

2012 International Association for Official Statistics: “Getting Our Message Across – Strategies and Best Practices to Ensure the Use of Statistics in Decision Making”

12-14 September. Kiev, Ukraine <http://iaos2012.ukrstat.gov.ua/>

Paper Session: Enhancing Access to Information

Title: Ten Ways IPUMS-International Adds Value to Census Microdata¹

Authors:

Robert McCaa, Minnesota Population Center, USA

Lara Cleveland, Minnesota Population Center, USA

Albert Esteve, Center for Demographic Studies, Autonomous University of Barcelona, Spain

Steven Ruggles, Minnesota Population Center, USA

and Matt Sobek, Minnesota Population Center, USA

“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)

ABSTRACT. Many statistical offices recognize the need for enhancing access to census microdata. High costs, challenging risks, and low rewards are substantial obstacles to “going-it-alone”. An economical, essentially cost-free solution, endorsed by more than 90 National Statistical Offices, is offered by the IPUMS-International project. This paper discusses ten ways the project enhances access and adds value to census microdata—grouped into four categories: 1. statistical confidentiality (security, disclosure protections, managing access), 2. Integration (comprehensive source documentation, integrated metadata, integrated, pooled microdata, IPUMS-I constructed variables), 3. Dissemination (free, trans-border access, custom-tailored extracts), and 4. ethics (statistical transparency, academic freedom, reduction of risks of fraud/mis-representation, and sharing of research findings). Statistical agencies not yet participating in the IPUMS-International initiative are invited to do so. Those already participating are encouraged to entrust 2010 round census microdata to the project in a timely manner.

INTRODUCTION.

Researchers and policy makers need access to census microdata to analyze challenging social, demographic, and economic transformations (McCaa and Esteve 2009). Most modern statistical agencies now understand the value of disseminating census microdata. Nonetheless some remain reluctant to do so because opening up access is costly, risky and often disappointing. Entrusting management of dissemination to a third party, such as IPUMS-International, eliminates additional costs, reduces risk to near zero, and assures outstanding user

¹Research for this paper was funded in part by the National Institutes of Health of the United States, grant HD047283 European and Asian census microdata harmonization project (IPUMS-EurAsia).

satisfaction. Statistical agencies that insist on going it alone should reflect on the considerable costs, risks to institutional trust-worthiness and likely user dissatisfaction.

- **Costs.** The statistical office must draw the sample, confidentialize the microdata, prepare the metadata, and manage access. These tasks amount to a lot of work, particularly when the number of requests by users may amount to no more than a dozen or so per year.
- **Risks.** Few statistical offices are experienced in confidentializing census microdata and with only one census per decade there is little opportunity to gain experience. Then too, experience is easily lost when specialists are hired away, re-assigned or otherwise unavailable. Managing access to microdata is particularly tricky because expert knowledge is required. Most official statisticians are not eager to accept responsibility for such time-consuming task of vetting users, monitoring access and providing user support.

Statistical agencies that insist on managing access to microdata should heed the scandal experienced by the United States Census Bureau (Alexander, Davern and Stevenson 2010). Due to a programming mistake by a Bureau statistician, age reporting of the elderly was egregiously corrupted for a large fraction of cases in the public use files of the American Community Survey, the 2000 Census and the Current Population Survey. In all over a period of eight years, 144 datasets were corrupted before the errors were detected. Researchers only discovered the mistakes by comparing the confidentialized age distribution against the full-count non-confidentialized microdata. Much to the embarrassment of the Census Bureau, shortly before the 2010 census got underway, the story of the mistakes found its way on page one of the New York Times with a headline blaring “**Can You Trust Census Data?**”² Unfortunately when the botched anonymization was fixed, the error was compounded (Cleveland, McCaa, Ruggles, and Sobek 2012). Nonetheless, the Census Bureau is to be praised for its transparency. By facilitating researcher access to the non-confidentialized microdata—under exceedingly strict security—the errors could be detected, even if they were not properly corrected.

- **Dissatisfaction.** Typically statistical offices err on the side of safety not only by excessively confidentializing the microdata but also by denying access to all but the most persistent or influential researchers. This strategy may promote safety, but it also increases incentives for bootlegging copies. Inevitably data utility is sacrificed, and dissatisfaction is widespread for both the statistical agency and researchers.

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.”

