IPUMS-International: Integrating and Disseminating High Precision Population Census Samples of the USA, Greece, Europe and the World.

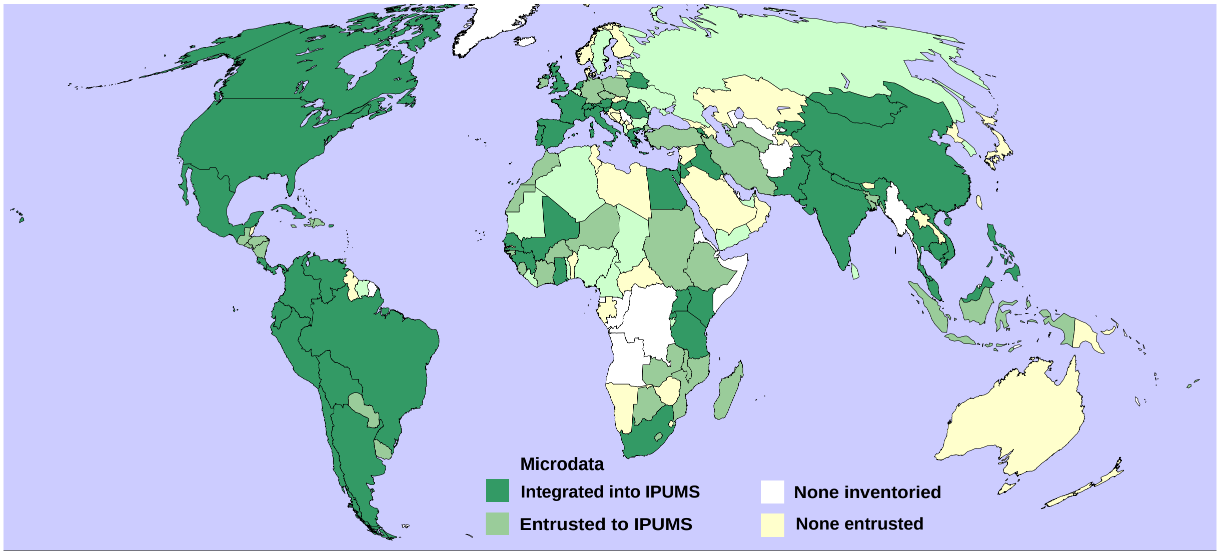
Robert McCaa
Minnesota Population Center
rmccaa@umn.edu

Introduction

The Minnesota Population Center's IPUMS-International project (www.ipums.org/international) integrates and disseminates high precision population census samples to researchers and policy makers world-wide at no cost (McCaa and Ruggles 2000). Begun in 1999 with a five year grant from the National Science Foundation (now extended to fifteen years), the initiative includes the United States, Greece, and more than 90 other countries (Figure 1). Regional projects are funded by the National Institutes of Health Europe (2004-2009), Eurasia (2009-2014) and Latin America (2003-2013). IPUMS-International encompasses over four-fifths of the world's population through a uniform Letter of Understanding negotiated with National Statistical Offices (NSO) almost identical to that endorsed in 2003 by the National Statistical Service of Greece (Appendix A). Greece is represented by ten percent samples for the 1971, 1981, 1991, and 2001 censuses.

Figure 1. The IPUMS-International World Map (June 2010)

Dark green = integrated; medium = integrating; light = negotiating; lightest = no data or interest



As of June 2010, microdata for 55 countries (159 samples, totaling 325 million person records) are integrated into the IPUMS-International database and are being disseminated to almost 4,000 registered researchers, educators, and policy-makers world-wide. Thirty-eight datasets for the 2000 census round are complemented by 44 for the 1990s, 37 for the 1980s, 27 from the 1970s and 13 from the 1960s. 128 samples are high precision with a density of five-ten percent. Of the 30 lower precision samples, many consist of all the surviving microdata for the respective censuses. More than 100 additional datasets have been entrusted to the Minnesota Population Center for integration over the next five years (see Appendix B).

Greece ranks among the first European signatories (preceded only by France), and Greek samples rank among the first European microdata integrated into the IPUMS (also preceded by samples of French censuses 1962, 1968, 1976, 1984, 1990, 1999, and to be followed next year by new recensement renovée named 2006, but covering the period 2003-2008). All the more remarkable is that, in terms of usage, the Greek samples place third among fourteen European countries represented in the IPUMS database. With 1,496 extracts, Greece is barely outpaced by Spain (1,514) and well ahead of Austria, Belarus, Hungary, Italy, the Netherlands, Portugal, Romania, etc. Greece is out-ranked by France (2,795), but not the United Kingdom (687 extracts). The UK ranks poorly because samples for only two censuses are integrated, and the sample designs are idiosyncratic and difficult to compare, even with each other.

The Greek samples are in great demand, not only due to an intrinsic interest in the population of Greece, but also because they are distinguished by meticulous, uniform construction, high precision, and rich content (see https://international.ipums.org/international/sample_designs/sample_designs_gr.shtml). The fastidiousness of the Greek census operations begins with the timing of the census—the third week of March. Only the 1981 census was held in April--due to a severe earthquake which struck in the Athens region a few days before the planned census date (see Michalopoulou, this volume). Greece is represented in the database by four magnificent, high precision household samples, totaling more than 1.2 million anonymized households and 3.7 million persons. The samples are not only consistent, but rich in content with 66 integrated variables common to all samples, 22 for dwellings and households and 44 for persons (Appendix C). European integrated variables include geography (NUTS1, NUTS2, and NUTS3), relationship to head or reference person, class of worker, and marital, educational attainment, and employment status (see http://www.iecm-project.org/). Due to the strong demand for the integrated Greek microdata, as soon as the sample for the 2011 census becomes available, it will be integrated into the IPUMS database for dissemination as quickly as possible.

Three papers on IPUMS-Greece microdata and metadata

Greek researchers are also among the first users to evaluate the microdata and metadata in the IPUMS. The three papers published here constitute analytical models for assessing not only the strengths and weaknesses of the integrated Greek samples, but also samples for other countries.

