United Nations ECE/CES/GE.41/2014/27

Economic and Social Council

Distr.: General
10 September 2014

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses organized jointly with Eurostat

Sixteenth Meeting
Geneva, 23-26 September 2014
Item 5 of the provisional agenda
Operational aspects of censuses

Statistical Coherence in Secondary Education Completed: Census Hub Hyper Cubes and IPUMS/IECM integrated census sample results compared

Note by IPUMS/IECM[1]

Summary

The IPUMS/IECM projects disseminate integrated census microdata samples to researchers across Europe and the world at no cost on a restricted access basis. For both researchers and census officials questions of quality, specifically statistical coherence, are fundamental. How do sample statistics from the microdata compare with provisional results disseminated from EUROSTAT's Census Hub? To answer this question, we begin by testing secondary education completed because it is a recognized milestone, measured by most European censuses and widely available in all IPUMS/IECM samples. We use the demographic concept of birth cohort to generate a series of estimates from each sample, for ages 22 to 89 years. Figures from the Census Hub and even successive samples, if they are coherent from one round to the next, should show the same or similar percentages completing secondary education, year-by-year. The results are quite promising, indicating a remarkable degree of coherence, even with the shift to population registers in some countries, such as Austria. We expect to extend this work checking the coherence of other variables as 2011 census samples are entrusted to the IPUMS/IECM projects.

[1]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).
Authors: Robert McCaa, rmccaa@umn.edu, Matthew Sobek, sobek@umn.edu, Lara Cleveland, clevelan@umn.edu, Minnesota Population Center, Minneapolis, MN USA; Albert Esteve aesteve@ced.uab.es, and Anthony Lopez, tlopez@ced.uab.es, Centre d'Estudis Demogràfics, Autonomous University of Barcelona, Spain.

I. Introduction

1. Census data are collected at great expense and have enormous capacity to inform public policy. They are among the most widely used data sources in the social sciences and are broadly employed by policy makers, journalists, and others. Given the societal investment in them and their utility, it is essential that these data be made available in a manner that maximizes their potential.

2. The Eurostat Census Hub and IPUMS/IECM microdata projects offer two complementary approaches to census data dissemination. The Census Hub provides easy access to full census tabulations and is likely to have a broad audience among planners and even the public. IPUMS/IECM aims at a more limited constituency: researchers who require access to the full detail of a sample of individual and household records to develop measures and models impossible to specify with summary tabulations. Nevertheless, even microdata researchers are likely to take advantage of the convenience of the Hub for exploratory work and to extract quick summary statistics. With both these important data sources capable of generating cross-national, national and subnational statistics on similar topics, it behooves us to explore their consistency with one another. There are reasons to expect some differences -- due, for example, to sampling and coding decisions -- but it benefits the entire statistical community if those differences are not arbitrary or due to quality problems with the underlying data or how they were manipulated.

3. This paper compares the statistics from the Census Hub hypercubes to those derived from the microdata samples in IPUMS/IECM. For this test we use secondary education completed, because it is a recognized milestone measured by most European censuses and widely available in IPUMS/IECM samples. We use the demographic concept of birth cohort to generate a series of estimates from each sample, for ages 22 to 89 years. Figures from the Census Hub and successive microdata samples, if they are coherent from one round to the next, should show the same or similar percentages completing secondary education, birth year-by-birth year.

II. IPUMS/IECM Projects: Access and Design

4. The IPUMS/IECM projects (henceforth "IPUMS") disseminate integrated census microdata free of cost to researchers regardless of country of birth, citizenship or residence under uniform protocols. 4. The IPUMS/IECM projects (henceforth "IPUMS") disseminate integrated census microdata free of cost to researchers regardless of country of birth, citizenship or residence under uniform protocols. Currently 20 European National Statistical Institutes are participating in the projects, encompassing over three-fourths the population of the continent—rising to almost nine-tenths, when Turkey and Ukraine are included and Russia excluded (Table 1). The IPUMS/IECM partnership disseminates, for each participating country, confidentialized samples of the entire series of extant census microdata using the legal, administrative and technical framework developed by Thorogood (1999). Technical anonymization procedures are applied in consultation with each NSI (McCaa and Esteve, 2006; McCaa, Muralidhar, Sarathy, Comerford, and Esteve, 2014). Dissemination is governed by a uniform Memorandum of Understanding (Appendix A; see also McCaa, Esteve and Lopez, 2012). More than one hundred national statistical offices have embraced the terms of the IPUMS memorandum of understanding. As the fifteenth anniversary approaches, the global database contains 258 samples representing 79 countries, and totalling over 560 million person records. The complete listing of samples is available here: https://international.ipums.org/international/samples.shtml. Each year roughly two dozen samples are added to the database, once the demanding task of integrating both the microdata and metadata is complete.