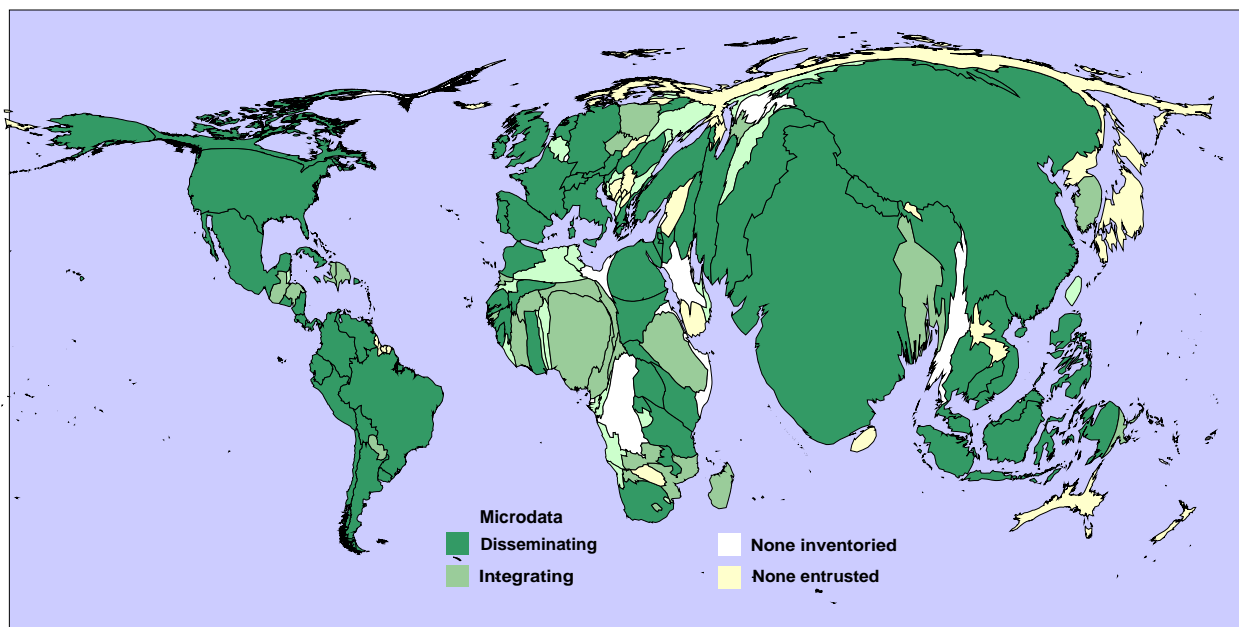
Some official statistical agencies, despite elaborate websites and microdata access policies which seemingly incorporate best practices, never even reply to requests for access. In a

² Wolfers, J. Can You Trust Census Data? Freakonomics blog. New York Times, February 2, 2010. <http://freakonomics.blogs.nytimes.com/2010/02/02/can-you-trust-census-data>.

virtual world, such sites may look great, but in reality they are doubly disappointing, particularly for researchers for whom “delay is the deadliest form of denial” (Parkinson, 1970).

IPUMS-International (www.ipums.org/international) offers a comprehensive, one stop solution to disseminating confidentialized microdata samples. Led by the Minnesota Population Center (MPC), the IPUMS-International initiative is a global collaboratory of National Statistical Offices (NSOs), universities, and major research centers. Regional partners, include the Integrated European Census Microdata project (Center for Demographic Studies, Autonomous University of Barcelona) and the African Integrated Census Microdata project (African Center for Statistics, Addis Ababa). The project provides state-of-the-art microdata and metadata services. Thanks to sustained funding by the National Science Foundation and National Institutes of Health – USA, IPUMS-International bears the costs—and assumes the responsibilities and risks—for recovering, archiving, confidentializing, integrating, and disseminating microdata worldwide (McCaa and Thomas 2009). A bona fide researcher—regardless of country-of-birth, residence or citizenship—with need to analyze microdata is typically granted access to the IPUMS-International website within five days of submitting a valid application. Microdata are disseminated free of cost to bona fide researchers and policy makers worldwide.

**Figure 1. Cartogram weighted by population size of countries participating in the IPUMS-International collaboratory, June 2012:
3 shades of green--disseminating (dark), integrating (medium), and negotiating (light).**



Begun in 1999, IPUMS-International enjoys the satisfaction of over 90 NSOs and more than 5,000 researchers in some 100 countries. Twenty-one European countries participate in the IPUMS initiative under a uniform memorandum of understanding (Appendix A and Figure 1). By means of a single licensing agreement bona fide researchers gain access to the microdata on a restricted basis for all countries participating in the project (Appendix B). As of June 2012, the project had completed the anonymization and integration of population microdata for 69 countries, totaling 212 samples and 480 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 first author of this paper. Participating statistical agencies are invited to entrust microdata for the 2010 round of censuses without undue delay to facilitate access to anonymized, integrated samples while the data remain timely. Doing so will increase stakeholders and, in the words of the President of Eurostat, open “up the value inherent in” census microdata (see epigraph).

