**Thanks to 70 years of Inter American Statistical cooperation,
the world's largest integrated census microdata dissemination site www.ipums.org/international**

ROBERT McCAA, University of Minnesota Population Center, USA rmccaa@umn.edu

ABSTRACT

Seventy years of Inter American Statistical cooperation, symbolized by the 70th anniversary of *Estadística*, made possible the construction of IPUMS-International, the world's largest integrated census microdata dissemination site, www.ipums.org/international. Currently, the site offers access to 234 census samples totaling over 540 million person records representing 74 countries. The Americas, which account for only about one-seventh of the world's population, amount to over one-third (36%) of the person records in the IPUMS-International database. Likewise, 35% of the citations in the IPUMS-International bibliography are for studies focused on Latin America, with about half of these analyzing a single Latin American country. This article discusses salient features of the IPUMS integration methods and system. National Statistical Institutes that have not yet entrusted 2010 census microdata to the initiative are invited to do so. Researchers and teachers are freely invited to use the data in analysis and teaching.

 Spanish translation here:

Key words: Census, sample, public statistics, microdata, metadata, dissemination, IPUMS

Introduction.

In 1940, the Inter American Statistical Institute (IASI) was founded and quickly established itself as the institutional dynamo for official statistics in the Western Hemisphere. A mere three years later, IASI's first and most long-lived initiative, the journal *Estadística*, was launched. In subsequent years, this was followed by the Census Of The Americas (COTA), the Commission on the Improvement of National Statistics (COINS), the United Nations Demographic Center for Latin America and the Caribbean (CELADE), the Inter American Center for the Teaching of Statistics (CIENES), and many other initiatives, too numerous to list here.

The over-representation of the Americas in the IPUMS-International census microdata and the associated citation database is due in large part to the spirit of cooperation of national statistical offices in the region, nurtured, initially, by IASI, *Estadística*, and, somewhat later, CELADE. Not only did the Americas conduct censuses more regularly and more comprehensively than Europe and other regions of the world, the anonymized digital person and dwelling records of the censuses were zealously archived at CELADE headquarters, thanks to the generous cooperation of the official statistical agencies of Latin America. Thus, nearly one-third of the world's stock of census microdata, 186 censuses, for the first quarter century of the digital era (1960-1985) was preserved by Latin American statistical offices and CELADE (see: http://www.hist.umn.edu/~rmccaa/IPUMSI/ census microdata inventory.htm). It is this digital legacy that made the IPUMS-International census microdata harmonization initiative possible in Latin America, and indeed, around the world (Ruggles 2013 and McCaa 2013).

The global nature of the project is evident in this "Google-Analytics" map of recent activity for the IPUMS-International website (Figure 1). Internet traffic from 125 countries and territories and more than 1,400 cities is depicted on the map. Among cities, Buenos Aires ranks first with more visitors to the site than even Minneapolis, the home of IPUMS.