5. As Table 1 shows European samples are in the bottom half of usage rankings. This is quite remarkable given their high rankings only a few years ago. In 2011, for example, five European samples ranked in the top quartile and only three in the bottom. Now, only one, France 2006 at #38, ranks above the median and six have fallen into the bottom quartile. In large part the poor showing is due to the delay in entrusting samples for the 2010 round of censuses. The CSO-Ireland was the first to entrust a 2011 census sample and it was the first to be integrated and disseminated. Recently 2011 samples have been received from Austria, the Czech Republic, France, Poland, Portugal and Spain while the statistical offices of Hungary, the Netherlands, Romania and Slovenia have communicated that samples to IPUMS specifications are in preparation.

Table 1. Rank of the Top Five plus Canada and 17 European Countries by Number of Extracts for Latest Round Census Microdata Available (2014 Jan-Aug)

Rank	Country	Sample %	Variables (n)	Census years of samples	IPUMS/ IECM Extracts
1	Brazil	5	106	1960, 70, 80, 91, 2000, 2010	1,193
2	Mexico	10	120	1960p, 70, 90, 95, 2000, 05, 2010	1,146
3	Colombia	10	120	1964p, 72, 85, 93, 2005	892
4	United States	5	92	1960, 70, 80, 90, 2000, 05, 10	763
5	Ghana	10	86	2000, 10	704*
21	Canada	2.5#	59	1971p, 81p, 91p, 2001p	503
38	France	33	94	1962, 68, 75, 82, 90, 99, 2006	385
42	Spain	5#	99	1981, 91, 2001	381
46	Portugal	5#	96	1981, 91, 2001	334
47	Austria	10#	75	1971, 81, 91, 2001	315
48	Greece	10#	89	1971, 81, 91, 2001	314
49	Ukraine	10#	35	2001	312*
50	Romania	10#	97	1976, 92, 2002	309
52	Switzerland	5#	79	1970, 80, 90, 2000	299
54	United Kingdom	3#	47	1991, 2001p	296
55	Italy	5#	81	2001	295
57	Ireland	10	86	1971, 79, 81, 86, 91, 96, 2002, 06, 11	280*
60	Turkey	5#	71	1985, 90, 2000	276
63	Hungary	5#	74	1970, 80, 90, 2001	262
73	Germany	5#	49	East: 1971, 81; West: 1970, 87	240
74	Netherlands	1#	33	1960p, 71p, 2001p	240
76	Belarus	10#	84	1999	227
78	Slovenia	10#	80	2002	207
Total samples extracted for 79 countries (258 samples)					7,890

*Extrapolated estimate based on one month usage
#2000 or earlier round census sample
Number of variables refers to integrated variables, including those constructed by IPUMS/IECM.
"p" = person sample; all other samples are of households
3 Countries participating but samples not integrated yet: Bulgaria, Czech Republic, Poland
22 Countries not participating: Albania, Belgium, Bosnia-Herzegovina, Croatia, Cyprus, Denmark, Estonia, Finland, Iceland, Kosovo, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Macedonia, Moldova, Norway, Russia, Serbia, Slovakia, Sweden

6. There are a number of key differences between IPUMS and the Census Hub. IPUMS disseminates individual microdata records, not summary data. The microdata are samples, and are therefore subject to sampling error, and various confidentiality measures have been imposed. The microdata are subject to restricted access requiring proof of scientific need; each request for access is reviewed by a staff member. IPUMS has broader geographic scope, with data from every continent. For most countries, there are microdata samples for multiple censuses, allowing analysis of change over time. IPUMS has a data tabulator, but most researchers download the microdata and analyze it on their desktops. This requires knowledge of statistical package software, unlike the user-friendly Census Hub system.

