



Data Resource Profile

Data Resource Profile: IPUMS-International

Kristen Jeffers,* Miriam King, Lara Cleveland and Patricia Kelly Hall

Minnesota Population Center, University of Minnesota, Minneapolis, MN, USA

*Corresponding author. Minnesota Population Center, University of Minnesota, 225 19th Ave S, 50 Willey Hall, Minneapolis, MN 55455, USA. E-mail: kjeffers@umn.edu

Accepted 14 October 2016

Data resource basics

The Integrated Public Use Microdata Series-International (IPUMS-International) disseminates high-precision census microdata samples from around the world. Since its inception in 1999, IPUMS-International has partnered with official statistical agencies to assemble the world's largest collection of publicly available census microdata. With over 100 national statistical office (NSO) partners, IPUMS-International currently disseminates integrated data on more than one-half billion persons, spanning five continents and provided at no charge to researchers and students worldwide. The data series includes data from 1960 to 2011, with multiple samples available for most countries. IPUMS-International reduces the barriers to comparative research across time and space by converting international census microdata into a uniform format, providing comprehensive documentation and making the data available to researchers through a Web-based access system.¹

The data series includes information on a broad range of population characteristics, including fertility, nuptiality, mortality, migration, disability, labour force participation, occupational structure, education, ethnicity and household composition. Variable coding schemes are standardized across samples (without loss of detail) to provide an integrated database that allows samples to be easily combined for comparisons across years or national boundaries. The IPUMS-International online data access system allows researchers to create customized data extracts that contain only the samples, variables and cases they require. The data access system is fully integrated with the variable and

sample documentation in a user-friendly online interface, so researchers can make informed decisions as they define their datasets. Other features include intra-household relationship pointer variables, spatiotemporally harmonized geographical variables and accompanying boundary files (shape files), and an online data tabulator.

Researchers who use census microdata disseminated through the IPUMS-International partnership are required to cite the NSOs that contributed their original data as well as IPUMS-International for harmonizing and disseminating the data. For each data extract, researchers receive an e-mail citation format which includes a list of the NSOs for each country in the extract.

Data collected

As of 2016, 277 anonymized microdata samples from 82 countries are available to researchers and students through the IPUMS-International online data dissemination system (Table 1). Truly global in its coverage, the series includes more than 50 samples each from Africa, Asia, Europe and the Americas. Most participating national statistical agencies have entrusted the country's full series of extant census microdata to the project, facilitating intra-national as well as international trend analysis. Future annual releases will incorporate data from newly participating countries: Benin, Botswana, Bulgaria, Cape Verde, Central African Republic, Cote d'Ivoire, Guinea Bissau, Honduras, Republic of Korea, Lesotho, Madagascar, Mauritius, Namibia, Niger, Papua New Guinea, Poland, Trinidad and Tobago, Tunisia and Turkmenistan.

IPUMS-International in a nutshell

- IPUMS-International integrates and disseminates high-precision census microdata samples from around the world. Microdata and metadata are fully integrated; data are disseminated as customized datasets that contain only the samples, variables and cases required by the user.
- Initiated in 1999, IPUMS-International has integrated 277 samples from 82 countries into a single database containing more than 600 million person records. Data from 1960 to the present are available.
- Participating national statistical offices generously provide source data. Nationally representative samples are systematically drawn from the total enumerated population by IPUMS-International or by the statistical offices of the country of origin.
- More than 700 harmonized variables on a broad range of population characteristics are available, including fertility, nuptiality, mortality, migration, disability, labour force participation, occupational structure, education, ethnicity and household composition. Most samples include low-level geographical detail.
- Microdata are available to researchers and students free of charge via an online data extraction system. Apply for access at [international.ipums.org].