Professor Catherine E. Michalopoulou (Panteion University), in the paper entitled "Comparability issues of the IPUMS microdata for Greece, 1971-2001," offers a nuanced critique of the comparability of concepts, definitions, field operations, and the construction of the microdata samples for each the four Greek censuses. The comparability analysis by Dr. Michalopoulou is a valuable synthesis. It is much more thorough than what can be offered by the IPUMS metadata. To compare, begin here: https://international.ipums.org/international-action/variables/group. To view only the variables for Greece, click "Select Samples", "Greece" and "Submit Sample Selections". To read the metadata for a specific variable click its name (e.g., "Citizen"). Click "Codes" to review the categories of the variable (e.g., citizenship) recorded in each integrated sample. To compare the source metadata in English for each census, click "Enumeration Text". To view original source documents produced by the census agency as image files in the official language, from the IPUMS-International homepage, click "Source Documents," then select the specific document for the country and census year to examine.

Prof. Stefanos G. Giakoumatos (Kalamata School of Management and Economics), in "Weighting Methods: An application to IPUMS-Greek Samples," demonstrates the power of four weighting methods to calibrate the IPUMS-Greek sample for the 2001 census. Prof. Giakoumatos compares the results from calibrations for a dozen household variables with the official statistics disseminated by the National Statistical Service and the simple frequencies produced by the design weights disseminated by the IPUMS. His analysis shows that indeed calibrated weights exactly reproduce the official statistics, while the design weights available from IPUMS differ, usually to a small degree, but for variables with considerable dispersion, such as central heating, the magnitudes can be substantial. It is important to emphasize, first, that the purpose of the IPUMS database is to facilitate analysis, not to generate official statistics, and second, that design weights are what they are--a general, all-round approximation for analytical purposes. The IPUMS-International database empowers researchers to generate their own weights and estimates of variance, within the constraints of the microdata provided by each national statistical office. Additional guidance is available at https://international.ipums.org/international/variance_estimation.shtml on sampling error and variance estimation (see also Cleveland, Davern and Ruggles, 2010).

Profs. Tsimbos (University of Pireaus), Verropoulou (University of Pireaus) and Bagavos (Panteion University), in "A comparison of IPUMS Microdata and Census Aggregated Data with a Special Reference to Native and Migrant Fertility in Greece," offer an excellent example of precisely applied analytical research. Their analysis of the IPIUMS sample for the 2001 census of Greece shows that educational attainment and work experience reduces fertility for Greek women, but for Albanians and Bulgarians not so much. On the contrary, for these immigrant groups, as little as six or more years of schooling is associated with markedly higher fertility. Few statistical offices provide ready access to such detailed data by means of their publications or official websites. The National Statistical Service of Greece offers generous access to tabular data. Nonetheless, only by means of microdata is it possible to conduct the fine, subtle analysis in the Tsimbos-Verropoulou-Bagavos paper.

Get data for Greece and dozens of other countries

To get data from IPUMS, registration is required, but there is no charge. Casual users are dissuaded from accessing the data by the registration form, which requests research bona-fides, as well as a brief project description demonstrating a need for census microdata for research or teaching. As the papers in this volume illustrate good use of IPUMS microdata requires expertise. However graduate and undergraduate students may readily gain expertise through classroom exercises (Meier, McCaa and Lam, in press). Student access is encouraged, but the consent of an instructor may be required.

Once the registration is approved the user is certified to download microdata from the IPUMS website. Before getting IPUMS microdata, researchers are encouraged to carefully study both the official source and IPUMS documentation available from the website. Since the data are so easily accessed, researchers are obligated to devote adequate time to studying the questionnaires, instructions to enumerators, and the IPUMS metadata for samples and variables, including the comparability discussions. The IPUMS metadata system makes it easy to navigate the integrated documentation (discussed below), census-by-census, variable-by-variable for the samples selected by the user.

The database is too huge for researchers to simply download everything (in fact there is a download limit to prevent frivolous downloads). Instead users request extracts custom-tailored to each researcher's needs. To request an extract, the user first signs in by entering the registered name and password then making a series of selections by means of point-and-click menus, specifying country (or countries), census year(s), sample(s), variables and, if desired, sub-populations, as well as metadata format (SAS, SPSS, or STATA are supported). Once the selections are complete, there is an opportunity to review or revise before final submission of the request. Then, once submitted, the IPUMS extract engine registers and queues the request. When the extract is ready (usually in a matter of hours, if not minutes), the researcher is notified by email that the data should be retrieved within 72 hours. A link is provided to a password-protected page for downloading the specific extract via SSL (Secure Sockets Layer) protocol. Microdata are transmitted using the 128-bit encryption standard, matching the level used by the banking and other industries where security and confidentiality is essential. The researcher may then securely download the file, decompress it and proceed with the analysis using the supplied integrated metadata consisting of variable names and labels provided in ASCII format. The IPUMS help-desk responds to user questions and queries and assembles copies of publications for transmission to the respective National Statistical Offices.

Get anonymized microdata on all Greeks in the IPUMS-International database

Greek researchers, indeed researchers of any nationality, may readily study their immigrant compatriots in any of the census samples where detailed country of birth is available. Almost 400,000 persons born in Greece are represented in the 2000 census samples currently in the IPUMS database (see https://international.ipums.org/international-action/variables/173813/codes). Only data for the United Kingdom and Australia are lacking. For the UK detailed country of birth is not available in the 2001 sample and for Australia, to date, the Bureau of Statistics has declined invitations to cooperate with the IPUMS initiative. Additional information on Greeks is available in integrated variables such as the country of previous residence (2,800 in Armenia in 2001), the country of residence five years ago (14,500 in the USA in 2000) and country of citizenship (Austria 2001: 1,800; Switzerland 2000: 6,500).

