

Disclosure Controls for One-Stop, Trans-Border Access to Census Microdata for 98 Countries via a Single License and Dissemination Point: the IPUMS-IECM Partnership*

Robert McCaa, University of Minnesota Population Center rmccaa@umn.edu
Albert Esteve, Center for Demographic Studies, Universitat Autònoma of Barcelona aesteve@ced.uab.es

ESSnet Workshop on
Statistical Disclosure Control (SDC) of Census data
Luxembourg, 19-20 April 2012

“You have to do due diligence, something to assure yourself that the people you’re giving your data to can be trusted.”
--<http://www.nytimes.com/2011/09/09/us/09breach.html?hp>

I. Introduction.

1. 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 International Union for the Scientific Study of Population (IUSSP) resides outside the country of birth. To require that researchers physically present themselves within national borders to access census microdata is old-fashioned, costly and inefficient. IPUMS-International disseminates cross-national and temporally integrated census microdata extracts to researchers world-wide at no cost. According to Google Analytics, <https://www.ipums.org/international> generated more than 45,000 visits (one-half million page views) over the past twelve months, representing 169 countries and territories. An analysis of pages visited and new applications for access suggests surprising insights on the effectiveness of IPUMS disclosure controls in restricting access to bona-fide researchers and excluding casual user, even the much cited, if rarely seen, privacy-threatening intruder.
2. IPUMS-International offers a comprehensive suite of disclosure control protections—statistical, legal and administrative—to minimize risk and maximize utility of integrated census microdata and metadata for researchers around the world. Would-be users must submit a [detailed, legally binding electronic application](#) both to establish research bona-fides and to explain need for access. The University of Minnesota Population Center (MPC) administers the dissemination of the microdata, assiduously policing access and enforcement of the licenses and agreements. As an example, this paper discusses the contract executed between the Federal Statistical Office of Switzerland and the University of Minnesota for disseminating five percent samples of four censuses, 1970-2000.
3. The IPUMS-International project has negotiated uniform agreements (Appendix A) with 98 national statistical offices (21 in Europe) to disseminate integrated census microdata to accredited researchers world-wide at no cost through a single licensing agreement

* Research 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) and Harmonizing Integrated European Census Microdata (HIECM), funded by the European Union, Research Infrastructures Action, FP6-026033.

(Appendix B). As of June 2011, 185 samples of integrated microdata and metadata for the censuses of 62 countries (397 million person records) are available to almost 5,000 registered researchers from more than 100 countries. Thanks to exceedingly generous cooperation from National Statistical Offices worldwide and sustained funding by the National Science Foundation, the National Institutes of Health (USA), and Research Infrastructures Action (EU), the database is expanding at the rate of 5-10 additional countries per year (see Table 1). The project (www.ipums.org/international) is led by the University of Minnesota Population Center (MPC) in partnership—for the censuses of Europe—with the Centre d'Estudis Demogràfics (CED), Autonomous University of Barcelona (www.iecm-project.org).

4. Usage statistics for 2011 reveal a year-to-year increase of 22% in registered researchers and 39% extract requests. In 2011, over 8,000 extracts were made. The average extract consisted of 5 samples and 35 variables. For a decade now, 40 variables account for 60% of the extract requests. Among these are sixteen IPUMS value-added variables, four of which are imputed from household information and, as such, are unique to IPUMS:
 - spouse's location in household,
 - mother's location,
 - father's location, and
 - rules for inferring locations.

II. Trans-Border access: legal, administrative and statistical disclosure controls

5. 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 is even higher.¹ For many demographers—and many social science researchers in general—trans-border access is essential if analysts are to research census microdata of their country of birth. All census microdata in the IPUMS-International system are accessible to bona fide researchers², regardless of country of birth, residence or citizenship.
6. To maximize effectiveness disclosure controls for access to census microdata must be legal, administrative and statistical (Thorogood 1999). Otherwise utility is sacrificed on the altar of risk. Access to the IPUMS-International microdata is restricted—despite the “P” (Public) in IPUMS—governed, on the one hand, by the letter of understanding endorsed by the University and the National Statistical Authority, and, on the other by the license agreement between the University, the researcher, and the researcher's institution. The letter of understanding grants the right to the University to disseminate microdata extracts electronically for teaching and research purposes, according to the authorization

¹ Statistics provided to the authors by the Secretariat of the International Union for the Scientific Study of Population, September 14, 2011.