The 36 European countries of one million or more inhabitants may be divided into 4 groups, according to cooperation with the IPUMS initiative (number in parentheses indicates microdatasets entrusted):

- **17 are integrated:** Armenia (1), Austria (4), Belarus (1), France (7), Germany (8—includes GDR, FRG and 4 microcensuses to be integrated), Greece (4), Hungary (4), Ireland (8), Italy (1), the Netherlands (3), Portugal (3), Romania (3), Slovenia (1), Spain (3), Switzerland (4), Turkey (3), and the *United Kingdom (6—4 are still being recovered).
- **2 are being integrated:** the Czech Republic (2), Poland (6—2 still being prepared).
- **2 are yet to entrust microdata:** Bulgaria and Ukraine.
- **15 are yet to cooperate:** Albania, Belgium (negotiating), Bosnia-Herzegovina, Croatia (negotiating), Denmark, Estonia (negotiating), Finland, Georgia, Latvia (negotiating), Lithuania, Macedonia, Moldova, Norway, Russia, Slovakia (negotiating), and Sweden.

The 2012 IPUMS launch integrated samples for Turkey (1985, 1990, and 2000) as well as six non-European nations: El Salvador, Indonesia, Mexico, Morocco, Nicaragua, and Uruguay. Annual launches are planned for successive years, integrating 2010 round census samples as expeditiously as they become available.

IPUMS-International seeks to enhance access to census microdata by 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 core of MPC staff—each with tens of thousands of hours experience in confidentializing, integrating, and managing access to microdata and metadata.

In the twenty-first century, passing along a copy of a census sample and codebook is *not* sufficient for high quality research. This paper discusses 10 value-added offered by IPUMS-International, IECM, and AICM to enhance research access to census microdata. The value-added 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 the confidentializing process is completed. Under Federal Law, MPC employees are subject to civil fine (up to US\$250,000) and 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 perhaps the first academic organization disseminating international census microdata to be 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 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 all participating statistical agencies as a group than for any single 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 is 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 a few 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 to the Colombian statistical agency (DANE). DANE not only reduced the geographical threshold to 20,000, but also harmonized the geo-statistical identifiers so that the entire series of census microdata samples for Colombia, 1964-2005, 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. Swapping is an exceedingly powerful method for assuring confidentiality at minimal cost in terms of utility. Swapping means that no one can allege with certainty that an individual or household has been identified. For the United Kingdom census of 2011, swapping is likely to be

an important method for protecting confidentiality not only of the microdata but also of tabular outputs (Frend et al, 2012).

In consultation with the national statistical office, we top-code some variables, global-recode others, selectively delete digits of hierarchical variables (occupation, industry, geography), or even suppress entire variables that are deemed too risky. 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. Managing 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 usage, protecting confidentiality, assuring security, enforcing strict rules of confidentiality, permitting scholarly publication, citing properly, threatening disciplinary action for violations, and reporting 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. A violation by a single user may suspend access to all users at that institution. Researchers at an embargoed institution may be sanctioned to undergo remedial training for the protection of human subjects so that the institution may regain its accreditation for handling sensitive microdata.

IPUMS-I procedures resolve 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.

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 users to jealously guard the microdata. Since complete datasets are not distributed on CDs or any other media, the temptation to share microdata with unauthorized individuals is greatly reduced.

Google Analytics suggests that the IPUMS registration form alone is a substantial deterrent to casual users. Over a recent twelve month period, 5,593 views of the registration page yielded only 1,057 completed applications. One reason for the large drop-off is that the registration form is a daunting deterrent to the statistically naive.

A qualified researcher, regardless of how much time is required to fill-out the form, will readily agree to these conditions and meticulously provide the requested information, while the unqualified—faced with identifying by name the Human Subjects Protection Committee of his or her institution, supervisor, and website listing the individual’s institutional affiliation as well as describing the research project for which the microdata are to be used—will not complete the form at all (see Appendix B). Incomplete forms are automatically rejected by the web page controls. It is impossible to submit an incomplete application for consideration.

Of the 1,057 completed applications noted above, a mere 46 were denied. Once the registration is submitted, applicants are carefully vetted to prevent access to researchers who are unqualified or who lack a research need. For a majority of denials the currently disseminated census microdata were not suitable for the proposed research. The daunting detail required to complete the form leads to self denial by the casual visitor. From this brief description of the IPUMS-International registration process it should be apparent that before an individual account is activated, as the epigraph advises, we do due diligence to confirm the identity and to establish the research bona fides of each applicant.

