United Nations ECE/CES/GE.41/2016/22

Economic and Social Council

Distr.: General
23 September 2016

English and Russian only

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Eighteenth Meeting
Geneva, 28 - 30 September 2016
Item 2 of the provisional agenda
Assessing costs and benefits of censuses

Life and value of data after the results are published: scholarly and policy research uses of census microdata from the IPUMS International partnership

Note by the University of Minnesota[1]

Summary

A census provides a valuable record of a country and its people. As the availability of census microdata samples has grown, scholarly research using the data has surged. The harmonization and dissemination system provided by IPUMS facilitates cutting edge research by minimizing researcher data discovery and preparation time and by providing extensive metadata in a user-friendly system. United Nations recommendations now encourage governments to provide access to microdata for research, and the 2030 Sustainability Agenda emphasizes disaggregation of indicators on a number of factors. Microdata are ideal for this purpose as they enable measurements across age, sex, geographic regions, disability status, and other individual or household characteristics. Census microdata are vital for understanding large-scale trends such as urbanization, economic development, fertility patterns and transitions, migratory activities, aging, educational development, and disability. The marginal cost of creating a microdata sample represents a tiny fraction of the overall census budget and is small compared to the benefits. To preserve our collective history and encourage scientific discovery, statistical offices should make provisions to ensure the creation of microdata samples in the course of census preparations.

[1] Prepared by Lara Cleveland and Patricia Kelly Hall ,IPUMS International.

GE.16-16418(E)
1616418

1. Like a great cathedral, iconic building, or important monument, a census provides a record of a people and their culture at a particular time in history. Like those large physical structures, censuses are enormously expensive—costing millions or even billions of dollars. Even the least expensive censuses can cost a government the equivalent of US $1 per person. Registering close to the top of the scale, the United States reportedly spent about $42 per person to count its people in 2010. Estimated costs for a number of recent censuses are listed in Appendix A. Planning, fielding, and analyzing the results of a census requires enormous commitment and investment. In the face of recent global financial events and tightening budgets, censuses are having to do more with less.

I. Importance of census microdata

2. SA census is arguably a more valuable a record of a country and its people than any building or monument, particularly when the information is preserved and made available for scientific research. For a small marginal cost, enormous benefit can be gained by preparing a census microdata sample for public use. In order to preserve our collective history and enable scientific discovery, we argue that statistical offices should make sure that the creation of microdata samples be considered a vital part of the census production process.

3. Access to census microdata is becoming more important than ever. The flexibility of microdata to enable social science exploration is noted in the recent United Nations report on Strengthening the demographic evidence base for the 2030 Agenda for Sustainable Development. In it, the Commission on Population and Development "encourages Governments to adopt open-data policies allowing the dissemination of public-use, geo-referenced and anonymized microdata from censuses, household surveys, civil registration, population registers, health information systems and other relevant administrative records with respect for confidentiality." (http://www.un.org/en/development/desa/population/pdf/commission/2016/documents/CPD49%20Resolution%202.pdf.) The recommendation recognizes the power and value of microdata for disaggregation and complex multivariate analysis.

4. The UN microdata dissemination recommendation coincides with the 2030 Agenda's shift away from singular targets of the Millennium Development Goals toward measurements of change and assessment of differences among subgroups in the Sustainable Development Goal indicators. The new goals also emphasize the tracking of progress by directing monitoring groups to establish baselines against which change can be measured. Microdata is flexible enough to fulfil the disaggregation requirements of SDG monitoring in the service of leaving no one behind. Census data, in particular, provides broad enough coverage and even a sample from a census provides sufficient case counts to enable disaggregation of measurements across age, sex, geographic regions, disability status, and so on. Census microdata also provide rich information about interrelationships among household, family and individual characteristics.

II. Microdata availability in the United States

5. The United States has more than 50 years of experience distributing census microdata samples to the public. The United States has long held a philosophy of open data. Public tax payer dollars pay for government data collection and creation. As such, data are considered the property of the people and should be made public, with proper respect for confidentiality of individuals. By 1962, technological advances made it possible for the United States Census Bureau to release the first census microdata sample for public use from the 1960 census (Ruggles et. al 2015). The 1 in 1000 sample of households and

persons was distributed on 13 Univac tapes or 18,000 punchcards. Following the 1970 census, the United States released 1 in 100 samples with compatible formats and coding schemes for both 1970 and 1960. The existence of the two similarly formatted files spanning a 10-year time frame led to an explosion in research on change over time. Sociologist Otis Dudly Duncan observed the importance of the availability of census microdata samples in these words, "all too often efforts to put information into an appropriate form are frustrated by the inadequacy of the published summary tables for the purpose at hand. With access to the unit records, the social scientist may specify in detail how variables are to be manipulated so as to produce an optimal estimate" (Duncan 1974; 5097).

6. In 1993, the Minnesota Population Center was funded to harmonize codes, record layouts and metadata for nine previously incompatible censuses, spanning the period 1880-1990. For academics old enough to remember, 7-track computer tape was the medium of dissemination with sixty million integrated person records packed on a single reel. Two years later, the first internet website went on-line for the Integrated Public Use Microdata Series (IPUMS), and dissemination by tape was quickly forgotten. The work done by IPUMS to standardize and harmonize data is a boon for researchers, who no longer need to search in multiple places, and through multiple documents to access and compile these data files. The popularity of the data documentation and access system has led to a number of additional grants to harmonize other census and survey files.