To make an extract of all persons born in Greece, first log into IPUMS. Second, click "select samples", "all samples", "submit sample selections". From the Variables selection screen, click each household and person variable desired (e.g., sex, age, etc.), including the country of birth variable (BPLCTRY). Once the selection of variables is completed click "Make data extract". Review the variables selected. If the selection is correct, then click "Continue to next step" twice—in succession unless you to wish change the extract options (there are only 2 options and the defaults are suitable for most researchers). The third step is to "Set variable options". Click the "Select Cases" box for BPLCTRY and any other variable that you may wish to use as a selection criteria. For example if males are the focus of the study, then click SEX here. On the next screen, specific options will be presented for each variable for which the corresponding "Select Cases" box is ticked. From this screen, researchers who favor working with "rectangular" datasets may wish to click "attach characteristics" for each of the desired variables (see below). When finished, click "Continue to next step". From the select cases screen, for BPLCTRY, scroll down to Greece (ISO code 43060) and click Greece. If SEX "Select Cases" was chosen, click either male or female, as desired. Click "Continue to next step". If you chose to "Attach Characteristics", the corresponding screen will appear. There are 4 choices for each variable selected: head, mother, father, spouse. Clicking the corresponding box will cause the extract engine to place the selected variable on the corresponding person record. For example, if you wish to place the educational attainment of the mother and father on a child's record, then for the educational attainment variable, click the mother and father boxes. Click "continue to next step". The fourth step is to "Customize sample sizes" in terms of megabytes. If the physical size of the dataset is not too huge for your computing resource, click "continue to next step". Unfortunately, at present, the estimated size does not take into account "Select Cases". Since Greek immigrants compromise only a tiny fraction of the database, click the "customize sample sizes" option to maneuver around the fact that the extract engine does not adjust the estimated size of the data file. The size will be vastly over-estimated, resulting in a cancelation of the request. This is easily remedied by customizing sample sizes. Since there are no more than 100,000 Greek residents enumerated in any census outside Greece, enter "100" in the "persons" for "All Samples." Scroll down to Greece, then enter 10 (percent) in the "density" column for each Greek census. Press "Calculate to obtain a revised estimate of database size. When finished, click "Continue to next step". For the fifth and final step, review the contents of the extract. Changes may be made by clicking options to select samples, variables, cases, sample sizes, etc. Once everything is acceptable, name the extract in the "Description" box with as much detail as desired so that if later you wish to revise an extract it will be easily identifiable by name. Click "Submit extract". When the extract is ready for download, an email will be sent to the registered address, along with instructions on how to obtain the microdata and metadata.

Tips on how to use the IPUMS-International microdata are posted in Arabic, English, French, Russian, and Spanish--but not Greek!—at: http://www.hist.umn.edu/~rmccaa/IPUMSI (click "Tips on using IPUMSi"). Tips 1-4 are briefly explained above (register, protect confidentiality, study documentation, extract judiciously). Tip 5, apply weights, is essential for most analytical purposes (see Appendix C, person variable 4, WTPER). Failure to apply weights may produce misleading results. Sophisticated users may wish to generate their own weights by using one or another of the calibration methods discussed in the paper by Giakoumatos in this volume. .

Constructing IPUMS-International integrated microdata and metadata

From the above discussion, it should readily be apparent that IPUMS-International is *not* simply a conduit for relaying census samples from National Statistical Offices to researchers. Instead, typically, two or more years of labor are invested by the IPUMS team in each country's censuses to prepare anonymized, integrated microdata and metadata for dissemination (Esteve and Sobek 2003 and Sobek and Esteve forthcoming). There are five steps to the IPUMS process before microdata are disseminated to researchers.

1. Confirm the integrity and validity of the source microdata and metadata for each census
2. Draw and anonymize high precision samples
3. Integrate the metadata (documentation)
4. Integrate the microdata
5. Confirm the integrity and validity of the integrated microdata samples and metadata

Steps one and two are conducted on the original source microdata entrusted to the Minnesota Population Center. These microdata are never disseminated to anyone or any institution—other than the corresponding National Statistical Office-owner. For this reason, at the MPC, access to these data is restricted to senior civil service staff thoroughly trained in protecting data security. These data are exceedingly sensitive and for that reason only seasoned, specially trained, full time staff with a need to complete the first two tasks of the IPUMS process have access to these data. MPC employees are subject to civil fine (up to US$250,000) and criminal prosecution for violation of security procedures. The University legal authority assumes responsibility for protecting the total confidentiality of these datasets. A complete review of these processes was conducted on-site by Mr. Dennis Trewin, the chairman of the UN-ECE joint-committee on Statistical Confidentiality and Microdata Access and President ex-officio of the International Statistical Institute. Mr. Trewin's report (2007) concludes:

> "Without question IPUMS International meets the four Core Principles outlined in CES [Conference of European Statisticians] (2007). It is cited in CES (2007) as a Case Study of good practice. This review confirms its status as good practice for Data Repositories. Indeed it is likely to provide the best practice for a Data Repository for international statistical data. … The security of the computing environment used by IPUMS-International is first class and appears to be of the standard of the best statistical offices."

Consider each of the five steps of the IPUMS integration process in more detail.

1. Confirm integrity and validity. The microdata are exhaustively evaluated by IPUMS senior staff to resolve issues of data integrity and validity. Note, however, to date the project does not perform data editing or imputation. Instead effort is focused on ensuring the household structure of records and confirming that sample statistics approximate official published figures. Since the purpose of the sample is to provide a dataset for analysis, there is no need to insure that samples exactly replicate published census results.

2. Draw and anonymize the sample. Geography is one of the most important stratifying variables in survey research and in drawing high precision census microdata samples. Geography is related to a great number of variables researchers are interested in studying and therefore increases the efficiency of stratified samples. Many of the IPUMS-International samples, including those for Greece, capitalize on *implicit* geographic stratification. The raw census files used to construct IPUMS samples are typically geographically organized within districts. Systematic random samples capitalize on this low-level geographic sorting. By ensuring a representative geographic distribution of sampled cases, they are equivalent to extremely fine geographic stratification with proportional weighting. Since many economic and demographic characteristics are highly correlated with geographic location, this implicit stratification yields substantially greater precision than would a simple random sample of households. As part of the IPUMS project, we are developing stratification variables that allow researchers to make reliable variance estimates from implicitly stratified samples (Cleveland, Davern and Ruggles, 2010).

