

**Needs Assessment Conference on Census Analysis in Asia (NACCA)  
23-25 November, 2011  
Bali, Indonesia**

**Paper Session: How third parties add value**

**Title: IPUMS-International Adds Value to Census Microdata in Countless Ways.**

**Author: Robert McCaa, Professor Emeritus of Population History;  
Ambassador, IPUMS-International, Minnesota Population Center**

**“Dissemination [means] opening up the value inherent in our data”**

Seminar on Emerging Trends in Data Communication and Statistics, New York Feb. 19, 2010  
Walter Radermacher (President, Eurostat) and Pieter Everaers (Director, Eurostat)

**INTRODUCTION.**

Researchers and policy makers need access to census microdata to analyze challenging social, demographic, and economic transformations (McCaa and Esteve 2009). Most statistical agencies wish to disseminate census microdata, but are reluctant to do so. Others disseminate microdata directly to users, but find it costly, risky and even unsatisfactory:

- **Costly**, because the statistical office must draw the sample, protect the statistical confidentiality of the microdata, prepare the metadata, and manage access. These tasks amount to a lot of work, particularly when the number of requests may not amount to more than a couple of dozen per year.
- **Risky**, because few statistical offices are experienced in anonymizing census microdata and with only one census per decade there is little opportunity to gain experience. Then too, experience is easily lost when the specialist is hired away, re-assigned or otherwise unavailable. Managing access to microdata is particularly tricky because an expert is required and few wish to assume the time-consuming task of vetting users, monitoring access and providing user support.

Would-be anonymizers of microdata should take heed of a recent, unfortunate episode at the United States Census Bureau’s public use files of the American Community Survey (Alexander, Davern and Stevenson 2010). Due to a programming mistake age reporting of the elderly was egregiously corrupted in a large fraction of cases. Researchers only discovered the error by comparing the anonymized age distribution against the full-count non-confidentialized microdata. Much to the embarrassment of the Census Bureau, the brouhaha found its way to the front pages of the New York Times shortly before the 2010 census got underway.

- **Unsatisfactory**, because typically statistical offices err on the side of safety by excessively anonymizing the microdata and denying access to all but the most persistent or influential researchers. This strategy assures safety, but utility is minimized. The result is widespread disappointment.

Consider the following complaints—addressed to a major statistical office with over two decades experience in disseminating census microdata:

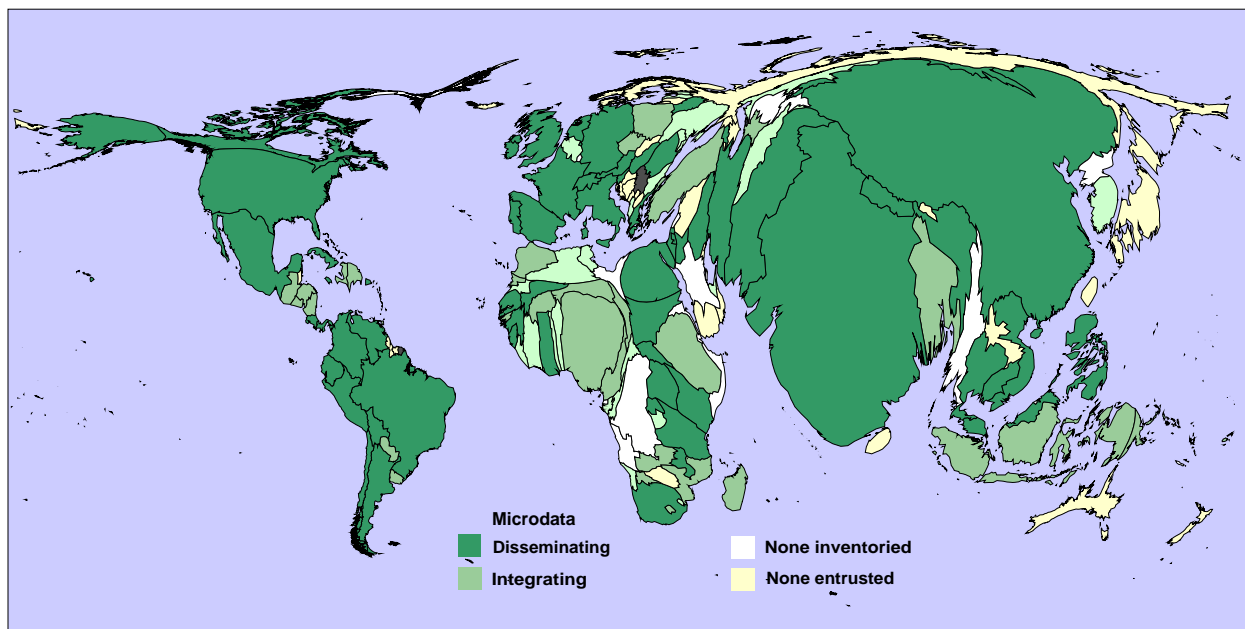
- “I haven't used the [microdata]; the bureaucracy was just too slow to get much use out of it.”
- “[Access] is unbelievably bureaucratic and difficult – this discourages people from using it. It took me 6 months to get the data.”

Some statistical agencies, despite elaborate websites and microdata access policies which seemingly incorporate best practices, never even reply to requests. In a virtual world, such sites may look great, but in reality they are doubly disappointing for the researcher who feels “delay is the deadliest form of denial.”

IPUMS-International ([www.ipums.org/international](http://www.ipums.org/international)) offers a one stop solution to disseminating anonymized microdata samples. Led by the Minnesota Population Center (MPC), the initiative is a global collaboratory of National Statistical Offices (NSOs), universities, and major research centers. The project provides services and assumes the responsibilities and risks for recovering, archiving, anonymizing, integrating, and disseminating microdata worldwide (McCaa and Thomas 2009). A bona fide researcher with need to access microdata on the IPUMS-International website is typically granted access within five days of submitting the completed application.

Begun in 1999, IPUMS-International enjoys the satisfaction of more than 5,000 researchers in some 100 countries and of over 90 NSOs. Thanks to sustained funding by the National Science Foundation and the National Institutes of Health (USA), the microdata are disseminated free of cost to bona fide researchers and policy makers, regardless of country of birth, residence, or nationality.

**Figure 1. Cartogram weighted by population size of countries participating in the IPUMS-International collaboratory, November 2011:  
3 shades of green--integrated (dark), integrating (medium), and negotiating (light).**



Twenty-two Asian and Pacific national statistical offices, encompassing 87% of the region's population, entrust census microdata and national surveys to IPUMS under a uniform memorandum of understanding (Appendix A and Figure 1). By means of a uniform licensing agreement bona fide researchers gain access to the microdata on a restricted basis (Appendix B).

As of June 2011, the project had completed the anonymization and integration of population microdata for 62 countries, totaling 185 samples and 397 million person records. Thanks to the exceedingly generous cooperation of NSOs worldwide, the IPUMS-International database is expanding at the rate of 5-10 additional countries per year. By 2015, it is likely that coverage will extend to 85 or more countries, encompassing four-fifths of the world's population.

**Invitation to participate and entrust 2010 round census microdata.** Statistical agencies not currently cooperating with the initiative are invited to contact the author of this paper to discuss questions and concerns. Participating statistical agencies are invited to entrust microdata for the 2010 round of censuses without undue delay to facilitate timely access to anonymized, integrated samples. Doing so will increase stakeholders and, in the words of the President of Eurostat (see epigraph), open “up the value inherent in” census microdata.

The fifty-two Asian and Pacific countries of 200,000 or more population may be divided into 3 groups, according to participation in the IPUMS initiative (microdatasets in parentheses):

**a. 16 are integrated:** Cambodia (1998, 2008), China (1982, 1990), India (5), Iran (2006), Iraq (2007), Israel (1972, 83, 95), Jordan (2004), Kyrgyz Republic (1999), Malaysia (1970, 80, 90, 2000), Mongolia (1989, 2000), Nepal (2001), Pakistan (1973, 81, 98), Palestine (1997, 2007), Philippines (1990, 95, 2000), Thailand (1970, 80, 90, 2000), and Vietnam (1989, 99, 2009).