IPUMS-International samples are individual-level subsets of full-count census data. The samples are systematically drawn from the total enumerated population by IPUMS-International or by the statistical offices of the country of origin according to a variety of sample designs. Where possible, IPUMS-International provides 10% samples of census data by selecting every 10th household after a random start. Nearly all samples available from IPUMS-International are cluster samples: they are samples of households rather than individuals. Individuals are sampled as parts of households because many important topics, such as fertility, household composition, and nuptiality, require information about multiple individuals within the same household. Some samples employ complex sampling techniques that may include geographical or social stratification (for example, different sampling fractions to administer census long forms in urban versus rural areas). Household and person weight variables that account for these complexities are attached to each record and are automatically included in every customized data extract. Detailed sample design information is available on the IPUMS-International website.

Unique individual, household, dwelling and subnational geographical identifiers allow researchers to select the level of analysis most suitable to their research. Geographical detail varies across samples (see [Table 1](#)). For most countries, the first and second administrative levels are identified; for some countries, smaller entities such as municipalities are specified. Most samples are truly nationally representative, including individuals living in group quarters such as prisons, nursing homes, children's homes and religious institutions, and thus providing information on population subgroups often excluded from household,

health and labour force surveys. Census and sample characteristics, including treatment of special populations, are documented on the IPUMS-International website.

Each year, 20–30 new census samples are harmonized and released via the IPUMS-International online data access system. The integration process consists of two steps. Integrated metadata are constructed by studying the original source documentation (such as census forms, instructions to enumerators and published census tables) and extensively analysing the raw data. Microdata are then integrated and documented, variable by variable, and retested until fully validated for dissemination to researchers. Samples for the latest round of censuses are given priority. Along with launching new samples, the annual data releases incorporate new integrated and technical variables that expand the topics covered by the database and improve precision of research results. For example, the 2014 data release added new variables related to variance estimation, and the 2015 and 2016 data releases are adding more geographical detail.

Measures and data enhancements

The data series includes information on a broad range of population and housing characteristics. The population questions address fertility, nuptiality, migration, disability, labour force participation, occupational structure, education, ethnicity and household composition. Housing questions cover economic indicators (such as dwelling ownership and building material), possession of amenities (such as a car or television) and utilities (such as water source, sewage disposal and cooking fuel), with the last group having obvious public health implications. In short, the censuses cover whatever national governments

Table 1. IPUMS-International samples available to researchers

Country	Sample years	Lowest geographical unit identified	Administrative level of lowest geographical unit
Argentina	1970, 1980, 1991, 2001, 2010	Department	Level 2
Armenia	2001	Province	Level 1
Austria	1971, 1981, 1991, 2001	NUTS3 region ^a	Level 2
Bangladesh	1991, 2001, 2011	Upazila	Level 3
Belarus	1999	Region	Level 1
Bolivia	1967, 1992, 2001	Province	Level 2
Brazil	1960, 1970, 1980, 1991, 2000, 2010	Municipality	Level 2
Burkina Faso	1985, 1996, 2006	Province	Level 2
Cambodia	1998, 2008	District	Level 2
Cameroon	1976, 1987, 2005	Arrondissement	Level 3
Canada	1971, 1981, 1991, 2001	Province	Level 1
Chile	1960, 1970, 1982, 1992, 2002	Municipality	Level 3
China	1982, 1990	City/Prefecture	Level 2
Colombia	1964, 1973, 1985, 1993, 2005	Municipality	Level 2
Costa Rica	1963, 1973, 1984, 2000	Canton	Level 2
Cuba	2002	Province	Level 1
Dominican Republic	1960, 1970, 1981, 2002, 2010	Municipality	Level 2
Ecuador	1962, 1974, 1982, 1990, 2002, 2010	Canton	Level 2
Egypt	1996, 2006	District	Level 2
El Salvador	1992, 2007	Municipality	Level 2
Ethiopia	1984, 1994, 2007	Wereda	Level 3
Fiji	1966, 1976, 1986, 1996, 2007	Province	Level 1
France	1962, 1968, 1975, 1982, 1990, 1999, 2006	Region	Level 1
Germany	1970, 1971 (DR), 1981 (DR), 1987	State	Level 1
Ghana	2000, 2010	District	Level 2
Greece	1971, 1981, 1991, 2001	Municipality	Level 2
Guinea	1983, 1996	Prefecture	Level 2
Haiti	1971, 1982, 2003	Arrondissement	Level 2
Hungary	1970, 1980, 1990, 2001	None	None
India	1983, 1987, 1993, 1999, 2004	Region	Level 2
Indonesia	1971, 1976, 1980, 1985, 1990, 1995, 2000, 2005, 2010	Regency	Level 2
Iran	2006	Sub-province	Level 2
Iraq	1997	District	Level 2
Ireland	1971, 1979, 1981, 1986, 1991, 1996, 2002, 2006, 2011	Region	Level 1
Israel	1972, 1983, 1995	Subdistrict	Level 2
Italy	2001	Region	Level 1
Jamaica	1982, 1991, 2001	Parish	Level 1
Jordan	2004	District	Level 2
Kenya	1969, 1979, 1989, 1999, 2009	District	Level 2
Kyrgyz Republic	1999, 2009	District	Level 2
Liberia	1974, 2008	District	Level 2
Malawi	1987, 1988, 2008	District	Level 1
Malaysia	1970, 1980, 1991, 2000	District	Level 2
Mali	1987, 1998, 2009	District	Level 3
Mexico	1960, 1970, 1990, 1995, 2000, 2005, 2010	Municipality	Level 2
Mongolia	1989, 2000	Province	Level 1
Morocco	1982, 1994, 2004	Province	Level 2
Mozambique	1997, 2007	Administrative post	Level 3
Nepal	2001	District	Level 2
Netherlands	1960, 1971, 2001	None	None
Nicaragua	1971, 1995, 2005	Municipality	Level 2
Nigeria (GHS)	2006, 2007, 2008, 2009, 2010	State	Level 1
Pakistan	1973, 1981, 1998	District	Level 3
Palestine	1997, 2007	Governorate	Level 1
Panama	1960, 1970, 1980, 1990, 2000, 2010	District	Level 2