Almost all the statistical agency partners of the IPUMS project, including the National Statistical Service of Greece, have endorsed the use of implicitly stratified samples of households. Thirty-seven national statistical offices have entrusted complete sets of census microdata to facilitate the drawing of implicitly stratified samples by the MPC. In Europe, almost all the statistical agencies have drawn new samples adopting IPUMS specifications. IPUMS sample densities typically range between 5 and 10%. Lower densities are provided by countries where privacy matters are a greater issue than quality (Netherlands, United Kingdom) or, as in the case of 1960 round of censuses, where low precision samples are constitute all the extant microdata.

In cases where fully anonymized samples are entrusted to the project, no further statistical confidentiality measures are imposed. However, in many cases, full datasets are provided to the project, including detail sufficient to pose a theoretical risk of re-identification. To minimize risk, statistical confidentiality edits are performed by the IPUMS project. The lowest level of geography to be released is identified (e.g., for European countries, typically "NUTS3") and all finer geographic variables are suppressed. Any technical variables that could be used to identify records or which have been identified as sensitive within the original data are also suppressed. Variables with very small population categories are recoded into larger groups (e.g., grouping a detailed occupation with its parent category) and top- or bottom-coding is performed where needed (e.g., income). Finally, the sequence of dwellings within the smallest geographic unit identified in the data is randomized, so geography cannot be inferred. An undisclosed fraction of cases is randomly swapped across geographic districts to add uncertainty about the origin of any particular record. Finally, a new serial number is generated to reflect the ordering of the file.

3. Metadata integration. Metadata integration—that is, integration of census documentation regarding definitions, concepts and codes—is essential if microdata integration is to succeed. Integrated metadata relieve researchers of the task of studying documentation *en toto* for every census to discover changes or deviations in concepts and definitions. Instead, before microdata are integrated into the IPUMS system, experts carefully consider all the documentation and analyze the microdata as a prelude to writing new, comprehensive documentation that spells out common practices and discusses significant differences and discrepancies. Once the data are released, researchers study the integrated metadata confident that their attention will be directed to issues of greatest salience for the research questions at hand.

The IPUMS eXtensible Markup Language (XML) tool facilitates navigation of both source and integrated metadata in any way desired by means of a few clicks. For example, to compare the wording of the employment status variable, select the countries and census years desired, then click "employment status", and "enumeration text". This allows the researcher to compare the precise wording, in English, of the question on the form as well as the instructions to the enumerators for all selected census.

a. Censuses and samples. IPUMS metadata offer detailed descriptions of each census in the database, listing the title, year, universe, de jure/de facto, enumeration unit, official census day, forms, field work period and type, respondent and estimates of undercount, if any. Images of census enumeration forms and instructions manuals are available in the official language and the transcribed text in English or 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, such as nomads, military personnel, non-citizens, etc.

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, enumeration text (source), and non-harmonized variables used for integration. This information is readily navigable through clickable hypertext on the IPUMS website. A general comparability discussion is provided for every variable, with country or census specific discussions focusing on departures from standard practice. The purpose of these discussions is to highlight important contrasts. Clicking "Enumeration text" leads to source questions and corresponding instructions for each selected census in English. 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 as the census was conducted by field workers.

Coding of variables in the IPUMS system may be viewed in either general or detailed versions using one of two views. The default view is a table in which "X" indicates the presence of the code in a specific census. An optional view (click "Case Count View") provides the exact, un-weighted case count in the integrated sample for each code. The "codes" table is handy for determining whether specific attributes are present in sufficient quantity for the contemplated research as well as planning recodes for specific analytical purposes.

4. Microdata Integration. The principal benefit of IPUMS to researchers and National Statistical Offices (NSOs) alike is integration—integration of both microdata and metadata. For decades, many NSOs have provided census samples for academic and policy research, but few statistical offices re-examine earlier samples to harmonize successive datasets or to draft new documentation to facilitate comparative analysis of two or more censuses. At best, as soon as the final data cleaning is complete, the more modern statistical offices construct a census sample and a data dictionary for researchers, Five or ten years later, with the ensuing census, the process is repeated with little guidance on enhancing the comparability of successive census microdatasets.

We must reiterate that the IPUMS project does *not* disseminate census files entrusted by national statistical offices. Instead high-precision census samples are anonymized (McCaa et. al. 2006) and integrated, variable-by-variable, using a composite coding system (Esteve and Sobek, 2003). Samples are integrated both chronologically and cross-nationally. Integrated metadata are constructed by means of meticulous study of comprehensive original source documentation and after extensive analysis of the microdata. Thousands of hours are devoted to analyze, discuss, debate, draft, test and re-test until the microdata integration is validated for dissemination to researchers. The process is repeated with each annual launch of additional samples into the IPUMS database.

As an example of the IPUMS method of integrating a variable, consider the seemingly simple concept "married". In recent decades, as the United Nations Statistics Division Principles and Recommendations have evolved, there is increasing precision in definitions. Nonetheless, in practice, some censuses refer to "married," while others report "formally married" according to civil law or religious convention or both. Still other censuses distinguish informal unions from formal ones. The IPUMS system seeks to retain all significant distinctions in the original microdata for every sample in the database. Thus for "married otherwise undefined", the IPUMS code is "200". Formally married is "210". Formal civil marriage is "211", which contrasts with religious marriage "212," both religious and civil "213", either "214", traditional "215", and polygamous "217". Consensual unions are coded "220". Depending on research need, a researcher may decide that the first digit provides sufficient detail and inore the trailing digits. Others may require the full 3-digit code. In any case, the composite coding scheme is easily understood and researchers may proceed with confidence that significant distinctions are retaining in the IPUMS codes. For complete details, see the IPUMS metadata for "Marital Status": https://international.ipums.org/international-action/codes.do?mnemonic=MARST Successful international integration must document these distinctions so that researchers may readily be informed of these and thousands of other details (Sobek and Esteve, forthcoming). As the next section indicates, the IPUMS integrated metadata describe these general details as well as subtle distinctions for specific countries and individual censuses.

IPUMS integration has enjoyed such success that some statistical agency partners prefer to use the integrated IPUMS microdata rather than their own original source data. For example, DANE-Colombia, the first statistical agency to participate in the IPUMS collaboratory, is using the five integrated Colombian census samples (12.3 million person records) to construct a nationally integrated dataset with metadata in the Spanish. With IPUMS assistance, DANE is simplifying internationally integrated datasets to a national system. It is more efficient to simplify an international integration because where a national integration is performed first; important details may be unwittingly sacrificed as trivial at the national level but are essential for successful international integration.

