Timely dissemination of integrated microdata and metadata for the 2010 round of censuses: The IPUMS-Latin America challenge[1] [2]

Robert McCaa[3]

> "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 [on-site] 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."
> —Dennis Trewin (2007), president emeritus International Statistical Institute
> www.hist.umn.edu/~rmccaa/ipums-global/trewin_report_2007.pdf

1. Summary.

1. Integrated, anonymized census microdata and metadata for 159 censuses encompassing 55 countries are presently being disseminated from the IPUMS-International web-site by the Minnesota Population Center (MPC). The Americas are represented by 66 censuses, 16 countries, and 158 million person records. 48.6% of the integrated person records available to researchers are from the Americas. All the Spanish speaking countries plus Brazil, Canada, Jamaica, Saint Lucia and the USA are participating in IPUMS-Latin America.   Research demand for Latin American census microdata is high.   58.8% of IPUMS-International users request integrated samples of Latin American countries—20% for Mexico and 14% for Brazil, compared with only 5% for France and 2% for Spain.

2. The purpose of this paper is, first, to invite statistical institutes of the Americas to continue to participate in the IPUMS-Latin American census microdata project, and, second, to suggest guidelines for preparing 2010 round census microdata and metadata for timely, efficient integration into the IPUMS-International database. Over the past decade, more than 300 sets of census microdata and the corresponding documentation, in a great profusion of forms, have been entrusted to the IPUMS-Latin America, Europe and International projects. Nonetheless, processing time is reduced and errors minimized when metadata in the official language fully document the microdata. Statistical institutes are strongly encouraged to complete a detailed form (see appendix A) to accompany each set of census microdata and metadata.

3. For maximum safety, microdata should be transmitted as REDATAM or ASCII files. If the later, the files should be encrypted and zipped, with the password emailed or faxed in a separate communication to the IPUMS-International project coordinator. Metadata may be transmitted as images, but should also be made available as ASCII, REDATAM, CSPro, SPSS, STATA, SAS, spreadsheet, or document files, using DDI (Data Document Initiative) hypertext or other emerging standards. Documentation in the official language(s) is essential. English translation should be provided, where available. Where necessary, translators—contracted and paid by the MPC—prepare unofficial English texts in simple ASCII format.

4. In a few short years, the complete integration of high-precision samples for all of Latin America will be accomplished, thanks to the early, enthusiastic support of the official statistical institutes of the region. Over the next four years, the IPUMS-International database is likely to double in size, and the Americas will lead the way by contributing microdata from the 2010 round of censuses without undue delay.

---

2. Introduction. IPUMS-International: "best practice".

4. Mr. Dennis Trewin's accolade "best practice" sums up his meticulous assessment of the IPUMS-International facilities, policies and procedures for archiving, processing and disseminating anonymized census microdata samples. Mr. Trewin, as the chair of the UNECE task force to produce guidelines on good practice for the release of microdata and the protection of confidentiality, is widely recognized as an authority in this field. His strongly positive evaluation of the data protections afforded by the IPUMS-International project assures producers and users alike that we are on the right path as we begin our second decade of activities. Readers unfamiliar with the IPUMS-International project's data protections and confidentiality measures are referred to our paper for the UNECE/Eurostat work session on statistical data confidentiality subsequently published in Monographs of Official Statistics ((www.unece.org/stats/documents/2005.11.confidentiality.htm see wp.5 and McCaa and Esteve, 2006).

5. 159 anonymized, integrated high-precision samples of population census microdata are presently available at no cost via www.ipums.org/international, the IPUMS-International web-site. The database is likely to double in size over the next five years, thanks to renewed major funding through 2014 by the National Science Foundation and National Institutes of Health (USA) and to the generous, efficient support of national statistical institute partners. More than 4,000 researchers representing 82 countries are accredited to access microdata through the IPUMS-International site. Researchers use integrated census microdata for comparative analysis across time and space. It is important to note that the IPUMS-International project disseminates only integrated, anonymized microdata—not official statistics nor the source files entrusted to the project. Researchers seeking official census statistics or official samples are directed to websites of our National Statistical Institute partners.

6. This massive data infrastructure already encompasses 55 countries, including 16 for the Americas: Argentina, Bolivia, Brazil, Canada, Chile, Colombia, Costa Rica, Cuba, Ecuador, Mexico, Panama, Peru, Puerto Rico, Saint Lucia, USA and Venezuela (see Table 1). The IPUMS-International database totals more than 325 million anonymized, integrated person records representing 86 million households. The 2011 release is scheduled to incorporate samples for two American countries—Jamaica and Nicaragua—plus eight countries from Europe, Africa and Asia. Over the next five years we propose to incorporate household samples from the 2010 round censuses as well as microdata from other countries, including for the Americas (McCaa, Esteve, Ruggles and Sobek 2006).

4. Need for succinct descriptions of Census and Microdata: form "A".

7. If the IPUMS-Latin America project is to disseminate 2010 round samples in a timely manner, cooperation of national statistical institute partners is essential. As academics, we understand that official statisticians are typically over-burdened with pressing demands from government, business, and the public for an ever increasing array of timely statistics. Therefore we are prepared to work, as we have over the past decade, with metadata and microdata in whatever form without special treatment or consideration.