² Note that we refer “bona fide researcher”, not the term cited by Duncan et al (2011, p. 139).

procedures stated in the agreement. Microdata may *not* be used for commercial purposes. Strict confidentiality of persons, households and other entities must be maintained. Alleging that a person or other entity has been identified is prohibited. The University is charged with assuring that users will guard against access to the microdata by unauthorized individuals.

7. 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.
8. 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. 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.
9. 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.
10. 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. 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. 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.
11. Where the statistical agency entrusts the task of implementing some or all of the statistical disclosure protections to the IPUMS project, we apply technical controls on a subjective, ad-hoc basis as negotiated with each country for each census. Contemporary microdata, from a census taken less than ten years ago, require greater protections than older, historical data.
12. The most important statistical control is the suppression of records by subsampling. All the values in the records outside the sample are suppressed. Second, is the suppression of names and geographical detail, such as place of birth or residence. Each statistical

authority balances the risk-utility trade-off by instructing the IPUMS project as to the minimum threshold for identifiable geographical units for the most recent census. In the case of many African and Latin American countries, the threshold is commonly set at 20,000 inhabitants in the latest census. Others set the floor as high as 100,000 (United States) or in the most extreme case (Netherlands) all administrative geography is suppressed. We are gratified that in some cases our statistical agency partners have reconsidered earlier, overly strict decisions, offering higher precision samples (Mexico 1990 increased from one to ten percent) and greater detail. In the case of Colombia, the geographical floor, initially set at 100,000, was reduced to 20,000 after Colombian researchers vigorously registered their dissatisfaction with the Colombian statistical agency, DANE. DANE not only reduced the threshold, but also harmonized the identifiers so that all the census microdata samples for Colombia could be disseminated with a single set of integrated geographical codes, in harmony with current DANE practice. When the sample for the 2005 census became available, confidentializing and harmonizing the current geographical identifiers was easily accomplished.

13. In consultation with the national statistical office, some variables may be top-coded, others may be subjected to global recoding, deletion of digits for hierarchical variables (occupation, industry, geography), or the suppression of a variable entirely. Decisions are made in consultation with the corresponding national statistical authority. Sensitive variables, if any, may also be wholly suppressed. Weight variables are usually not a confidentiality risk because most samples adopt the IPUMS design and are therefore implicitly stratified with a single weight, such as “10” for a ten percent sample.
14. An additional, iron-clad guarantee of protection is provided by randomly ordering the records and swapping the geographical identifiers of an undisclosed number of households. This means that no one can assert with certainty that an individual or household has been identified.
15. For a concrete example of the IPUMS process, consider the case of the 1970-2000 censuses of Switzerland. The Federal Statistics Office (FSO) prepared five percent household samples for each census according to the IPUMS standard design. Households were ordered geographically from highest to lowest administrative level. A random seed was chosen for the first household, then every 20th household was selected. The samples were entrusted to the MPC to apply standard IPUMS statistical disclosure control protections, as follows:
 - 1) Detailed geographical codes were suppressed. Canton was selected as the lowest level geography identifiable in each sample.
 - 2) The code for any canton with fewer than 20,000 individuals was aggregated with a co-terminus canton; a new label was formed to include both cantons.
 - 3) The code for any social characteristic of individuals (categorical variables such as place of birth, occupation, etc.) with fewer than 250 individuals in the population was recoded as missing, suppressed or aggregated. For the 2000 sample, 68 variables had one or more codes suppressed.
 - 4) 12 continuous variables (such as income, size of rooms, etc.) were top/bottom coded to prevent identification of individuals or other entities with unique characteristics.
 - 5) The geographical identifiers of a fraction of households were re-coded from the actual canton of residence to a false canton. The error introduced provides absolute protection against allegations of positive identification of individuals or households.