7. Since the mid-1990s, the IPUMS family of data integration projects has new United States censuses and surveys to the harmonized database and has continued to improve the data distribution system. The U.S. family of IPUMS projects includes contemporary U.S. Census data samples, full-count historical censuses, the American Community Survey, the National Health Interview Survey, Current Population Survey, and American Time Use Survey. International census data and the Demographic and Health Surveys have also been added to the IPUMS. Today, IPUMS-USA disseminates custom-tailored extracts of samples for censuses of the USA from 1850 to 2010. Each extract, regardless of the number of censuses requested, is pooled into a single data file. Annually, within a week or two after release by the US Census Bureau, the latest American Community Survey (ACS) sample is integrated into the IPUMS-USA database. The updates include triennial and quinquennial versions of the ACS as well as annual. Countries and available samples sizes of the international census data project are available in Table 1.

III. The IPUMS approach to research-ready harmonized microdata

8. IPUMS offers a means of disseminating microdata which complements the dissemination activities of National Statistical Offices. NSOs disseminate official statistics and official statistical products to a large number of publics—citizens, officials, the media, analysts, etc. IPUMS-International disseminates microdata on a restricted access basis to a tiny, but important constituency—researchers who require detailed data on individuals and households to measure and analyze complex relationships, often making comparisons over time and between nations.

9. IPUMS never disseminates the original, raw source census data files. Instead the microdata are transformed, anonymized, harmonized, and integrated such that any single concept, such as primary schooling completed, has the same code in every sample through-out the entire database. Yet, entire datasets are never disseminated. Instead each researcher constructs a bespoke "extract," similar to a table with multiple variables, for a particular research question. The researcher uses the IPUMS data dissemination tool's electronic menus to submit a data extract request that is tailored as to country(ies), census year(s), subpopulation(s), sample densities and variables, according to the individual needs of the

researcher. Each extract request produces a single pooled dataset that is registered to facilitate replicability and to guard against fraud. This method provides strong incentives for users to jealously guard their microdata extracts and to comply fully with the conditions of use. Since complete datasets are not distributed on DVDs or any other medium, the temptation to share microdata with unauthorized individuals is greatly reduced.

10. The classifications in the original source microdata files are inconsistent across census years and countries. Reconciling these codes is a major part of the IPUMS project. IPUMS samples retain all the detail provided in the original samples, except where confidentiality edits are needed. At the same time, we provide a truly integrated database, in which identical categories in different samples receive identical codes. Appendix B lists harmonized variables common to most samples in the IPUMS database.

11. We employ several strategies to achieve the competing goals of maximizing comparability and retaining detail. For simple variables, such as age and sex, the original variables are compatible, and recoding them into a common classification is straightforward. For more complicated variables, it is impossible to construct a single serial 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 hierarchical or composite coding schemes. The first digit of the code provides 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).

12. To illustrate the IPUMS approach, consider the classification scheme for employment status, EMPSTAT (see Table 2). Each column represents a sample. The rows represent concepts, each with its hierarchical code. Each cell indicates the number of cases for the corresponding concept and sample. Composite codes offer an intuitive means of understanding relationships between concepts. With the IPUMS design, the first digit of employment status has three categories consistently available in all samples: 1) active, 2) unemployed, and 3) inactive. Within each of these three categories, various concepts are identified with second and third digits. For example, among the inactive, distinctions are quite complicated. At the second digit of inactive (3x) the IPUMS coding scheme distinguishes between housework, disabled, in school, retired, elderly, institutionalized, intermittent workers, other income recipients, and others. The third and final digit add even more detail, even though a particular concept may occur in relatively few censuses. Retirees (34x) are among the most complicated and varies greatly across samples. The third digit is made up of nine distinct types, although, for brevity, only four are reported in Table 2. Finally, note that the IPUMS coding scheme always assigns "0" to "not in universe (NIU) and 9, 99, 999, etc., to unknown, with the number of 9s determined by the maximum number of digits for the particular variable.

13. The principal advantage of the IPUMS integration scheme is its reconciliation of census-specific variable codes to produce datasets that integrate records across time and space. The basic goal of variable harmonization is to make data suitable for comparative analysis by applying comparable codes for each variables across all samples in the database. Microdata are integrated so that identical concepts have identical codes.

14. In addition to the integrated codes, the IPUMS metadata offer detailed documentation by means of "tabs" for general descriptions, comparability discussions, statements of universe, availability of concepts, detailed wording of the original texts and links to the source documents in the official language and English translation.

15. Appendix C summarizes IPUMS integrated value added variables that are available for each household sample in the IPUMS database. Most of these variables are unique to IPUMS and are rarely available in samples disseminated by National Statistical Offices.

16. Integrated microdata and metadata facilitate informed comparative research, but changes in administrative boundaries pose a major challenge for the spatio-temporal analyses required to accurately monitor demographic and social change. Holding space constant is critical in measuring progress toward goals at sub-national levels; units that have changed boundaries cannot be compared across time in any meaningful way. Consistent spatial geographic units are necessary for accurate measures of change over time involving contextual or spatial elements (Sarkar et al. 2015). IPUMS has developed a method for creating spatially consistent units in the microdata, starting with the first and second administrative units identified in census samples. GIS boundary files corresponding to place of residence, birth place, and other geographic variables are also available for download..