8. Form "A" (see Appendix A) should be used to succinctly describe each census and its corresponding metadata and microdata. Form A should be completed by a census expert of the respective National Statistical Institute. Microdata are disseminated according to the terms of the standard IPUMS-International Letter of Agreement (see Appendix B). An example of completed forms for four censuses of Argentina is reproduced as Appendix C. Additional examples may be viewed at https://international.ipums.org/international/samples.shtml by clicking the name of a country.

9. Form "A" is organized into four categories: description of the census, characteristics of the microdata, units identified in the microdata and unit definitions.

    1) <u>Description of the census.</u> The following elements are requested:
       i. official title,
      ii. agency that conducted the census,

    iii. population universe (note if special populations are omitted, such as nomads, foreigners, etc),
    iv. de jure or de facto,
    v. census day(s),
    vi. field work period,
    vii. number and type of enumeration forms,
    viii. type(s) of field work,
    ix. respondent and
    x. coverage.

2) <u>Characteristics of the microdata</u>:
    i. source (usually the National Statistical Institute, National Data Archive or University Research Organization),
    ii. sample design (preferably every tenth household after a random start),
    iii. sample unit (household for private entities; individual for collective or group quarters),
    iv. sample fraction (10% for both private households and group quarters because these may differ—see below),
    v. sample size (number of person records), and
    vi. brief description of sample weights, when standard IPUMS protocols are not used.

3) <u>Units identified in the microdata</u> (indicate yes/no and add any comments desired):
    i. dwellings,
    ii. vacant dwellings,
    iii. households,
    iv. individuals,
    v. group quarters,
    vi. lodging,
    vii. smallest identified geographical unit (name),
    viii. settled/unsettled/special populations identified in the microdata
    ix. special household modules (mortality, emigration, agriculture, health, disability, etc.).

4) <u>Unit definitions</u>:
    i. dwellings,
    ii. private households,
    iii. group quarters, and
    iv. settled/unsettled or special populations.

10. Additional items may be added to the form as necessary (e.g., details for modules regarding mortality, emigration, fertility, agriculture, etc.). The form should be submitted to the MPC in the official language. If form "A" is already posted on the IPUMS-International website for the country of expertise (see "samples.shtml" web link above), please check entries for each census to confirm that the information is correct and email any suggestions, corrections or comments to ipumsi@pop.umn.edu.

5. Metadata needs.

11. Metadata serve a number of purposes within the IPUMS-International system. Much of the basic metadata is required to accurately process and assess the microdata as they are incorporated into the database and to support the harmonization work conducted on specific variables. Comprehensive and complete metadata is essential if the integration is to succeed and researchers are to make best use of the microdata (Statistics Canada 2008; see also McCaa and Thomas 2009). Metadata may be transmitted as images, but should also be made available as ASCII, REDATAM, CSPro, SPSS, STATA, SAS, spreadsheet, document, or hypertext files. We are happy to receive more than one version as well. When documents are *not* available in electronic form, the project will scan them, for posting on the IPUMS-International website, organized by country and census year, so that they are easily accessible. Copies of census documentation scanned by the MPC are also made available on CD/DVD to the respective statistical agency as well as national and international research organizations.

12. We have three goals with respect to metadata.

13. First, researchers must have ready access to the original census documentation in the official language. At a minimum, census questionnaires, enumerator instructions or training manuals, data dictionaries and codebooks are required. Additional metadata regarding the organization, preparation, and actual census taking are also valuable to the IPUMS-International project and are catalogued and archived with all other documents received. Original hardcopy or PDF documents are preferred for published metadata materials. Our goal is to provide an archived collection of high-quality PDF files for all forms of metadata pertaining to census microdata. Census outputs of the following metadata are requested from the National Statistical Institutes:

1) Census enumeration forms.
2) Census enumerator instructions (sometimes referred to as training manuals).
3) "Codebooks" or "Data Dictionaries" for each dataset (definitions of record structures, column location of variables and labels for codes, such as the U.S. Census Bureau "IMPS" data dictionary files), including administrative geography, occupations, etc.
4) Correspondence tables indicating the equivalence between coding schemes in two or more censuses or between a census and an international standard (ISCO, ISCED, etc.) These tables are especially helpful to harmonize changes in administrative geography and in the integration of occupation, industry, and educational attainment variables.
5) Basic tables of official results as they are published on a website, book, or CD.
6) Technical and methodological reports on census operations, concepts, nomenclatures, comparability, quality, post-enumeration surveys, etc.
7) Where microdata are provided as samples, the sample design should be described in detail. Where the standard IPUMS-International design of every n^{th} household after a random start is employed, no additional documentation is needed (see microdata specifications below). Otherwise, it would be helpful to receive estimates of sampling errors for a scale of absolute or relative frequencies (for example, where sample percent = 2, 5, 10, 15, 20, 25, 30, 35, 40, and 50), and for key variables, such as age, relationship to reference person, education, and employment status. It should be noted that, to date, the National Institute of Statistics of Mozambique has provided the most comprehensive documentation on sample design and errors (Megill 2007).
8) Boundary files corresponding to the administrative geography coded in the microdata (corresponding to the European standard of NUTS1, NUTS2 and NUTS3) and suitable for dissemination to researchers. If boundary files are not provided, we plan to construct unofficial files from readily available sources.