IPUMS disseminates extracts, custom-tailored to the precise research needs of each user. The average IPUMS extract last year consisted of a mere 35 variables, including 6 technical variables that are automatically included with each extract.

This contrasts with the practices of most statistical offices where microdata for a census are disseminated as single set, consisting of a data dictionary and an entire sample containing all variables and all person records. Typically, under this old-fashioned 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 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 who 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 essential to make good use of the microdata. Additional metadata regarding the organization, preparation work, and actual census fieldwork 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 succinct 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, questionnaire text, and non-harmonized variables used for

integration. Access to this information is through clickable hypertext on the IPUMS website. Comparability discussions summarize the most important similarities and differences in definitions for each variable, including country or census specific information signaling departures from standard practice. The purpose of these discussions is to highlight important contrasts. Clicking “Questionnaire 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 census questions and instructions to field workers for any combination of countries and census years. The MPC integration team applies XML tags to the census documents, associating variables in the microdata with the concepts in the text. The tagged material is then imported into the 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 analytical power became apparent, we harnessed the dynamic metadata system to the web-site, to permit public access to the metadata.

6. Integrated, pooled microdata. The principal benefit of IPUMS to researchers and NSOs alike is the integration of a complete, chronological series of microdata samples for each country—typically beginning with the earliest census for which microdata exist or are recoverable and continuing through the 2010 census round. Many NSOs provide 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. Most statistical offices, at best, construct a census sample and a data dictionary. Five or ten years later, at the next census, the process is repeated with little guidance to users on how to compare microdatasets from successive censuses.

IPUMS pools the microdata for each individual request into a single file, regardless of the number of censuses selected. Metadata are custom-tailored to each researcher’s request and available also as a single, searchable file in DDI and/or text format. Thus, the researcher is empowered to analyze a series of census years and even multiple countries in a single data file, facilitating comparative analysis over time and space.

We must reiterate that the IPUMS project does *not* disseminate copies of census files as they are entrusted by national statistical offices. Instead high-precision census samples are confidentialized (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. This 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 microdata and metadata while losing no meaningful information. This is challenging because to make the microdata 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 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 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, for children too young to attend or others to whom the question was not addressed, a zero, not in universe code, is assigned. 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. The third digit distinguishes between technical and general or other tracks common to two or more countries. 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 15 countries, represented by the most recent census integrated into the IPUMS-I database. A similar table for any combination of countries or censuses is generated from the IPUMS-I home page by first clicking “Browse and Select Data”. At this point, click “Select Samples,” and click and submit your selections, if you wish to avoid any annoyance from displaying metadata for all samples in the database. To select the education attainment variable, mouse-over “Person,” click “Education,” and on the “EDATTAN” line click “Codes”. The display will indicate the general, one digit codes for the “Educational Attainment, international recode” variable for all countries and census years selected. “X” indicates that the corresponding code is available for the country and census indicated. Clicking “detailed codes” generates the detailed, three digit codes in Appendix C. To obtain simple, un-weighted frequencies for each code in the selected samples, click “Case-count view” as seen in Appendix C. All IPUMS-I metadata are open to everyone. No registration is required.

7. IPUMS-I Value-added variables. The IPUMS team has decades of experience in using census microdata. Based on this experience we have developed more than thirty value-added variables to facilitate research use. These may be grouped into three types of variables: technical, summary and pointers.

a. Technical variables consist of the following: 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 identify co-resident spouses (SPLOC) and parents (MOMLOC, POPLOC) of children: 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). Pointer variables are exceedingly useful to facilitate the study of characteristics of an individual by those of a co-resident spouse or parent (for an example, see Esteve, Garcia and McCaa, 2011).

III. Dissemination

8. 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 rises to two-fifths.³ For many professional demographers—and many social science researchers and policy analysts in general—trans-border access is essential if they are to conduct research using census microdata for 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.

Some national statistical agencies attempt to satisfy user demand for microdata by offering access to an on-line tabulator. It should be readily apparent why tabulators are no substitute for trans-border access to microdata. Tabulators lack tools for statistical analysis. Few tabulators offer any means for recoding or transforming the data. None have tools for generating statistics based on matches between parents and children, household reference person and householders, or spouses (see section "7.c", above).

9. Custom-tailored extracts. With IPUMS-I no two extracts are alike. Each extract is custom-tailored by the researcher. Selections are made by a series of point-and-click screens. To facilitate the selection process, metadata are readily available to surf the documentation in any sequence desired without interrupting the extraction process (see section "6" above). To construct an extract 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 both the individual as well as all co-resident persons enumerated in the selected household)
- and sample density (either as a percent or number of cases).

