin, Dennis. (2007). A review of IPUMS-International. Unpublished.

Table 1. IPUMS-International: Country list of 249 census microdatasets entrusted indicating sample design and density



Sample density			Country	Sample design	Census decade				
10%	~5%	<=4%			2000s	1990s	1980s	1970s	1960s
<i>Anonymized, integrated and disseminating 2002-2009: 130 censuses, 77 million households, 279 million person records (44 countries)</i>									
4			Argentina	IPUMS	2001	1991	1980	1970	1960
1			Armenia	IPUMS	2001		1989	1979	1970
4			Austria	IPUMS	2001	1991	1981	1971	1961
1			Belarus	IPUMS		1999	1989	1979	1970
3			*Bolivia	IPUMS	2001	1992		1976	
5			Brazil	IBGE (clustered)	2001	1991	1980	1970	1960p
1			Cambodia	IPUMS		1998			1962
		4	Canada	STATSCAN	2001	1991-6	1981-6	1971-6	1961, 6
4		1	*Chile	IPUMS	2002	1992	1982	1970	1960p
		2	China	NBS	2000	1990	1982		1964
3		2	*Colombia	DANE/IPUMS	2005,6	1993	1985	1973	1964p
3	1		*Costa Rica	IPUMS	2000		1984	1973	1963
4		1	*Ecuador	IPUMS	2001	1990	1982	1974	1962p
2			Egypt	CAPMAS/IPUMS	2006	1996	1986	1976	1964
	6		France	IPUMS	1999	1990	1982	1975	1968, 2
2			*Ghana	IPUMS	2000		1984	1970	
4			Greece	IPUMS	2001	1991	1981	1971	1961
2			*Guinea, Conakry	IPUMS		1996	1983		1960
	4		Hungary	IPUMS	2001	1990	1980	1970	
		5	India (microcensuses)	NSSO	2005m	1993,9m	1983,7m		
1			*Iraq	IPUMS		1997	1987	1977	1967
3			Israel	IPUMS		1995	1983	1972	1961,7
	1		Italy	IPUMS	2001	1991	1981	1971	1961
1			Jordan	IPUMS	2004	1994	1979		
	3		Kenya	IPUMS	1999	1989	1979	1969	
1			Kyrgyz Republic	IPUMS		1999			

		4	Malaysia	IPUMS	2000	1991	1980	1970	1960
3		3	Mexico	INEGI/IPUMS	2000,5	1990,5	1980	1970	1960p
2			*Mongolia	IPUMS	2000		1989	1979	1970
		3	Netherlands	CBS/IPUMS	2001m			1971	1960
1			Palestine	CBS		1997			
5			*Panama	IPUMS	2000	1990	1980	1970	1960
3			*Philippines	IPUMS	2000	1990	1980	1970	1960p
	3		Portugal	IPUMS	2001	1991	1981	1970	1960
3			Romania	IPUMS	2001	1992		1977	1965
2			*Rwanda	IPUMS	2002	1991			
1			Slovenia	IPUMS	2001	1991	1981		
3			South Africa	StatsSA	2001,7m	1996-1	1985-0	1970	1960
	3		Spain	IPUMS	2001	1991	1981	1970	1960
2			*Uganda	IPUMS	2002	1991	1980		1969
		2	United Kingdom	ONS	2001p	1991	1981	1971	1966,1
	6		United States	USCB	2000,5	1990	1980	1970	1960
4			*Venezuela	IPUMS	2001	1990	1981	1971	1961
	2		Vietnam	IPUMS		1999	1989	1979	
<i>Asia and Oceania (21 countries, 51 censuses—including samples for 13 countries noted above)</i>									
2			*Bangladesh	IPUMS	2001	1991	1981	1974	1961
3			*Fiji Islands	IPUMS	2007	1996	1986	1976	1966
7			Indonesia (translate documents)	IPUMS	2000	1990	1980	1971	1961
			Korea, Republic of (negotiating)	KNSO	2000,5	1990,5	1980,5	1970,5	1960,6
1			Nepal	CBS	2001	1991?	1981?	1971	1961
3			*Pakistan	SD/IPUMS		1998	1981	1973	1961
		4	Thailand (translate documents)	NSO	2000	1990	1980	1970	1960
1			Turkmenistan	IPUMS		1995	1989	1979	1970
<i>Europe (22 countries, 76 censuses—including samples for 11 countries noted above)</i>									
			Bulgaria	-	2001	1992	1985	1975	1965
	2		Czech Republic	IPUMS	2001	1991	1980	1970	1961
1			Germany	IPUMS/FSO	2001m	1991m	1981-7	1970-1	1961
			Ireland (negotiating)	-	2001	1991	1981	1971	
			Latvia (negotiating)	-	2000		1989	1979	
			Poland (negotiating)	-	2001		1988	1970-8	1960
			Russia (negotiating)	-	2002		1989	1979	1970
	4		Switzerland	IPUMS	2000	1990	1980	1970	1960
			Turkey (in process)	TurkSTAT	2000	1990	1980-5	1970-5	1960, 5
1			Ukraine (in process)	IPUMS	2001		1989	1979	1970
<i>North America and the Caribbean (14 countries, 45 censuses—including samples for 5 countries noted above)</i>									
1			Cuba	IPUMS	2002		1981	1970	
1	1	2	*Dominican Republic	IPUMS	2003	1993	1981	1970	1960p
1			*El Salvador	IPUMS	2007	1992		1971	1961
2		3	*Guatemala	IPUMS	2002	1994	1981	1973	1964
2			*Haiti	IPUMS	2003		1982	1971	
3		1	*Honduras	IPUMS	2000		1988	1974	1961
2		1	*Nicaragua	IPUMS	2005	1995		1971	1963
	4		Puerto Rico	USCB	2000	1990	1980	1970	1960
2			Saint Lucia	IPUMS	2001	1991	1980	1970	1960