14. Second, we construct a dynamic metadata system for every variable, integrated as well as non-harmonized, to make it easy to compare both the phrasing of a particular question and the corresponding instructions to the enumerators, in English, for any combination of countries and censuses.

15. Third, from the original source documentation, we write integrated metadata describing each variable as follows:
1) brief definition and description of the selected variable,
2) availability (list of countries and census years with the variable),
3) general comparability (nuances of varying definitions),
4) universe (population to which the question is addressed),
5) reference period (e.g., for economic activity, seven days, last month, a year, etc.),
6) variations in definitions of specific attributes (e.g., "employed"), and
7) comparability discussions for specific censuses organized by country.
The researcher views these pages by simply clicking the variable name. The pages are constructed on demand by the dynamic metadata system. Only the comparability discussions for the currently selected censuses are displayed.

16. Electronic copies of source documentation are preferred. Nonetheless, paper publications or photocopies are also welcome. Electronic files may be emailed as attachments or sent by courier service on CDs. Where English translations are needed, professional translators will be contracted

and unofficial translations produced in simple text format. To avoid loss of paper or CD materials and to economize effort, the entire collection should be assembled in a single package, and sent by courier at project expense.

17. For structured metadata (data dictionaries, code lists, definitions, forms, etc.) the use of emerging standards—such as the Data Documentation Initiative (www.icpsr.umich.edu/DDI/codebook/) found in NESSTAR and the Microdata Toolkit developed by the International Household Survey Network (http://www.surveynetwork.org/home/) and WorldBank—facilitates the transfer of information into the IPUMS-International processing system. DDI is a mark-up structure using Extensible Markup Language (XML) which identifies specific elements commonly found in the codebook accompanying a data file. It covers identifying information on the data file, census or survey characteristics, sample characteristics, unit definitions, methodology, file structures, variable content and structure, question content and relationship to variables, code lists, and related materials either in-line or through reference to external documents.

18. New versions of DDI, available since 2008, expand coverage to support capturing and relaying information about the complex harmonization process used to construct integrated variables. Soon, we expect to offer to accredited researchers who request microdata extracts the corresponding customized codebooks constructed from the metadatabase underlying the IPUMS-International interface and extraction system.

6. Microdata needs.

19. For microdata we have two main goals: first, to permanently archive original source files on behalf of the National Statistical agency partner, and second, to disseminate high-precision, anonymized, integrated and customized household sample extracts to accredited researchers. We prefer that National Statistical Institutes entrust confidentialized copies (names, addresses, and identification numbers suppressed) of complete source files (i.e., 100% microdata) so that we may draw samples consistently, efficiently, and with a minimum of burden on statistical agency partners. Moreover, should imperfect records be encountered, such problems may be resolved easily by replacement, rather than imputation. It should be noted that all microdata source files entrusted to the Minnesota Population Center are archived under total security ("Icebox") and are never reproduced for any person or institution under any circumstances. As the Trewin report notes the Minnesota Population Center seeks to maintain a perfect, unblemished record of security.

20. Additional goals, under consideration, are:
   1) Develop an on-line tabulator to offer integrated tabulations for multiple countries and census years. Preferably the tabulator would be harnessed to 100% microdata, but for anonymization purposes, low-level geography would be suppressed. A prototype is already functioning for a dozen European countries.
   2) Develop a GIS product with microdata linked to harmonized areal units of a minimum of 20,000 population at the most recent census.
   3) Over-sample important, but infrequently occurring events (maternal mortality) or characteristics (disabilities). For example, from the 100% microdata we propose to include households with all maternal deaths to provide the highest possible precision to analyze this difficult to measure phenomena (see Garenne, McCaa, and Nacro 2008). We have developed a user-friendly method for supplying over-samples without compromising our strong anonymization protections. Moreover our method ensures that researchers use the proper expansion factors.

21. Four modalities for entrusting microdata have emerged over the first decade of IPUMS-International partnerships (bulleted items are examples):

   1) The task of archiving 100% microdata source files and producing samples is entrusted to the Minnesota Population Center (38 national statistical institutes).
   2) Samples produced entirely by the national statistical institute according to IPUMS-International specifications where 100% microdata are available (25 countries).

- Federal Statistical Office—Germany: All work performed by FSO, including the 1970 and 1987 censuses of the Federal Republic of Germany and the 1971 and 1981 censuses of the German Democratic Republic.
- Statistics Netherlands (SN). 1960 and 1971 and a register based sample for 2001—all work performed by SN.
- Federal Statistics Office (FSO)—Switzerland: 1971, 1981, 1991, and 2001 – prepared by the FSO.