5. Validation and Certification. Before samples are made available to researchers, the entire database is checked for consistency and accuracy. This requires verification of hundreds of thousands of coding decisions. The process is facilitated by the fact that the database contains both integrated and non-harmonized variables. Verification is performed by cross-tabulating each integrated variable by its corresponding non-harmonized version of the variable. Initially IPUMS-International focused on a number of harmonized variables that were common across nations and over time. As historians, comparisons over time as well as space are vitally important. However, after five years of providing only integrated microdata, it became clear that there was a demand to retain both the content of and access to the unharmonized variables that are specific to each census sample. The importance of facilitating access to the unharmonized variables is two-fold. First, it provides the source material from which the harmonized variables are constructed, allowing researchers a fuller understanding of the harmonization process. Second it preserves the original census structure and content, providing a reflection of changes in a country's census series over time and differences between nations concerning what concepts are covered and how they are expressed in national censuses. In 2006 over 5000 unique sample-specific non-harmonized variables were added to the IPUMS database. Each variable has its universe documented and empirically verified. Although some variables are suppressed for confidentiality reasons or obvious data errors, the goal is to provide as complete a picture of the original census as possible.

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

We have learned that integration is difficult, but the IPUMS system has streamlined the process so that it is now possible to harmonize a dozen or two census samples per year to a high degree of precision and to the satisfaction of National Statistical Office-owners of the microdata as well as for users—academic researchers, educators and policy makers alike.

Disseminating microdata and metadata to accredited researchers world-wide

To date, more than 4,000 researchers are registered to access the database, representing 76 countries, and hundreds of universities and international organizations. Recall that researchers must first be approved before access to any microdata on the IPUMS-International website is permitted. Access is obtained by submitting a detailed application and agreeing to each condition of use as required by the project Letter of Understanding. Once approved, microdata are provided in the form of extracts, custom tailored to each researcher's need. The IPUMS database is not distributed *en toto* and the ability to reassemble subsets into a replication of the whole is effectively curtailed by the IPUMS dissemination system.

Conclusions

Thanks to widespread support by official statistical agency partners like the National Statistical Service of Greece and the constructive feed-back from researchers such as the papers by Michalopoulou, Giakoumatos, Tsimbos, Verropoulou and Bagavos, IPUMS has already become the world's largest demographic database. Researchers and educators who wish to use the microdata are cordially invited to visit the website, register, construct an extract and download the data. The considerable demand for census microdata is growing rapidly. Indeed, in the case of the United States, IPUMS-USA is the single most frequently cited data-source in the premier journal of population studies, *Demography.* If IPUMS-International is successful, it is likely to become one of the most widely used sources for academics and policy makers requiring census samples for analysis.

REFERENCES

Cleveland, Lara L., Michael Davern, and Steven Ruggles. (2010). "Drawing Statistical Inferences from International Census Data," *Joint Statistical Meetings*. Vancouver, BC, Canada.

Conference of European Statisticians. (2007). *Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines on Good Practice*. http://www.unesce.org/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf Geneva: United Nations Economic Commission for Europe.

Esteve, Albert and Matthew Sobek. (2003). Challenges and Methods of Census Harmonization. *Historical Methods* 36: 66-79.

Esteve, A., Cabré, A., Valls, M., Garcia, J., (2008) "IECM: Integration European Census Microdata", *4th Conference of Social and Economic Data,* German Federal Office of Statistics, Wiesbaden, Germany, June 19-20.

Giakoumatos, Stefanos G. (this volume). "Weighting Methods: An application to IPUMS-Greek Samples."

McCaa, Robert, and Steven Ruggles (2000). "IPUMS-International: A Global Project to Preserve Machine-Readable Census Microdata and Make Them Usable." In *Handbook of International Historical Microdata*, ed. By Patricia Kelly Hall, Robert McCaa, and Gunnar Thorvaldsen, 335-346. Minnesota Population Center https://international.ipums.org/international/microdata_handbook.shtml .

McCaa, Robert, Steven Ruggles, Michael Davern, Tami Swenson, Krishna Mohan Palipudi. (2006). "IPUMS-International High Precision Population Census Microdata Samples: Balancing the Privacy-Quality Tradeoff by Means of Restricted Access Extracts," Privacy in Statistical Databases. Berlin: Springer, pp. 375-382.

Meier, Ann, Robert McCaa and David Lam. (in press). "Creating Statistically Literate Global Citizens: The Use of IPUMS-International Integrated Census Microdata in Teaching," Statistical Journal of the IAOS (International Association of Official Statisticians.