IV. IPUMS microdata users and usage

17. Since the dissemination of researcher-ready harmonized data files began, use of these data for scholarly research has exploded. To date, more than 100 thousand unique users (Figure 1) have registered to use one or more data sources identified above. These users download 2.6 terabytes of data per week from IPUMS websites (Figure 2), producing more than 2,000 scholarly citations per year as listed in Google Scholar (Figure 3). In fact, 13% of articles in the top ranked journal, Demography, are now based on data disseminated through the IPUMS project websites (Figure 4).

18. In 1999, IPUMS went International and within five years disseminated pooled extracts of confidentialized, integrated microdata for eight countries: Brazil (1960-2000), China (1982-1990), Colombia (1962-1993), France (1962-1996), Kenya (1989-1999), Mexico (1960-2000), the United States (1960-2000) and Vietnam (1989-1999). Ten years later the site offers a ten-fold increase in in the quantity of microdata and the number of countries available to researchers. Usage grew even faster, doubling every two or three years. Now, more than 600 million integrated microdata records, encompassing four-fifths of the world's population, are disseminated to over 10,000 researchers in more than 100 countries. Per agreements with National Statistical Offices, only vetted applicants with verifiable bona fides and a credible research project are approved to access the international microdata samples.

19. IPUMS International data users have produced well over 1,000 articles and working papers. Users consist of economists (about 40%), demographers (20%), sociologists (10%), and other social scientists. Statisticians are a growing and newer group of IPUMS International thanks to increased project outreach to members of the professional International Statistical Institute affiliate groups. The majority of users request data to conduct research for the purposes of writing an article or thesis (Table 2). Recent journal articles covering a wide range of topics have appeared in Journal of Applied Economics, Journal of Developmental Economics, Journal of Population Research, International Migration Review, Migration Studies, Population and Development Review, and a host of others. Researchers have modeled internal migration flows (Garcia et al 2014); the influence of gender, marital status, and immigration on labor force participation (Donato et al 2014); unmarried cohabitation in Latin America (Esteve et al 2012); and diversity in household structures (Demont & Heuveline 2008). Two of these studies pair census data with health information to examine AIDS mortality and the labor market (Chicoine 2011) or to analyze a quasi-experimental situation resulting from a particular malaria treatment (Bleakley 2010). It is difficult to put a price tag on the value of research stemming from microdata access, but it vastly outweighs the marginal cost of drawing a preparing a sample from the data prepared for census reports.

V. Conclusion

20. International census microdata, thanks in part to IPUMS, has become a vital part of our shared scientific infrastructure because it provides a unique laboratory for the analysis of economic and social processes and offers the empirical foundation we need for developing and testing theoretical models. Microdata are vital for understanding large-scale trends such as urbanization, economic development, fertility patterns and transitions, migratory activities, aging, educational development, and disability. The data are ideal for assessing consequences of social, economic and demographic change. New methods are enabling the assessment of relationships between human population, climate and the environment. Census microdata are rich and unique living artifacts of our shared social and cultural condition. Preserving this rich history enables continuous reexamination of what makes us grow. Consider again the level of investment in census operations. If published census results provide an important snapshot in time, microdata samples are a living history available for continual exploration. The ability to revisit our rich history and make new scientific discoveries make the marginal cost of drawing a microdata sample from the census well worth the investment.

VI. References

Bleakley, Hoyt. 2010. Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure. *American Journal of Applied Economics* 2(2): 1-45.

Chicoine, Luke. 2011. Aids mortality and its effect on the labor market: Evidence from South Africa. *Journal of Development Economics* 98: 256-269.

Demont, Floriane and Patrick Heuveline. 2008. Diversity and Change in Cambodian Households, 1998-2006. *Journal of Population Research* 25(3): 287-313.

Donato, Katherine M., Bhumika Piya, and Anna Jacobs. 2014. The Double Disadvantage Reconsidered: Gender, Immigration, Marital Status, and Global Labor Force Participation in the 21st Century. *International Migration Review* 48: S335-S376. doi: 10.1111/imre.12142

Duncan, O. D. 1974. Developing social indicators. *Proceedings of the National Academy of Sciences* 71, 5096-5102.

Esteve, A., R. Lesthaeghe and A. López-Gay. 2012. The Latin American cohabitation boom, 1970-2007. *Population and Development Review* 38: 55-81.

Esteve, A. and M. Sobek. 2003. Challenges and methods of international census harmonization. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 36 (2): 66-79.

Garcia, Andres J., Deepa K. Pindolia, Kenneth K. Lopiano, and Andrew Tatem. 2014. Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies* 2014: 1-22.

Ruggles, S., 2014. Big microdata for population research. *Demography*, *51*(1), pp.287-297.

Ruggles, S., L. Cleveland, R. McCaa and M. Sobek. 2015. The IPUMS Collaboration: Integrating and Disseminating the World's Population Microdata, *Journal of Demographic Economics*, 81(2): 203-216.