3) Public or restricted use microdata samples entrusted to researchers are also entrusted to IPUMS-International with or without payment of license fee (12 countries):
- National Bureau of Statistics, China (license fee invoiced for 1982; not 1990)
- National Statistical Survey Organization, India (standard license fee invoiced for 5 samples)
- Statistics Canada (no license fee invoiced)
- Office of National Statistics, United Kingdom (no license fee invoiced)
- Statistics South Africa (no license fee invoiced)

4) The task of producing anonymized samples is entrusted to an institution or individual expert under supervision of the national statistical authority (6 countries)
- INSEE—France: 1962, 1968, 1975, 1982, 1990 and 1999 – prepared by an individual researcher working within the INSEE under contract with the Minnesota Population Center and with INSEE oversight.
- INSSE—Romania: Work performed by a university research institute for the censuses of 2002, 1992, and 1977 under contract with the MPC and with INSSE oversight.

22. Each national statistical institute determines the modality to be used. The project is always amenable to considering other arrangements. Regardless of modality, the project offers a fee of US$5,000 to license microdatasets numbering 1 million or more person records as well as to offset the costs of assembling microdata and documentation.

23. "High precision" is typically defined as samples of ten percent or higher (83 of 159 datasets currently integrated), followed by 5% (n=36). Of the 40 samples that are less than 5%, twenty-one include all extant microdata, many of which are historical samples.

24. Where 100% microdata cannot be entrusted, systematic random samples are preferred according to the following simple protocol:
1) Sort the microdata files by major and minor administrative divisions down to the census tract level, dwelling, household, family and person.
2) After a random start, select every n^{th} private dwelling (every tenth for a 10% sample).
3) For institutional households—or large private households that could be identifiable solely because of their size—after a random start, draw every n^{th} person using the same density as for private dwellings.

25. Systematic random samples capitalize on 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. To the extent the strata used to draw a high precision sample are associated with the variables of interest (e.g., orphanhood, poverty, unemployment, etc.), the resulting estimates of these variables will have lower standard errors than what would have resulted had a simple random sample of records been drawn (Davern, et. al., 2009).

26. One of the major advantages of using census microdata is its geographical power, which allows sub-national analysis without compromising statistical significance. Due to confidentiality constraints, fine geographical detail must be excluded from census microdata, even when disseminated on a restricted access basis, as in the case of the IPUMS project. Typically only the first two levels of geographic detail is provided, such as province and commune, state and county, NUTS1 and NUTS2, etc. In addition, a size of locality variable is preferred because it would facilitate a consistent measure

of urban-rural residence across samples. Size of place categories for Germany and France are as follows:

Germany (preferred)	France
1) 1 to 2,499 persons	1 to 4,999
2) 2,500 to 9,999	5 to 9,999
3) 10,000 to 49,999	10 to 19,999
	20 to 49,999
4) 50,000 to 99,999	50 to 99,999
5) 100,000 to 499,999	100,000 to 1,999,999
6) 500,000 or 1,999,999	2,000,000 or more
7) 2,000,000 or more	

27. Anonymization may be performed by the statistical institute or, upon request, by the Minnesota Population Center. Microdata extracts are disseminated to accredited researchers under strict legal and administrative controls (McCaa and Esteve 2006; McCaa, Ruggles, et. al. 2006). While we concur with Anderson and Fienberg (2001) that sampling of datasets alone "provides the additional uncertainty needed to protect many data releases…," we do not stop there. We employ six layers of technical protections. First, we suppress place of enumeration, residence, work or schooling codes for geographical units that fall below a threshold of 20,000 persons in the most recent census. (Some statistical institutes set the threshold higher, such as the UK, where the number is 65,000). Second, for categorical variables, any value with a population frequency of less than 250 is likewise suppressed (FSO-Germany is applying a threshold of 2,500). Such values are recoded as "other," "missing," or in the case of composite codes, the right most digit is coded zero (and the process repeated). Third, for continuous variables, such as income or size of dwelling, top and bottom coding is used to truncate the tails of distributions as they begin to "thin". Fourth, certain sensitive variables that are particularly susceptible for identifying individuals, such as birth-date, are suppressed. Fifth, a small fraction of households are "swapped" from the geographical unit reported to a neighbouring one to contribute an additional degree of uncertainty. Finally, households are assigned a unique random number and re-sorted.

7. Conclusions.

28. If we are to bridge the gap between producers and users, new information and communication technologies make census microdata dissemination not only feasible, but easy. The IPUMS project requests a formidable range and amount of metadata and microdata. Nonetheless these are easy to prepare and the return on the investment is substantial   By entrusting census microdata to the IPUMS project, statistical institutes are relieved of the far more burdensome, indeed risky, tasks and responsibilities of disseminating microdata to researchers. Moreover, by relying on the standard IPUMS procedures, which are now used by a majority of the world's statistical institutes, there is safety in numbers. The isolated statistical office that disseminates microdata on an ad hoc basis incurs substantial risks and responsibilities as well as significant human resource and material costs, for a relatively small return with respect to number of users.   The IPUMS project offer substantial economies of scale with the highest standards of security and disseminates integrated metadata and microdata that greatly facilitates sound scientific research.   Interactive tabulation of integrated variables offers a vast increase in the number of users and usage of census data with no additional cost to the National Statistical Institute.

29. Statistical institutes participating in the IPUMS-Latin America initiative are invited to entrust metadata and microdata for the 2010 census round at their earliest convenience.

