Distr.: General
30 April 2012

English only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Fourteenth Meeting
Geneva, 24-25 May 2012
Item 3 of the provisional agenda
Access to microdata

IPUMS-International and Integrated European Census Microdata Projects Reduce Risks of Managing Trans-border Access and Add Significant Value[1]

Note by the Minnesota Population Center, Minneapolis, MN USA and Centre d'Estudis Demogràfics, Autonomous University of Barcelona, Barcelona, Spain

"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)

I. Introduction

1.   Researchers and policy makers need access to census microdata to analyse social, demographic, and economic change in the twenty-first century (McCaa and Esteve 2009). Most modern statistical agencies now understand the value of disseminating census microdata.   Nonetheless some remain reluctant to do so because undertaking sole responsibility is risky, costly and often disappointing.   Entrusting management of dissemination to a third party, such as IPUMS-International and its partner the Integrated European Census Microdata (IECM) project, eliminates costs, reduces risk to a minimum, and maximizes user satisfaction.   Statistical agencies that insist on "going it alone" should reflect on the risks, costs, and user satisfaction/dissatisfaction.

- **Risks of "going it alone".** 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 the specialist is hired away, re-assigned or otherwise unavailable. Managing access to microdata is particularly tricky because expert knowledge is required. Few official statisticians are eager to accept responsibility for the time-consuming task of vetting users, monitoring access and providing user support. Statistical agencies that insist on managing access to microdata should heed the lessons from the scandal caused 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 2000 Census, American Community Survey, and the Current Population Survey. In all 144 microdatasets were corrupted. Researchers were able to discover the errors 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 scandal found its way to the front pages of the New York Times. The Census Bureau is to be praised, however, for its transparency. By facilitating access to the non-confidentialized microdata—under exceedingly strict security—these errors could be detected.

- **Costs of "going it alone".** 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.

- **Dissatisfaction from "going it alone".** Typically statistical offices err on the side of safety by excessively confidentializing the microdata and denying access to all but the most persistent researchers. This strategy assures safety, but sacrifices utility. The result is widespread dissatisfaction for both the statistical agency and researchers. Consider the following complaints—addressed to a major statistical office (ONS-UK) with over two decades experience in disseminating census microdata:

  o "I haven't used the data; the bureaucracy was just too slow to get much use out of it."

  o "[Access] is unbelievably bureaucratic and difficult – this discourages people from using it. It took me 6 months to get the data."

2.   Some official statistical agencies, despite elaborate websites and microdata access policies which seemingly incorporate best practices, neglect to reply to requests for weeks or even months, if at all. In a virtual world, such sites may look great, but in reality they are doubly disappointing, particularly for those for whom "delay is the deadliest form of denial."

3.   **IPUMS-I (www.ipums.org/international) and IECM (www.iecm-project.org) offer a comprehensive, one stop solution to managing access to 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. For Europe, the Centre d'Estudis Demogràfics (CED), Autonomous University of Barcelona (www.iecm-project.org) produces optimized metadata and organizes networking activities for the region. The IPUMS and IECM projects provide 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.

4.   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/IECM initiatives under a uniform memorandum of understanding (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.   As of this writing, the project has 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 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.