The IPUMS extract engine fulfils the request by generating a single, pooled dataset containing the microdata, the corresponding set of DDI compatible metadata as well as user-selected SPSS, SAS or STATA system files and their corresponding codebooks. 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.

³ Statistics communicated to the author by the Secretariat of the International Union for the Scientific Study of Population, September 14, 2011.

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 media, the inclination of researchers 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, with a single click by the user, the extract engine will include all individuals co-resident in the same household with the selected individual.

b. **Attach characteristics.** By simply clicking the attach characteristics “change” option, the extract engine attaches variables 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, October 2011), it was easy to generate a dataset from the IPUMS-I website for 51 countries representing 12 million foreign born individuals married or in union with co-resident native born spouses (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 number or percent of persons or households. 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 of the desired size. The appropriate weights (expansion factors) are automatically computed on-the-fly and included in the extract.

d. **Usage.** IPUMS-International liberates microdata from the confines of statistical offices to permit statisticians, demographers, economists, social scientists and policy makers to extract un-dreamed of value from the microdata without cost (Meyer, McCaa, and Lam, 2011). Many national researchers were unable to access the microdata of their country until IPUMS-International opened the door. Other researchers faced troublesome delays, exorbitant fees, etc.

In 2011, 8,048 extracts were made from the IPUMS-International website, totaling 40,142 samples and 281,640 variables. The average number of extracts per country was almost 150 samples for the 55 countries represented in the database for the full year (Table 2). Nonetheless, usage by country varied greatly. The smallest number of extracts, 127, was registered for the 1997 census of Palestine. The greatest number, 712, was registered for the sample of the 2000 census of Brazil.

Table 3 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). The fact that the top most requested variable was downloaded almost six times more frequently than the 32nd most common variable shows that researchers are indeed carefully constructing their extracts, instead of willy-nilly requesting all variables for

every sample. As a matter of fact, the huge size of the samples and the requirement that users must click a box to select a variables force researchers to make choices.

A most striking finding is that 40 account for 60% of the variables requested and among these are sixteen IPUMS constructed variables, four of which, the pointer or “LOC” variables, are unique to IPUMS. Pointer 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 as well as 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 IPUMS-I household sample (Sobek and Kennedy, 2009).

The IPUMS-I top 33 institutions in terms of data usage includes many of the world’s premier universities and research organizations (see Appendix D), scattered across fourteen countries. In 46 countries, we find a total of 501 institutions with researchers making ten or more extracts. (In addition, for the United States, there are 295 institutions at the same 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 a dozen countries with microdata in the IPUMS-I database but as yet no national researchers use them. The 12 are: Armenia, Belarus, Guinea, Iraq, Jordan, Kyrgyzstan, Mali, Mongolia, Nepal, Rwanda, Saint Lucia, and Slovenia. It should be noted that 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. The NSO shoulders the responsibility for any such dissemination.

IV. Enhancing ethical access to census microdata.

10. Enhancing ethical access through statistical transparency, academic freedom, reduction of fraudulent research findings and facilitating access to research results.

10a. Statistical transparency. Entrusting census microdata to IPUMS-International demonstrates to the world that the NSO has nothing to hide and is willing to subject its most treasured microdata to even the most rigorous, external scrutiny.

10b. Academic freedom. Microdata disseminated by IPUMS-International are under no prior restraint for purposes of publication or dissemination of results—other than for reasons of statistical confidentiality. NSOs that limits access should not be surprised that researchers may be inclined to slant their results to curry favor and assure continuation of privileged access. Principled researchers shun such tainted data not only to avoid being perceived as tamed but, more importantly, to promote good science and to reduce fraud.

10c. Reduce fraud or the temptation to exaggerate research findings. Since all researchers have equal access to the microdata disseminated by IPUMS-International, results may readily be replicated to test the findings of others. IPUMS maintains a registry so that if

there is any doubt about the veracity of results, an extract may be replicated so that a third party may check a researcher's analysis. This is a powerful stimulus to promote good quality analysis and reduce, if not eliminate, reporting of fraudulent results. These issues are of grave concern to the National Science Foundation and the National Institutes of Health.

10d. Research results. Researchers post links to their findings in the IPUMS open-access bibliography: <http://bibliography.ipums.org/>. The bibliography currently lists over 500 entries under the IPUMS-International tag.

Reflections.