7. Each IPUMS record represents an individual and is composed of variables describing that person's characteristics as collected by the relevant census. For the great majority of samples, individuals are organized into households, and within households one can identify family interrelationships. This data structure provides substantially more power than would a simple sample of individuals. Thus, a researcher has access not only to the person's characteristics, but to all the characteristics of the people with whom they lived, family and non-family. This household data structure allows the construction of new variables drawing from information across individual person records, such as the number of wage earners in a family, or whether a mother has children under age five.

8. The signature activity of IPUMS is harmonization of data across space and time. The same codes mean the same things for all times and countries in the database. For complicated variables, it is impossible to construct a single uniform classification without losing information. Some censuses provide more detail than others, so the lowest common denominator of all samples inevitably loses important information. In these cases, we construct composite coding schemes. The first one or two digits of the code provide information available across all samples. The next one or two digits provide additional information available in a broad subset of samples. Finally, trailing digits provide detail only rarely available (Esteve and Sobek 2003; Ruggles 2006). IPUMS makes the unrecoded source data available as well, to ensure no information is lost.

9. Table 2 illustrates the complexity of coding educational attainment in censuses around the world, although the examples are limited to Europe. The frequencies are the raw cell counts in each sample to show the simple frequencies. These metadata help a researcher plan the analysis. Thus if technical education is of interest, the codes page indicates that the 2011 census of Ireland and the 2001 of Ukraine did not measure technical education at the secondary level and thus are not useful for analysing the topic.

10. Data are useful only when researchers understand what they mean. Accordingly, we have developed comprehensive harmonized documentation on each variable and sample. This documentation covers enumeration procedures and instructions; definitions of households, dwellings, group quarters, and other enumeration units; and scanned images of original-language versions of the questionnaires. We also provide detailed descriptions of each variable, including question wording and instructions (in the original and translated into English), universe definitions, frequency distributions, and variable codes. Comparability discussions describe any deviations of particular censuses from the standard variable definition and address differences over time and across countries. Seven types of metadata are organized in a user-friendly tabbed variable description on the web. The general discussion of cross-temporal and cross-country comparability continues for 553 words and is then followed by an explanation for each country and census sample currently selected by the user. The comparability text for Austria is 114 words, while that for France is 181 and Germany 339.

11. The IPUMS data are accessed via a web-based dissemination system. The data access system allows users to design datasets that are customized to their particular

research problem, by merging data across time periods and countries, selecting population subsets, selecting variables, and defining new variables that capitalize on the hierarchical structure of the data. Users design their dataset in a rich informational environment that describes each sample and variable, with special attention to the comparability of particular items across time and space. The IPUMS system supports SAS, SPSS and STATA, and we plan to add R in the near future. The goal is to remove the logistical barriers to cross-national studies, allowing researchers to focus on the substantive matters of interest to them.

Table 2. Educational Attainment (EDATTAN) 3-digit detailed code with frequencies for 8 European Countries

Code	Label	Austria 2001	Greece 2001	Ireland 2011	Italy 2001	Romania 2002	Spain 2001	Switzerland 2000	Ukraine 2001
000	NIU (not in universe)	135,480	60,222	117,404	160,897	223,986	9,429	61,858	249,197
100	LESS THAN PRIMARY COMPLETED	·	34,748	·	317,651	·	·	·	·
110	No schooling	·	·	·	·	78,667	473,420	12,776	341,478
120	Some primary	·	124,318	·	·	1,456	·	·	·
130	Primary (4 years)	·	·	·	·	384,871	·	·	637,966
	PRIMARY COMPLETED, LESS THAN SECONDARY								
	Primary completed								
211	Primary (5 years)	·	·	·	719,220	·	502,742	·	·
212	Primary (6 years)	·	300,134	51,896	·	·	·	·	·
	Lower secondary completed								
221	General and unspecified track	237,957	108,576	64,702	850,317	528,362	462,072	·	708,507
222	Technical track	·	·	·	·	·	·	·	·
	SECONDARY COMPLETED								
	General or unspecified track								
311	General track completed	34,593	213,311	123,420	600,717	122,104	199,858	77,527	1,544,926
312	Some college/university	·	·	·	19,065	7,717	·	·	·
320	Technical track	·	·	·	·	·	·	·	·
321	Secondary technical degree	342,559	19,591	·	129,727	576,739	159,802	134,987	·
322	Post-secondary technical education	14,942	46,434	15,876	10,513	56,970	114,012	27,674	782,292
400	UNIVERSITY COMPLETED	37,940	121,550	84,414	182,632	128,632	117,939	23,033	608,088
999	UNKNOWN/MISSING	·	·	16,823	·	28,463	·	26,231	16,834