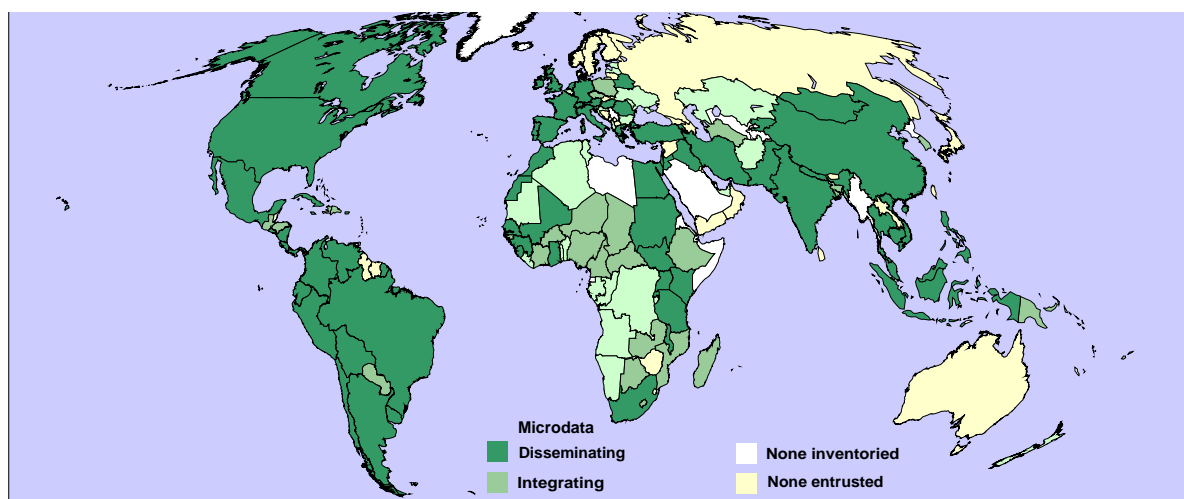
- 6) Households with more than 15 persons were re-classified as group dwellings and residents were sampled as individuals (n=2,294 for the 2000 sample).
 - 7) Finally, the FSO required the University to certify the destruction of the original, raw microdata files upon completion of the technical disclosure control process.
16. IPUMS privacy protection protocols provide powerful disclosure control protections at modest cost in terms of biases introduced into the data—a growing concern of researchers as more statistical agencies undertake the task of developing statistical disclosure controls (see Reiter 2011). Uniform protocols enhances comparative research and minimizes infelicities due to variations in confidentiality procedures and errors due to programming mistakes, such as the embarrassment experienced recently by the United States Census Bureau with respect to the microdata files of the American Community Survey (Alexander, Davern and Stevenson 2010).
17. Census agencies that confidentialize microdata must take heed of this unfortunate episode. Due to a programming mistake, age reporting of the elderly was egregiously corrupted in a large fraction of cases in the ACS samples. Researchers could not prove the error until they were able to compare confidentialized samples against the full-count non-confidentialized microdata available through the Census Bureau’s Research Data Center. The story broke on the front pages of the New York Times, shortly before field work for the 2010 census got underway. In the case of IPUMS statistical disclosure controls, both our Statistical Agency partners and researchers are reassured that microdata disseminated by IPUMS are carefully checked for coherence and robustness not only before the microdata are disseminated to researchers but also before the microdata are integrated into the IPUMS database.

III. IPUMS-International: a massive, widely used, global resource for restricted access to census microdata

18. IPUMS-International (www.ipums.org/international) archives, integrates, and disseminates high precision, richly detailed microdata from national population and housing censuses. This massive data infrastructure—currently totalling more than 397 million anonymized, integrated person records representing more than 100 million households—will expand to 69 countries and 212 censuses with the upcoming release in June 2012 (see Figure 1).
19. Twenty-one European countries currently participate in the IPUMS/IECM initiatives. The status of microdata is as follows (number of sets of microdata entrusted in parentheses):
- **16 are integrated (53 samples):** Armenia (1), Austria (4), Belarus (1), France (7), Germany (4—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), and the *United Kingdom (2—to be expanded to 6).
 - **5 are being integrated:** Bulgaria (0), the Czech Republic (2), Poland (2 to be expanded to 5), Turkey (3 to be launched in 2012), and Ukraine (0)
 - **In several European countries,** national statistical offices continue to consider participation. Others that previously declined invitations to be founding members of the initiatives are invited to reconsider.

20. The 2012 IPUMS launch is scheduled to incorporate samples for Turkey (1985, 1990, 2000)—as well as five other non-European nations (El Salvador, Indonesia, Morocco, Mexico, and Uruguay). Additional launches are planned for successive years, integrating 2010 round census samples as expeditiously as they become available.