(Continued)

Table 1. Continued

Country	Sample years	Lowest geographical unit identified	Administrative level of lowest geographical unit
Paraguay	1962, 1972, 1982, 1992, 2002	District	Level 2
Peru	1993, 2007	Province	Level 2
Philippines	1990, 1995, 2000	Municipality	Level 3
Portugal	1981, 1991, 2001	Sub-region	Level 1
Puerto Rico	1970, 1980, 1990, 2000, 2005 (PRCS)	100 000+ PUMAS ^b	Level 1
Romania	1977, 1992, 2002	County	Level 1
Rwanda	1991, 2002	Province	Level 1
Saint Lucia	1980, 1991	None	None
Senegal	1988, 2002	Department	Level 2
Sierra Leone	2004	Chiefdom	Level 2
Slovenia	2002	Region	Level 1
South Africa	1996, 2001, 2007	Municipality	Level 3
South Sudan	2008	County	Level 2
Spain	1981, 1991, 2001	Municipality	Level 3
Sudan	2008	County	Level 2
Switzerland	1970, 1980, 1990, 2000	Canton	Level 1
Tanzania	1988, 2002	District	Level 2
Thailand	1970, 1980, 1990, 2000	Province	Level 1
Turkey	1985, 1990, 2000	District	Level 2
Uganda	1991, 2002	County	Level 2
Ukraine	2001	Raion	Level 2
UK	1991, 2001	SARs region ^c	Level 1
USA	1960, 1970, 1980, 1990, 2000, 2005 (ACS), 2010	100 000+ PUMAS ^b	Level 1
Uruguay	1963, 1975, 1985, 1996, 2006, 2011	Department	Level 1
Venezuela	1971, 1981, 1990, 2001	Municipality	Level 2
Vietnam	1989, 1999, 2009	District	Level 2
Zambia	1990, 2000, 2010	Constituency	Level 3

^aEuropean Union's Nomenclature of Territorial Units for Statistics 3.

^bPublic Use Microdata Areas containing 100 000 or more residents.