Ruggles, Steven, Miriam L. King, Deborah Levison, Robert McCaa, and Matthew Sobek. 2003. "IPUMS-International." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 36 (2): 60–65. doi:10.1080/01615440309601215.

Sarkar, Sula, Lara Cleveland, Majory Silisyene, and Matthew Sobek. 2015. "Harmonized census geography and spatio-temporal analysis: gender equality and empowerment of women in Africa." Paper presented at the Annual Meeting of the Population Association of America, San Diego, CA in April.

Sobek, M., L. Cleveland, S. Flood, M. King, S. Ruggles and M. Schroeder. 2011. Big Data: Large-Scale Historical Infrastructure from the Minnesota Population Center, *Historical Methods* 44: 61-68. doi:10.1080/01615440.2011.564572.

https://www.ons.gov.uk/file?uri=/methodology/classificationsandstandards/sexualidentityguidanceandprojectdocumentation/2011censusassessmentsexualorientation_tcm77-181168.pdf

ONS, 2009a. Measuring sexual identity – a guide for researchers.

http://www.ons.gov.uk/ons/guide-method/measuring-equality/equality/sexual-identity-project/guidance/measuring-sexual-identity--a-guide-for-researchers.pdf

ONS, 2009b. 'Trans Data Position Paper'.

http://www.ons.gov.uk/ons/guide-method/measuring-equality/equality/equality-data-review/trans-data-position-paper.pdf

ONS, 2016a. 2021 Census topic consultation.

https://www.ons.gov.uk/census/censustransformationprogramme/consultations/the2021censusinitialviewoncontentforenglandandwales

Specific reports on sexual identity and gender identity
Sexual identity topic report and the Research plan for sexual identity
Gender identity topic report and the Research plan for gender identity

ONS, 2016b. Results from household surveys.
https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/sexuality/bulletins/integratedhouseholdsurvey/2015-10-01.

Appendix A. Cost of Doing the Census

Country	Year	Cost	Cost/Person
Australia[1]	2011	$445,000,000 AUD	$19.75/Person
Cambodia[2]	2018	11.8M	
Canada[1]	2011	$630,373,000 CAD	$18.25/Person
China[3]	2011	$1.34 billion	$1/person
Finland[3]	2011	Around €1m	
Germany[4]	2011	€710 million	
India[5]	2010	US $446M	
Ireland[1]	2011	€63,519,000	€14.12/Person
Kenya[6]	2009	$80M	
Mozambique[7]	2017	US$75Mil	
Nepal[8]	2011	US 15M	
Netherlands[9]		1.4M Euros	
New Zealand[1]	2011	$90,332,087 NZD	$20.41/Person
Poland[10]	2010	$138,184,507	
Scotland[1]	2011	£57,035,027	£10.84/Person
Uganda[11]	2014	About $76 million	
United Kingdom[1]	2011	£482 million	£8.66/Person
USA[3]	2010	$13 billion	$42.11/person

Sources

1- Statistics New Zealand (2012). *Transforming the New Zealand Census of Population and Dwellings: Issues, options, and strategy.* Wellington: Statistics New Zealand.

2- Deervew Marc GL (2016) The Cambodia Population Census: Leaving no one behind. *The Phnom Pen Post.* July 6. http://www.phnompenhpost.com/analysis-and-op-ed/cambodia-population-census-leaving-no-one-behind

3- *The Economist.* (2011). "Costing the count." June 2. http://www.economist.com/node/18772674

4- Der Spiegel Online. (2011). "Germany launches first census since reunification." May 9. http://www.spiegel.de/international/germany/the-count-germany-launches-first-census-since-reunification-a-761497.html

5- Chandramouli, D. (2015). Census of India: Lessons learnt and the way ahead. Presentation at the UN EGM on Strengthening the Demographic Evidence Base for the Post-2015 Development Agenda. October 5-6. http://www.un.org/en/development/desa/population/events/pdf/expert/23/Presentations/EGM-S2-Chandramouli%20presentation.pdf

6- The Coast Week. "Kenyan government kicks off preparations for census." September 13, 2016. http://www.coastweek.com/3933-Kenyan-government-kicks-off-preparations-for-census.htm

7- 2017 Census to cost US75 million. http://clubofmozambique.com/news/mozambique-2017-population-census-to-cost-us75-million/

8- Malla, Uttam Narayan (2011). "Managing population census 2011 of Nepal." Nepal Constitution Foundation. http://www.ncf.org.np/upload/files/658_en_2441.pdf

9- Statistics Netherlands (2014). Dutch Census 2011: Analysis and Methodology. The Hague: Statistics Netherlands. https://www.cbs.nl/NR/rdonlyres/5FDCE1B4-0654-45DA-8D7E-807A0213DE66/0/2014b57pub.pdf

10- Dygaszewicz, Janusz (2014). "Modern Census Polish Case Study." GUS e-census presentation. Washington, D.C. July 31-August 1. http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_088794.pdf

11- New Vision staff reporter (2014). "National census to cost sh190b." May 19. http://www.newvision.co.ug/new_vision/news/1340887/national-census-cost-sh190b

Appendix B. Selected Topical Coverage of Harmonized IPUMS Variables

HOUSEHOLD CHARACTERISICS

Geography
- First administrative level
- Second administrative level
- Urban-rural status