**Figure 1. Status of census microdata for countries participating in the IPUMS-International collaboratory, June 2012:
3 shades of green--Integrated (dark), Integrating (medium), and Negotiating (light).**



21. Although access to the IPUMS-International microdata is free of cost, usage is restricted to bona-fide researchers who agree to abide by stringent conditions of use (Appendix B). IPUMS disseminates pooled extracts containing many samples in a single dataset, custom-tailored to the precise research needs of each user. This contrasts with the practices of most statistical offices where census microdata are disseminated as a single sample containing all variables and all person records. Each research receives exactly the same sample so there is a temptation to copy the data for an unauthorized person. Given the massive size of the IPUMS-International database, disseminating the full set of variables and unvarying size of samples is impractical. Most importantly, IPUMS disseminates pooled microdata with multiple samples and a varying selection of variables for each extract request. 83% of extracts contain more than one sample. This is possible because both microdata and metadata are integrated for all censuses and all countries. Nonetheless 88% of requests are for microdata from a single country.
22. With IPUMS no two extracts are alike. Each extract is custom-tailored. The researcher, by means of a series of choices, places an order, selecting:
- 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 along with all other co-resident persons in the selected household)
 - and sample density (either as a percent or number of cases).
21. The IPUMS extract engine fulfils the request by generating a dataset containing only the requested microdata and the corresponding set of DDI compatible metadata as well as a codebook suitable for constructing a system file in SPSS, SAS or STATA. Copies of original source metadata are available from the web-site. Most importantly the integrated

metadata are always readily available in interactive form from the IPUMS-International web-site.

22. At the UN-ECE Expert Group Meeting on Statistical Data Confidentiality, November 2005, we explained the IPUMS-International data dissemination procedure 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 extract, decompress it and proceed with the analysis using the supplied integrated metadata consisting of variable names and labels.

23. This method of dissemination has weathered the test of time, and indeed as usage soars, the rapid acceleration of internet transmission speeds has validated the IPUMS approach.
24. 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.

Rank	Country	Sample %*	Variables (n)*	Years of census 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					

25. Brazil, Mexico and Colombia predominate in usage not only because their samples offer lots of variables and a long series covering a half century of dramatic demographic transformations, but also due to the fact that many Latin American emigrants reside in the United States and thus it is possible to analyze these populations in a single integrated database, whether the researcher resides in the country of birth or in the two most important countries of emigration—the USA and Spain. In addition, all the Latin American samples, as well as those for the United States and Spain, are high precision, household samples with richly detailed, extensive information on migration, economic, social and demographic variables of both individuals and households.
26. For the year 2011, 1,011 researchers qualified for access to the IPUMS-International database, representing 98 countries. The IPUMS-I “Top 33” institutions in terms of data usage represents fourteen countries and includes many of the world’s premier universities and research organizations (see Appendix C).
27. Future initiatives: 1. Cooperate with the European Union funded Data without Boundaries Service Center for Official Statistics (SC-OS). As stated by the DwB project document:

“The overarching goal of the Service Center for official statistics (SC-OS) is to remove barriers to access to official statistics data in Europe and stimulate research in order to make better use of the wealth of high quality data available. To this aim the SC-OS will provide researchers with comprehensive services for European official statistics micro-data. This will entail extensive data documentation and tools for data preparation as well as training in the use of OS micro-data. Moreover the SC-OS shall seek to facilitate exchange among researchers, data producers and data archives throughout Europe and work towards reducing boundaries to data access.

...

“Another major aspect of this web service will be integrating the existing Integrated European Census Microdata (IECM) database which was established by the Centre d'Estudis Demogràfics (CED) in Barcelona in close collaboration with the Integrated Public Use of Microdata Series-International project (IPUMS) lead by the Minnesota Population Center. The IPUMS/IECM disseminates integrated and anonymised European census micro-data free of charge to registered users and provides extensive metadata for these censuses both at the study and variable level.”

28. Future initiatives: 2. Trans-border on-line tabulator. The Center for Demographic Studies, in collaboration with the Minnesota Population Center, is designing new tools to foster trans-border dissemination of European census microdata. As part of this effort, the IECM online tabulator of integrated variables for European census samples is currently under development. This will add a fast and convenient tool to explore the contents of the database before making an extract. Researchers will be able to do exploratory analysis or obtain some basic figures without the need to download microdata. The IECM tabulator is built upon REDATAM, a trusted and secure UN software to provide access to census microdata (www.eclac.org/redatam/). Researchers may access a single sample or multiple samples, as research needs dictate. Access will be restricted to registered researchers. The microdata will be identical to the integrated samples disseminated by IPUMS-International and will be subjected to the same statistical disclosure controls.