Figure 1. The Americas (and Argentina) lead the world in traffic to the IPUMS-International website, Aug 1-Sep 18 (Google Analytics, City Statistics: n=12,360).
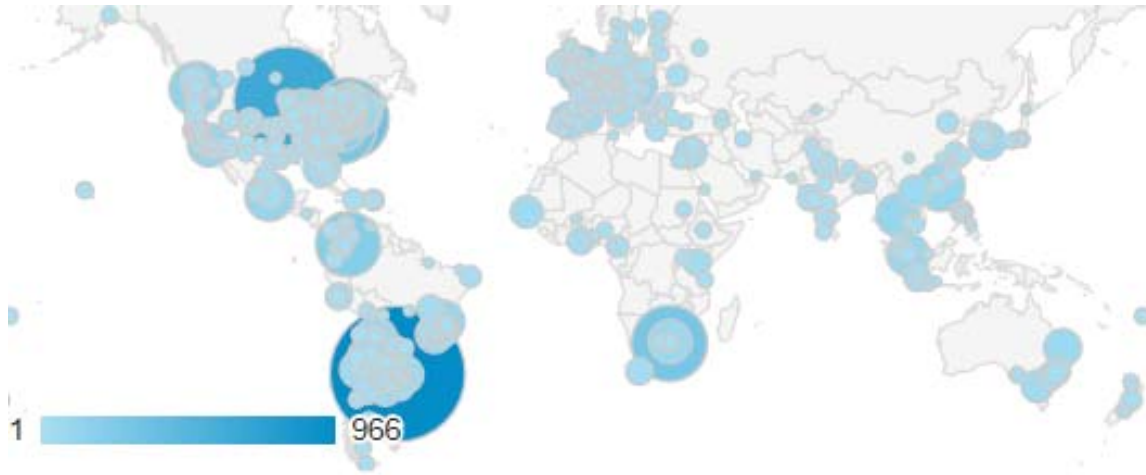
The graphic shows Buenos Aires encompassing much of the Southern Cone of South America. This intense activity was sparked by the release of the integrated sample of the 2010 population census of Argentina. The big circle for Buenos Aires does not entirely conceal the considerable interest by Argentine researchers and policy makers in accessing the series of integrated Argentine census samples stretching over five decades. The Argentine cities of Cordoba, La Plata, Salta, Santa Fe, San Miguel de Tucuman, Bahia Blanca, Mar del Plata, Neuquen, Quilmes, Corrientes, and Resistencia along with Buenos Aires rank among the top 100 cities worldwide for traffic to the IPUMS site. While Argentina contains half of all Latin American cities ranking in the IPUMS-International top 100, the heavy traffic from Latin America depicted on the map is quite remarkable.

Latin America researchers are also over-represented in the IPUMS citation database. Many researchers using census microdata disseminated by IPUMS-International register their publications via the bibliography link on the home page. Almost 800 citations were registered as of September 2013, and the number increases annually by one-fifth or more. One-sixth of the citations are general or comparative covering much of the world, including Latin America. Of the geographically focused publications, 35% deal with Latin America. Half of these focus on a single country, most commonly Mexico, followed by Brazil, Colombia, Argentina, Chile, etc. All Latin American countries with integrated census microdata in the IPUMS-International database are represented in the bibliography with at least one citation, and often by a half dozen or more.

Studies of migration are the single most common topic, followed by education, technical/methodological studies, marriage, inequality, ageing, fertility, children, labor, health, gender, family, demography, transitions, mortality, economics, poverty, life-course, housing, race, etc. Most citations encompass three or more topics of research. The references for this paper include a half dozen citations of published research using microdata disseminated by IPUMS-International and focused on Latin America (Van Hook and Glick, 2007; Lam and Marteleto, 2008; Rosero-Bixby, Castro-Martin and Martin-Garcia, 2009; Medina and Posso, 2010; Manacorda and Rosati, 2011; Esteve, Lesthaeghe, and López-Gay, 2012).

Why IPUMS?

The purpose of IPUMS is to facilitate comparative population analysis using anonymized, integrated census microdata disseminated to researchers world-wide at no cost. Researchers download custom tailored extracts of the microdata for analysis. Recently an on-line data analysis tool was implemented to analyze microdata in real time for a single census, a single country, or an entire region. While rudimentary, the on-line system is a helpful supplement to the IPUMS extract system.

Most researchers are likely to continue to download extracts as the principal mode of access. Both modes are free of cost, but registration is required.

It is noteworthy that in the 1960s, CELADE began an ambitious project of constructing standardized samples of Latin American censuses, "Operación de Muestras de Censos"(OMUECE), but the project was never adequately funded and dissemination by computer tape proved impractical. Nonetheless several important comparative studies on fertility, migration, and population growth were published using the OMUECE microdata (McCaa and Jaspers-Faijer 2000).

In the 1980s, the OMUECE project wound down just as microcomputers opened the way to a revolution in the analysis of census data. In the 1990s, as microcomputers became ever more powerful, demand exploded for census microdata to analyze social, demographic, and economic change over time and place. CELADE responded by developing REDATAM (www.redatam.org), a powerful on-line tabulator that is widely used by statistical offices in the Americas and beyond to facilitate access to census and survey data. As output, REDATAM produces publication-ready tables, graphs and thematic maps. REDATAM is available in Spanish, English, French, and several other languages. An important limitation is that each session with the program analyzes a single dataset. An important advantage is that REDATAM is often harnessed to the complete count census microdata. The United Nations Population Division is a major user of IPUMS-International samples, but relies on REDATAM datasets, where possible, to avoid working with samples and the resultant problem of sampling error (Gerland et al 2013).