Source: https://international.ipums.org/international-action/variables/EDATTAN/#codes_section

III. Census Quality and Coherence

12. Baffour and Valente (2012), in a recent overview, define census quality as "fitness for use" and argue that it is characterized by six elements or dimensions: relevance, accuracy, timeliness, accessibility, interpretability, and coherence (p. 122). The Eurostat Census Hub offers metadata on each of these dimensions and an overall section on "Quality management" (#11-#17 https://ec.europa.eu/CensusHub2[2]; see also Eurostat 2014). In this

[3] Note that Census Hub figures are preliminary. The User cannot access the statistics without clicking a box indicating that the statistics are preliminary (see Appendix C).

paper we are concerned with coherence, although accuracy and coherence are obviously interrelated.

A. Coherence defined.

13. The Sixteenth Meeting of the UNECE Group of Experts on Population and Housing Censuses defines coherence as follows (see UNECE 2014, p. 4, Section B.4.f):

> Coherence reflects the degree to which census information can be successfully brought together with other statistical information within a broad analytical framework and over time. The use of standard concepts, definitions, and classifications—possibly agreed at the international level—promotes coherence.

Baffour and Valente identify two types of coherence: internal (whole census results are coherent within themselves) and external (check against prior censuses). To achieve statistical coherence, definitions, concepts, frameworks and classifications must be clear and consistent at the national and international levels. When these change, explanatory text is essential to explain similarities and differences between the old and the new. The authors conclude that "ideally the [census] questions should keep the historical formulation to facilitate longitudinal comparison," and any unusual trends or inconsistencies in the data should be explained (p. 126).

14. The Eurostat census hub metadata section on coherence (#17) speaks to both external (17.1 "cross domain") and internal (17.2) types, as indicated by the following entry for Ireland (https://ec.europa.eu/CensusHub2/ --after checking the "agree" box, then click metadata, choose "Ireland" then scroll down to "17 Coherence"):

> **17.1 Coherence - cross domain**
> Prefilled text by Eurostat explain how census data relate to other social statistics.
> **17.2 Coherence - internal**
> Prefilled text by Eurostat explain how the EU transmission program for the 2011 census is conceived to obtain coherent data.

Continuing with Ireland as an example, the Eurostat Census Hub offers extensive information on quality assurance, but little or none on quality assessment. This is true for almost all the countries for which metadata are available with the notable exception of the United Kingdom, where there is a link to several reports on the analysis conducted.

B. The intracohort comparison method.

15. Population censuses are embedded with the demographic history of a people. Successive, high quality, coherent censuses should tell similar stories. The population historian's tool kit includes methods for comparing birth cohorts.

16. For internal coherence we ask the simple question: For the 2011 census of Austria, is the proportion completing secondary school of those born in, say, 1950 the same for the sample as the figure returned by the Census Hub? We can extend the series to include all years of birth, beginning at twenty-two years before the census (very few individuals complete secondary school at a more advanced age) and extending back in time until the absolute frequencies become too small to be reliable.