IV. Conclusion.

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

REFERENCES

- Alexander, J.T.; Davern, M.; and Stevenson, B. 2010. "Inaccurate Age and Sex Data in the [United States] Census PUMS Files: Evidence and Implications," *Public Opinion Quarterly*, 10 (Aug 10), pp. 1-10. doi: 10.1093/poq/nfq033
- Conference of European Statisticians. 2007. "Annex 1.23 Case study: Access to anonymized census microdata samples via the IPUMS-International and the Integrated European Census Microdata websites," *Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines on Good Practice*. Geneva: United Nations Economic Commission for Europe. See online edition: <http://www.unece.org/stats/publications/> pp. 98-104.
- Duncan, G.T.; Elliot, M.; and Salazar-Gonzalez, J.-J. 2011. *Statistical Confidentiality: Principles and Practices*. New York: Springer.
- McCaa, Robert and Albert 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, Robert, Steven Ruggles and Matthew L. 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.
- Reiter, J.P. 2011. "Statistical Approaches to Protecting Confidentiality for Microdata and Their Effects on the Quality of Statistical Inferences," *Public Opinion Quarterly*, Vol. 76, No. 1, Spring 2012, pp. 163–181
- Thorogood, D. 1999. 'Statistical Confidentiality at the European Level.' Paper presented at: Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, March.

Appendix A. Example of Uniform Memorandum of Understanding between the University of Minnesota and National Statistical Offices (Italy, 2006).

Letter of Understanding	
<p>Integrated Public Use Microdata Series International and L'ISTITUTO NAZIONALE DI STATISTICA (ISTAT)</p>	
<p><u>Purpose.</u> 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"> <u>Ownership.</u> 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. <u>Use.</u> 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. <u>Authorization.</u> 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. <u>Restriction.</u> 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. <u>Confidentiality.</u> Users will maintain the absolute confidentiality of persons and households. Any attempt to ascertain the identity of a person, family, household, dwelling, organization, business or other entity from the microdata is strictly prohibited. Alleging that a person or any other entity has been identified in these data is also prohibited. <u>Security.</u> Users will implement security measures to prevent unauthorized access to microdata acquired from Integrated Public Use Microdata Series International or its partners. <u>Publication.</u> 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"> <u>Violations.</u> 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. <u>Sharing.</u> 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. <u>Jurisdiction.</u> Disagreements which may arise shall be settled by means of conciliation, transaction and friendly composition. Should a settlement by these means prove impossible, a Tribunal of Settlement shall be convened which will rule upon the matter under law. This Tribunal shall be composed of an 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. <u>Order of Precedence.</u> 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>2/2/06</u></p> <p>Signed: <u>[Signature]</u></p> <p>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></p> <p><small>Rev. Jan. 27, 2005</small></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> <p>Application to Use Restricted Microdata</p> <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 [omitted]</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> <p>Name of institution or employer <input type="text"/></p> <p>Your email address at institution (*) <input type="text"/></p> <p>Web link showing your affiliation with institution (*) <input type="text"/></p> <p>Email address of employer, supervisor, or instructor (*) <input type="text"/></p> <p>Phone number of institution (*) <input type="text"/></p> <p>Does your institution have an Institutional Review Board (IRB), or Office for Human Subject Protections, Professional Conduct or similar committee?</p> <p style="padding-left: 40px;"><input type="radio"/> No <input type="radio"/> Yes; Name of board or office <input type="text"/></p> <p>RESEARCH PROJECT</p> <p>Please provide at least 75 words <i>in English</i> describing your research project or educational use for the data. This description will be used to evaluate your application.</p> <p style="text-align: center;">.....</p> <p>If your research is funded by someone other than your employer, indicate the name of the granting institution, title of grant, and other pertinent information. (*) <input type="text"/></p>	<p style="text-align: center;">USAGE LICENSE</p> <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 conditions 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>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> <p>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> <p>Commercial use and redistribution of the microdata is strictly prohibited.</p> <p>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> <p>Use of the microdata must follow strict rules of confidentiality.</p> <p>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> <p>The microdata must always be safely secured.</p> <p>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.</p> <p>Scholarly publications are permitted, and must be cited appropriately.</p> <p>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.</p> <p>Any violation of this license agreement will result in disciplinary action, including possible loss of employment.</p> <p>Violation of this agreement will lead to revocation of this license, recall of all microdata acquired, a motion of censure to the relevant professional organization(s) and civil prosecution under national or international statutes, at the discretion of the Regents of the University of Minnesota and the official statistical agencies. Sanctions likewise may be taken against the institution with which the violator is affiliated.</p> <p><input type="checkbox"/> User agrees to notify ipums@pop.umn.edu regarding errors in the data.</p>
--	--

**Appendix C. IPUMS-I Top 33 University/Research Institutions by Number of Extracts
– 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 of Demographic Studies (Barcelona, Spain)	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's home, the University of Minnesota)