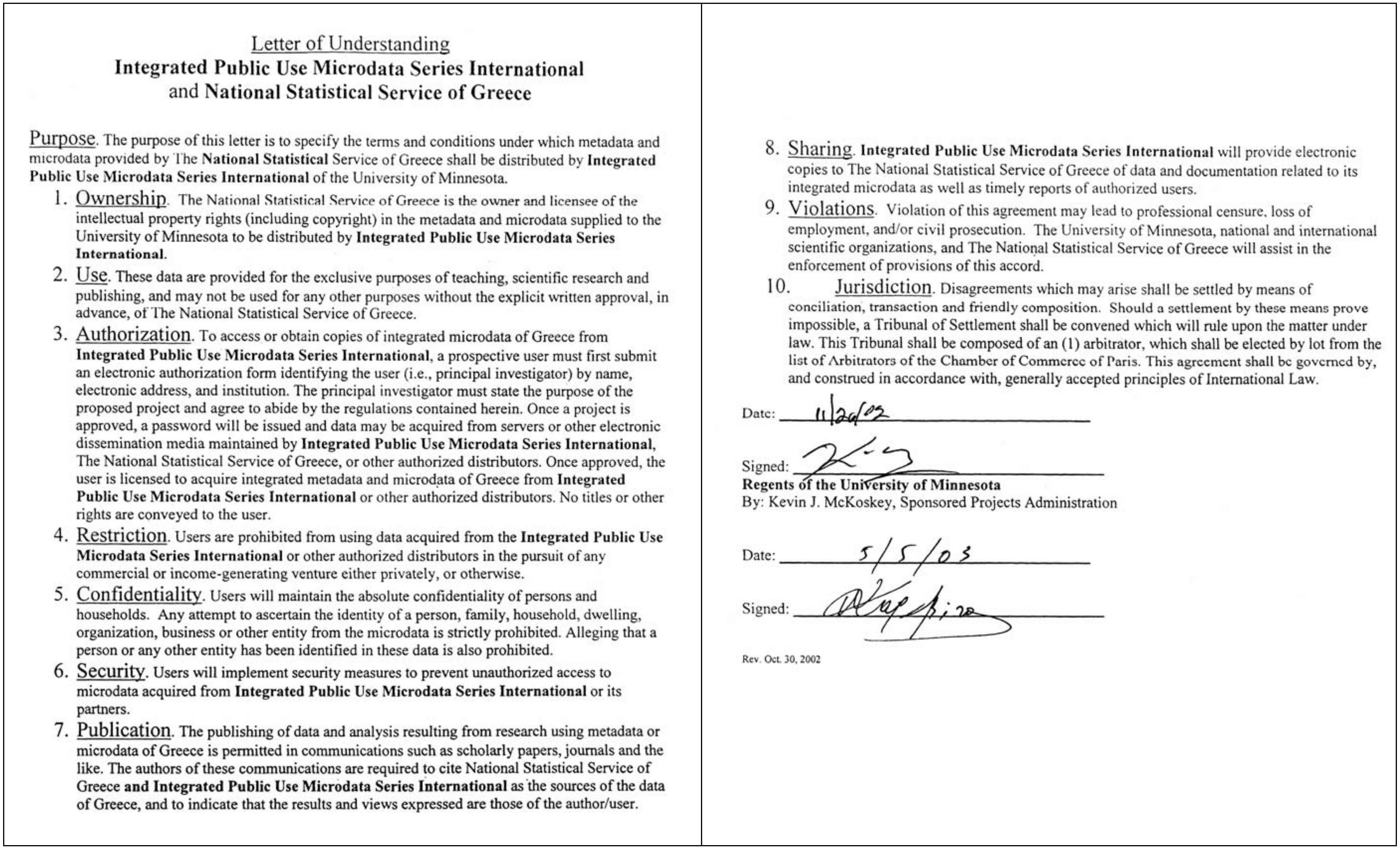
Michalopoulou, Catherine E.   (this volume). "Comparability issues of the IPUMS-International microdata for Greece, 1971-2001."

Sobek, A., Esteve, A. (forthcoming) 'The harmonization of international census microdata for demographic research: the IPUMS project', Institut National d'Études Démographiques.

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

Tsimbos, Cleon, Georgia Verropoulou, and Christos Bagavos (in this volume) "A comparison of IPUMS Microdata and Census Aggregated Data with a Special Reference to Native and Migrant Fertility in Greece."

Appendix A. Letter of Understanding
between the National Statistical Services of Greece and the University of Minnesota.



Letter of Understanding
Integrated Public Use Microdata Series International
and National Statistical Service of Greece

Purpose. The purpose of this letter is to specify the terms and conditions under which metadata and microdata provided by The **National Statistical** Service of Greece shall be distributed by **Integrated Public Use Microdata Series International** of the University of Minnesota.

1. Ownership. The National Statistical Service of Greece is the owner and licensee of the intellectual property rights (including copyright) in the metadata and microdata supplied to the University of Minnesota to be distributed by **Integrated Public Use Microdata Series International.**
2. Use. These data are provided 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 The National Statistical Service of Greece.
3. Authorization. To access or obtain copies of integrated microdata of Greece 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**, The National Statistical Service of Greece, or other authorized distributors. Once approved, the user is licensed to acquire integrated metadata and microdata of Greece from **Integrated Public Use Microdata Series International** or other authorized distributors. No titles or other rights are conveyed to the user.
4. 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.
5. 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.
6. Security. Users will implement security measures to prevent unauthorized access to microdata acquired from **Integrated Public Use Microdata Series International** or its partners.
7. Publication. The publishing of data and analysis resulting from research using metadata or microdata of Greece is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite National Statistical Service of Greece **and Integrated Public Use Microdata Series International** as the sources of the data of Greece, and to indicate that the results and views expressed are those of the author/user.
8. Sharing. **Integrated Public Use Microdata Series International** will provide electronic copies to The National Statistical Service of Greece of data and documentation related to its integrated microdata as well as timely reports of authorized users.
9. Violations. Violation of this agreement may lead to professional censure, loss of employment, and/or civil prosecution. The University of Minnesota, national and international scientific organizations, and The National Statistical Service of Greece will assist in the enforcement of provisions of this accord.
10. 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 (1) arbitrator, which shall be elected by lot from the list of Arbitrators of the Chamber of Commerce of Paris. This agreement shall be governed by, and construed in accordance with, generally accepted principles of International Law.

Date: 11/26/02

Signed:

Regents of the University of Minnesota
By: Kevin J. McKoskey, Sponsored Projects Administration

Date: 5/5/03

Signed:

Rev. Oct. 30, 2002

Appendix B. IPUMS-International, June 2010:
Microdatasets entrusted by country and census decade,
with sample densities (%) and confidentiality protocols