Figure 1.   IPUMS-International cartogram (weighted by population size) disseminating (dark green), integrating (medium), and negotiating (light).
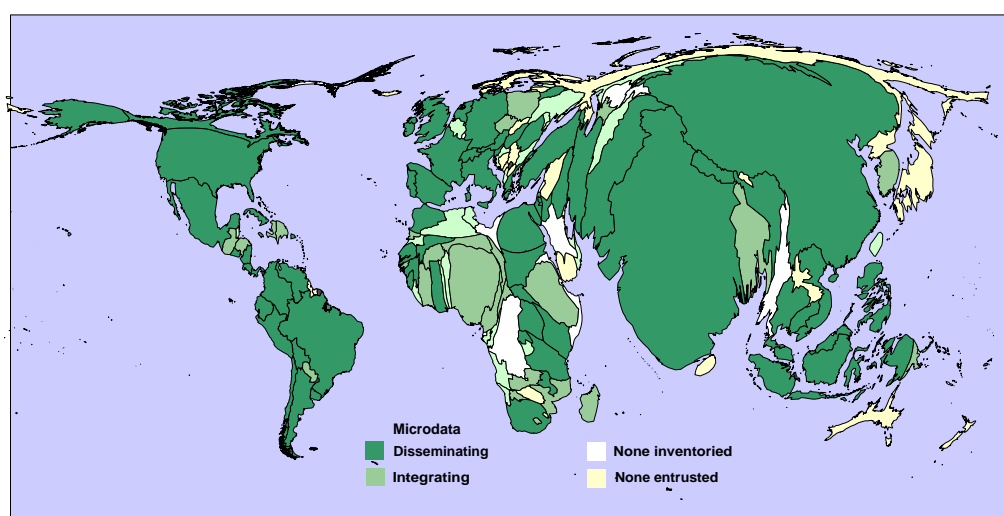
5.   **Invitation to participate and entrust 2010 round census microdata.**   Statistical agencies not currently cooperating with the initiative are invited to contact the author of this paper to discuss questions and concerns.   Agencies already participating are invited to entrust microdata for the 2010 round of censuses without undue delay.   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 status of 36 most populous European countries regarding participation in the IPUMS/IECM initiatives is as follows:

- **17 are integrated**: Armenia (1 census), Austria (4), Belarus (1), France (7), Germany (8—includes GDR, FRG and 4 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 (4—2 still being recovered).

- **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, Lithuania, Macedonia, Moldova, Norway, Russia, Slovakia (negotiating), and Sweden.

6.   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 core of MPC staff with tens of thousands of hours each in confidentializing, integrating, and managing access to microdata and metadata.

7.   In the twenty-first century, passing along a copy of the source microdata and a codebook is _not_ sufficient for high quality research. This paper discusses twelve value-addeds offered by IPUMS-International and the IECM project

II.   Statistical Confidentiality and Security

A.   Microdata Security and Statistical Confidentiality

8.   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.   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 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. (Trewin 2007).

B.   Statistical disclosure control protections

9.   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.

10.   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.

11.   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 state with certainty that an individual or household has been identified. For the United Kingdom census of 2011, swapping is the principal means of protecting confidentiality not only of the microdata but also for tabular outputs (Frend et al, 2012).

12.   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 a entire variable. 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.

C. Managing Restricted Access to Census Microdata

13. 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, individually, 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.

14. Google Analytics suggests that the IPUMS registration form alone is a substantial deterrent to unqualified 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. The applicant must agree to abide by each of ten stringent restrictions on conditions of use—prohibiting redistribution, restricting to scholarly use, prohibiting commercial user, protecting confidentiality, assuring security, enforcing strict rules of confidentiality, permitting scholarly publication, citing the microdata properly, threatening disciplinary action for violations, and reporting of any errors discovered.

15. 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. Incomplete forms are automatically rejected by the web page controls. It is impossible to submit an incomplete application for consideration.

16. A mere 46, of the 1057 completed applications. were denied. 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. Once the registration is submitted, applicants are carefully vetted to prevent access to researchers who are unqualified or who lack a research need. The application 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.

17. 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. A violation by a single user will suspend access to all users at that institution, until researchers undergo remedial training the protection of human subjects and the institution regains its accreditation for handling sensitive microdata. IPUMS-I is the only academic organization disseminating international census microdata that is cited as good practice by the Conference of European Statisticians Task Force on Managing Statistical Confidentiality and Microdata Access (2007).

18. 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. The average IPUMS extract in 2011 consisted of a mere 35 variables, including six technical variables that are automatically included with each extract.

19.   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 the 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

20.   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

III.   Integration

D.   Comprehensive source metadata

21.   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 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.

E.   Integrated, DDI compatible metadata

22.   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. Integrated metadata in English 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.