Instead of single datasets, academic researchers want to pool microdata for multiple censuses to use sophisticated econometric, statistical and demographic models to analyze dramatic social transformations. For the USA, a team of historians at the University of Minnesota constructed an integrated database of census samples, which they called "IPUMS" (Integrated Public Use Microdata Series). The team keyed microdata from microfilm copies of the original manuscripts of United States censuses. Ultimately the series reached back to 1850. In 1995, the on-line version of IPUMS-USA was launched with an integrated series of census samples stretching from 1880 to 1990 (Ruggles, Hacker and Sobek 1995). The project is still on-going with the addition of samples for the 1920, 1930 and 2000 censuses, as well as, since 2005, annual, triennial and quinquennial samples from the American and Puerto Rican Community Surveys (Ruggles 2013).

Like IPUMS-USA, IPUMS-International integrates both census microdata and metadata to disseminate custom tailored extracts according to each individual user's request. Unlike the USA site, which is fully public, to gain access to the international microdata the researcher must agree to stringent conditions of use to protect privacy and statistical confidentiality and to use the data solely for statistical purposes of teaching and research. Once approved, researchers have access to the entire database—currently, 234 census samples totaling over 540 million person records representing 74 countries. The database is updated annually with the addition of a couple of dozen samples from around the world. Chronologically, IPUMS-International is limited to censuses that were processed by computer, essentially 1960 to the present. Samples for the 2010 round of censuses are fast-tracked for launch within one year of the data being entrusted to the project. It is remarkable that more 2010 round samples for Latin America are already integrated into the database than for any other region of the world.

Table 1 near here.

The Americas, including Latin America, Canada, Jamaica, Saint Lucia and the USA, is represented by 89 samples totaling almost 200 million person records (Table 1). This amounts to 37% of the samples in the database and 36% of the individual person records (n=194,752,993). The samples for the USA and Puerto Rico in the International database are integrated according to international norms, while the USA database closely tracks the norms of the United States Census Bureau. Most Latin American researchers will favor the International version. However, researchers interested exclusively in immigration to the USA will find variables and details unique to the American Community Survey represented in IPUMS-USA (and Puerto Rico) that are not included in the International database.

Over the next few years, samples for the remaining Latin American countries not currently represented in the database—Dominican Republic, Guatemala, Honduras, and Paraguay—will be added. Their absence is *not* due to a lack of cooperation. On the contrary, the statistical offices of these countries endorsed the project Memorandum of Understanding without hesitation, and in due course, entrusted the necessary census microdata and documentation. Nonetheless, because the project is global, not all the microdata for all countries can be integrated at once. By 2018, it is expected that all the Latin American countries will be represented in the database, including samples for the 2010 round of censuses. Unfortunately, cooperation from the English-speaking statistical offices of the Caribbean region is not as widespread as for Latin America, despite not inconsiderable effort to address questions and concerns and to resolve obstacles.

IPUMS Integration.

Most Latin American microdata integrated into the IPUMS-International website are high precision, ten percent household samples. This uniformity does not extend to other characteristics of the microdata, as hinted at in Table 1. There are substantial differences in geographic detail and whether residence is actual as of census night (de facto) or judicial (de jure). These incompatibilities highlight the challenges of integrating census microdata after-the-fact. Researchers must study the documentation to inform and guide their analysis so that they are informed of incomparabilities, both major and minor. For example, de-facto censuses may yield higher levels of migration simply because travelers and visitors may be counted as migrants when in fact they are only temporarily absent from their usual place of abode. Some de facto censuses include a "usual residence" question to obtain better indications of migration and similarly affected questions.