Dwelling
- Number of rooms
- Toilet access
- Construction materials
- Age of structure
- Living area

Utilities
- Electricity
- Water
- Sewage
- Fuel
- Heating

Amenities
- Automobiles
- Washer
- Television
- Computer
- Phone

Other
- Home or land ownership
- Number of deaths
- Number of international migrants
- Family and household composition

PERSON CHARACTERISTICS

Core Demographic
- Age/Year of Birth
- Sex
- Marital status
- Age at marriage
- Relationship to householder

PERSON CHARACTERISTICS

Migration
- Previous residence
- Years in current locality

Fertility/Mortality
- Children ever born
- Children surviving
- Parental mortality

Nativity/Ethnicity
- Place of birth
- Country of birth
- Citizenship
- Year of immigration
- Religion
- Race
- Ethnic group
- Language spoken
- Mother tongue

Education
- School attendance
- Literacy
- Educational attainment
- Years of schooling

Labor Force
- Employment status
- Occupation
- Industry
- Class of worker
- Hours worked
- Total income
- Wage and salary income
- Source of livelihood

Disability
- Disability status
- Type of disability
- Cause of disability

Appendix C. IPUMS-International Value-Added Variables www.ipums.org/international

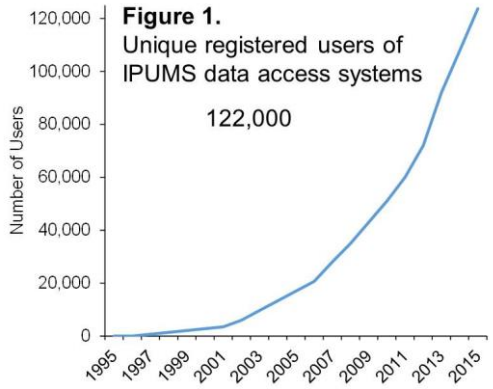
The IPUMS team, with decades of experience in using microdata, developed more than thirty value-added variables to augment each sample. These augmented variables may be grouped into four types: technical, geographic, summary, and pointer.

Technical variables: Record type, Country, Year, IPUMS sample identifier, Household serial number, Number of person records in household, Household weight, Subsample number, Group quarters status, Continent, Region of country, Residence at first and second administrative levels, and Expansion factors (sample weights—for households and persons).

Geographic variables: Urban/rural, Region of the world, Geographic Sub-National Level 1, Geographic Sub-National Level 2, and GIS Boundary files plus consistent geographic units harmonized by name or spatial area, according to researcher preference. See: https://international.ipums.org/international/gis.shtml

Summary household and family variables: Household classification, Number of families in household, Number of married couples in household, Number of mothers in household, Number of fathers in household, Head's location in household, Number of unrelated persons, Family unit membership, Number of own family members in household, Number of own children in household, Number of own children under age 5 in household, Age of eldest own child in household, and Age of youngest own child in household.

Pointer variables to identify co-resident spouses, children and their parents: Mother's, Father's and Spouse's location in household, Rule for linking parent(s) and spouse(s), Probable stepmother, Probable stepfather, Man with 2 or more wives linked and Second or higher order wife, etc. The availability of pointer variables is the one of the most valuable addition to IPUMS integrated microdata because pointer variables readily facilitate analysis of children by the characteristics of their mothers or fathers as well as husbands by the characteristics of their wives and vice-versa (Sobek and Kennedy 2009). Own-child fertility analysis is made easy because every IPUMS household dataset already links mothers to their co-resident children (MOMLOC) and the "Attach Characteristics" feature of the IPUMS extract system can be used to place mother's characteristics on the record of each child. For biological mothers it is important to filter with the STEPMOM=0 variable, also provided by IPUMS. https://international.ipums.org/international-action/variables/MOMLOC#description_section.



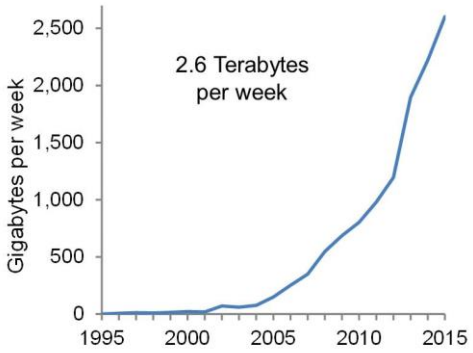


Figure 2. IPUMS Data Dissemination

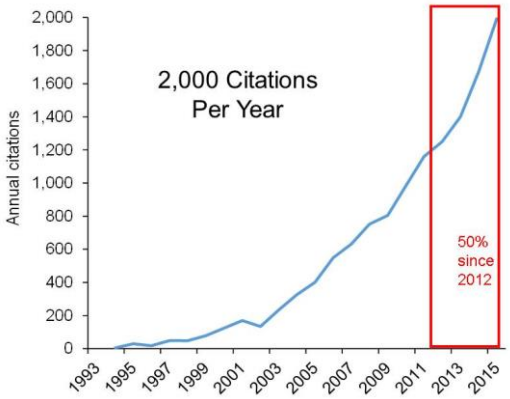


Figure 3. Citations of IPUMS Data in Google Scholar, 1994-2015

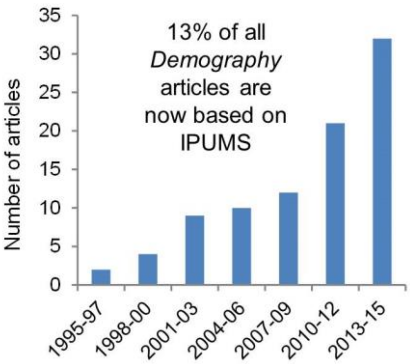


Figure 4. *Demography* articles based on IPUMS Data

Table 1. IPUMS Integrated Microdata Samples https://international.ipums.org - 86 countries 675 million person records (+ = 2016 launch)