**b. 6 are being integrated:** Bangladesh (1981, 91), Fiji Islands (1966, 76, 86, 96, 2007), Indonesia (1971, 76, 80, 85, 90, 95, 2000, 05, 10), Papua New Guinea (1980, 90, 2000), Turkey (1985, 90, 2000), and Turkmenistan (1995).

**c. 30 are *not* participating:** Afghanistan, Australia, Azerbaijan, Bahrain, Bhutan, Brunei, Georgia, India (Census Commission), Japan, Kazakhstan, Korea DPR, Korea RO, Kuwait, Laos, Lebanon (no census data), Maldives, Myanmar, New Zealand, Oman, Qatar, Saudi Arabia, Singapore, Sri Lanka, Syria, Taiwan, Tajikistan, Timor Leste, United Arab Emirates, Uzbekistan, and Yemen.

The 2012 IPUMS launch is scheduled to integrate samples for Indonesia (1971, 1976, 1980, 1985, 1990, 1995, 2000, 2005, and 2010) and Turkey (1985, 1990, and 2000) as well as five nations beyond the region (El Salvador, Morocco, Mexico, Nicaragua, and Uruguay). Additional launches are planned for successive years, integrating 2010 round census samples as expeditiously as they become available.

IPUMS-International seeks to open the value inherent in census microdata for academic researchers, policy makers and public officials. IPUMS-I is *not* simply a conduit for passing census samples from NSOs along to researchers. Instead, typically, two or more years of labor are invested by a sizeable group of experienced MPC personnel to prepare anonymized, integrated microdata and metadata for dissemination. In the twenty-first century, handing along a copy of the source microdata and a codebook is *not* sufficient for high quality research. This paper discusses 16 value-added offered by IPUMS-International. These may be categorized into four groups: statistical confidentiality, integration, dissemination and ethics.

## I. Statistical Confidentiality and Security.

**1. Microdata Security and Statistical Confidentiality.** Neither the MPC nor its employees ever disseminate original source microdata. These data are exceedingly sensitive and for that reason only seasoned, specially trained, full time researchers are allowed to work with the data until they are anonymized. MPC employees are subject to civil fine (up to US\$250,000), criminal prosecution and loss of employment for violating University of

Minnesota security procedures. The University legal authority assumes responsibility for protecting the total security and confidentiality of data entrusted to the MPC.

IPUMS-I is one of only a handful of academic organization disseminating census microdata that has been audited by external review. The reviewer, Mr. Dennis Trewin, the respected Australian statistician and chair of the UNECE Task Force Managing Statistical Confidentiality and Microdata Access, cited the computing environment of the MPC as “best practice,” “the standard of the best statistical offices”, etc. Mr. Trewin conducted a week-long, on-site inspection of the MPC and concluded his report as follows (Trewin 2007):

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 [emphasis added].

At the UN-ECE Expert Group Meeting on Statistical Data Confidentiality, November 2005, we explained the IPUMS-International data dissemination security procedures as follows (McCaa and Esteve 2005):

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 site for downloading the specific extract. The data are encrypted during transmission using 128-bit SSL (Secure Sockets Layer) encryption standard, matching the level used by the banking and other industries where security and confidentiality are 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.

This method of dissemination continues to weather the test of time, and indeed as usage soars, the rapid acceleration of internet transmission speeds has validated IPUMS-I security protocols.

**2. Statistical disclosure control protections.** The microdata are subjected to strong, uniform legal, administrative, and technical statistical disclosure controls providing greater protections for the group as a whole than for any single statistical office that chooses to “go it alone” (McCaa, Ruggles and Sobek 2010). The most important technical statistical disclosure control is the suppression of records by the use of sub-sampling. All the values in the records outside the sample are suppressed. Second comes the suppression of names and low level geographical detail.

Each statistical authority balances the confidentiality/utility trade-off by instructing the IPUMS-I project as to the minimum threshold for identifiable geographical units. For many countries, the threshold is commonly set at 20,000 inhabitants. Others place it as high as 100,000 (United States) or in the most extreme case (Netherlands) all administrative geography is suppressed. We are gratified that in some cases our statistical agency partners have reconsidered earlier decisions, offering higher precision samples (Mexico 1990 increased from one to ten percent) and greater detail. In the case of Colombia, the geographical threshold, initially set at 100,000, was reduced to 20,000 after Colombian geographers vigorously complained. The Colombian statistical agency not only reduced the threshold, but also harmonized the geo-statistical identifiers so that all the census microdata samples for Colombia could be disseminated with a single set of geographical codes.

Additional statistical disclosure protections are provided by randomly ordering the records and swapping the geographical identifiers of an undisclosed number of households. This means that no one can state with certainty that an individual or household has been identified.

In consultation with the national statistical office, some variables may be top-coded, others may be subjected to global recoding, deletion of digits for hierarchical variables (occupation, industry, geography), or the suppression of a variable entirely. Decisions are made in consultation with the corresponding national statistical authority. Weight variables and expansion factors are usually not an issue because most of the samples are implicitly stratified so that all records carry an identical weight.

**3. Restricted access.** Access to the IPUMS-International microdata is restricted—despite the “P” in IPUMS. Would-be users must submit a [detailed electronic application](#) both to establish research bona-fides and to explain need for access. An essential part of the process is to agree to ten stringent restrictions on condition of use—prohibiting redistribution, restricting to scholarly use, prohibiting commercial user, protecting confidentiality, assuring security, enforcing strict rules of confidentiality, permitting scholarly publication, citing properly, threatening disciplinary action for violations, and the reporting of errors. In other words, the IPUMS-I is a “trusted user” access system.

Agreeing to the conditions of use binds both the researcher and the researcher’s institution. The Legal Counsel of the University of Minnesota is poised to strike at the first indication of misuse. Both the individual researcher and the researcher’s institution are responsible for maintaining security and enforcing the license agreement. Violations are likely to lead to sanctions against both the individual and the offending researcher’s institution.

IPUMS-I resolves the conundrum of managing the broadest possible access to sensitive microdata while protecting statistical confidentiality. Many statistical agencies have long wanted to make census microdata available to researchers, but lack the substantial material and human resources required to implement and manage secure systems. IPUMS-I is the only academic organization disseminating census microdata that is cited as good practice by the Conference of European Statisticians Task Force on Managing Statistical Confidentiality and Microdata Access (2007).

Although access to the IPUMS-International microdata is free of cost, usage is restricted to bona-fide researchers who agree to abide by stringent conditions of use (Appendix B). IPUMS disseminates extracts, custom-tailored to the precise research needs of each user. The average IPUMS extract consists of a mere 10 variables.

This contrasts with the practices of most statistical offices where census microdata are disseminated as complete sets, consisting of a data dictionary and an entire sample containing all variables and all person records. Typically, under the traditional approach, when requests are fulfilled, each researcher receives exactly the same set of data and documentation. Given the massive size of the IPUMS-International database, disseminating the full set of variables and unvarying size of samples is impractical

Despite the the stringent conditions of use and restrictions more than five thousand researchers—representing over 100 countries and 900 institutions—are approved for access to the IPUMS-I database. More than one-third of IPUMS-I trusted users request access to microdata for a single country. A large fraction of these are resident abroad and seek access to data for their own country of identity.

## II. Integration

**4. Comprehensive source metadata.** Researchers must have ready access to the original census documentation in the official language. At a minimum, census questionnaires, enumerator instructions or training manuals, and codebooks are required. Additional metadata regarding the organization, preparation work, and actual census taking is also valuable to the IPUMS-International project and is catalogued and archived with all other documents received. Original hardcopy or PDF documents are preferred for published metadata materials. Our goal is to provide an archived collection of high-quality PDF files for all forms of metadata pertaining to census microdata. Source documents are made available to researchers in English and the official language of the original.

**5. Integrated, DDI compatible metadata**—cross national and over time. To facilitate the research process microdata are integrated for all censuses and countries, including detailed descriptions of each census, sample and variable.