The challenge of integrating census microdata is, first, to synthesize the definitions used in questions and concepts; second, to highlight comparability issues; and third, to apply coding schemes that intuitively communicate commonalities and differences. With regard to comparability, for example, consider economic activity. The questionnaire for the 1990 census of Mexico contains the word "principal" in referring to activity. This word had never been used before nor since nor does it appear, in the instructions to field enumerators (in regard to the activity question). Nonetheless, analysts need to be alerted to this subtle change in wording as well as others even more significant modifications such as the change from "activity" with no period specified to "activity last week", which is commonly used in the 2010 round of censuses.

Table 2 near here.

Table 2, taken from the IPUMS-International metadata, indicates the availability of 26 integrated variables (of 1,677 total) for 17 Latin American countries in the latest census sample to be integrated into the IPUMS-International database. Country is represented in the column head by a two-digit ISO 3166 code, and census year by a two-digit year code with century omitted. "X" indicates that the variable is available for the census sample identified by the corresponding column head. Thus, number of persons in the household is identified in every sample, as is age, sex, relationship to household head (reference person), nativity status, educational attainment, and employment status. Second administrative level of geography is available for 14 of the samples, but not for Cuba 2002, Panama 2010, or Uruguay 2006. The availability of variables is determined by what questions were asked in the corresponding census. If a question was not asked then the associated variable cannot appear in the census microdata. Occasionally statistical confidentiality may require that a variable be suppressed, such as census tract and other low-level geography.

As an example of this method of integrating variables, consider the concept "educational attainment," the single most widely used variable in the IPUMS-International database. Most census microdata with information on this measure indicate whether the respondent completed primary, secondary or higher schooling or no schooling at all. Thus the first digit of the IPUMS-International composite code consists of four categories (1-4), plus codes for missing data (9) and "not in universe" (0—for children too young to attend or others to whom the question was not addressed). Many census samples contain further information indicating, for example, those who attended primary, secondary or even

tertiary schooling, but did not complete the course of study. The second digit captures this information. Zero is used as a place-holder, signifying no information. The third digit distinguishes between technical and general or other tracks common to two or more countries. Successful international integration must document such distinctions so that researchers may readily be informed of these and thousands of other details.

Table 3 near here.

Table 3 illustrates the current IPUMS-International coding scheme for the educational attainment variable. The array of codes displays the considerable variability from country-to-country in the level of specificity regarding the various tracks of schooling completed. The frequencies listed are the raw, un-weighted counts in the corresponding sample. These metadata are readily available from the IPUMS-International documentation system. Researchers find them helpful in guiding extract selection. Thus, if a researcher wishes to study Lower Secondary Technical Track, it is readily apparent that few of these samples have more than 10,000 cases. In five instances no cases are specified, suggesting that general and technical are represented by a single code.

At the primary level, in Brazil, four years of schooling was once considered adequate—a historical legacy from a distant past—and thus is so identified in Brazilian censuses. We have coded this as "less than primary completed," code 130. A researcher might decide to recode this as primary completed, according to the needs of a specific research question. In contrast, in most Latin American censuses, six years of schooling is universally considered "primary completed." Excepts are in the case of Colombia, where five years has been the standard, and in Bolivia, if only for a fraction of the population. At the lower secondary level, all samples, except Peru 2007, identify completion. At the secondary completed level (general code 3xx), there is greatest diversity. All samples identify general track completed (311), and all but one (Cuba 2002) identify whether some university or college studies were attempted but not completed (312). Technical track codes (321, 322) reveal considerable incomparability in samples. Researchers interested in this topic should study the documentation carefully, including source documents in the official language.

IPUMS-International codes are unofficial and somewhat arbitrary, but they are harmonized across samples. In contrast, the original sources codes are a Tower of Babel. For most census questions, aside from sex and age, no two censuses share the same coding scheme. We have sought to substitute for the original, usually serially ordered codes, a system of hierarchical, composite codes, to provide guidance toward meaning, with the idea that researchers will want to recode to suit specific research needs. For the curious or fastidious researchers, the original source codes are readily available under the rubric "Unharmonized variables" on the IPUMS-International variable selection page. Unharmonized variables are handy for checking integrated variables for errors.