^cSamples of Anonymized Records region.

considered essential topics to include during their enumeration (Table 2). As described in further detail below, IPUMS-International integrates the original material from each sample and supplies additional material, including documentation about each variable, within-household relationship pointer variables and geographical information system (GIS) boundary files. Researchers then access the data by building customized datasets with the online extract system or using the online data tabulator.

Variable harmonization

Along with supplying unique access to these nationally representative datasets, the principal advantage of IPUMS-International is its replacement of sample-specific variable codes with new integrated codes consistent across time and space. This 'variable integration' ensures that identical concepts always have identical codes, which simplifies comparative analysis of multiple samples. Over 700

integrated variables are included in the IPUMS-International database, and the website displays at a glance which variables are included in each sample.

For some uncomplicated variables, such as sex, harmonization simply requires imposing the same codes across all samples (e.g. 1 for male and 2 for female). For other variables, the issue is complicated by different response categories across censuses. Variable integration in IPUMS-International retains all original detail by using composite coding. The first digit, called the 'general code', provides information available across all samples (the lowest common denominator data). The second digit provides information available in a substantial subset of the samples, and trailing digits supply additional detail only rarely available.

As an example of IPUMS-International's composite coding, consider the EDATTAIN variable on 'educational attainment', the single most widely used variable in the database. The first digit of EDATTAIN's composite code consists of four broadly available categories (1–4)

Table 2. Most commonly requested IPUMS-International variables ranked by number of requests. Availability (number of samples) in brackets

Person record	Household record
Employment status [246]	Ownership of dwelling [230]
Marital status [273]	Urban-rural status [187]
Educational attainment [266]	Number of person records in household [276]
Age [276]	Group quarters status [276]
Sex [276]	Water supply [193]
Relationship to household head [269]	Number of families in household [246]
Class of worker [244]	Household classification [245]
School attendance [222]	Number of rooms [210]
Occupation [235]	Toilet [192]
Years of schooling [165]	Electricity [178]
Literacy [192]	Number of married couples in household [246]
Member of an indigenous group [31]	Sewage [143]
Religion [130]	1st subnational geographical level [261]
Children ever born [194]	Number of mothers in household [246]
Nativity [231]	Telephone availability [124]
Industry [240]	Head's location in household [223]
Number of own children in household [246]	Number of fathers in household [246]
Mother's location in household [249]	Television set [107]
Country of birth [174]	Wall or building material [123]
Spouse's location in household [249]	Floor material [109]
Father's location in household [249]	Cooking fuel [119]
Number of own family members in household [246]	Radio in household
Children surviving [146]	Automobiles available [99]
Age of eldest child [246]	Refrigerator [86]
Age of youngest child [246]	Roof material [97]
Total income [36]	Kitchen or cooking facilities [113]
Migration status 5 years ago [101]	Trash disposal [55]
Citizenship [142]	Computer [59]
Race [39]	Bathing facilities [106]
Hours worked per week [53]	Cell phone availability [42]

Source: IPUMS-International User Statistics Database, April 2016.

distinguishing between 'less than completed primary school', 'completed primary, less than secondary school completed', 'secondary school completed' and 'university completed' 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). The second digit of EDATTAIN captures frequently, but not universally, available information on whether the person attended school without completing the course of study, and the third digit distinguishes between technical and general education tracks. [Table 3](#) illustrates the values available for EDATTAIN for 16 countries (represented by two-digit ISO codes) and their associated census year (with x's representing the presence of the value in a given sample). As this example shows, the first digit code supports cross-country comparisons and the second and third digits summarize information only sporadically available but nonetheless essential to some researchers.