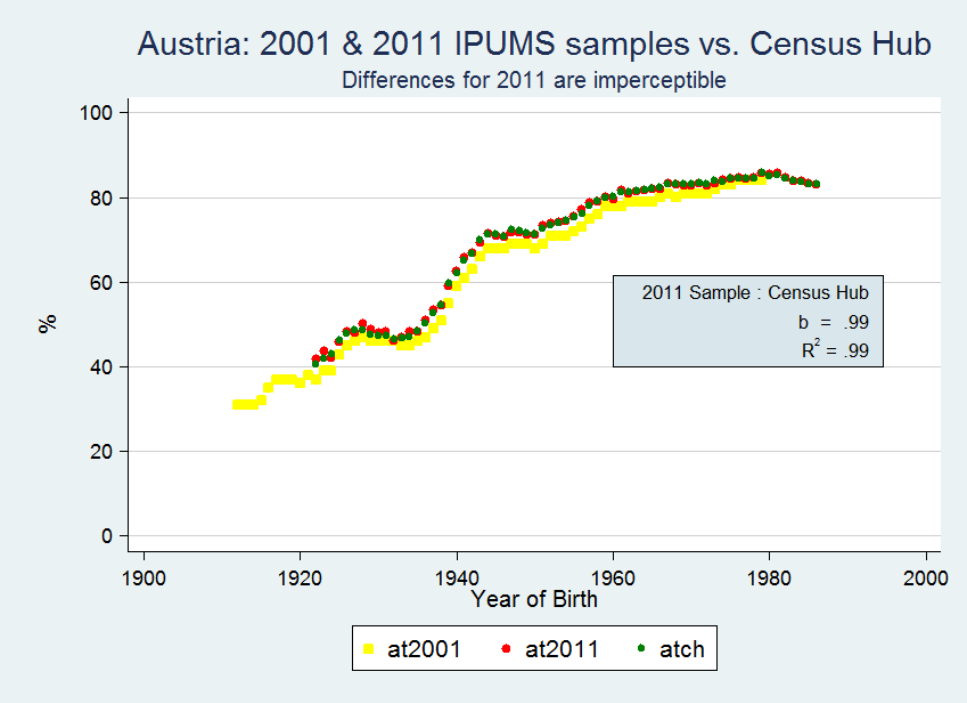
17. For "cross-domain" or external coherence the question becomes: is the proportion reported in the 2011 census similar to that in the 2001 sample, for each birth year? There

are a number of caveats for assessing external coherence. The method assumes that there are no differentials in mortality or migration by level of educational attainment. Where the less educated suffer from higher risks of dying then this will introduce a systematic bias. Likewise where the likelihood of out-migration is associated with educational attainment, then lack of coherence will be exaggerated. For additional details on the method see Feeney and McCaa (2014).

IV. Comparing Census Hub and Census Microdata: Secondary Schooling Completed

18. Results for the Austrians censuses of 2011 and 2001 are depicted in Figure 1—where the red points indicate the sample drawn by Statistics Austria for IPUMS/IECM, Census Hub figures are portrayed in green dots, and yellow the sample for the 2001 census. The curves reveal astonishing coherence, with both R2 and b at .99. Perhaps there should be little surprise that the results are so nearly identical because both sets of statistical data are produced by a single statistical agency. Nevertheless, the underlying data in each case were processed and coded, providing opportunities for errors which were apparently avoided. For researchers, the coherence between the sample and Census Hub sources is comforting.

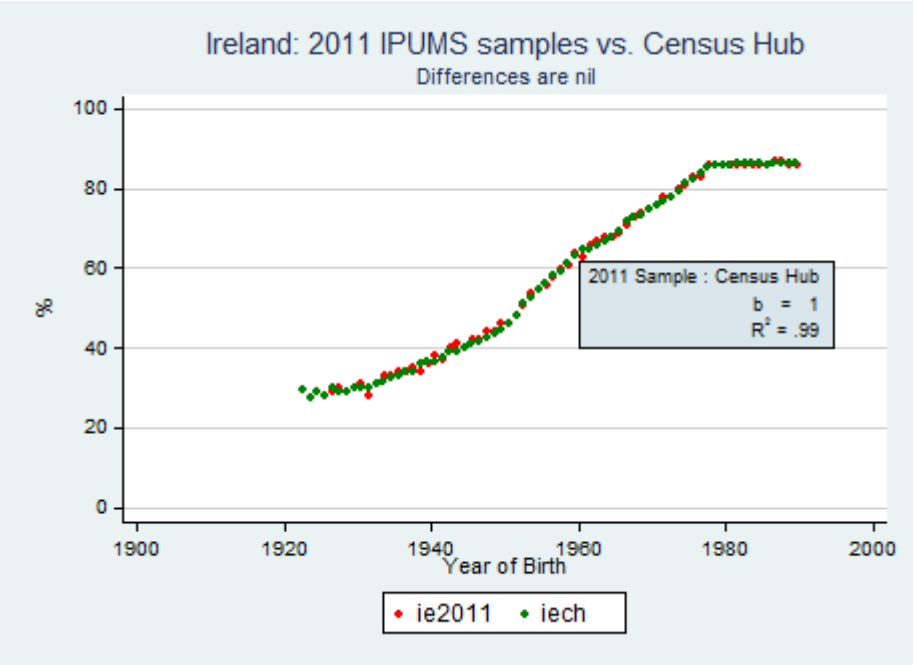
Figure 1. Austria 2001 and 2011 Censuses.