References

Anderson, Margo and Stephen E. Fienberg. (2001). "*U.S. census confidentiality: Perception and reality*," International Statistical Institute Biennial Meeting (Seoul). (unpub.)

CES (2007), "Managing statistical confidentiality and microdata access: Principles and guidelines on good practice", published by the Conference of European of Statisticians:
http://www.unece.org/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Davern, Michael, Steven Ruggles, Tami Swenson, J. Trent Alexander and J. Michael Oakes. (2009) "*Drawing statistical inferences from historical census data, 1850-1950*," Demography, 46(3):589-603.

Garenne, Michel; Robert McCaa and Kourtoum Nacro. (2008) *"Maternal Mortality in South Africa in 2001: From Demographic Census to Epidemiological Investigation,"* Population Health Metrics, 6:4(Aug) 1-13.

McCaa, Robert and Albert 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, Robert; Albert Esteve, Steven Ruggles and Matt Sobek. (2006) "*Using integrated census microdata for evidence-based policy making: the IPUMS-International global initiative*." African Statistical Journal, 2:83-100

McCaa, Robert; Steven Ruggles, Michael Davern, Tami Swenson, and 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 (New York: Springer), 375-382.

McCaa, Robert and Wendy Thomas. (2009) "*IPUMS-International: lessons from 10 years of archiving and disseminating census microdata,*" 57th Session International Statistical Institute, Durban, South Africa (unpub.) www.hist.umn.edu/~rmccaa/ipums-global (scroll to "IPM 100 Microdata Session" and click "IPUMS"

Megill, David. (2007). "*Technical documentation for public use microdata samples files for the 1997 Mozambique census of population and housing,*" Instituto Nacional de Estatística, Maputo (unpub.). www.hist.umn.edu/~rmccaa/ipums-africa (click "electronic holdings," scroll to Mozambique, and click "sample design" for the 1997 census.

Statistics Canada. (2008) "*Metadata requirements for archiving structured data,*" Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS), Luxembourg.

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

Table 1. IPUMS-América Latina in global perspective (June 1, 2010)

Microdata
Integrated into IPUMS — None inventoried
Entrusted to IPUMS — None entrusted

Country	2010 Round	Notes	Census Decade				
			2000s	1990s	1980s	1970s	1960s
Anonymized samples integrated and completed during IPUMS-LA I and II (2003-10)							
Argentina	2011	Complete; 1960 not integrated	**2001**	**1991**	**1980**	**1970**	*1960*
Bolivia	2011	Complete	**2001**	**1992**		**1976**	
Brasil	2011	Complete; 1960 "B" sample not validated	**2001**	**1991**	**1980**	**1970**	**1960**
Canadá	2011	Complete	**2001**	**1991**	**1981**	**1971**	1961
Colombia	2015	Complete	**2005/6**	**1993**	**1985**	**1973**	**1964**
Costa Rica	2010	Complete	**2000**		**1984**	**1973**	**1963**
Cuba	2012	Complete	**2002**				
Chile	2012	Complete	**2002**	**1992**	**1982**	**1970**	**1960**
Ecuador	2011	Complete	**2001**	**1990**	**1982**	**1974**	**1962**
Estados Unidos	2010	Complete	**2000,5**	**1990**	**1980**	**1970**	**1960**
México	2010	Complete; 1980 recovered, not integrated	**2000,5**	**1990,5**	*1980*	**1970**	**1960**
Panamá	2010	Complete	**2000**	**1990**	**1980**	**1970**	**1960**
Perú	2007	Complete, 1981 recovered, not integrated	**2007**	**1993**	*1981*	1972	1961
Puerto Rico	2010	Complete	**2000**	**1990**	**1980**	**1970**	1960
Saint Lucia	2011	Complete	**2001**	**1991**	**1980**	1970	1960
Venezuela	2011	Complete	**2001**	**1990**	**1981**	**1971**	1961
Additional samples to be integrated in IPUMS-LA II (-2013)							
El Salvador	2007	*2007: lack database*	**2007**	**1992**		1971	1961
Guatemala	2012	Complete; to be released in 2012	**2002**	**1994**	**1981**	**1973**	**1964**
Jamaica	2011	Complete; to be released in 2011	**2001**	**1991**	**1980**		
Haití	2013	*2003: lack database*	**2003**		**1982**	**1971**	
Honduras	2010	Complete; to be released in 2012	**2000**		**1988**	**1974**	**1961**
Nicaragua	2005	Complete; to be released in 2011	**2005**	**1995**		**1971**	1963
República Dominicana	2012	*2002: lack database and manual*	**2002**	1993	**1981**	**1970**	**1960**
Paraguay	2012	Complete; to be integrated in 2013	**2002**	**1992**	**1982**	**1972**	**1962**
Surinam	2011	*2004: lack database and manual*	**2004**			**1980**	
Uruguay	2010	*2004: lack database, form and manual*	**2004**	**1996**	**1985**	**1975**	**1963**

Note: **bold year** = microdata confirmed. "Manual" refers to instructions to the enumerators.
For more information, please contact rmccaa@umn.edu

Apéndice A. Plantilla "A" para describir el censo y sus microdatos

Instrucciones: Para agilizar la homologación en IPUMS-International, favor de describir brevemente el censo de la ronda 2010 y su base de datos correspondiente.   No hay porque preocuparse del formato de esta plantilla.