In addition to codes, the IPUMS-International metadata offers 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. The goal is to facilitate informed analysis of the microdata by providing as much essential information as possible—all readily accessible from the website by means of a few clicks. Metadata pages are generated dynamically by the user (or peruser—the metadata are open to all, no registration required). The user may select a single sample or any combination of samples to peruse the metadata in any order desired.

Conclusion.

Over the past seven decades, the Inter American Statistical Institute, *Estadística*, and numerous other regional statistical institutions have nurtured various initiatives that have yielded a flowering of official statistics throughout the Americas. Census microdata, its preservation and dissemination, are a small, but significant development that is directly attributable to the region-wide spirit of professionalism and cooperation cultivated by IASI and *Estadística*.

Researchers are invited to access integrated census microdata from the IPUMS-International web-site as the need arises. Educators are invited to "Teach with IPUMS" by enrolling an entire class to simplify the registration process for students.

Two new initiatives are currently under-development at the Minnesota Population Center: 1) Terra-Populus – an integrated GIS project to make microdata interoperable with other spatiotemporally referenced data, such as satellite imagery; and 2) Remote Data Enclave – integrated full count census and other microdata that are too detailed and therefore too risky to disseminate in the form of individual person records. Over the coming years, National Statistical Institutes of the region will be invited to participate in these initiatives. Meanwhile, the IPUMS-International project would be pleased to integrate, without undue delay, 2010 round census microdata produced by Latin American statistical offices.

References.

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

Esteve, Albert, Ron Lesthaeghe, Antonio López-Gay. 2012. "The Latin American Cohabitation Boom, 1970-2007." *Population and Development Review* 38: 55-81.

Gerland, Patrick, Thomas Spoorenberg, Jorge Bravo, Pablo Lattes, Cheryl Sawyer, Vladimira Kantorova, and Mun Sim Lai. 2013. "Uses of Census Microdata by the United Nations Population Division," XXVII IUSSP International Population Conference, Aug. 24.

Lam, David, Letícia Marteleto. 2008. "Stages of the Demographic Transition from a Child's Perspective: Family Size, Cohort Size, and Children's Resources." *Population and Development Review* 34: 225-252

Manacorda, Marco, Furio Rosati. 2011. "Industrial Structure and Child Labor Evidence from the Brazilian Population Census." *Economic Development and Cultural Change* 59: 753-776.

McCaa, Robert. 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):69-88.

McCaa, Robert, Dirk J. Jaspers_Faijer. 2000. "The Standardized Census Sample Operation (OMUECE) of Latin America, 1959-1982 [1995]: a project of the Latin American Demographic Center (CELADE)," in Patricia Kelly Hall, Robert McCaa and Gunnar Thorvaldsen, eds., *Handbook of International Historical Microdata for Population Research*, Minneapolis: Minnesota Population Center: 287-302.

McCaa, Robert, Steven Ruggles. 2002. "The census in global perspective and the coming microdata revolution." *Scandinavian Population Studies* 13: 7-30

Medina, Carlos, Christian Posso. 2010. "Technical Change and Polarization of the Labor Market: Evidence for Brazil, Colombia and Mexico." *Borradores de Economia*: 1-51.

Meier, Ann, Robert McCaa, David Lam. 2011. "Creating statistically literate global citizens: The use of IPUMS-International integrated census microdata in teaching." *Statistical Journal of the IAOS* 27: 145-156

Ruggles, Steven, J. David Hacker, and Matthew Sobek. 1995. "Order out of chaos: General design of the Integrated Public Use Microdata Series." *Historical Methods* 28: 33-39.

Ruggles, Steven. 2013. "Big Microdata for Population Research," *Demography* DOI 10.1007/ s13524-013-0240-2.

Rosero-Bixby, Luis, Teresa Castro-Martin, Teresa Martin-Garcia, Teresa. 2009. "Is Latin America Starting to Retreat from Early and Universal Childbearing?" *Demographic Research* 20: 169-194.