19. External coherence is all the more remarkable because the 2001 census was conducted using traditional face-to-face interviews in the field while the 2011 census is drawn from administrative registers, where characteristics of individuals are composed by means of a identification number to link the information drawn from several different administrative streams (Berka, Humer, Lenk, Moser, Rechta, and Schwerer 2010). The fact that the 2001 proportions are systematically, if only very slightly lower, at every age may suggest that the less educated have slightly worse survival chances than the better educated and therefore the proportions completing secondary tend to rise over time. There may also be an upward bias in reporting events more distant in the past.

20. For a second test, we have chosen Ireland, because in this case the sample data are heavily perturbed to protect statistical confidentiality. The CSO agreed to an experiment as a means of facilitating researcher access to single years of age and a number of other variables heavily requested by microdata researchers. A controlled shuffling method was applied to the data to protect confidentiality yet retain the highest possible utility. For this reason, a very high proportion of ages were shuffled. The shuffling was performed without taking into account educational attainment (McCaa, Muralidhar, Sarathy, Comerford, and Esteve, 2014). As Figure 2 indicates, the birth year comparisons are even better than for Ireland, if only slightly. Note that no comparison can be made with an earlier sample because only five-year age banded data are available. Thanks to the success of the shuffling technique, the CSO will entrust single years of age data for all earlier samples.

Figure 2. Ireland 2011 Census



21. Additional internal comparisons for other countries cannot be made for a lack of integrated 2011 census samples. However, as an experiment we did intracohort comparisons between Census Hub figures and the IPUMS/IECM samples for earlier censuses and found that the results were quite promising with regression coefficients ranging from .81 (Portugal 2001) to .99 (France 2006). For Portugal, the low figure may be explained by disproportionately high out-migration by the less educated.

22. Comparisons cannot be made for samples where age is banded, such as in five year age groups. Unfortunately a number of European samples offer banded microdata. These samples are used much less frequently by researchers, because of their greatly reduced utility for many kinds of analysis.

V. Reflections

23. Census microdata and summary tabulations are complementary formats for disseminating this fundamental statistical information to analysts and the broader public. The IPUMS/IECM data system and the Eurostat Census Hub serve different core audiences, but it is important that they are consistent. Some researchers will even use them in combination, perhaps tabulating older sample data to add historical depth to the Hub

results from the modern census round. It is therefore reassuring that our preliminary cohort analysis reveals a high degree of correspondence in the two data sources for one of the key variables amenable to this method, educational attainment. In the future we will conduct further tests to assure quality control in the microdata.

24. The Census Hub is a notable technical and organizational achievement, but it cannot satisfy the needs of most population researchers, who require microdata for individual-level analysis. But microdata pose challenges for statistical offices with other priorities and a large public with limited use for such specialized information. The IPUMS/IECM partnership for disseminating census microdata offers substantial returns at minimal risk. Statistical offices are relieved of many of the most burdensome tasks and responsibilities. IPUMS-International is now relied upon by most of the world's statistical offices—98 of some 157 countries numbering more than one million inhabitants (plus two smaller countries—Fiji Islands and Saint Lucia). The isolated statistical office that disseminates microdata on an ad hoc basis incurs substantial risks as well as significant costs in human resource—all for a relatively small return with respect to number of users. The IPUMS/IECM projects offer substantial economies of scale with the highest standards of quality: accuracy, timeliness, accessibility, interpretability, and coherence.

References.