Nombres:_____ email: _____ fecha: _____

Favor de revisar las descripciones para las muestras censales de su país en *https://international.ipums.org/international/samples.shtml*

Cualquiera duda o pregunta debe dirigirse a Robert McCaa: rmccaa@umn.edu Al llenar la plantilla *envíesela a la misma dirección:*

Características censales (país): _____

1. **Título oficial del censo**	
2. **Nombre del instituto responsible**	
3. **Población a que se dirije (universo)**	
4. **De jure o de facto**	
5. **Unidad de enumeración**	
6. **Día oficial del censo**	
7. **Período de trabajo en el campo**	
8. **Cuestionarios censales**	
9. **Tipo de trabajo de campo**	
10. **Respondiente**	
11. **Cobertura**	

Características de la base de datos

12. **Fuente de los microdatos**	
13. **Sample design**	
14. **Sample unit**	
15. **Sample fraction**	
16. **Sample size (person records)**	
17. **Sample weights (describe)**	

Units identified ("yes" = unit identified; else enter "No")

18. **Dwellings**	
19. **Vacant units**	
20. **Households**	
21. **Individuals**	
22. **Group quarters**	
23. **Settled/Unsettled Population**	
24. **Special populations**	
25. **Smallest geography in microdata**	
26. **Special modules (mortality, etc.)**	

Unit definitions

27. Dwellings	
28. Private Households	
29. Group Quarters	
30. Unsettled population	
31. Special populations	
Metadata entrusted (list file names of electronic or titles of paper copies)	
32. Census forms	
33. Enumerator instructions/manuals	
34. Data Dictionary	
35. Codebooks (education, occupation, industry, geography, etc.)	
36. Correspondence tables (education)	
37. Official results	
38. Technical, Methodological Reports	
39. Post-Enumeration Survey Report	
40. Sample design, sampling errors	
41. Boundary files (if any)	

Apéndice B. Ejemplo de los principios de acuerdo del IPUMS-International

Principios de Acuerdo

Integrated Public Use Microdata Series International

y **El Instituto Nacional de Estadística y Censos de la República Argentina.**

Objetivo. Esta carta tiene por objetivo especificar los términos y condiciones bajo los cuales los metadatos y microdatos proporcionados por el **Instituto Nacional de Estadística y Censos de la República Argentina** serán distribuidos por **Integrated Public Use Microdata Series International** de la Universidad de Minnesota.

1. **Propiedad**. El **Instituto Nacional de Estadística y Censos de la República Argentina** es el titular de los derechos de propiedad intelectual (incluido el copyright) de los metadatos y microdatos proporcionados a la Universidad de Minnesota para ser distribuidos por **Integrated Public Use Microdata Series International**.

2. **Uso**. Los metadatos y microdatos son proporcionados con las finalidades exclusivas de docencia, investigación académica y publicaciones, y no pueden ser usados con ningúna otra finalidad sin la previa autorización escrita del **Instituto Nacional de Estadística y Censos de la República Argentina**.

3. **Autorización**. El potencial usuario que quiera acceder u obtener copias de los microdatos integrados de Argentina desde **Integrated Public Use Microdata Series International** deberá, en primer lugar, solicitar electrónicamente la autorización identificándose por su nombre (ej. el investigador principal), dirección electrónica e institución. El investigador principal debe exponer el objetivo del proyecto propuesto y aceptar cumplir las regulaciones contenidas en éste. Una vez aprobado el proyecto, se expedirá una contraseña con la cual los datos podrán ser adquiridos desde los servidores u otros medios de distribución electrónicos mantenidos por **Integrated Public Use Microdata Series International**, el **Instituto Nacional de Estadística y Censos de la República Argentina**, u otros distribuidores autorizados. Una vez aprobado, el usuario está autorizado a adquirir los microdatos de **Integrated Public Use Microdata Series International** u otros distribuidores autorizados. Ningún título u otro derecho será otorgado al usuario.

4. **Restricción**. A los usuarios se les prohibe usar los datos argentinos adquiridos desde **Integrated Public Use Microdata Series International** u otro distribuidor autorizado con finalidades comerciales lucrativas, privadas o de otra naturaleza.

5. **Confidencialidad**. Los usuarios respetarán la absoluta confidencialidad de las personas y los hogares. Cualquier intento por averiguar la identidad de las personas u hogares a partir de los microdatos queda estrictamente prohibido. Afirmar que una persona u hogar ha sido identificado en estos datos queda también prohibido.

6. **Seguridad**. Los usuarios implementarán medidas de seguridad para prevenir el acceso no autorizado a los microdatos argentinos adquiridos desde **Integrated Public Use Microdata Series International** o sus colaboradores.

7. **Publicación**. La publicación de los datos y análisis resultantes de la investigación con los metadatos y microdatos argentinos está autorizada en aquellas comunicaciones como trabajos de investigación, revistas científicas y semejantes. Los autores de estas comunicaciones deberán citar al **Instituto Nacional de Estadística y Censos de la República Argentina** e **Integrated Public Use Microdata Series International** como fuentes de los microdatos, e indicar que los resultados y comentarios expresados en éstos son exclusivamente del autor/usuario.