Census	%	Persons	Census	%	Persons	Census	%	Persons	Census	%	Persons
1970 Argentina	2	466,892	**1960 Chile**	1	88,184	**1962 France**	5	2,320,901	**2005 Indonesia-cont.**	0.5	1,090,892
1980	10	2,667,714	1970	10	890,481	1968	5	2,487,778	2010	10	23,603,049
1991	10	4,286,447	1982	10	1,133,062	1975	5	2,629,456	**2006 Iran**	2	1,299,825
2001	10	3,626,103	1992	10	1,335,055	1982	5	2,631,713	*2011+*	*2*	*1,482,000*
2010	10	3,966,245	2002	10	1,513,914	1990	4.2	2,360,854	**1997 Iraq**	10	1,944,278
2001 Armenia	10	326,560	**1982 China**	1	10,039,191	1999	5	2,934,758	**1971 Ireland**	10	296,878
2011	10	301,831	1990	1	11,835,947	2006 (RR)	33	19,973,287	1979	10	337,686
1971 Austria	10	749,894	*2000+*	*1*	*11,804,000*	2011	33	20,541,337	1981	10	344,291
1981	10	756,556	**1964 Colombia**	2	349,652	**1970 Germany**	5	3,094,845	1986	10	355,020
1991	10	780,512	1973	10	1,988,831	1971 DR	25	4,089,856	1991	10	353,149
2001	10	803,471	1985	10	2,643,125	1981 DR	25	4,278,563	1996	10	365,323
2011	10	839,501	1993	10	3,213,657	1987	5	3,160,224	2002	10	410,688
1991 Bangladesh	10	10,580,904	2005	10	4,006,168	**1984 Ghana**	15	1,309,352	2006	10	440,314
2001	10	12,442,115	**1963 Costa Rica**	6	82,345	2000	10	1,894,133	2011	10	474,535
2011	5	7,205,720	1973	10	186,762	2010	10	2,466,289	**1972 Israel**	10	315,608
1999 Belarus	10	990,706	1984	10	241,220	**1971 Greece**	10	845,483	1983	10	403,474
2009+	*10*	*941,000*	2000	10	381,500	1981	10	923,108	1995	10	556,365
1976 Bolivia	10	461,699	2011	10	430,082	1991	10	951,875	**2001 Italy**	5	2,990,739
1992	10	642,368	**2002 Cuba**	10	1,118,767	2001	10	1,028,884	**1982 Jamaica**	10	223,667
2001	10	827,692	**1960 Dominican Rep**	6.6	201,556	*2011+*	*10*	*1,057,000*	1991	10	232,625
1981+ Botswana	*10*	*97,000*	1970	6.8	272,090	**1983 Guinea**	10	457,837	2001	10	205,179
1991+	*10*	*133,000*	1981	8.5	475,829	1996	10	729,071	**2004 Jordan**	10	510,646
2001+	*10*	*169,000*	2002	10	857,606	**1971 Haiti**	10	434,869	**1969 Kenya**	3.3	659,310
2011+	*10*	*202,000*	2010	10	943,784	1982	2.5	128,770	1979	5	1,033,769
1960 Brazil	5	3,001,439	**1962 Ecuador**	3	136,443	2003	10	838,045	1989	5	1,074,098
1970	5	4,953,759	1974	10	648,678	**1970 Hungary**	5	515,119	1999	5	1,407,547
1980	5	5,870,467	1982	10	806,834	1980	5	536,007	2009	10	3,841,935
1991	5.8	8,522,740	1990	10	966,234	1990	5	518,240	**1999 Kyrgyzstan**	10	476,886
2000	6	10,136,022	2001	10	1,213,725	2001	5	510,502	2009	10	564,986
2010	5	9,693,058	2010	10	1,448,233	*2011+*	*5*	*497,000*	**1974 Liberia**	10	150,256
1985 Burkina Faso	10	884,797	**1986+ Egypt**	*15*	*6,799,000*	**1983 India - NSSO**	0.1	623,494	2008	10	348,057
1996	10	1,081,046	1996	10	5,902,243	1987	0.1	667,848	**1987 Malawi**	10	798,669
2006	10	1,417,824	2006	10	7,282,434	1993	0.1	564,740	1988	10	991,393
1998 Cambodia	10	1,141,254	**1992 El Salvador**	10	510,760	1999	0.1	596,688	2008	10	1,341,977
2008	10	1,340,121	2007	10	574,364	2004	0.1	602,833	**1970 Malaysia**	2	175,997
1976 Cameroon	10	736,514	**1984 Ethiopia**	10	3,404,306	*2010+*	*0.1*	*460,000*	1980	2	182,601
1987	10	897,211	1994	10	5,044,598	**1971 Indonesia**	0.5	634,642	1991	2	347,892
2005	10	1,772,359	2007	10	7,434,086	1976	0.2	281,170	2000	2	435,300
1971 Canada	1	214,019	**1966 Fiji Islands**	10	47,579	1980	5	7,234,577	**1987 Mali**	10	785,384
1981	2	486,875	1976	10	57,214	1985	0.3	605,858	1998	10	991,330
1991	3	809,654	1986	10	72,158	1990	0.5	912,544	2009	10	1,451,856
2001	2.5	801,055	1996	10	77,382	1995	0.3	718,837			
2011+	*3*	*926,000*	2007	10	84,323	2000	10	20,112,539	(continued)		