**a. Censuses and samples.** IPUMS metadata offer detailed descriptions of each census in the database, listing the title, year, universe, de jure/de facto, enumeration unit, official 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 the text in 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: availability by census, universe of the variable or question, codes, source (enumeration) text, and non-harmonized variables used for integration. Access to this information is through clickable hypertext on the IPUMS website. A general comparability discussion is provided for every variable, with country or census specific discussions focusing on departures from standard practice. The purpose of these discussions is to highlight important contrasts. Clicking “Enumeration text” leads to source questions and corresponding instructions in English for each selected census. Additional clicks yield views of the original documentation in image form so that researchers may study lay-out and actual wording in the official language.

Researchers navigate the integrated metadata to quickly examine the actual wording of census questions and instructions to enumerators for any combination of countries and census years. The MPC integration team applies XML tags to the census documents, associating the variables in the microdata with the concepts in the text. The tagged material is then imported into a database. Once this step is completed, metadata may be retrieved dynamically for any combination of countries and census years, variable-by-variable. Initially this tool was developed to speed the work of the integration team. Once its utility became apparent, we harnessed the dynamic metadata system to the web-site, to permit open access to the metadata. Metadata are custom-tailored to each researcher’s request and available as a single, searchable file in DDI or text format.

**6. Integrated microdata.** The principal benefit of IPUMS to researchers and NSOs alike is integration of multiple microdata samples for each country—typically beginning with the earliest census for which microdata exist or are recoverable. For decades, many NSOs have provided a sample for the most recent census, but few re-examine earlier censuses to harmonize successive datasets or to draft new documentation to facilitate comparative analysis of two or more censuses. Modern statistical offices construct a census sample and a data dictionary for

national researchers. Five or ten years later, with the ensuing census, the process is repeated with little guidance on enhancing the comparability of successive census datasets.

We must reiterate that the IPUMS project does *not* disseminate census files entrusted by national statistical offices. 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 chronologically and cross-nationally. Integrated metadata are constructed by means of meticulous study of comprehensive original source documentation and after extensive analysis of the microdata. Thousands of hours are devoted to analyze, discuss, debate, draft, test and re-test until the microdata integration is validated for dissemination to researchers. The process is repeated with each annual launch of additional census samples into the IPUMS database.

The basic goal of the IPUMS-I harmonization effort is to simplify the use of the data while losing no meaningful information. This is challenging because to make the data simple for comparative analysis across time and space, it is necessary to develop comparable coding schemes. Microdata are integrated so that identical concepts (variables, categories) have identical codes. To avoid the loss of important information for those samples that have even more detail, IPUMS-I uses a composite coding strategy to retain all original detail, and at the same time provide comparable codes across samples. With composite codes, researchers may easily compare across time and space, yet nuances in changes of meaning are readily discernible. The first digit, which we call the “general code,” provides information that is available across all samples (the lowest common denominator data). The next one or two digits provides additional information available in a substantial subset of the samples. Trailing digits provide detail that is only rarely available. Where information is not available for a particular sample, a zero placeholder is assigned to that digit.

As an example of the IPUMS method of integrating variables, consider the concept “educational attainment,” the single most widely used variable in the IPUMS-International database. Most census microdata with information on this measure indicate whether the respondent completed primary, secondary or higher schooling or no schooling at all. Thus the first digit of the IPUMS-International composite code consists of four categories (1-4), plus a missing data code (9) and not-in-universe code (0, for children too young to attend or others to whom the question was not addressed). Some census samples contain further information indicating, for example, those who attended, primary, secondary or even tertiary schooling, but did not complete the course of study. The second digit captures this information.

Successful international integration must document such distinctions so that researchers may readily be informed of these and thousands of other details. Appendix C illustrates the detailed and general coding schemes for the educational attainment variable for 16 Asian countries (represented by its two-digit ISO 3166 code) and censuses (represented by a two-digit year code with century omitted).

**7. Integrated boundary files.** IPUMS-International provides boundary files (shapefiles) to facilitate national and international data mapping. The researcher downloads the boundary files and manipulates them with software as desired. Two levels of GIS boundary files are available: a world file with country level boundaries, and country level files with top administrative level boundaries (states, provinces, etc.). Users can create maps with IPUMS-International data using a statistical software program and ArcMap (a GIS mapping software). For more information about linking IPUMS-International data to world and select country maps, see: <https://international.ipums.org/international/gis.shtml> .

**8. IPUMS-I Value-added variables.** Appendix D and E lists integrated household and person variables for the most recent census of 16 Asian countries. There are three types of IPUMS-I value-added variables:

**a. Technical variables:** Record type, Country, Year, IPUMS sample identifier, Household serial number, Number of person records in household, Household weight, Subsample number, Group quarters status, Continent, region of country, Residence at first administrative level, expansion factors (sample weights—for households and persons), etc.

**b. Summary household and family variables:** Household classification, Number of families in household, Number of married couples in household, Number of mothers in household, Number of fathers in household, Head's location in household, Number of unrelated persons, Family unit membership, Number of own family members in household, Number of own children in household, Number of own children under age 5 in household, Age of eldest own child in household, Age of youngest own child in household, etc.

**c. Pointer variables** to identify co-resident spouses, children and their parents: Mother's, Father's and Spouse's location in household, Rule for linking parent(s) and spouse(s), Probable stepmother, Probable stepfather, Man with 2+ wives linked, Second+ order wife, etc. (Sobek and Kennedy, 2009).

### III. Dissemination

**9. Trans-border access:** Trans-Border access to microdata is essential in today's global world, where researchers are highly mobile. Consider, for example, the field of demography, where one-fifth of the membership of the global professional society, the International Union for the Scientific Study of Population (IUSSP), resides outside their country of birth. For the 506 members of the IUSSP resident in the USA, thirty percent were not born there. Of Chinese born demographers, almost one-third reside outside China. For German and Dutch born IUSSP members the fraction stands at two-fifths.<sup>1</sup> For many professional demographers—and many social science researchers in general--trans-border access is essential if they are to conduct research using census microdata of their country of birth, whether comparative or not. The IPUMS-International portal provides a uniform experience for accessing samples for all countries and census years, regardless of the researcher's country of birth, residence or nationality.

**10. Custom-tailored extracts.** With IPUMS-I no two extracts are alike. Each extract is custom-tailored by the researcher by means of a series of point-and-click selection screens. While making selections, metadata are readily available to “surf” the documentation in any order desired. To place an order the researcher selects:

- country (or countries)
- census year(s)
- variables (age, sex, educational attainment, etc.)
- sub-populations (e.g., female heads of households aged less than twenty five years old. Note that there is an option for selecting the individual or all co-resident persons in the selected household)
- and sample density (either as a percent or number of cases).

---

<sup>1</sup> Statistics communicated to the author by the Secretariat of the International Union for the Scientific Study of Population, September 14, 2011.



The IPUMS extract engine fulfils the request by generating a dataset containing only the requested microdata and the corresponding set of DDI compatible metadata as well as a codebook suitable for constructing an SPSS, SAS or STATA file. Copies of original source metadata are available from the web-site, as well as integrated metadata in interactive form. These also may be downloaded freely.

The fact that IPUMS-International distributes microdata electronically as custom extracts, tailored as to country(ies), census year(s), subpopulation(s), and variables, according to the individual needs of the researcher, provides additional incentives for jealously guarding extracts. Since complete datasets are not distributed on CD or other medium, the inclination to share data with unauthorized individuals is greatly reduced, if not completely eliminated.