8. **Compartir**. **Integrated Public Use Microdata Series International** entregará al **Instituto Nacional de Estadística y Censos de la República Argentina** copias electrónicas de los datos, tanto metadatos como microdatos, así como informes periódicos de los usuarios autorizados.

9. **Violaciones**. La violación de este contrato puede acarrear reprobación profesional y/o un proceso civil.

10. **Jurisdicción**. Cualquier discrepancia que pueda surgir debe ser resuelta mediante conciliación, tramitación y amistosa redacción. Cuando el acuerdo por estos medios resulte imposible, un Tribunal de Conciliación sera convocado para regular sobre el asunto según la ley. Dicho tribunal constará de un árbitro a elegir por sorteo de la lista de Árbitros de la Cámara de Comercio de París. Este acuerdo será regido por, e interpretado de acuerdo a, los principios generalmente aceptados del Derecho Internacional.
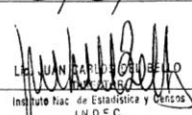
Fecha: 3/15/02

Firma:

Regentes de la Universidad de Minnesota.

By. Kevin McKoskey, Grants Manager, Sponsored Projects Administration.

Fecha: 03/05/2002

Firma:

Instituto Nacional de Estadística y Censos de la República Argentina

Apéndice C. Características de las muestras de Argentina. **https://international.ipums.org/international/sample_designs/sample_designs_ar.shtml**			
Census characteristics			
	1970	**1980**	**1991**	**2001**
Title of census	National Population, Family, and Housing Census, 1970	National Population and Housing Census, 1980	National Population and Housing Census, 1991	National Population, Households, and Dwellings Census, 2001
Census agency	Argentine National Institute of Statistics and Censuses	Argentine National Institute of Statistics and Censuses	Argentine National Institute of Statistics and Censuses	Argentine National Institute of Statistics and Censuses
Population universe	All the population in the national territory at the moment the census is carried out.	All the population in the national territory at the moment the census is carried out.	All the population in the national territory at the moment the census is carried out.	All the population in the national territory at the moment the census is carried out. This includes everyone (of any nationality) that spent the night of November 16-17 in any Argentine diplomatic embassy abroad; all the sailors or fishermen that spent the night of November 16-17 in ships with an Argentine flag or a foreign one docked in Argentine waters; and all Argentine workers that are abroad performing missions for the national government.
De jure or de facto	De facto	De facto	De facto	De facto
Enumeration unit	Particular dwelling, group quarters, census household, and population	Dwelling, quarter dwelling, census household, and population	Dwelling, collective quarters, household, and population	Collective quarters, household, and population
Census day	30-Sep-70	22-Oct-80	15-May-91	November 17 and 18, 2001
Field work period	Interviews take place on the expected census day in each state.	The interviews take place on the expected census day. In some areas the enumeration took place the following day because of access problems due to heavy rains.	The interviews take place on the expected census day.	The interviews take place on the expected census day. In some areas, the enumeration continued a week after the originally expected date, and, in flooded areas, the operation was performed afterwards. In some cases of the flooded areas like Rivadavia in the State of Buenos Aires, the interviews could not be completed until May 2002.

Enumeration forms used	(1) Dwelling questionnaire (2) Population questionnaire (both questionnaires make up a single booklet).	Short form questionnaire: (1) Dwelling questionnaire (2) Population questionnaire (both questionnaires made up a single booklet). Long form questionnaire: (1) Dwelling questionnaire (2) Population questionnaire (both questionnaires make up a single booklet).	Short form questionnaire: (1) Dwelling questionnaire (2) Population questionnaire (both questionnaires made up a single booklet). Long form questionnaire: (1) Dwelling questionnaire (2) Population questionnaire (both questionnaires make up a single booklet).	(1) Household questionnaire (2) Population questionnaire (both questionnaires are part of the same booklet).
Type of field work	Direct enumeration via house-to-house visits and personal interviews performed by teachers called upon to perform census activities as public obligation.	Direct enumeration via house-to-house visits and personal interviews performed by teachers called upon to perform census activities as public obligation.	Direct enumeration via house-to-house visits and personal interviews performed by teachers called upon to do census tasks as a public obligation (even though an economic compensation was established in order to cover expences). In bordering areas, islands, and ports, the participation of the security forces (the National Guard and Navy) was required.	Direct enumeration via house-to-house visits and personal interviews performed by teachers called upon to do census tasks as a public obligation (even though an economic compensation was established in order to cover expences). In addition, employees of the Ministry of Economics were called upon by Ministerial Resolution Number 690/01, Article 14, Law Number 17.622 as census interviewers in cases where replacements were needed due to a conflict by the Teacher's Union. In bordering areas, islands, and ports, the participation of the security forces (the National Guard and Navy Prefecture) was required.
Respondent	Anyone in the dwelling who is able to answer interview questions	Anyone in the dwelling who is able to answer interview questions	Anyone over 14 years old in the dwelling who is able to answer interview questions	Anyone over 14 years old in the dwelling who is able to answer interview questions
Coverage	97.21%	98.99%	98.94%	97.25%