To meet the needs of researchers and policy makers on a global scale, the IPUMS-International project requires a formidable range and amount of metadata and microdata. Nonetheless the return on investment is substantial. Statistical offices are relieved of many of the most burdensome tasks and responsibilities for disseminating microdata. Moreover, by entrusting the responsibility for managing access to IPUMS-International, National Statistical Offices gain safety in numbers. IPUMS-International is now relied upon by most of the world's statistical offices—96 of some 157 countries numbering more than one million inhabitants (plus two smaller countries—Fiji Islands and Saint Lucia). The isolated statistical office that disseminates microdata on an ad hoc basis incurs substantial risks and responsibilities as well as significant costs in human resources—all for a relatively small return with respect to number of users. The IPUMS-I project offers substantial economies of scale with the highest standards of security.

When we began a decade ago, we dreamed of integrating samples for 21 countries over a period of ten years. Thanks to the generous cooperation of National Statistical Offices and undreamed of technological innovations, the number of countries with integrated microdata has more than tripled to 69, while work continues on those not yet completed. The number of users and the amount of use also far exceeds our expectations.

For the next 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 continued success. Microdata for 2010 round of censuses are fast-tracked for integration without due delay. Microdata received by September 1 are processed for launch in nine months, for June 1st of the following year. To date, all the IPUMS-International partners are eager to entrust microdata as soon as 2010 round census data processing is successfully concluded. Samples from the 2010 censuses of Indonesia and Mexico were launched in June 2012—less than two years from enumeration day.

Participating statistical agencies are invited to entrust metadata and microdata for the 2010 census round at their earliest convenience. Those not yet participating in the IPUMS initiative 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 as 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

Cleveland, L; McCaa, R.; Ruggles, S.; and Sobek, M. 2012. "When Excessive Perturbation Goes Wrong and Why IPUMS-International Relies Instead on Sampling, Suppression,

Swapping, and Other Minimally Harmful Methods to Protect Privacy of Census Microdata,” in J. Domingo-Ferrer and E. Magkos (Eds.): *Privacy in Statistical Databases 2012*. Springer, Heidelberg.

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.

Frend, J., C. Abrahams, A. Forbes, P. Groom, K. Spicer, C. Tudor and P. Youens. 2012. “Statistical Disclosure Control in the 2011 UK Census: Swapping Certainty for Safety.” *ESSnet Workshop on Statistical Disclosure Control of Census Data*, Luxembourg, Apr. 19-20.

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 Databases 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>

Parkinson, C. Northcote. 1970 *The Law of Delay: Interviews and Outerviews*. London: John Murray.

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 .

Trewin, D. 2007. “A Review of IPUMS-International.” Unpub. http://www.hist.umn.edu/~rmccaa/IPUMSI/trewin_ipums_report.pdf

Table 1. Rank of the Top Five and all European Countries plus Canada and the USA by Number of Extracts for the 2000 round census (statistics for calendar year 2011)

Rank	Country	Sample %*	Variables (n)*	Census years of samples	Extracts
1	Brazil	5	106	1960, 70, 80, 91, 2000	712
2	Mexico	10	120	1960p, 70, 90, 95, 2000, 05	626
3	United States	5	92	1960, 70, 80, 90, 2000, 05	554
4	Colombia	10	120	1964p, 72, 85, 93, 2005	516
5	South Africa	10	108	1996, 2001, 2007	428
7	Canada	2.5	59	1971p, 81p, 91p, 2001p	409
9	France	33	94	1962, 68, 75, 82, 90, 99, 06	380
10	Spain	5	99	1981, 91, 2001	366
13	Greece	10	89	1971, 81, 91, 2001	327
18	Austria	10	75	1971, 81, 91, 2001	310
25	Italy	5	81	2001	285
26	Portugal	5	96	1981, 91, 2001	283
29	Romania	10	97	1976, 92, 2002	272
30	Switzerland	5	79	1970, 80, 90, 2000	266
32	United Kingdom	3	47	1991, 2001p	263
38	Hungary	5	74	1970, 80, 90, 2001	222
42	The Netherlands	1	33	1960p, 71p, 2001p	211
45	Slovenia	10	80	2002	185
48	Belarus	10	84	1999	179
Total samples extracted for 55 countries (162 samples) available from January 1, 2011.					8,048
*2000 round census; refers to all integrated variables, including IPUMS constructed variables. “p” = person sample; all other samples are of households					

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 (Italy, 2006).