The IPUMS extract engine adds even more value to each extract by means of three unique tools:

a. **Select cases** (see “sub-populations” above). Filter the samples to select precisely the cases of research interest. For example, if the researcher wishes to research only economically active females, aged 15-19, born in a country different from the current country of residence, the IPUMS extract engine will generate a dataset precisely to these characteristics. Moreover the extract engine will include in the extract all individuals co-resident in the same household with the selected individual.

b. **Attach characteristics.** The extract engine attaches characteristics of mothers, fathers, spouses and household heads to co-resident individuals. This feature facilitates the analysis of children by characteristics of their mothers, fathers, and/or household heads. The feature is useful for analyzing own-child fertility, marital homogamy, and a host of other topics where the joint-characteristics of two or more members of the household are required. For example, for the recent international seminar on “Cross Border Marriages” (Seoul, Republic of Korea), it was easy to generate a dataset from the IPUMS-I website representing 12 million foreign born individuals married or in union with co-resident native born spouses for 51 countries (Esteve, Garcia and McCaa, 2011; see also Parker, 2011).

c. **Customize sample size.** The extract system offers a tool to customize the size of each sample in terms of the absolute number of persons or households or in percentages. If the researcher desires a sample of only 50,000 households, simply enter “50” in the corresponding table and the extract engine will construct a systematic sample. The appropriate weights (expansion factors) are automatically computed on-the-fly and included in the extract.

**11. Usage.** Microdata are liberated from the confines of statistical offices to permit some of the brightest statisticians, demographers, economists, social scientists and policy makers on the planet to extract un-dreamed of value from the data without cost (Meyer, McCaa, and Lam, 2011). For, national researchers many were never able to use the microdata of their country until IPUMS-International opened the door. Other researchers faced troublesome, sometimes exorbitant fees.

As of April 2011, 85,505 samples (24,699 extracts) had been downloaded from the IPUMS-International site, averaging over 1,500 samples per country for the 55 countries represented in the database at that time (Table 1). Nonetheless, usage by country varies greatly. The smallest number of extracts, 153, was registered for the 2001 census of Nepal. The greatest number, 9,338 was registered for the six censuses of Mexico. The top ranked sample was Mexico 2000, with 2,945 extracts.

The mean number of extracts per user was 5.5 and the mean number of samples per extract was 3.5. The fact that the average extract consisted of a mere ten variables shows that IPUMS-International users are researchers, not hoarders. Their extract requests are parsimonious, limited to specifically what is needed to address well-defined research questions.

Table 2 lists the 32 most commonly extracted variables. The top eight encompass four demographic variables (marital status, relationship to head, age and sex), two economic (employment status and class of worker), and one each social (educational attainment) and technical (person weight).

A most striking finding is that 40 variables account for 60% of those requested and among these are sixteen IPUMS constructed variables, four of which, the “LOC” variables, are unique to IPUMS-I. The LOC variables are imputed from household information: spouse’s location in household, mother’s location, father’s location, and the rules for inferring locations. Researchers exploit these variables to study the joint characteristics of spouses and characteristics of parents relative to their children. The variables are constructed by imputation from the relationship to head variable, age, sex, marital status, order of individuals listed in the household, and a few other variables. The heavy usage of the “LOC” variables indicates their great importance for analyzing individuals in relation to characteristics of their spouses, mothers and fathers. The LOC variables constitute an exceedingly important value-added to each household sample (Sobek and Kennedy, 2009).

The IPUMS-I “top 40” institutions in terms of data usage includes many of the world’s premier universities and research organizations (see Appendix F), scattered across fourteen countries. In 46 countries, we find a total of 501 institutions with researchers making ten or more extracts. (In addition, in the United States, there are 295 institutions at this level of usage.) A surprising number of extracts are made by researchers from countries with no microdata in the IPUMS-I system. The top 10 of these are: Singapore (494 extracts), Belgium (250), Australia (229), Japan (170), Russian Federation (58), Republic of Korea (45), Czech Republic (42), Sweden (41), Hong Kong SAR (40), and New Zealand (40). On the opposite side of the coin are 14 countries with microdata in the IPUMS-I database but as yet no national researchers use them. The 14 are: Armenia, Belarus, Ghana, Guinea, Iraq, Jordan, Kyrgyzstan, Mali, Mongolia, Nepal, Peru, Rwanda, Saint Lucia, and Slovenia. Of course, researchers from these countries—instead of accessing microdata electronically from the IPUMS-I website—may acquire copies of the integrated samples on CDs supplied by the Minnesota Population Center to the corresponding National Statistical Office. We advise NSO partners to register any such users and admonish them to respect the IPUMS-I conditions of use, but there is no obligation to do so because the NSO is responsible for any such dissemination.

**12. 2010 round census samples.** Microdata for 2010 round of censuses are “fast-tracked” processing by IPUMS-International without due delay. Microdata received by September 1 are processed for launch June 1<sup>st</sup> of the following year. To date, all the IPUMS-International partners are eager to entrust microdata as soon as data processing is completed. Samples from the 2010 censuses of Indonesia and Mexico will be launched in 2012. Samples of the 2009 census of Vietnam and 2008 of Cambodia were launched in 2011.

#### IV. Ethics.

**13. Statistical transparency.** Entrusting census microdata demonstrates to the world that the Official Statistical agency “has nothing to hide” and is willing to subject its most treasured microdata to external scrutiny.

**14. Academic freedom.** Microdata disseminated by IPUMS-International are under no prior restraints on publication or dissemination of results.

**15. Reduce fraud** and the temptation to exaggerate research findings. Since no researcher has exclusive access to the microdata, results may readily be replicated to test the findings of other researchers.

**16. Research results.** Researchers post links to their findings in the IPUMS-International open-access bibliography: <http://bibliography.ipums.org/>. The bibliography for IPUMS-International lists over 500 entries.

### **Reflections.**

As census outputs to meet user needs, the IPUMS/IECM projects request a formidable range and amount of metadata and microdata. Nonetheless the return on the investment is substantial. Statistical offices are relieved of many of the most burdensome tasks and responsibilities in disseminating microdata to researchers. Moreover, by relying on standard procedures used by a majority of the world's statistical offices, there is safety in numbers. The isolated statistical office that disseminates microdata on an ad hoc basis incurs substantial risks and responsibilities as well as significant human resource and material costs, for a relatively small return with respect to number of users. The IPUMS/IECM projects offer substantial economies of scale with the highest standards of security and disseminates integrated metadata and microdata that greatly facilitates sound scientific research.

When we began a decade ago, we dreamed of integrating samples for 21 countries in ten years. Thanks to the generous cooperation of National Statistical Offices and undreamed of technological innovations, that number has tripled (including South Sudan as no. 63), and work continues. The number of users and the amount of use also far exceeded our expectations. For the second decade, we dream of doubling the number of users and doubling again the number of samples. High precision samples for the 2010 round of censuses will be crucial to our success. Participating statistical agencies are invited to entrust metadata and microdata for the 2010 census round at their earliest convenience. Those agencies not yet participating in the IPUMS initiatives are invited to consider doing so. Researchers who have yet to access the IPUMS microdata are invited to peruse the metadata and use the microdata should their research needs require.

## References.

Alexander, J.T.; Davern, M.; and Stevenson, B. 2010. "Inaccurate Age and Sex Data in the [United States] Census PUMS Files: Evidence and Implications," *Public Opinion Quarterly*, 10 (Aug 10), pp. 1-10. doi: 10.1093/poq/nfq033

Conference of European Statisticians. 2007. "Annex 1.23 Case study: Access to anonymized census microdata samples via the IPUMS-International and the Integrated European Census Microdata websites," *Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines on Good Practice*. Geneva: United Nations Economic Commission for Europe. See online edition: <http://www.unece.org/stats/publications/> pp. 98-104.

Esteve, A., J. Garcia and R. McCaa. 2011. "Comparative perspectives on Marriage and International Migration, 1970-2000: findings from IPUMS-International census microdata samples," *Seminar on Global Perspectives on Marriage and International Migration*, Seoul, South Korea: IUSSP Scientific Panel, Oct. 20-21.

McCaa, R. and A. Esteve. 2005. "[IPUMS-Europe: Confidentiality measures for licensing and disseminating restricted access census microdata extracts to academic users](#)," *Joint UNECE/Eurostat Work Session on Statistical Confidentiality*, Geneva, Nov. 9-11.

McCaa, R. and A. Esteve. 2009. "[Entrusting census microdata and metadata for timely integration and dissemination via the IPUMS-EurAsia and IECM initiatives, 2010-2014](#)," *Census Outputs to Meet User Needs*. Geneva: United Nations Economic Commission for Europe, Oct. 28-30.