Microdata sample characteristics

Microdata source	Argentine National Institute of Statistics and Censuses (INDEC)	Argentine National Institute of Statistics and Censuses (INDEC)	Argentine National Institute of Statistics and Censuses (INDEC)	Argentine National Institute of Statistics and Censuses (INDEC)
Sample design	Systematic sample of every 50th household with a random start. The sample was elaborated by INDEC.	In locations where the sample fraction is 10 percent, 100 percent of households are chosen. In locations where the sample fraction is 20 percent, every 2nd household is chosen with a random start. For the remaining locations were there is not a sample, every 10th	The microdata base was obtained in the following way. In those locations in which the sample fraction was 10 percent, it was chosen 100 percent of the households. In the locations in which the sample fraction was 20 percent it was chosen a sample of 1 out of two	Systematic sample of every 10th private household and collective quarters with a random start. The sample was elaborated by INDEC from the microdata of 100 percent of households.

		household is chosen. In group quarters, the sample was not applied in the enumeration. They are chosen as a systematic sample of every 10th person with a random start. The sample was elaborated by INDEC.	households with random start. for the remaining locations were there was not a sample it was chosen one of every 10 households. In group quarters the sample was not applied in the enumeration. It was chosen a systematic sample of every 10 people with random start. The sample was elaborated by INDEC.	
			NOTE: Roughly half the population lacks data for several important write-in variables, including occupation, industry, birthplace, and previous residence. An alternative weight variable, AR91A434 (WTPERC), properly adjusts the statistics, and it should be applied to any analyses involving the affected variables.	
Sample unit	Household	Occupied dwelling/household	Household	Household
Sample fraction	2%	10%	10%	10%
Sample universe	2% of households and population in private homes.	100% of the population and households enumerated. The 1980 census used a sample in the data gathering for the application of the long questionnaire form. The fraction of the sample varies according to the size of location.	100% of the population and households enumerated. The 1991 census used a sample in the data gathering for the application of the long questionnaire form. The fraction of the sample varies according to the size of location.	100% of the population and household enumerated.
Sample size	466,892 (person records)	2,667,714	4,143,727	3,626,103
Sample weights	Self-weighting (expansion factor = 50)	Computed by census agency and should be used for most types of analysis.	Computed by census agency and should be used for most types of analysis.	Self-weighting (expansion factor = 10)
Units Identified in Microdata				
Dwellings	Yes	Yes	Not available in sample	Not available in sample
Vacant units	Not available in microdata sample	Not available in microdata sample	Not available in microdata sample	Not available in microdata sample
Households	Yes	Yes	Yes	Yes
Individuals	Yes	Yes	Yes	Yes
Group quarters	Yes	Yes	Yes	Yes

Special populations				Special populations include the identification of households where at least one member has a disability and households that self-identify or descend from Indigenous people. Based on this information, additional information was gathered via the National Disability Survey and the Survey of Indigenous People (appendices to the 2001 Census).
Geography	Department	Department	Department	Department
Unit Definitions (Census)				
Dwellings	Any lodging fixed or movable that has been built or adapted to be inhabited by people. It can be occupied by one or more particular census households or a group quarters census household and one or more private households. A private dwelling is one that is occupied by one or several private census households. A group quarters dwelling is the one that is occupied by a group quarters census household.	Any lodging fixed or movable that has been built or adapted to be inhabited by people. A private dwelling is occupied by private households. A dwelling inhabited by six or more lodgers is considered to be a group quarters household.	Any place destined to lodge people or a place not originally destined to lodge people but on the census day it was used for that purpose. Private dwellings are those occupied by one or more private households. The Inquilinato houses are those which made up a single dwelling in which one or more households are situated. Non-touristic hotels with a housing capacity inferior to 15 rooms in the federal capital and 10 rooms for the rest of the country are considered private dwellings.	A place for lodging structurally separated and adapted to be inhabited by people, and those which were not originally built or adapted to be inhabited by people but used for that purpose at the moment when the census was being carried out. Each Inquilinato room and each hotel room or pension not used for touristic purposes are considered as private dwellings.
Households	A group of people, relatives or not, that live under the same roof and regularly share their food.	A person or a set of people, relatives or not, which occupy the same dwelling.	A person or group of people, relatives or not that live under the same roof according to a family arrangement, meaning they share their food expenses.	A person or a group of people that live under the same roof and share food expenses.
Group quarters	A group of people that share the same dwelling under non-family reasons due to military, work, health, discipline, religion, punishment, etc. A family household with more than six boarders is also a group quarters home.	A dwelling occupied by a group of people that share the dwelling under a non-family arrangement due to the following reasons: work, military, health, religion, discipline.	A place for lodging groups, or the one that was not primarily designed for that purpose but used for that purpose on the census day. The following are considered group quarters: a home for the elderly, a home for the under-aged, boarding schools, camping/working sites, hospitals, prisons, policemen, firemen, and military barracks, religious households and hotels.	A place destined to lodge people that live under a non-family arrangement by norms of living and being these include places that are administrative, military, religious, health, punishment, work-related, etc.