<i>South America (9 countries, 39 censuses—including samples for 5 countries noted above)</i>									
4		1	*Paraguay	IPUMS	2002	1992	1982	1972	1962
1			*Peru	IPUMS	2007	1993	1981?	1972	1961
4			*Uruguay	IPUMS		1996	1985	1975	1963
<i>Africa (22 countries, 49 censuses—including samples for 6 countries noted above)</i>									
3			*Botswana	IPUMS	2001	1991	1981	1971	1964
2			*Ethiopia	IPUMS	2007	1994	1984		
			Lesotho (in process)	IPUMS	2006	1996	1986	1976	1966
1			*Madagascar	IPUMS		1993			
2			*Malawi	IPUMS	2008	1997	1987	1977	1967
3			*Mali	IPUMS		1998	1987	1976	
2			*Mauritius	IPUMS	2000	1990	1983	1972	1962
			Morocco (in process)	IPUMS	2004	1994	1982	1970	1960
			Mozambique (in process)	IPUMS	2007	1997	1980?		
3			*Senegal	IPUMS	2002		1988	1976	
1			*Sierra Leone	IPUMS	2004		1985?	1974	1963
3			*Sudan	IPUMS	2008	1993	1983	1973	
2			*Tanzania	NBS/IPUMS	2002		1988	1978	1967
2			*Zambia	IPUMS	2000	1990	1980	1969	1963

Note: **bold country** = Memorandum of Understanding signed with Regents of the University of Minnesota;
 IPUMS = systematic household sample(s) drawn to IPUMS specifications: every nth household stratified by enumeration district.
 Year = census conducted; **bold year** = microdata survive; * = **100% microdata entrusted**, where extant; m = microcensus; p = person sample

**Table 2. IPUMS-International Integrated Samples:
 44 countries, 130 censuses, 279 million person records
 Note: Samples in bold will become available June 2009**

Census	Sample Fraction (%)	Households	Persons	Weighted
Argentina 1970	2	129,728	466,892	–
Argentina 1980	10	672,062	2,667,714	yes
Argentina 1991	10	1,148,351	4,143,727	yes
Argentina 2001	10	1,040,852	3,626,103	–
Armenia 2001	10	81,929	326,650	–
Austria 1971	10	264,655	749,894	–
Austria 1981	10	283,693	756,556	–
Austria 1991	10	310,099	780,512	–
Austria 2001	10	341,035	803,471	–
Belarus 1999	10	385,508	990,706	–
Bolivia 1976	10	121,378	461,699	–
Bolivia 1992	10	177,926	642,368	–
Bolivia 2001	10	239,475	827,692	–
Brazil 1960	5	613,273	3,001,439	–
Brazil 1970	5	1,022,207	4,953,759	yes
Brazil 1980	5	1,343,377	5,870,467	yes
Brazil 1991	5.8	2,012,276	8,522,740	yes
Brazil 2000	6	2,652,356	10,136,022	yes
Cambodia 1998	10	223,513	1,141,254	–