McCaa, R., S. Ruggles and M. Sobek. 2010. "[IPUMS-International statistical disclosure controls: 159 census microdata samples in dissemination, 100+ in preparation](#)," in J. Domingo-Ferrer and E. Magkos (Eds.): *Privacy in Statistical Data 2010*, LNCS 6344. Springer, Heidelberg, pp.74-84.

McCaa, R. and W. Thomas. 2009. "[IPUMS-International: lessons from 10 years of archiving and disseminating census microdata](#)," *International Statistical Institute IPM100*. Durban, South Africa.

Minnesota Population Center. 2011. *Integrated Public Use Microdata Series – International: Version 6.0*. Minneapolis: University of Minnesota: <https://www.ipums.org/international>.

Meier, A., R. McCaa and D. Lam. 2011. "[Creating statistically literate global citizens: The use of IPUMS-International integrated census microdata in teaching](#)". *Statistical Journal of the IAOS* 27(3):145-156.

[Parker, J.] 2011. "Herr and Madame, Señor and Mrs. Research at last begins to cast some light on the extent, causes and consequences of cross-border marriages," *Economist*, Nov. 12. <http://www.economist.com/node/21538103>

Sobek, M and S. Kennedy. 2009 The development of family interrelationship variables for international census data, Minnesota Population Center. [https://international.ipums.org/international/resources/misc\\_docs/pointer\\_working\\_paper\\_2009.pdf](https://international.ipums.org/international/resources/misc_docs/pointer_working_paper_2009.pdf) .

Trewin, D. 2007. "A Review of IPUMS-International." Unpub. [http://www.hist.umn.edu/~rmccaa/IPUMSI/trewin\\_ipums\\_report.pdf](http://www.hist.umn.edu/~rmccaa/IPUMSI/trewin_ipums_report.pdf)

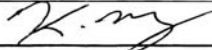
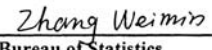
**Table 1. Rank of the Top Five and all Asian Countries  
by Number of IPUMS-International Extracts (through April 18, 2011)**

Rank	Country	Sample %*	Variables (n)*	Years of census samples	Extracts
1	Mexico	10	120	1960, 70, 90, 95, 2000, 05	9,338
2	Brazil	5	106	1960, 70, 80, 91, 2000	6,889
3	United States	5	92	1960, 70, 80, 90, 2000, 05	6,171
4	Colombia	10	120	1964, 72, 85, 93, 2005	4,629
5	France	33	94	1962, 68, 75, 82, 90, 99, 06	3,861
17	India	0.1	56	1983, 87, 93, 99, 2004	1,700
18	China	1	63	1982, 90	1,635
19	Philippines	10	89	1990, 95, 2000	1,632
22	Vietnam	15	92	1989, 99, 2009	1,447
26	Malaysia	2	88	1970, 80, 90, 2000	1,275
27	Israel	10	81	1972, 83, 95	1,086
32	Thailand	1	83	1970, 80, 90, 2000	622
37	Cambodia	10	97	1998, 2008	443 <sup>a</sup>
42	Mongolia	10	76	1989, 2000	364
46	Palestine	10	85	1997, 2007	334 <sup>a</sup>
47	Iraq	10	102	1997	301
48	Pakistan	10	59	1973, 81, 98	297
51	Jordan	10	90	2004	235
52	Kyrgyz Republic	10	78	1999	216
55	Nepal	11	100	2001	153
<sup>a</sup>	Iran	2	97	2006	<sup>a</sup>
Total samples extracted for 55 countries (159 samples)					24,699
*2000 round census; refers to all integrated variables, including IPUMS constructed variables.					
“ <sup>a</sup> ” = most recent sample launched June 2011.					

**Table 2. 32 most popular variables in IPUMS-International (85,505 Sample Extracts)**

<u>Rank</u>	<u>Label</u>	<u>Extracts</u>	<u>Mnemonic</u>	<u>Comment</u>
1	Educational attainment	19,307	EDATTAN	
2	Age (single years to 85+)	19,009	AGE	Grouped age n=3,838
3	Employment status	18,490	EMPSTAT	
4	Marital status	18,214	MARST	
5	Person weight	17,511	WTPER	Technical variable
6	Relationship to head	15,783	RELATE	
7	Sex	14,595	SEX	
8	Class of work	12,583	CLASSWK	
9	Ownership of dwelling	8,050	OWNRSHP	
10	Occupation ISCO recode	8,004	OCCISCO	
11	School attendance	7,919	SCHOOL	
12	Years of schooling	7,576	YRSCHL	
13	Literate	7,290	LIT	
14	Urban/rural	7,098	URBAN	
15	Industry-general code	7,044	INDGEN	
16	Household weight	6,656	WTHH	Technical variable
17	Children ever born	6,363	CHBORN	
18	Nativity (native/foreign born)	6,332	NATIVTY	
19	Occupation	6,246	OCC	
20	Country of birth	6,153	BPLCTRY	
21	Religion	6,075	RELIG	
22	Industry	5,670	IND	
23	Location of spouse in household	5,007	SPLOC	IPUMS unique
24	Rule for locating spouse	4,171	SPRULE	IPUMS unique
25	Location of mother in household	4,153	MOMLOC	IPUMS unique
26	Number of children surviving	4,074	CHSURV	
27	Place of residence 5 years ago	4,064	MGRATE5	
28	Location of father in household	3,983	POPLOC	IPUMS unique
29	Total household income	3,965	INCTOT	Household variable
30	Earned income	3,655	INCEARN	
31	Number of rooms	3,465	ROOMS	
32	Consensual union	3,443	CONSENS	

## Appendix A. Example of Uniform Memorandum of Understanding between the University of Minnesota and National Statistical Offices (China, 2002).

<p>IPUMS-International agreement      November 18, 2001      1</p> <p style="text-align: center;"><u>Letter of Understanding</u></p> <p style="text-align: center;"><b>Integrated Public Use Microdata Series International</b></p> <p style="text-align: center;"><b>and the National Bureau of Statistics of China</b></p> <p><u>Purpose.</u> The purpose of this letter is to specify the terms and conditions under which metadata and microdata provided by the <b>National Bureau of Statistics of China</b> shall be distributed by <b>Integrated Public Use Microdata Series International</b> of the University of Minnesota.</p> <ol style="list-style-type: none"> <li><u>Ownership.</u> The <b>National Bureau of Statistics of China</b> is the owner and licensee of the intellectual property rights (including copyright) in the metadata and microdata supplied to the University of Minnesota to be distributed by <b>Integrated Public Use Microdata Series International</b>.</li> <li><u>Use.</u> These data are provided for the exclusive purposes of teaching, academic research and publishing, and may not be used for any other purposes without the explicit written approval, in advance, of the <b>National Bureau of Statistics of China</b>.</li> <li><u>Authorization.</u> To access or obtain copies of integrated microdata of China from <b>Integrated Public Use Microdata Series International</b>, 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 <b>Integrated Public Use Microdata Series International</b>, the <b>National Bureau of Statistics</b>, or other authorized distributors. Once approved, the user is licensed to acquire integrated metadata and microdata of China from <b>Integrated Public Use Microdata Series International</b> or other authorized distributors. No titles or other rights are conveyed to the user.</li> <li><u>Restriction.</u> Users are prohibited from using data acquired from the <b>Integrated Public Use Microdata Series International</b> or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise.</li> <li><u>Confidentiality.</u> 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.</li> <li><u>Security.</u> Users will implement security measures to prevent unauthorized access to microdata acquired from <b>Integrated Public Use Microdata Series International</b> or its partners.</li> <li><u>Publication.</u> The publishing of data and analysis resulting from research using metadata or microdata of China is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite <b>National Bureau of Statistics of China</b> and <b>Integrated Public Use Microdata Series International</b> as the sources of the data of China, and to indicate that the results and views expressed are those of the author/user.</li> </ol>	<p>IPUMS-International agreement      November 18, 2001      2</p> <ol style="list-style-type: none"> <li><u>Sharing.</u> <b>Integrated Public Use Microdata Series International</b> will provide electronic copies to the <b>National Bureau of Statistics of China</b> of documentation and data related to its integrated microdata as well as timely reports of authorized users.</li> <li><u>Violations.</u> Violation of this agreement may lead to professional censure and/or civil prosecution.</li> <li><u>Jurisdiction.</u> 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.</li> </ol> <p>Date: <u>1/22/02</u></p> <p>Signed: <u></u></p> <p><b>Regents of the University of Minnesota</b> By: Kevin McKoskey Grants Manager, Sponsored Projects Administration</p> <p>Date: <u>3/5/2002</u></p> <p>Signed: <u></u> <b>National Bureau of Statistics</b></p>
--	--