Letter of Understanding	
Integrated Public Use Microdata Series International and L'ISTITUTO NAZIONALE DI STATISTICA (ISTAT)	
<p>Purpose. The purpose of this letter is to specify the terms and conditions under which metadata and microdata produced by L'ISTITUTO NAZIONALE DI STATISTICA shall be distributed by Integrated Public Use Microdata Series International of the University of Minnesota.</p> <ol style="list-style-type: none"> Ownership. ISTAT is the owner and licensee of the intellectual property rights (including copyright) in the metadata and microdata of Italy acquired by the University of Minnesota to be distributed by Integrated Public Use Microdata Series International. Use. These data are for the exclusive purposes of teaching, scientific research and publishing, and may not be used for any other purposes without the explicit written approval, in advance, of ISTAT. Authorization. To access or obtain copies of integrated microdata of Italy from Integrated Public Use Microdata Series International, a prospective user must first submit an electronic authorization form identifying the user (i.e., principal investigator) by name, electronic address, and institution. The principal investigator must state the purpose of the proposed project and agree to abide by the regulations contained herein. Once a project is approved, a password will be issued and data may be acquired from servers or other electronic dissemination media maintained by Integrated Public Use Microdata Series International, ISTAT, or other authorized distributors. Once approved, the user is licensed to acquire integrated metadata and microdata of Italy from Integrated Public Use Microdata Series International or other authorized distributors. No titles or other rights are conveyed to the user. Restriction. Users are prohibited from using data acquired from the Integrated Public Use Microdata Series International or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise. Confidentiality. Users will maintain the absolute confidentiality of persons and households. Any attempt to ascertain the identity of a person, family, household, dwelling, organization, business or other entity from the microdata is strictly prohibited. Alleging that a person or any other entity has been identified in these data is also prohibited. Security. Users will implement security measures to prevent unauthorized access to microdata acquired from Integrated Public Use Microdata Series International or its partners. Publication. The publishing of data and analysis resulting from research using metadata or microdata of Italy is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite ISTAT 	<p>and Integrated Public Use Microdata Series International as the sources of the data of Italy, and to indicate that the results and views expressed are those of the author/user.</p> <ol style="list-style-type: none"> Violations. Violation of the user license may lead to professional censure, loss of employment, and/or civil prosecution. The University of Minnesota, national and international scientific organizations, and ISTAT will assist in the enforcement of provisions of this accord. Sharing. Integrated Public Use Microdata Series International will provide electronic copies to ISTAT of documentation and data related to its integrated microdata as well as timely reports of authorized users. Jurisdiction. Disagreements which may arise shall be settled by means of conciliation, transaction and friendly composition. Should a settlement by these means prove impossible, a Tribunal of Settlement shall be convened which will rule upon the matter under law. This Tribunal shall be composed of an arbitrator, which shall be selected by the ICC International Court of Arbitration. This agreement shall be governed by, and construed in accordance with, generally accepted principles of International Law. Order of Precedence. In the event of a conflict between a term or condition of this Letter of Understanding and a term or condition of any Contract, to which this Letter of Understanding is attached, the term or condition in this Letter of Understanding shall prevail. <p>Date: <u>2/21/06</u></p> <p>Signed: <u>[Signature]</u> Regents of the University of Minnesota By: Kevin J. McKoskey, Sponsored Projects Administration</p> <p>Date: <u>23.01.2006</u></p> <p>Signed: <u>[Signature]</u> Rev. Jan. 27, 2005</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 Page 1 of 1</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>Legal notice: Submission of this application constitutes a legally binding agreement between the applicant, the applicant's institution, the University of Minnesota, and the relevant official statistical authorities. Submitting false, misleading or fraudulent information constitutes a violation of this agreement. Misusing the data by violating any of the conditions detailed below also constitutes a violation of this agreement and may lead to professional censure, loss of employment, or civil prosecution under relevant national and international laws, and to sanctions against your institution, at the discretion of the University of Minnesota and the official statistical authorities.</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;">PERSONAL INFORMATION</p> <p style="text-align: center;">.....</p> <p style="text-align: center;">INSTITUTIONAL AFFILIATION</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. Note: The license is valid for one year and may be renewed.</p> <p style="text-align: center;">Data must not be redistributed without authorization.</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>The microdata are intended only for scholarly research and educational purposes.</p> <ul style="list-style-type: none"> <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. <p>Commercial use and redistribution of the microdata is strictly prohibited.</p> <ul style="list-style-type: none"> <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. <p>Use of the microdata must follow strict rules of confidentiality.</p> <ul style="list-style-type: none"> <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. <p>The microdata must always be safely secured.</p> <ul style="list-style-type: none"> <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.
--	---

Name of institution or employer	<input type="text"/>
Your email address at institution (*)	<input type="text"/>
Web link showing your affiliation with institution (*)	<input type="text"/>
Email address of employer, supervisor, or instructor (*)	<input type="text"/>
Phone number of institution (*)	<input type="text"/>
Does your institution have an Institutional Review Board (IRB), or Office for Human Subject Protections, Professional Conduct or similar committee?	
	<input type="radio"/> No
	<input type="radio"/> Yes; Name of board or office
	<input type="text"/>
RESEARCH PROJECT	
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.	
.....	
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. (*)	
<input type="text"/>	

Scholarly publications are permitted, and must be cited appropriately.