Canada 1971	1	n.a.*	214,019	–
Canada 1981	2	n.a.*	486,875	–
Canada 1991	3	n.a.*	809,654	–
Canada 2001	2.5	n.a.*	801,055	yes
Chile 1960	1	n.a.*	88,184	–
Chile 1970	10	199,041	890,481	–
Chile 1982	10	282,356	1,133,062	–
Chile 1992	10	373,964	1,335,055	–
Chile 2002	10	486,115	1,513,914	–
China 1982	1	2,428,658	10,039,191	–
China 1990	1	3,152,818	11,835,947	–
Colombia 1964	2	n.a.*	349,652	–
Colombia 1973	10	349,853	1,988,831	–
Colombia 1985	10	571,046	2,643,125	yes
Colombia 1993	10	774,321	3,213,657	–
Colombia 2005	10	1,054,901	4,117,607	yes
Costa Rica 1963	5	n.a.*	82,345	–
Costa Rica 1973	10	36,323	186,762	–
Costa Rica 1984	10	56,186	241,220	–
Costa Rica 2000	10	106,973	381,500	–
Ecuador 1962	3	n.a.*	136,443	–
Ecuador 1974	10	145,902	648,678	yes
Ecuador 1982	10	195,401	806,834	–
Ecuador 1990	10	243,898	966,234	–
Ecuador 2001	10	354,222	1,213,725	–
Egypt 1996	10	1,270,787	5,902,243	yes
France 1962	5	748,917	2,320,901	–
France 1968	5	815,699	2,487,778	–
France 1975	5	915,624	2,629,456	–
France 1982	5	969,632	2,631,713	–
France 1990	4.2	949,893	2,360,854	–
France 1999	5	1,219,323	2,934,758	–
Ghana 2000	10	397,097	1,894,133	–
Greece 1971	10	249,350	845,483	–
Greece 1981	10	294,323	923,108	–
Greece 1991	10	320,387	951,875	–
Greece 2001	10	367,438	1,028,884	–
Guinea (Conakry) 1983	10	110,777	457,837	–
Guinea (Conakry) 1996	10	108,793	729,071	–
Hungary 1970	5	172,831	515,119	–
Hungary 1980	5	211,355	536,007	–
Hungary 1990	5	219,389	518,240	–
Hungary 2001	5	227,252	510,502	–
India 1983	.1	120,847	623,494	yes
India 1987	.1	129,060	667,848	yes

India 1993	.1	115,409	564,740	yes
India 1997	.1	120,578	596,688	yes
Iraq 1997	10	265,402	1,944,278	–
Israel 1972	10	89,190	315,608	–
Israel 1983	10	124,610	403,474	–
Israel 1995	10	177,412	556,365	–
Italy 2001	5	1,168,044	2,990,739	–
Jordan 2004	10	97,286	510,646	–
Kenya 1989	5	224,861	1,074,098	–
Kenya 1999	5	317,106	1,407,547	–
Kyrgyz Republic 1999	10	110,285	476,886	–
Malaysia 1970	2	37,903	175,897	–
Malaysia 1980	2	38,049	182,601	–
Malaysia 1991	2	78,192	347,892	–
Malaysia 2000	2	95,564	435,300	–
Mexico 1960	1.5	n.a.*	502,800	–
Mexico 1970	1	82,856	483,405	–
Mexico 1990	10	1,648,280	8,118,242	–
Mexico 1995	0.4	72,277	332,061	–
Mexico 2000	10.6	2,312,035	10,099,182	yes
Mexico 2005	10	2,546,985	10,284,550	–
Mongolia 1989	10	42,783	190,631	yes
Mongolia 2002	10	55,495	243,725	–
Netherlands 1960	1.2	n.a.*	143,251	yes
Netherlands 1971	1.2	n.a.*	159,203	yes
Netherlands 2001	1.2	n.a.*	189,725	yes
Palestine 1997	10	40,753	259,191	yes
Panama 1960	5	11,869	53,553	–
Panama 1970	10	31,755	150,473	–
Panama 1980	10	47,726	195,577	yes
Panama 1990	10	61,458	232,737	–
Panama 2000	10	84,346	284,081	–
Philippines 1990	10	1,156,126	6,013,913	yes
Philippines 1995	10	1,362,190	6,864,758	–
Philippines 2000	10	1,511,890	7,417,810	yes
Portugal 1981	5	179,409	492,289	–
Portugal 1991	5	214,155	491,755	–
Portugal 2001	5	258,843	517,026	–
Romania 1977	10	619,904	1,937,021	–
Romania 1992	10	728,846	2,238,578	–
Romania 2002	10	732,016	2,137,967	–
Rwanda 1991	10	153,041	742,918	–
Rwanda 2002	10	191,719	843,392	–
Slovenia 2002	10	63,637	179,632	–
South Africa 1996	10	993,801	3,621,164	yes

South Africa 2001	10	991,543	3,725,655	yes
South Africa 2007	2	345,170	1,047,657	yes
Spain 1981	5	n.a.*	2,084,221	yes
Spain 1991	5	592,276	1,931,458	yes
Spain 2001	5	714,473	2,039,274	–
Uganda 1991	10	339,166	1,548,460	yes
Uganda 2002	10	529,271	2,497,449	–
United Kingdom 1991	1	215,761	541,894	–
United Kingdom 2001	3	n.a.*	1,843,525	–
United States 1960	1	579,212	1,799,888	–
United States 1970	1	744,475	2,029,666	–
United States 1980	5	4,711,341	11,343,120	–
United States 1990	5	5,527,406	12,501,046	yes
United States 2000	5	6,184,438	14,081,466	yes
United States 2005	1	1,245,246	2,878,380	yes
Venezuela 1971	10	284,336	1,158,527	yes
Venezuela 1981	10	323,321	1,441,266	–
Venezuela 1990	10	468,808	1,803,953	yes
Venezuela 2001	10	646,080	2,306,489	–
Vietnam 1989	5	534,223	2,626,985	yes
Vietnam 1999	3	534,139	2,368,167	yes
TOTAL		77,469,216	279,464,934	
<i>n.a. * = Not applicable; sample is of persons only, not households or dwellings.</i>				