Guidelines from organizations like the United Nations and the International Labour Organization have encouraged consistency in census question wording and coding, but each country's statistical office ultimately decides the subjects covered, the question wording, who was asked a question (i.e. the question universe) and the response categories included in their national census. For UN guidelines and recommendations for population and housing censuses, see [<http://unstats.un.org/unsd/demographic/sources/census/census3.htm>]; for ILO standards and guidelines, see: [<http://www.ilo.org/global/statistics-and-databases/standards-and-guidelines/lang-en/index.htm>]. Inevitably then, other issues of comparability not covered by IPUMS-International's composite coding schemes arise for researchers doing comparative analysis of census data. The sample descriptions and variable-specific documentation on the IPUMS-International website are designed to highlight

Table 3. IPUMS integrates census variables to capture common concepts while preserving detail. Educational attainment harmonized codes for a selection of IPUMS-I samples ('x' indicates that the code is present in the respective sample)

		Country (ISO code)																
Code	Variable label	Sample year	BR	CN	EG	FR	DE	IN	IR	MX	PK	PH	ZA	ES	SD	TH	US	VN
			00	90	06	06	87	04	06	06	98	00	07	01	08	00	05	09
General (1 digit) codes and labels																		
0	NIU (not in universe)		x	x	x	x	x	.	x	x	x	x	x	x	x	x	x	x
1	Less than primary completed		x	x	x	x	.	x	x	x	x	x	x	x	x	x	x	x
2	Primary completed		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
3	Secondary completed		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4	University completed		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
9	Unknown/missing		.	.	x	.	x	x	x	x	x	x	x	.	x	x	.	.
Detailed (3 digit) codes and labels																		
0	NIU (not in universe)		x	x	x	x	x	.	x	x	x	x	x	x	x	x	x	x
100	Less than primary completed		.	.	x
110	No schooling		x	x	.	x	.	x	x	x	x	x	x	x	x	x	x	x
120	Some primary		x	x	.	x	.	x	x	x	x	x	x	.	x	x	x	x
130	Primary (4 years)		x
Primary completed, less than secondary																		
Primary completed																		
211	Primary (5 years)		x	x	x	.	.	.	x
212	Primary (6 years)		x	x	x	x	x	.	.	x	x	x	x	.	x	x	x	.
Lower secondary completed																		
221	General and unspecified track		x	x	x	x	x	x	x	x	.	.	x	x	x	x	x	x
222	Technical track		.	.	.	x	.	.	.	x
Secondary completed																		
General or unspecified track																		
311	General track completed		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
312	Some college/university		x	x	.	.	.	x	x	x	.	x	.	.	.	x	x	x
320	Technical track	
321	Secondary technical degree		.	x	.	.	x	.	x	x	.	.	.	x	.	x	.	x
322	Post-secondary technical education		.	x	x	.	x	x	.	x	x	x	.	x	x	x	.	.
400	University completed		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
999	Unknown/missing		.	.	x	.	x	x	x	x	x	x	x	.	x	x	.	.

Source: [https://international.ipums.org/international-action/variables/EDATTAIN#codes_section].

possible comparability problems, so researchers can make informed judgments or adjustments and avoid inadvertent errors. The online documentation for every variable shows with a few clicks the codes and unweighted frequencies, the universe, the question wording and instructions to enumerators (translated into English) and a discussion of major comparability issues for each country/sample. Because researchers generally care about a subset of countries and/or years, the documentation can be easily limited to show only the sample(s) of interest.

Constructed variables

The characteristics of other family members, especially parents and spouses, are empirically related to outcomes for individuals (for example, an association between maternal

education and child health). Fortunately, IPUMS-International has created individual-level family inter-relationship variables that help researchers use information about household structure implicit in the census data samples. Data provided by national statistical agencies indicate the relationship of each person to the head of household, but relationships among other household members are rarely identified. The IPUMS-International 'pointer' variables identify each household member's co-resident mother, father and spouse (if present). These constructed variables make it easy for researchers to automatically attach individual-level variables representing the characteristics of co-resident persons, such as occupation of spouse, age of mother, educational attainment of father or sex of household head. Other constructed variables describe household composition (such the individual's number of own children in the household and age of youngest own child).