Sample %			Country	Protocols	Census decade				
10%+	~5%	<5%			2000s	1990s	1980s	1970s	1960s
Integrated and Disseminating 2002-2010: 55 countries, 159 censuses, 87 million households and 325 million person records									
4			**Argentina**	INDEC	**2001**	**1991**	**1980**	**1970**	1960
1			**Armenia**	SCS	**2001**		1989	1979	1970
4			**Austria**	IPUMS	**2001**	**1991**	**1981**	**1971**	1961
1			**Belarus**	IPUMS		**1999**	1989	1979	1970
3			***Bolivia**	IPUMS	**2001**	**1992**		**1976**	
5			**Brazil**	IBGE	**2001**	**1991**	**1980**	**1970**	**1960p**
2			**Cambodia**	IPUMS	**2008§**	**1998**			1962
		4	**Canada**	STATSCAN	**2001p**	**1991p**-6	**1981p**-6	**1971p**	1961
4		1	***Chile**	IPUMS	**2002**	**1992**	**1982**	**1970**	**1960p**
		2	**China**	NBS	**2000**	**1990**	**1982**		1964
3		2	***Colombia**	IPUMS	**2005**	**1993**	**1985**	**1973**	**1964p**
3	1		***Costa Rica**	IPUMS	**2000**		**1984**	**1973**	**1963**
1			**Cuba**	IPUMS	**2002**		1981	1970	
4		1	***Ecuador**	IPUMS	**2001**	**1990**	**1982**	**1974**	**1962p**
3			**Egypt**	IPUMS	**2006§**	**1996**	**1986**	1976	1964
1	6		**France**	INSEE	**2006§**	**1990,9**	**1982**	**1975**	**1968,2**
2			***Ghana**	IPUMS	**2000**		**1984**	1970	
4			**Greece**	IPUMS	**2001**	**1991**	**1981**	**1971**	1961
2			***Guinea, Conakry**	IPUMS		**1996**	**1983**		1960
	4		**Hungary**	CSO	**2001**	**1990**	**1980**	**1970**	
		5	**India**	NSSO	**2005m**	**1993,9m**	**1983,7m**		
1			***Iraq**	IPUMS		**1997**	1987	1977	1967
5			**Israel**	CBS	**2008**	**1995**	**1983**	**1972**	**1961,**7
	1		**Italy**	ISTAT	**2001**	**1991**	**1981**	1971	1961
1			**Jordan**	IPUMS	**2004**	**1994**	1979		
	3		**Kenya**	IPUMS	**1999**	**1989**	**1979**	**1969**	
1			**Kyrgyz Republic**	IPUMS	**2009**	**1999**	1989		
		4	**Malaysia**	IPUMS	**2000**	**1991**	**1980**	**1970**	1960
3			***Mali**	IPUMS	**2008**	**1998**	**1987**	**1976**	
4		3	**Mexico**	INEGI	**2000,5**	**1990,5**	**1980**	**1970**	**1960p**
2			***Mongolia**	IPUMS	**2000**		**1989**	1979	**1956**
1			**Nepal**	CBS	**2001**	1991?	1981	1971	1961
		3	**Netherlands**	CBS	**2001pm**			**1971p**	**1960p**

2			**Palestine**	CBS	**2007§**	**1997**			
3			***Pakistan**	IPUMS		**1998**	**1981**	**1973**	1961
5			***Panama**	IPUMS	**2000**	**1990**	**1980**	**1970**	**1960**
2			**Peru**	IPUMS	**2007**	**1993**	**1981**	1972	1961
3			***Philippines**	IPUMS	**2000**	**1990**	**1980**	**1970**	**1960p**
	3		**Portugal**	INE	**2001**	**1991**	**1981**	1970	1960
	4		**Puerto Rico**	USCB	**2000**	**1990**	**1980**	**1970**	1960
3			**Romania**	IPUMS	**2001**	**1992**		**1977**	1965
2			***Rwanda**	IPUMS	**2002**	**1991**			
2			***Saint Lucia**	IPUMS	**2001**	**1991**	**1980**	1970	1960
3			***Senegal**	IPUMS	**2002**		**1988**	**1976**	
1			**Slovenia**	SORS	**2001**	1991	1981		
6		1	**South Africa**	StatsSA	**2001,7**	**1996-1**	**1985-0**	**1970**	1960
	3		**Spain**	INE	**2001**	**1991**	**1981**	1970	1960
	4		**Switzerland**	IPUMS	**2000**	**1990**	**1980**	**1970**	1960
2			***Tanzania**	IPUMS	**2002**		**1988**	1978	1967
		4	**Thailand**	NSO	**2000**	**1990**	**1980**	**1970**	1960
2			***Uganda**	IPUMS	**2002**	**1991**	1980		1969
		2	**United Kingdom**	ONS	**2001p**	**1991**	**1981**	**1971**	**1966,1**
	6		**United States**	USCB	**2000,5**	**1990**	**1980**	**1970**	**1960**
4			***Venezuela**	IPUMS	**2001**	**1990**	**1981**	**1971**	1961
	2		**Vietnam**	IPUMS	**2009**	**1999**	**1989**	1979	
Countries where samples are being integrated (%), in preparation (bold) or negotiations underway *Europe*									
			Albania	-	**2001**	1989	1979	1969	1960
			Bulgaria	-	**2001**	**1992**	**1985**	1975	1965
			Belgium	-	**2001**	**1991**	**1981**	**1970**	1961
	2		**Czech Republic**	IPUMS	**2001**	**1991**	**1980**	**1970**	1961
			Estonia	-	**2000**	**1989**	1979	1970	1959
4			**Germany §**	FSO	**2001m**	**1991m**	**1981-7**	**1970,1**	1961
8			**Ireland §**	CSO	**2002, 6**	**1991, 6**	**1981, 6**	**1971,9**	
			Latvia	-	**2000**		**1989**	1979	
			Poland	-	**2001**	**1995**	**1988**	**1970,8**	1960
			Russia	-	**2002**		**1989**	1979	1970
			Turkey	TurkSTAT	**2000**	**1990**	**1985**, 0	1975,0	1960
			Ukraine	IPUMS	**2001**		1989	1979	1970
North and South America and the Caribbean									
1	1	2	***Dominican Republic**	IPUMS	**2003**	1993	**1981**	**1970**	**1960p**
1			***El Salvador**	IPUMS	**2007**	**1992**		1971	1961