  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. 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. 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 "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.

c) **Example: "Educational Attainment".** 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. Table 1 illustrates the detailed three-digit coding schemes for the educational attainment variable for 12 European countries (represented by its ISO 3166 code) and census year (century omitted). 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 zero ("not-in-universe") 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. The third digit distinguishes between technical and general or other tracks common to two or more countries

F.   Integrated, pooled microdata

23.   The principal benefit of IPUMS to researchers and NSOs alike is the integration of a complete series of microdata samples for each country—typically beginning with the earliest census for which microdata exist or are recoverable. IPUMS "pools" the microdata for each individual request into a single file. Metadata are custom-tailored to each researcher's request and available as a single, searchable file in DDI or text format. Thus, the researcher is empowered to analyze multiple census years and even multiple countries in a single data file, facilitating comparative analysis over time and space.

G.   Integrated boundary files

24.   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.

H. IPUMS-I Value-added variables

25.   . 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, grouped into three types: technical, summary and pointer 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).

IV. Managing Trans-border access to census microdata

I. Trans-border access

26.   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.[2] 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.

J. Custom-tailored extracts

27.   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. To order an extract the

---

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

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), and sample density (either as a percent or number of cases).

28. 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 codebooks for SPSS, SAS or STATA system files. 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.

29. 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.

30. The IPUMS extract engine adds even more value to each extract by means of three unique tools: select cases (filter the extract to select precisely the cases of interest), attach characteristics (attach any variable of the father, mother, spouse or household head to the record of the corresponding person), and customize sample size (reduce samples from the maximum to any percentage or absolute number of households or persons).

K. Usage.

31. 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.

32. In 2011, 8,048 extracts were made from the IPUMS-International website, totalling 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 sample of the 2000 census of Brazil.

33. Of the 32 most commonly extracted variables. The top eight consist of 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).

34. 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. 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 heavy usage of the "LOC" variables indicates their great importance for analyzing individuals in relation to characteristics of their spouses, mothers and fathers—an exceedingly important value-added to each household sample (Sobek and Kennedy, 2009).

35. The IPUMS-I "top 40" institutions in terms of data usage includes many of the world's premier universities and research organizations, scattered across fourteen countries.

36.   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 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).

L.   2010 round census samples.

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

V.   Reflections

38.   The IPUMS project offers substantial returns at minimal risk. Statistical offices are relieved of many of the most burdensome tasks and responsibilities in disseminating microdata. 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 countries with fewer—Fiji Islands and Saint Lucia). The isolated statistical office that disseminates microdata on an ad hoc basis incurs substantial risks as well as significant costs in human resource—all for a relatively small return with respect to number of users. The IPUMS project offers substantial economies of scale with the highest standards of security.

39.   Statistical agencies not yet participating in the IPUMS initiative are invited to consider doing so. Participating agencies are invited to entrust metadata and microdata for the 2010 census round at their earliest convenience.

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.

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. 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: UNECE, 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 .

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

Table 1. Educational attainment: IPUMS-I harmonized codes for 12 European Countries most recent sample
"X" indicates that the code is present in the respective sample

Code	Label	BY 99	FR 06	DE 87	GR 01	HU 01	IE 06	IT 01	PT 01	RO 02	SI 02	ES 01	CH 00
0	NIU (not in universe)	X	X	X	X	X	X	X	·	X	X	X	X
100	LESS THAN PRIMARY COMPLETED	X	·	·	X	X	·	X	X	·	X	·	·
110	No schooling		X	·	·	·	·	·	·	X	·	X	X
120	Some primary		X	·	X	X	·	·	·	X	·	·	·
130	Primary (4 years)		·	·	·	·	·	·	X	X	·	·	·
	PRIMARY COMPLETED, LESS THAN SECONDARY												
	Primary completed												
211	Primary (5 years)	X	·	·	·	·	·	X	·	·	·	X	·
212	Primary (6 years)		X	X	X	X	X	·	X	·	·	·	·
	Lower secondary completed												
221	General and unspecified track		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
312	Some college/university		·	·	·	·	·	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
400	UNIVERSITY COMPLETED	X	X	X	X	X	X	X	X	X	X	X	X
999	UNKNOWN/MISSING	X	·	X	·	·	X	·	·	X	X	·	X

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

Table 2. Rank of the Top Five and all European Countries plus Canada and the USA by Number of Extracts for the 2000 round census (calendar year 2011)

Total samples extracted for 55 countries (162 samples) available 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