- The publishing of research results based on IPUMS-International microdata is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite Integrated Public Use Microdata Series-International 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.
- Any violation of this license agreement will result in disciplinary action, including possible loss of employment.**
- 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.
 - User agrees to notify ipums@pop.umn.edu regarding errors in the data.**

Appendix C. Educational attainment: IPUMS integrated codes for most recent census sample: 15 countries																
Case counts indicate the absolute number of persons in the sample with the corresponding code																
Code	Label	Belarus 1999	Canada 2001	France 2006	Germany 1987	Greece 2001	Hungary 2001	Ireland 2006	Italy 2001	Portugal 2001	Romania 2002	Slovenia 2002	Spain 2001	Switzer 2000	Turkey 2000	USA 2000
0	NIU (not in universe)	195,221	155,094	3,500,985	889,475	60,222	35,708	104,603	160,897	.	223,986	26,186	9,429	61,858	398,619	570,821
100	LESS THAN PRIMARY COMPLETED	2,888	14,056	.	.	34,748	29,783	.	317,651	136,635	.	10,175	.	.	1,029,257	.
110	No schooling	.	.	327,422	78,667	.	473,420	12,776	.	655,973
120	Some primary	.	.	2,223,735	.	124,318	64,846	.	.	.	1,456	1,243,973
130	Primary (4 years)	143,511	384,871
PRIMARY COMPLETED, LESS THAN SECONDARY																
Primary completed																
211	Primary (5 years)	288,939	719,220	.	.	.	502,742	.	1,129,667	.
212	Primary (6 years)	.	49,390	2,425,484	1,198,197	300,134	145,691	57,791	.	142,932	85,730	1,429,549
Lower secondary completed																
221	General and unspecified track	.	138,894	1,577,008	332,591	108,576	.	71,586	850,317	.	528,362	39,227	462,072	.	212,962	1,488,337
222	Technical track	.	.	3,504,160	7,389
SECONDARY COMPLETED																
General or unspecified track																
311	General track completed	207,853	91,053	2,687,840	159,502	213,311	108,812	95,777	600,717	57,048	122,104	41,637	199,858	77,527	312,107	3,534,504
312	Some college/university	.	28,393	19,065	3,391	7,717	2,955,657
320	Technical track	177,645
321	Secondary technical degree	.	22,605	.	123,152	19,591	78,940	.	129,727	.	576,739	41,066	159,802	134,987	99,849	.
322	Post-secondary technical education	.	163,114	.	211,963	46,434	.	33,809	10,513	.	56,970	7,732	114,012	27,674	.	.
400	UNIVERSITY COMPLETED	112,692	138,456	3,726,653	128,093	121,550	46,722	60,668	182,632	33,509	128,632	11,822	117,939	23,033	168,195	2,202,652
999	UNKNOWN/MISSING	5,468	.	.	117,251	.	.	16,080	.	.	28,463	1,787	.	26,231	681	.
Source: https://international.ipums.org/international-action/variables/EDATTAN#codes_section																

Appendix D. IPUMS-I Top 33 University/Research Institutions in 2011

<u>Institution</u>	<u>Extracts</u>
Columbia University	558
University of Hong Kong (Hong Kong, SAR)	309
University of Michigan	296
Arizona State University	256
Institute for Health Metrics & Evaluation (Seattle)	245
Universidad Nacional de La Plata (Argentina)	205
Harvard University	205
National University of Singapore (Singapore)	198
Centre for Demographic Studies (Barcelona, Dartmouth College)	185
Dartmouth College	177
Baruch College	152
University of Virginia	141
Vienna Institute of Demography (Austria)	138
United Nations-Habitat (Nairobi, Kenya)	134
University of Stirling (Scotland, UK)	110
University of Pennsylvania	108
Institute of Political Studies (Paris, France)	107
Vanderbilt University	105
University of Colorado at Boulder	90
University of Cape Town (South Africa)	88
University of Chicago	86
University of Guelph (Canada)	85
World Bank	81
University of California at Berkeley	77
Brigham Young University	76
Pompeu Fabra University (Barcelona, Spain)	75
Pew Research Center	74
Inter-American Development Bank	73
University of Groningen (Netherlands)	72
Kenyon College	70
London School of Economics (UK)	62
Bocconi University (Milan, Italy)	59
Indiana University	57

Source: IPUMS-International User Statistics Database, January 1, 2012 (list excludes IPUMS' home, the University of Minnesota)