**Appendix B. Snippets of Application Form to Use Restricted Microdata disseminated by IPUMS-International. See: <https://international.ipums.org/international-action/register/0>**

<p>IPUMS International <span style="float: right;">Page 1 of 1</span></p> <h3 style="text-align: center;">Application to Use Restricted Microdata</h3> <p>IPUMS-International microdata are available free of charge, but their use imposes responsibilities upon the user. To access the data, a prospective user must submit an electronic authorization form (this form) identifying the user by name, electronic address, and institutional affiliation.</p> <p>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.</p> <p>Once a user 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.</p> <p><b>Legal notice:</b> Submission of this application constitutes a legally binding agreement between the applicant, the applicant's institution, the University of Minnesota, and the relevant official statistical authorities. Submitting false, misleading or fraudulent information constitutes a violation of this agreement. Misusing the data by violating any of the conditions detailed below also constitutes a violation of this agreement and may lead to professional censure, loss of employment, or civil prosecution under relevant national and international laws, and to sanctions against your institution, at the discretion of the University of Minnesota and the official statistical authorities.</p> <p>Information provided on this form will be kept confidential. All information on this form is required for registration unless otherwise indicated by an asterisk.</p> <p style="text-align: center;"><b>PERSONAL INFORMATION</b></p> <p style="text-align: center;">.....</p> <p style="text-align: center;"><b>INSTITUTIONAL AFFILIATION</b></p> <p>IPUMS-International staff must confirm the identity of prospective users. To speed the processing of your application, please provide as much of the following</p>	<h3 style="text-align: center;">USAGE LICENSE</h3> <p>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 condition under which integrated microdata samples distributed by Integrated Public Use Microdata Series International of the University of Minnesota may be used. <b>Note: The license is valid for one year and may be renewed.</b></p> <p style="text-align: center;"><b>Data must not be redistributed without authorization.</b></p> <p>All data extracted from the IPUMS-International database are intended solely for the use of the licensee. Under IPUMS-International agreements with collaborating agencies, redistribution of the data to third parties is prohibited. Each member of a research team using the data must apply for access and be licensed individually.</p> <p><b>The microdata are intended only for scholarly research and educational purposes.</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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 from the relevant official statistical authority.</li> </ul> <p><b>Commercial use and redistribution of the microdata is strictly prohibited.</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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.</li> </ul> <p><b>Use of the microdata must follow strict rules of confidentiality.</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 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. Statistical results that might reveal the identity of persons or entities may not be reported or published in any form.</li> </ul> <p><b>The microdata must always be safely secured.</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Users will implement security measures to prevent unauthorized access to microdata acquired from Integrated Public Use Microdata Series International, its partners or authorized distributors. Upon the completion of this research, data may be retained only if they can be safely secured. If security cannot be guaranteed, the microdata must be destroyed.</li> </ul>
--	---



<p>Name of institution or employer <input type="text"/></p> <p>Your email address at institution (*) <input type="text"/></p> <p>Web link showing your affiliation with institution (*) <input type="text"/></p> <p>Email address of employer, supervisor, or instructor (*) <input type="text"/></p> <p>Phone number of institution (*) <input type="text"/></p> <p>Does your institution have an Institutional Review Board (IRB), or Office for Human Subject Protections, Professional Conduct or similar committee?</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Yes; Name of board or office <input type="text"/></p> <p><b>RESEARCH PROJECT</b></p> <p>Please provide at least 75 words <i>in English</i> describing your research project or educational use for the data. This description will be used to evaluate your application.</p> <p style="text-align: center;">.....</p> <p>If your research is funded by someone other than your employer, indicate the name of the granting institution, title of grant, and other pertinent information. (*)</p> <input type="text"/>	<p><b>Scholarly publications are permitted, and must be cited appropriately.</b></p> <ul style="list-style-type: none"> <li>■ 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 and the relevant official statistical authority as the source of the microdata, and to indicate that the results and views expressed are those of the author. Users are requested to provide the IPUMS-International staff with a full citation for any publications resulting from their work with these data.</li> </ul> <p><b>Any violation of this license agreement will result in disciplinary action, including possible loss of employment.</b></p> <p>Violation of this agreement will lead to revocation of this license, recall of all microdata acquired, a motion of censure to the relevant professional organization(s) and civil prosecution under national or international statutes, at the discretion of the Regents of the University of Minnesota and the official statistical agencies. Sanctions likewise may be taken against the institution with which the violator is affiliated.</p> <ul style="list-style-type: none"> <li>■ User agrees to notify <a href="mailto:ipums@pop.umn.edu">ipums@pop.umn.edu</a> regarding errors in the data.</li> </ul>
---	---

**Appendix C. Educational attainment harmonized codes: 16 Asian Countries, most recent census sample in IPUMS-I**

“X” indicates that the code is present in the respective sample

Code	Label	Country	KH	CN	IN	IR	IQ	IL	JO	KG	MY	MN	NP	PK	PS	PH	TH	VN
		Census Year	08	90	04	06	97	95	04	99	00	00	01	98	07	00	00	09
0	NIU (not in universe)		X	X	.	X	X	X	X	X	.	X	X	X	X	X	X	X
100	LESS THAN PRIMARY COMPLETED		.	.	.	.	X	.	X	.	.	X	.	.	X	.	.	.
110	No schooling		X	X	X	X	.	X	.	X	X	.	X	X	.	X	X	X
120	Some primary		X	X	X	X	.	X	X	X	X	.	X	X	.	X	X	X
130	Primary (4 years)		.	.	.	.	.	.	.	X	.	X	.	.	.	.	.	.
PRIMARY COMPLETED, LESS THAN SECONDARY																		
Primary completed																		
211	Primary (5 years)		.	.	X	X	.	.	.	.	.	.	X	.	.	.	.	X
212	Primary (6 years)		X	X	.	.	X	X	X	.	X	.	.	X	X	X	X	.
Lower secondary completed																		
221	General and unspecified track		X	X	X	X	X	X	X	X	X	X	X	X	X	.	X	X
222	Technical track		.	.	.	.	.	.	.	.	.	X	.	.	.	.	.	.
SECONDARY COMPLETED																		
General or unspecified track																		
311	General track completed		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
312	Some college/university		.	X	X	X	.	.	X	X	.	.	.	.	X	X	X	X
320	Technical track		.	.	.	.	.	.	.	.	.	X	.	.	.	.	.	.
321	Secondary technical degree		X	X	.	X	X	.	X	X	.	.	.	.	.	.	X	X
322	Post-secondary technical education		X	X	X	.	X	X	.	.	X	.	X	X	.	X	X	.
400	UNIVERSITY COMPLETED		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
999	UNKNOWN/MISSING		X	.	X	X	X	X	X	.	X	.	X	X	X	X	X	.