Baffou, B. and P. Valente. 2012. "An evaluation of census quality," *Statistical Journal of the IAOS* 28:121-135. DOI 10.3233/SJI-2012-0752

Berka, C., S. Humer, M. Lenk, M. Moser, H. Rechta, E. Schwerer. 2010. "Administrative Data Sources using the Example of the Austrian Census 2011," *Austrian Journal of Statistics* 39(4):299-308.

Esteve, A. and M. Sobek. 2003. "Challenges and methods of international census harmonization ." *Historical Methods* 36: 66-79

Eurostat. 2014. 2011 Census Hub: https://ec.europa.eu/CensusHub2 Download from HyperCube 44, educational attainment statistics by year of birth for each country, 26.08.2014 04:31:01: https://ec.europa.eu/CensusHub2/query.do?step=selectHyperCube&qhc=false For metadata, see: https://ec.europa.eu/CensusHub2/metadata.do

Feeney, G. and R. McCaa 2014. "Evaluating the accuracy of population census and survey statistics by intracohort comparison." Unpub. Ms.

McCaa, R. 2013. "The Big Data Revolution: IPUMS-International. Trans-Border Access to Decades of Census Microdata Samples for Three-fourths of the World and more." *Revista de Demografía Histórica*, 30(1): 69-87.

McCaa, R. and A. Esteve. 2006. "IPUMS-Europe: Confidentiality measures for licensing and disseminating restricted access census microdata extracts to academic users," *Monographs of official statistics: Work session on statistical data confidentiality*. Luxembourg: Office for Official Publications of the European Communities, pp. 37-46.

McCaa, R., K. Muralidhar, R. Sarathy, M. Comerford, and A. Esteve. 2014. "Controlled shuffling, statistical confidentiality and microdata utility: a successful experiment with a 10% household sample of the 2011 population census of Ireland for the IPUMS-International database," J. Domingo-Ferrer (ed.): *Privacy in Statistical Data, LNCS vol. 8744:* 326-337.

McCaa, R., A. Esteve and T. Lopez. 2012. "Safety in numbers--The IPUMS/IECM trans-border, integrated census microdata dissemination system minimizes risks and maximizes benefits," *EUROSTAT Working group on Demography and Census.* Luxembourg: Sep. 17.

Ruggles, S. 2006. The Minnesota Population Center data integration projects: Challenges of harmonizing census microdata across time and place. *Proceedings of the American Statistical Association, Government Statistics Section.* Alexandria, VA: American Statistical Association, pp. 1405-1415. http://umn.edu/~ruggles/jsmx9.pdf

Thorogood, D. (1999). 'Statistical Confidentiality at the European Level.' Paper presented at: Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, March.

United Nations European Economic Commission (UNECE). Group of Experts on Population and Housing Censuses. 2014. "Quality Management – Draft Text for the Conference of European Statisticians Recommendations for the 2020 Census Round." Geneva. Sep. 23-26.

Appendix A. Example of a Letter of Understanding negotiated in 2003.

Letter of Understanding
Integrated Public Use Microdata Series International
And National Statistical Institute Of Bulgaria

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

1. Ownership. The National Statistical Institute of Bulgaria 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 exclusive purposes of teaching, scientific research. The data analysis results may be used for publishing. Data can not be used for any other purposes without the explicit written approval, in advance, of The National Statistical Institute of Bulgaria.
3. Authorization. To access or obtain copies of integrated microdata of Bulgaria 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,** or authorized distributor. Once approved the user is licensed to acquire integrated anonymous metadata and microdata of Bulgaria 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 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 analysis, resulting from research using metadata or microdata of Bulgaria is permitted in communications such as scholarly papers, journals and the like. The authors of this communications are required to cite the National Statistical Institute of Bulgaria and the **Integrated Public Use Microdata Series International** as sources of the data of Bulgaria, and to indicate that the results and views expressd are those of the author/user.
8. Sharing. **Integrated Public Use Microdata Series International** will provide electronic copies to the National Statistical Institute of Bulgaria of data and

Appendix A (continued).

documentation related to its integrated microdata as well as timely reports of authorized user.

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 Institute of Bulgaria 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: 5/4/05

Signed:

Regents of the University of Minnesota

By: Kevin J. McKoskey, Sponsored Projects Administration

Date: 09.03.05

Signed:

By: Alexander Hadjiiski, President