Table 1 (continued) IPUMS Integrated Microdata Samples https://international.ipums.org - 86 countries 675 million person records (+ = 2016 launch)

Census	%	Persons	Census	%	Persons	Census	%	Persons	Census	%	Persons
1960 Mexico	1.5	502,800	**1993 Peru**	10	2,206,424	**2008 Sudan**	17	5,066,530	**1963 Uruguay**	10	256,171
1970	1	483,405	2007	10	2,745,895	**1970 Switzerland**	5	312,538	1975	10	279,994
1990	10	8,118,242	**1990 Philippines**	10	6,013,913	1980	5	317,803	1985	10	295,915
1995	0.4	332,061	1995	10	6,864,758	1990	5	342,797	1996	10	315,920
2000	11	10,099,182	2000	10	7,417,810	2000	5	364,086	2006	6	256,866
2005	10	10,284,550	***1978+ Poland***	*10*	*3,577,000*	**1988 Tanzania**	10	2,310,424	2011	10	328,425
2010	10	11,938,402	*1988+*	*10*	*3,894,000*	2002	10	3,732,735	**1971 Venezuela**	10	1,158,527
2015+	*10*	*11,292,000*	*2002+*	*10*	*3,824,000*	*2011+*	*10*	*4,497,000*	1981	10	1,441,266
1989 Mongolia	10	190,631	*2011+*	*5*	*2,000,000*	**1970 Thailand**	2	772,169	1990	10	1,803,953
2000	10	243,725	**1981 Portugal**	5	492,289	1980	1	388,141	2001	10	2,306,489
1982 Morocco	5	1,012,873	1991	5	491,755	1990	1	485,100	**1989 Vietnam**	5	2,626,985
1994	5	1,294,026	2001	5	517,026	2000	1	604,519	1999	3	2,368,167
2004	5	1,482,720	2011	5	528,870	***1970+ Trinidad & T***	*10*	*69,000*	2009	15	14,177,590
1997 Mozambique	10	1,551,517	**1970 Puerto Rico**	1	27,212	*1980+*	*10*	*105,000*	**1990 Zambia**	10	787,461
2007	10	2,047,048	1980	5	160,219	*1990+*	*10*	*113,000*	2000	10	996,117
2001 Nepal	11	2,583,245	1990	5	177,655	*2000+*	*10*	*112,000*	2010	10	1,321,973
1960 Netherlands	1.2	143,251	2000	5	189,828	*2011+*	*10*	*117,000*	**Candidates for 2017 and beyond:**		
1971	1.2	159,203	2005 (PRCS)	1	35,416	**1985 Turkey**	5	2,554,364	**2010 round censuses**		
2001	1.2	189,725	2010	1	36,032	1990	5	2,864,207	**More countries/places**		
1971 Nicaragua	10	189,469	**1977 Romania**	10	1,937,021	2000	5	3,444,456	Angola		
1995	10	435,728	1992	10	2,238,578	**1991 Uganda**	10	1,548,460	Australia		
2005	10	515,485	2002	10	2,137,967	2002	10	2,497,449	Belgium		
2006 Nigeria - GHS	0.1	83,700	*2011+*	*10*	*1,990,000*	**2001 Ukraine**	10	4,889,288	Benin		
2007	0.1	85,183	**1991 Rwanda**	10	742,918	**1991 UK**	1	541,894	Bulgaria		
2008	0.1	107,425	2002	10	843,392	2001	3	1,843,525	Cote d' Ivoire		
2009	0.1	77,896	*2012+*	*10*	*1,038,369*	**1960 USA**	1	1,799,888	Finland		
2010	0.1	72,191	**1980 Saint Lucia**	10	11,451	1970	1	2,029,666	Guatemala		
1973 Pakistan	2	1,453,332	1991	10	13,382	1980	5	11,343,120	Guinea Bissau		
1981	10	8,433,058	**1988 Senegal**	10	700,199	1990	5	12,501,046	Honduras		
1998	10	13,102,024	2002	10	994,562	2000	5	14,081,466	Japan		
1997 Palestine	10	259,191	**2004 Sierra Leone**	10	494,298	2005 (ACS)	1	2,878,380	Korea, Republic of		
2007	10	227,067	**2002 Slovenia**	10	179,632	2010	1	3,061,692	Madagascar		
1960 Panama	5	53,553	**1996 South Africa**	10	3,621,164	**1963 Uruguay**	10	256,171	Mauritius		
1970	10	150,473	2001	10	3,725,655	1975	10	279,994	Myanmar		
1980	10	195,577	2007	2	1,047,657	1985	10	295,915	Namibia		
1990	10	232,737	2011	10	4,418,594	1996	10	315,920	Niger		
2000	10	284,081	**2008 South Sudan**	7	542,765	2006	6	256,866	Nigeria PES		
2010	10	341,118	**1981 Spain**	5	2,084,221	2011	10	328,425	Papua New Guinea		
1962 Paraguay	5	90,236	1991	5	1,931,458	**1971 Venezuela**	10	1,158,527	Russia		
1972	10	233,669	2001	5	2,039,274	1981	10	1,441,266	Tunisia		
1982	10	301,582	2011	10	4,107,465	1990	10	1,803,953	Turkmenistan		
1992	10	415,401				2001	10	2,306,489	Yemen		
2002	10	516,083							Zimbabwe, etc.		