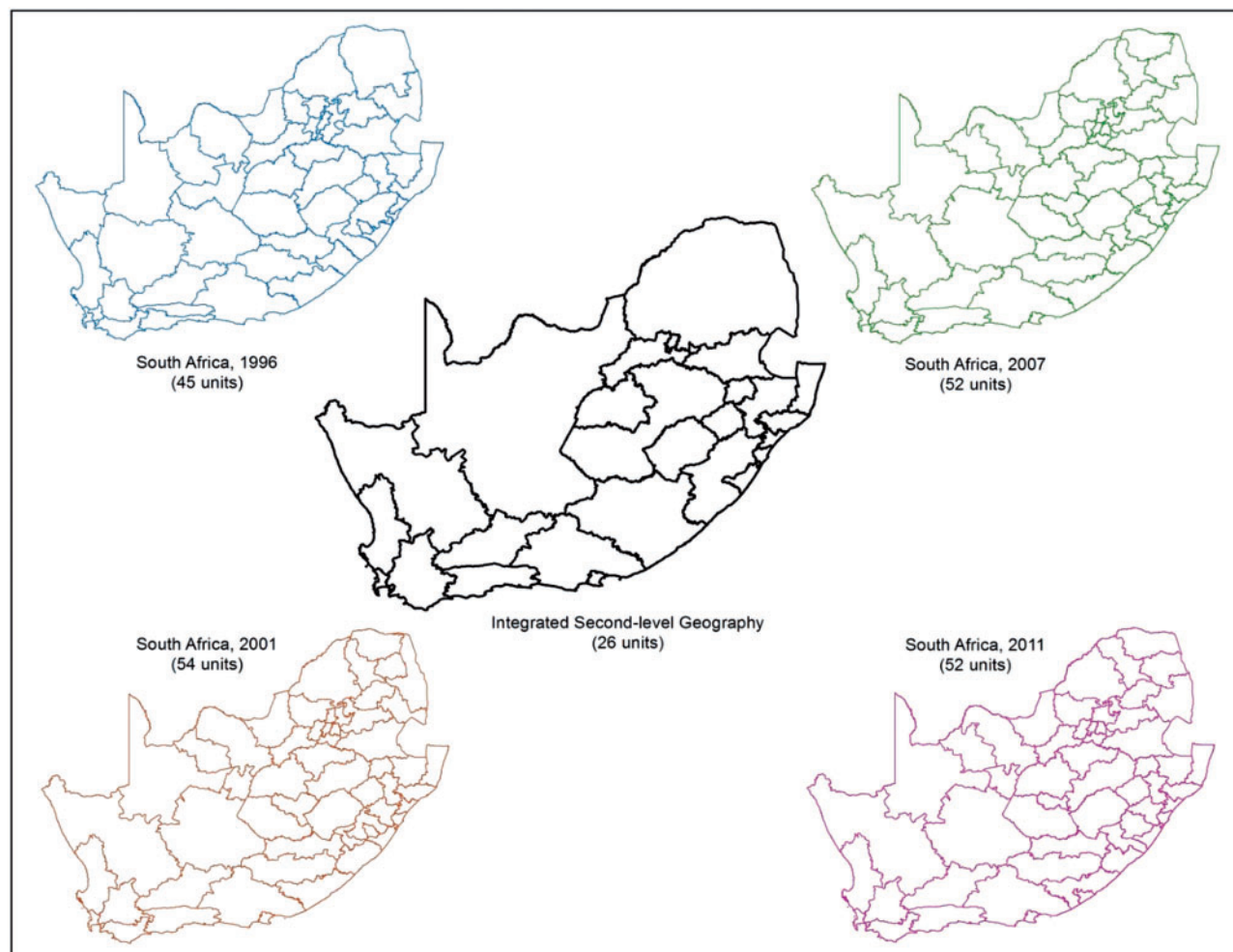


Figure 1. Year-specific and harmonized second-level geographical boundaries, South Africa 1996–2011.
Source: IPUMS-International.