2		3	***Guatemala**	IPUMS	**2002**	**1994**	**1981**	**1973**	**1964**
2			***Haiti**	IPUMS	**2003**		**1982**	**1971**	
3		1	***Honduras**	IPUMS	**2000**		**1988**	**1974**	**1961**
3			***Jamaica§**	IPUMS	**2001**	**1991**	**1982**	1970	1960
2		1	***Nicaragua §**	IPUMS	**2005**	**1995**		**1971**	1963
4		1	***Paraguay**	IPUMS	**2002**	**1992**	**1982**	**1972**	**1962**
4			***Uruguay**	IPUMS		**1996**	**1985**	**1975**	**1963**
Africa									
			Benin		**2002**	**1990**		**1979**	
3			***Botswana**	IPUMS	**2001**	**1991**	**1981**	1971	1964
			Burkina Faso		**2006**	**1996**	**1985**	1975	
			Burundi		**2008**	1990?	1979?	1970?	
			Cameroon		**2005**		**`1987**	**1976**	
			Cape Verde	IPUMS	**2000**	**1990**	1980	1970	1960
			Central African Rep.		**2003**		**1988**	1974	
			Chad		**2008**	**1993**	**1989**		1969
			Côte d'Ivoire		**2009**	**1998**	**1988**	1975	
2			***Ethiopia**	IPUMS	**2007**	**1994**	**1984**		
			Gabon		**2003**	**1993**	1980		1969
			Guinea-Bisssau	IPUMS	**2009**	**1991**		1979	
2			**Lesotho**	IPUMS	**2006**	**1996**	**1986**	**1976**	1966
			Liberia		**2008**		1984	**1974**	
1			***Madagascar**	IPUMS		**1993**			
2			***Malawi**	IPUMS	**2008**	**1997**	**1987**	1977	1967
			Mauritania		2001		1988	1977	
2			***Mauritius**	IPUMS	**2000**	**1990**	**1983**	1972	1962
	3		**Morocco**	IPUMS	**2004**	**1994**	**1982**	1971	1960
1			**Mozambique**	IPUMS	**2007**	**1997**	1980		
2			***Niger**	IPUMS	**2001**		**1987**	**1977**	
			Nigeria	NatPopCom	**2006**	**1991**		1973	1963
1			***Sierra Leone §**	IPUMS	**2004**		1985	1974	1963
3			***Sudan**	IPUMS	**2008**	**1993**	**1983**	**1973**	
			Togo		2010		**1981**	**1970**	**1958**
2			***Zambia**	IPUMS	**2000**	**1990**	1980	1969	1963
Asia and Oceania									
1		1	***Bangladesh**	IPUMS	**2001**	**1991**	**1981**	1974	1961
5			***Fiji Islands**	IPUMS	**2007**	**1996**	**1986**	**1976**	**1966**
8			**Indonesia §**	BPS/IPUMS	**2000, 5**	**1990, 5**	**1980, 5**	**1971,6**	1961
1			Iran §	SCI	**2006**	**1996**	**1986**	**1976**	**1966**

			Korea, Republic of	KOSTAT	**2005, 0**	**1995, 0**	**1985, 0**	**1975**	**1960,6**
			Sri Lanka	DCS	**2001**		**1981**	**1971**	1960
1			**Turkmenistan**	IPUMS		**1995**	1989	1979	1970
			United Arab Emirates		**2005**	1995	1985, 0	975	1968

bold country = Memorandum of Understanding with Regents of the University of Minnesota;

IPUMS = systematic household sample: every n^{th} household stratified by enumeration district; confidentiality specifications.

Year = census conducted; **Bold year** = microdata survive; §= samples in preparation for launch June, 2011

*** =** 100% microdata entrusted, where extant; m = microcensus; p = person sample

Appendix C. 66 Integrated variables in four Greek census samples (1971, 1981, 1991, and 2001) available from IPUMS-International

	Name	Label	Name	Label	Name	Label
Household Variables			**Person Variables, 1-22**		**Person Variables, 23-44**	
1	GQ	Group quarters status	SAMPLEP	IPUMS sample identifier [person ver.]	SEX	Sex
2	UNREL	Number of unrelated persons	SERIALP	Household serial number [person ver.]	MARST	Marital status
3	REGIONW	Continent and region of country	PERNUM	Person number	EMARST	Marital status, Europe
4	DEPTGR	Department, Greece	WTPER	Person weight	BIRTHYR	Year of birth
5	MUNIGR	Municipality, Greece	MOMLOC	Mother's location in household	CITIZEN	Citizenship
6	ENUTS1	NUTS1 Region, Europe	POPLOC	Father's location in household	NATION	Country of citizenship
7	ENUTS2	NUTS2 Region, Europe	SPLOC	Spouse's location in household	LIT	Literacy
8	ENUTS3	NUTS3 Region, Europe	PARRULE	Rule for linking parent	EDATTAN	Educational attainment, international
9	OWNRSHP	Ownership of dwelling	SPRULE	Rule for linking spouse	EDUCGR	Educational attainment, Greece
10	ELECTRC	Electricity	STEPMOM	Probable stepmother	EEDATTA	Educational attainment, Europe
11	WATSUP	Water supply	STEPPOP	Probable stepfather	EMPSTAT	Employment status
12	SEWAGE	Sewage	POLYMAL	Man with more than one wife linked	EEMPSTA	Employment status, Europe
13	ROOMS	Number of rooms	POLY2ND	Woman is second or higher order wife	OCCISCO	Occupation, ISCO
14	KITCHEN	Kitchen or cooking facilities	FAMUNIT	Family unit membership	OCC	Occupation, un-recoded
15	TOILET	Toilet	FAMSIZE	Number of own family members in hh.	INDGEN	Industry, general recode
16	BATH	Bathing facilities	NCHILD	Number of own children in household	IND	Industry, un-recoded
17	HHTYPE	Household classification	NCHLT5	Number of own children < 5 years in hh.	CLASSWK	Class of worker
18	NFAMS	Number of families in household	ELDCH	Age of eldest own child in household	ECLASWK	Class of worker, Europe
19	NCOUPLS	No. of married couples in hh.	YNGCH	Age of youngest own child in household	HRSWRK1	Hours worked per week
20	NMOTHRS	Number of mothers in hh.	RELATE	Relationship to household head	HRSWRK2	Hours worked per week, categorized
21	NFATHRS	Number of fathers in household	ERELATE	Relationship to head, Europe	MGRATE5	Migration status, 5 years
22	HEADLOC	Head's location in household	AGE	Age	MIGGR2	Department of residence 5 years ago

Note: Variable names are hyperlinked. Click the name to view the integrated variable for the Greek samples or click the source link below.

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