Source: <https://international.ipums.org/international-action/codes.do?mnemonic=EDATTAN>

**Appendix D. 40 (of 59) IPUMS Integrated Household Variables: Availability in Most Recent Sample for 16 Asian Countries**

Variable	Variable Label	Country Census Year	Sum	Country															
				KH 2008	CN 1990	IN 2004	IR 2006	IQ 1997	IL 1995	JO 2004	KG 1999	MY 2000	MN 2000	NP 2001	PK 1998	PS 2007	PH 2000	TH 2000	VN 2009
1	RECTYPE	Record type	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
2	CNTRY	Country	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
3	YEAR	Year	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
4	SAMPLE	IPUMS sample identifier	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
5	SERIAL	Household serial number	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
6	PERSONS	No. person records in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	WTHH	Household weight	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
8	SUBSAMP	Subsample number	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
9	GQ	Group quarters status	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
10	REGIONW	Continent, region of country	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
11	composite	Residence: 1st admin. level	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
12	HHTYPE	Household classification	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
13	NFAMS	Number of families in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
14	NCOUPLS	No. of married couples in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
15	NMOTHS	Number of mothers in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
16	NFATHRS	Number of fathers in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
17	HEADLOC	Head's location in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
18	UNREL	Number of unrelated persons	14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
19	URBAN	Urban-rural status	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
20	composite	Residence: 2nd admin. level	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
21	WATSUP	Water supply	13	1	.	.	1	1	.	1	1	1	1	1	1	1	1	1	
22	ELECTRC	Electricity	12	1	.	.	1	1	.	1	1	1	1	1	1	1	1	1	
23	PHONE	Telephone availability	12	1	.	.	1	1	1	1	1	1	1	.	.	1	1	1	
24	OWNRSH	Ownership of dwelling	11	1	.	.	1	1	1	1	.	.	.	1	1	1	1	1	
25	TOILET	Toilet	11	1	.	.	1	1	.	1	.	1	1	1	1	1	1	1	
26	ROOMS	Number of rooms	10	1	.	.	1	1	1	1	1	1	1	1	1	1	1	1	
27	TV	Television set	9	1	.	.	.	.	1	1	.	1	.	1	1	1	1	1	
28	FUELCK	Cooking fuel	8	1	.	.	1	.	.	.	.	.	1	1	1	1	1	1	
29	RADIO	Radio in household	8	1	.	.	.	.	.	1	.	1	.	1	1	1	1	1	
30	WALL	Wall or building material	8	.	.	.	1	1	.	1	.	1	.	1	.	1	1	1	
31	AUTOS	Automobiles available	7	1	.	.	1	.	1	1	.	1	.	.	1	.	1	.	
32	COMPUTR	Computer	7	1	.	.	1	.	1	1	.	1	.	.	1	.	1	.	
33	WASHER	Clothes washing machine	7	.	.	.	.	.	1	1	.	1	.	.	1	1	1	1	
34	BATH	Bathing facilities	7	.	.	.	1	1	1	1	.	1	.	1	1	.	.	.	
35	SEWAGE	Sewage	6	.	.	.	1	1	.	1	.	1	.	1	.	1	.	.	
36	REFRIG	Refrigerator	6	.	.	.	.	.	.	1	.	1	.	.	1	1	1	1	
37	KITCHEN	Kitchen or cooking facilities	6	.	.	.	1	1	.	1	.	.	1	.	1	1	.	.	
38	INTRNET	Internet access	5	1	.	.	1	.	.	1	.	1	.	.	1	.	.	.	
39	AIRCON	Air conditioning	5	.	.	.	1	.	1	.	.	1	.	.	.	.	1	1	
40	STRAG2	Age of structure, intervals	5	.	.	.	1	.	1	1	.	.	.	1	.	.	.	1	

Source: <https://international.ipums.org/international-action/variables/group/cons> .

**Appendix E. 114 IPUMS Integrated Person Variables: Availability in Most Recent Sample for 16 Asian Countries**

Variable	Variable Label	Country Census Year Sum	Country															
			KH 2008	CN 1990	IN 2004	IR 2006	IQ 1997	IL 1995	JO 2004	KG 1999	MY 2000	MN 2000	NP 2001	PK 1998	PS 2007	PH 2000	TH 2000	VN 2009
1	PERNUM	Person number	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	WTPER	Person weight	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	MOMLOC	Mother's location in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	POPLOC	Father's location in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	SPLOC	Spouse's location in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	PARRULE	Rule for linking parent	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	SPRULE	Rule for linking spouse	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	STEMMOM	Probable stepmother	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	STEPPOP	Probable stepfather	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	POLYMAL	Man with 2+ wives linked	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	POLY2ND	Woman is second+ order wife	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	FAMUNIT	Family unit membership	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	FAMSIZE	No. of own family members in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	NCHILD	No. of own children in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	NCHLT5	No. own children under age 5 in	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	ELDCH	Age of eldest own child in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	YNGCH	Age of youngest own child in hh	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	RELATE	Relationship to hh head	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
19	AGE	Age	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	AGE2	Age, grouped into intervals	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	SEX	Sex	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
22	MARST	Marital status	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
23	EDATTAN	Educational attainment, intl.	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	LIT	Literacy	15	1	1	1	1	1	.	1	1	1	1	1	1	1	1	1
25	OCC	Occupation, unrecoded	15	1	1	1	1	1	1	1	1	1	1	.	1	1	1	1
26	INDGEN	Industry, general recode	15	1	1	1	1	1	1	1	1	1	1	.	1	1	1	1
27	IND	Industry, unrecoded	15	1	1	1	1	1	1	1	1	1	1	.	1	1	1	1
28	OCCISCO	Occupation, ISCO general, 1-digit	14	1	1	1	1	1	.	1	1	1	1	.	1	1	1	1
29	CLASSWK	Class of worker	14	1	.	1	1	1	1	1	1	1	1	.	1	1	1	1
30	SCHOOL	School attendance	13	1	.	1	1	.	1	1	1	1	1	.	1	1	1	1
31	EMPSTAT	Employment status	13	1	1	1	1	1	1	1	1	1	1	.	1	.	.	1
32	NATIVTY	Nativity status	11	1	.	.	1	1	1	.	1	1	1	1	.	1	1	1
33	CHBORN	Children ever born	10	1	1	.	1	1	1	.	1	.	1	.	1	.	1	1
34	CITIZEN	Citizenship	10	.	.	.	1	1	.	1	1	1	1	1	1	.	1	1
35	CHSURV	Children surviving	9	1	1	.	1	1	.	1	.	1	.	1	.	1	.	1
36	BPLCTRY	Country of birth	9	1	.	.	.	1	1	.	1	1	1	1	.	.	1	1
37	composite	Administrative area of birth	9	1	.	.	.	1	1	.	1	1	1	1	.	1	.	1
38	RELIG	Religion	9	1	.	1	1	.	1	.	.	1	.	1	.	1	1	.
39	BIRTHYR	Year of birth	8	.	1	.	1	1	.	1	1	.	1	.	.	.	1	1
40	BIRTHMO	Month of birth	8	.	1	.	1	1	.	1	1	.	1	.	.	.	1	1