Van Hook, Jennifer, Jennifer Glick. 2007. "Immigration and Living Arrangements: Moving Beyond Economic Need versus Acculturation." *Demography* 44: 225-249

Table 1. 89 census microdata samples of the Americas disseminated from www.ipums.org/international

Country	Years	Samples Number (total)	% (latest)	Households (total)	Persons (total)	Administrative Geography	de facto/de jure (latest sample)
Argentina	1970-2010	5	10	4,259,459	15,013,401	Department	de facto
Bolivia	1976-2001	3	10	538,779	1,931,759	Province	de facto
Brazil	1960-2010	6	5	10,550,767	42,177,485	Municipality	de jure
Canada	1970-2010	4	3	n.a.	2,311,603	Province	de jure
Chile	1960-2002	5	10	1,341,476	4,960,696	Municipality	de facto
Colombia	1970-2010	5	10	2,750,032	12,201,433	Municipality	de facto
Costa Rica	1963-2000	4	10	199,482	891,827	Canton	de jure
Cuba	2002	1	10	371,878	1,118,767	Province	de jure
Ecuador	1962-2010	6	10	1,326,367	5,220,147	Canton	de facto
El Salvador	1992-2007	2	10	297,707	1,085,124	Municipality	de jure
Haiti	1971-2003	3	10	390,076	1,632,429	Arrondissement	de jure
Jamaica	1981-2001	3	10	181,134	661,471	Parish	de jure
Mexico	1960-2010	7	10	9,566,074	41,758,642	Municipality	de jure
Nicaragua	1971-2005	3	10	238,217	1,140,682	Municipality	de jure
Panama	1960-2010	6	10	332,733	1,257,609	District	de facto
Peru	1993-2007	2	10	1,386,440	4,952,319	Province	de facto
Puerto Rico	1970-2000	5	1	207,671	590,330	100,000+ population	de jure
Saint Lucia	1980-1991	2	10	6,068	24,833	n.a.	de jure
United States	1960-2010	8	1	20,389,907	47,695,258	100,000+ population	de jure
Uruguay	1963-2096	5	10	504,368	1,416,943	Department	de jure
Venezuela	1971-2000	4	10	1,722,545	6,710,235	Municipality	de jure
Total Americas		89		56,561,180	194,752,993		
Total Database		238		142,117,825	543,968,021		
Americas/Total		37.4%		39.8%	35.8%		

Source: https://international.ipums.org/international/samples.shtml

Table 2. 26 Most Common IPUMS Integrated Variables: 17 Latin American Countries (latest sample)

Name	Label	AR	BO	BR	CL	CO	CR	CU	EC	SV	HT	MX	NI	PA	PE	PR	UY	VE
		10	01	10	02	05	00	02	10	07	03	10	05	10	07	05	06	01
GEOLEV	Residence: 2nd subnational geographic level	X	X	X	X	1st	X	X	X	X	X	X	X	.	X	X	X	X
BPLLEV1	Birthplace: 1st subnational geographic level	.	X	X	X	X	X	X	X	X	X	X	X	X	X	.	X	X
OWNRSHP	Ownership of dwelling	X	X	X	X	X	X	.	X	X	X	X	X	X	X	X	X	X
ELECTRC	Electricity	.	X	X	X	X	X	X	X	X	X	X	X	X	X	.	X	X
WATSUP	Water supply	X	X	X	X	X	X	X	X	X	X	X	X	X	X	.	X	X
SEWAGE	Sewage	X	X	X	X	.	X	X	X	X	.	X	X	X	X	.	X	X
ROOMS	Number of rooms	X	X	X	X	X	X	X	X	X	X	X	.	X	X	X	X	X
FLOOR	Floor material	X	X	.	X	X	X	.	X	X	X	X	X	X	X	.	X	X
HHTYPE	Household classification	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RELATE	Relationship to household head	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
AGE, SEX	Age, Sex	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MARST	Marital status (de jure)	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
CONSENS	Consensual union (de facto)	.	X	X	X	X	X	X	X	X	X	X	X	X	X	.	X	X
CHBORN	Children ever born	.	X	X	X	X	X	.	X	X	X	X	X	X	X	.	X	X
CHSURV	Children surviving	.	X	X	X	X	X	.	X	X	X	X	X	X	X	.	.	X
NATIVTY	Nativity status	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
BPLCTRY	Country of birth	X	X	X	X	X	X	X	X	.	X	X	X	X	X	.	.	X
SCHOOL	School attendance	X	X	X	.	X	X	.	X	X	X	X	X	X	X	X	X	X
EDATTAN	Educational attainment, international recode	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
YRSCHL	Years of schooling	X	X	.	X	X	X	.	X	X	X	X	X	X	X	.	X	X
EMPSTAT	Employment status	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
OCCISCO	Occupation, ISCO general	.	X	X	X	.	X	X	X	X	X	X	X	X	X	X	X	X
INDGEN	Industry, general recode	.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
CLASSWK	Class of worker	.	X	X	X	X	X	.	X	X	X	X	X	X	X	X	X	X
DISABLE	Disability status	.	.	X	X	X	X	.	X	X	X	X	.	X	.	.	X	X
URBAN	Urban-rural status	.	X	X	X	X	X	.	X	X	X	.	X	X	X	.	.	X