Spatiotemporally harmonized geographical variables

The large samples distributed by IPUMS-International—most commonly, 10% of all enumerated households—make it possible to study small subpopulations (e.g. occupational or ethnic subgroups) and subnational regions of countries. Because public health policies often differ across regions of a country or are put in place incrementally across territory, this geographical detail supports comparative analyses and natural experiments in public health and public policy.²

To account for changing boundaries of administrative units over time, IPUMS-International offers two kinds of integrated geography variables: a version that harmonizes geographical units to have consistent boundaries over time and a set of year-specific geographical units identified in the census. [Figure 1](#) depicts changes in second-level boundaries across census years in South Africa. The map at the centre of the image displays the harmonized boundaries

constructed by IPUMS-International to account for these changes across time. These geography variables and the associated GIS boundary files (.shp files) are available at the first and second administrative levels for most countries. Users can thus easily create thematic maps with IPUMS-International data using a statistical software program and GIS mapping software. Boundary files available from IPUMS-International are in .shp format and can be used in ArcGIS and certain open source software applications, such as QGIS: [www.international.ipums.org].

Pooled, customized datasets

IPUMS-International disseminates pooled extracts containing many samples in a single dataset, tailored to the research needs of the user. By contrast, most statistical offices disseminate separate files that contain all variables and person records in each sample. The IPUMS-International

data dissemination approach is more convenient for researchers, who are not burdened with irrelevant material and not required to merge multiple files for comparative analyses. To create a customized file, the researcher 'shops' online for the free dataset, selecting:

- the country (or countries);
- census year(s);
- variables (age, sex, educational attainment, etc.).

The IPUMS-International extract engine fulfils the request by generating a dataset containing the requested microdata and the corresponding set of DDI (Document Data Initiative) compatible metadata, including a code-book suitable for constructing a system data file in SPSS, SAS or Stata. Other optional features include case selection, which allows users to limit their dataset to contain only records with specific values for selected variables (e.g. women age 15 to 49, employed persons, etc.), and custom sample densities, which keep file sizes manageable.

Online data tabulator

Quick tabulations can be made with the IPUMS-International Online Data Analysis System. The IPUMS-International online analysis system uses high-speed tabulation software developed at UC-Berkeley's Computer-assisted Survey Methods Program. Researchers registered with IPUMS-International can specify samples and variables of interest to get quick calculations output to their computer screen or mobile device. The tabulator is very flexible, allowing the user to create new recoded variables or exclude specified values (such as missing and not-in-universe cases). Along with supplying quick summary results to sophisticated analysts, the tabulator can support data exploration and hypothesis-testing by students who have not yet mastered use of a statistical package.

Data resource use

More than 10 000 registered IPUMS-International users represent a variety of disciplines including economics, demography, sociology, statistics, geography, public policy, public health, medicine, government and media. International research organizations such as the World Health Organization, International Labour Organization and United Nations Population Division have used the data extensively. In addition to academic research, IPUMS-International data can be used to produce reliable customized national and sub-national statistics for use in policy formation and evaluation.³ IPUMS-International data have also been used to track progress towards to Sustainable

Development Goals and other measures of economic development.^{4,5} Among more than 500 citations recorded in the IPUMS-International bibliography are nearly 50 books, a dozen World Bank studies, several dozen dissertations and more than 100 journal articles⁶. As a condition of the licence agreement, IPUMS asks that users supply the title and full citation for any publication, research report, or educational material that makes use of IPUMS data or documentation, at [<https://bibliography.ipums.org/>].

Among the 13 broad classifications offered by the online bibliography, six account for the majority of citations: labour force and occupational structure; migration and immigration; family and marriage; education; methodology and data collection; and fertility and mortality.⁴ Researchers often use IPUMS-International microdata in conjunction with other data sources. With regard to health research, IPUMS-International data are particularly well-suited for studies concerning fertility, mortality, ageing, union and family formation, sanitation, disability and social determinants of health.

In 2015, 11 000 customized datasets were created by more than 2000 unique users using the IPUMS-International online data extract system. Data extracts include five samples on average. Single-country cross-temporal analyses and multi-country comparative research are equally common. Each of the 82 countries represented in the database was included in at least 200 unique data extracts in 2015. Nonetheless, use varies greatly by sample. Over half of the citations in the bibliography focus on six countries that have been included in the database for several years: Mexico, Brazil, South Africa, Colombia, Chile and China.