Variable	Variable Label	Country Census Year Sum	KH	CN	IN	IR	IQ	IL	JO	KG	MY	MN	NP	PK	PS	PH	TH	VN
			2008	1990	2004	2006	1997	1995	2004	1999	2000	2000	2001	1998	2007	2000	2000	2009
41	NATION	Country of citizenship	8	.	.	1	1	.	1	1	1	1	1	.	.	.	1	.
42	EMPSECT	Sector of employment	8	1	.	1	.	1	.	1	.	.	.	.	1	1	.	1
43	MGYRS1	Years residing in current	8	1	.	.	1	1	.	1	.	1	1	.	.	.	1	.
44	DISABLE	Disability status	8	1	.	.	1	1	.	1	.	1	1	.	.	1	.	1
45	DISEMP	Employment disability	8	.	1	1	.	.	.	1	.	1	1	1	.	.	.	1
46	DISBLND	Blind or vision-impaired	8	1	.	.	1	1	.	1	.	1	1	.	.	1	.	1
47	BRTHLYR	Number of births last year	7	1	1	.	1	1	.	.	.	.	1	.	1	.	1	.
48	ISCO88A	Occupation, ISCO-1988, 3-digit	7	1	.	.	1	.	.	1	.	1	1	.	.	1	1	.
49	MGRATE5	Migration status, 5 years	7	.	1	.	.	.	1	.	.	1	1	1	.	.	1	.
50	DISDEAF	Deaf or hearing-impaired	7	1	.	.	1	1	.	.	.	1	1	.	.	1	.	1
51	Composite	Place of residence 5/10 years	7	.	1	.	.	1	.	.	1	1	1	.	.	1	.	1
52	Composite	Ethnicity (national definition)	6	.	1	.	.	.	.	.	1	1	1	.	.	1	.	1
53	Composite	Mother tongue (national)	6	1	.	.	.	1	.	1	.	.	1	1	.	1	.	1
54	MGCAUSE	Reason for migration	6	1	1	.	1	1	.	.	.	.	.	.	1	.	1	.
55	DISMUTE	Mute	6	1	.	.	1	1	.	.	1	.	1	.	.	1	.	.
56	DISMNTL	Mental disability	6	1	.	.	1	.	1	.	1	.	.	.	.	1	.	1
57	Composite	Place of previous residence	6	1	.	.	.	1	.	1	.	1	.	.	1	.	1	.
58	CHBORNF	No. female children ever born	5	1	1	.	1	1	.	.	.	.	1	.	.	.	.	.
59	CHBORNM	No. male children ever born	5	1	1	.	1	1	.	.	.	.	1	.	.	.	.	.
60	CHSURVF	No. female children surviving	5	1	1	.	1	1	.	.	.	.	1	.	.	.	.	.
61	CHSURVM	No. male children surviving	5	1	1	.	1	1	.	.	.	.	1	.	.	.	.	.
62	YRSCHL	Years of schooling	5	1	.	.	.	.	.	.	.	.	1	.	.	1	1	1
63	MGRATEP	Migration status, previous	5	1	.	.	.	1	.	1	.	1	.	.	.	.	1	.
64	MGCTRY1	Country of previous residence	5	1	.	.	.	1	.	1	.	1	.	.	.	.	1	.
65	MARRNUM	Number of marriages or unions	4	.	.	.	1	1	1	.	.	.	1	.	.	.	.	.
66	CHDEAD	Number of children dead	4	1	.	.	.	.	.	.	.	.	1	.	.	.	1	1
67	DISLOWR	Disability affecting lower	4	.	.	.	1	1	.	.	.	.	.	.	.	1	.	1
68	RESIDENT	Residence status: de facto, de	3	.	.	.	.	1	.	.	.	1	.	1	.	.	.	.
69	CONSENS	Consensual union	3	.	.	.	.	.	.	1	.	1	.	.	.	1	.	.
70	AGEMARR	Age at first marriage or union	3	.	.	.	.	1	.	.	.	.	1	.	1	.	.	.
71	LSTBSEX	Sex of last birth	3	.	1	.	.	.	.	.	.	.	1	.	.	.	.	1
72	AWAYCHD	Number of own children living	3	.	.	.	.	.	.	1	.	.	1	.	.	.	.	1
73	SPKENG	Speaks English	3	1	.	.	.	.	.	1	.	.	.	.	.	1	.	.
74	WRKMTHS	Months worked last year	3	1	.	.	.	1	.	.	.	.	1	.	.	.	.	.
75	MGRATE0	Migration status, 10 years	3	.	.	.	1	1	.	.	.	.	.	.	.	1	.	.
76	MGCTRY2	Country of residence 5 years	3	.	.	.	.	.	.	.	1	1	1	.	.	.	.	.
77	DISUPPR	Disability affecting upper	3	.	.	.	1	1	.	.	.	.	.	.	.	1	.	.
78	POLYGAM	Polygamous union	2	.	.	.	.	1	.	.	.	.	1	.	.	.	.	.
79	BTHSURV	Children surviving from births	2	.	.	.	1	.	.	.	.	.	.	.	1	.	.	.
80	DEADFEM	Number of female children	2	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.

Variable	Variable Label	Country Census Year Sum	KH	CN	IN	IR	IQ	IL	JO	KG	MY	MN	NP	PK	PS	PH	TH	VN
			2008	1990	2004	2006	1997	1995	2004	1999	2000	2000	2001	1998	2007	2000	2000	2009
81	DEADMAL	Number of male children dead	2	1	.	.	.	.	.	.	.	.	1	.	.	.	.	.
82	HOMECHD	Number of own children in hh	2	.	.	.	.	.	.	.	.	.	1	.	.	.	.	1
83	YRIMM	Year of immigration	2	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.
84	YRSIMM	Years since immigrated	2	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.
85	Composite	2 <sup>nd</sup> . Language (national	2	.	.	.	.	.	.	1	.	.	1	.	.	.	.	.
86	HRSWRK2	Hours worked per week,	2	.	.	.	.	1	1	.	.	.	.	.	.	.	.	.
87	INCWAGE	Wage and salary income	2	.	.	1	.	1	.	.	.	.	.	.	.	.	.	.
88	DISPSYC	Psychological disability	2	.	.	.	.	1	.	.	.	.	.	.	.	1	.	.
89	REFUGEE	Refugee status	2	.	.	.	.	.	.	1	.	.	.	.	1	.	.	.
90	MGFUTPH	Intended residence in 5 years,	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
91	MARRYR	Year of first marriage	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
92	MARRYR2	Year of first marriage,	1	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
93	LSTBMTH	Month of last birth	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
94	LSTBYR	Year of last birth	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
95	MORTMOT	Mortality status of mother	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
96	MORTFAT	Mortality status of father	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
97	HOMEFEM	Number of own female children	1	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
98	HOMEFEM	Number of own male children	1	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
99	AWAYFEM	Number of own female children	1	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
100	AWAYMAL	Number of own male children	1	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.
101	YRIMM2	Year of immigration,	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
102	YRSIMM2	Years since immigrated,	1	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.
103	YRSCHL2	Years of schooling, intervalled	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
104	LABORIL	Annual labor force, Israel	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
105	ISCO68A	Occupation, ISCO-1968, 3-digit	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
106	ESTABSZ	Size of work establishment	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
107	HRSFULL	Full-time or part-time work	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
108	HRSWRK1	Hours worked per week	1	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.
109	TRNWRK	Transportation to work or	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
110	OVERSEA	Overseas worker, Philippines	1	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
111	INCSELF	Self-employment income	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
112	INCWEL	Income from anti-poverty or	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
113	MGCTRY5	Country residence 10 yrs ago	1	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
114	DISORIG	Origin of disability	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Source: <https://international.ipums.org/international-action/variables/group/cons> .

**Appendix F. IPUMS-I Top 40 University/Research Institutions by Number of Extracts**

Rank	Institution	N	Rank	Institution	N
1	University of Michigan	742	21	University of North Carolina – Chapel Hill	203
2	Columbia University	701	22	Universite Montesquieu-Bordeaux IV, France	196
3	Universitat de Barcelona, Spain	615	23	University of California - San Diego	189
4	Harvard University	589	24	University of Utah	189
5	Inter - American Development Bank	499	25	World Health Organization, Switzerland	183
6	Arizona State University	495	26	University of Virginia	182
7	National University of Singapore, Singapore	467	27	Michigan State University	178
8	World Bank	408	28	Intl. Institute for Applied Systems Analysis, Austria	165
9	University of California - Berkeley	362	29	University of Sussex, U.K.	158
10	Universidade Federal de Minas Gerais, Brazil	314	30	London School of Economics, U.K.	157
11	University of Chicago	285	31	Dartmouth College	155
12	Universidad del Valle, Colombia	270	32	University of Guelph, Canada	148
13	Institute for Health Metrics & Evaluation	260	33	Institut de Recherche pour le Developpement, France	148
14	Princeton University	237	34	Banco de la Republica, Colombia	145
15	University of Wisconsin - Madison	234	35	Yale University	143
16	Brown University	229	36	University of Tübingen, Germany	143
17	University of Vienna, Austria	229	37	Org. of Economic Cooperation & Development, Fr.	140
18	University of Pittsburgh	227	38	Catholic University Leuven, Brussels	139
19	University of Delaware	213	39	Brigham Young University	138
20	El Colegio de México, México	214	40	University of Queensland, Australia	136

Source: IPUMS-International User Statistics Database, April 18, 2011 (excludes IPUMS's home, the University of Minnesota)