Source: https://international.ipums.org/international-action/variables/ (Note: first select the corresponding country and census samples of column heads)

Table 3. Educational Attainment: Example of IPUMS integrated variable (Latin American census samples).

Code	Label	AR	BO	BR	CL	CO	CR	CU	EC	SV	HT	MX	NI	PA	PE	PR	UR	VE
		2010	2001	2010	2002	2005	2000	2002	2010	2007	2003	2010	2005	2010	2007	2005	2006	2001
0	NIU (not in universe)	29,655	86,324	1,068,918	115,808	248,552	37,858	86,273	145,998	55,694	97,953	728,140	60,853	26,349	160,623	1,047	·	148,498
100	LESS THAN PRIMARY COMPLETED	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
110	No schooling	462,905	124,607	665,972	99,930	668,929	35,740	157,823	121,397	112,584	312,204	2,140,964	110,425	34,928	351,154	1,750	44,405	325,181
120	Some primary	780,196	229,111	2,214,459	276,514	1,044,071	96,460		305,970	167,728	201,903	2,924,954	149,129	55,382	646,203	4,683	58,983	502,076
130	Primary (4 years)	·	·	1,133,137	·	·	·	·	·	·	·	·	·	·	·	·	·	·
	PRIMARY COMPLETED																	
	Primary completed																	
211	Primary (5 years)	·	43,053	·	·	869,800	·	·	·	·	·	·	·	·	·	·	·	·
212	Primary (6 years)	1,052,586	107,668	1,784,100	319,479	·	117,157	194,407	318,147	133,423	94,741	2,446,262	93,804	78,658	574,487	5,340	67,002	560,540
	Lower secondary completed																	
221	General and unspecified track	573,001	86,545	623,820	194,572	355,585	25,967	278,766	172,613	26,534	67,971	2,081,938	33,169	40,012	·	4,080	38,452	230,289
222	Technical track	·	1,338	·	69,874	14,743	3,121	18,258	·	580	·	35,081	669	3,248	·	·	40	2,787
	SECONDARY COMPLETED																	
	General or unspecified track																	
311	General track completed	410,078	64,833	1,332,365	183,846	403,290	21,892	143,378	147,032	40,960	9,549	648,284	29,391	50,002	420,224	7,257	10,438	243,259
312	Some college/university	202,948	22,175	341,406	89,677	40,744	15,153	·	113,091	12,954	4,069	171,210	11,389	14,407	118,826	6,227	7,151	169,452
320	Technical track	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
321	Secondary technical degree	·	887	·	25,074	3,471	3,504	132,890	·	·	·	78,856	4,156	·	·	·	·	22,019
322	Post-secondary technical ed.	250,877	31,046	·	87,711	94,270	4,913	·	14,149	9,301	45,929	177,225	2,278	3,355	287,488	·	23,570	78,879
400	UNIVERSITY COMPLETED	203,999	20,278	528,881	51,429	147,950	19,735	71,282	76,152	14,606	3,726	416,996	17,292	31,270	186,890	5,032	6,825	2,954
999	UNKNOWN/MISSING	·	9,827	·	·	114,763	·	35,690	33,684	·	·	88,492	2,930	3,507	·	·	·	20,555

Source: https://international.ipums.org/international-action/variables/EDATTAN/#codes_section (first select samples)