Table 2. IPUMS Integrated variable: Employment Status (EMPSTAT) - Detailed codes, case count view for 8 selected samples

	ISO Country Code	BD	KH	CN	IN	ID	IR	KG	VN
	Year	2011	2008	1990	2004	2010	2006	2009	2009
Code	Label								
0	**NIU** (not in universe)	1,117,354	·	3,264,675	59,373	4,587,786	214,837	173,245	3,746,089
	ACTIVE (in labor force)								
100	**EMPLOYED**, not specified	2,123,058	692,961	6,775,149	·	·	·	222,282	·
110	At work	·	·	·	226,309	10,272,827	339,173	·	7,890,417
111	At work, and 'student'	·	·	·	·	·	·	·	·
112	At work, and 'housework'	·	·	·	·	·	·	·	·
113	At work, and 'seeking work'	·	·	·	·	·	·	·	·
114	At work, and 'retired'	·	·	·	·	·	·	·	·
115	At work, and 'no work'	·	·	·	·	·	·	·	·
116	At work, and other situation	·	·	·	·	·	·	·	·
117	At work, family holding, not specified	·	·	·	·	·	·	·	·
118	At work, family holding, not agricultural	·	·	·	·	·	·	·	·
119	At work, family holding, agricultural	·	·	·	·	·	·	·	·
120	Have job, not at work in reference period	·	·	·	2,901	257,792	21,262	·	45,629
130	Armed forces	·	·	·	·	·	·	·	·
131	Armed forces, at work	·	·	·	·	·	·	·	·
132	Armed forces, not at work in reference	·	·	·	·	·	·	·	·
133	Military trainee	·	·	·	·	·	·	·	·
140	Marginally employed	·	·	·	·	·	·	·	·
200	**UNEMPLOYED**, not specified	56,753	·	61,553	10,672	287,300	62,495	25,880	56,105
201	Unemployed 6 or more months	·	2,780	·	·	·	·	·	·
202	Worked fewer than 6 months, permanent job	·	·	·	·	·	·	·	·
203	Worked fewer than 6 months, temporary job	·	·	·	·	·	·	·	·
210	Unemployed, experienced worker	·	·	·	·	·	·	·	·
220	Unemployed, new worker	·	9,090	·	·	·	·	·	·
230	No work available	·	·	·	·	·	·	·	61,460
240	Inactive unemployed	·	·	·	1,138	1,026,714	·	·	·
300	**INACTIVE** (not in labor force)	2,057,472	·	·	·	7,170,630	·	143,579	·
310	Housework	1,851,082	47,041	729,891	105,647	·	305,437	·	557,280
320	Unable to work/disabled	·	·	295,175	·	·	·	·	151,033
321	Permanent disability	·	·	·	4,876	·	·	·	·
322	Temporary illness	·	·	·	539	·	·	·	·
323	Disabled or imprisoned	·	·	·	·	·	·	·	·
330	In school	·	330,375	436,317	155,130	·	222,874	·	816,753
340	Retirees and living on rent	·	4,523	·	6,662	·	·	·	·
341	Living on rents	·	·	·	·	·	·	·	·
342	Living on rents or pension	·	·	·	·	·	55,332	·	·
343	Retirees/pensioners	·	·	·	·	·	·	·	·
344	Retired	·	·	230,000	·	·	·	·	·
345	Pensioner	·	·	·	·	·	·	·	·
346	Non-retirement pension	·	·	·	·	·	·	·	·
347	Disability pension	·	·	·	·	·	·	·	·
348	Retired without benefits	·	·	·	·	·	·	·	·
350	Elderly	·	·	·	·	·	·	·	·
351	Elderly or disabled	·	·	·	·	·	·	·	·
360	Institutionalized	·	·	·	·	·	·	·	·
361	Prisoner	·	·	·	·	·	·	·	·
370	Intermittent worker	·	·	·	·	·	·	·	20,418
371	Not working, seasonal worker	·	·	·	·	·	·	·	·
372	Not working, occasional worker	·	·	·	·	·	·	·	·
380	Other income recipient	·	·	·	·	·	·	·	·
390	Inactive, other reasons	·	1,937	43,187	29,586	·	65,781	·	800,359
391	Too young to work	·	·	·	·	·	·	·	·
392	Dependent	·	251,414	·	·	·	·	·	·
999	**UNKNOWN/MISSING**	.	·	·	·	·	12,634	·	32,047