Strengths and weaknesses

The greatest contributions of the IPUMS-International database are: (i) freely distributing large nationally representative samples of population data unavailable elsewhere; and (ii) consistently naming and coding the variables to facilitate analyses across time and space. Other features that add value to the raw data include (as described above): extensive integrated metadata, within-household relationship pointer variables, GIS boundary files, a user-friendly data access system that allows users to build customized datasets, and an online data tabulator. An experienced user support team will answer questions and troubleshoot problems for free if contacted by e-mail at [ipums@umn.edu].

Special features of the data access system make IPUMS-International particularly valuable as a teaching resource. Classroom accounts give students expedited access to the extract-builder and online data tabulator, and allow

instructors to share datasets directly with students through the IPUMS-International website. Instructors can easily save and modify extracts for use in subsequent courses or teaching terms. This is particularly useful for complex classroom exercises or exams where data extracts can be re-used by modifying the data request with a different country or year.⁷ IPUMS-International invites instructors that do register their classes to share data exercises that others might find useful in their classrooms. Please send data exercises or other curriculum materials to [ipums@umn.edu]. If IPUMS publishes your materials, IPUMS will credit you and your institution with their development. A number of exercises are currently available online; see [www.pop.umn.edu/data-user-resources/data-support] for data exercises.

From the researcher's point of view, the primary shortcoming of IPUMS-International data is that they are cross-sectional; individuals cannot be linked across censuses. Notwithstanding, large sample sizes and harmonized variables facilitate precise cross-temporal analyses.

Epidemiologists will note that national censuses collect limited material specifically about health. Indeed, the content of national censuses is closer to labour force surveys than to health inquiries such as the Demographic and Health Surveys (DHS). Nonetheless, as noted some health topics, such as fertility, mortality, ageing, union and family formation, sanitation and disability, are covered by censuses. In addition, researchers can fruitfully combine IPUMS-International data with other health data for their research. New geography variables available from IPUMS-DHS match those available in IPUMS-International data. The variables correspond to the primary level of geography in both IPUMS-DHS and IPUMS-International. The spatially-consistent variables in the two databases allow researchers to summarize DHS data and attach them as contextual information to the census samples or vice versa.

Even when IPUMS-International variables are given consistent names and coding schemes, such integrated variables may incorporate subtle differences across samples for example, in the definition of disability. Researchers thus need to be attentive to underlying variations in question wording, instructions to enumerators and question universes. Fortunately, the IPUMS-International variable-specific online documentation is designed to highlight such differences.

Although more than 100 national statistical offices have agreed to disseminate samples of their census microdata through IPUMS-International, some countries (such as Russia and Japan) have chosen not to participate, and others (such as Congo-DR and Afghanistan) lack any census microdata. Still, with data on 614 million persons in 82 countries and 277 censuses, the current IPUMS-

International database represents a truly global resource for health research.

Data resource access

Access to the online documentation is freely available without restriction; however, users must apply for access to the data (as a downloadable microdata file or through the online tabulator). IPUMS-International's agreements with participating national statistical offices specify that access is limited to non-profit use (e.g. by scholars, policy makers, teachers and students). To ensure that these agreements are honoured, the application system requires a description of an applicant's proposed research and asks for the user's institutional affiliation and other information to verify identity. Every application is individually reviewed by project staff. Access to the system enables a user to extract data from any country in the database; registrations to use the data expire after 1 year and can be renewed. To apply for access, visit [international.ipums.org].

Funding

The IPUMS-International project is a collaboration of the Minnesota Population Center, national statistical offices and international data archives. Major funding is provided by the U.S. National Science Foundation and the Demographic and Behavioral Sciences Branch of the National Institute of Child Health and Human Development. Additional support is provided by the University of Minnesota Office of the Vice President for Research and the Minnesota Population Center.

Conflict of interest: None declared.

References

1. Ruggles S, King ML, Levison D, McCaa R, Sobek M. IPUMS International. *Historical Methods* 2010;**36**:60–65.
2. See, for example: Bleakley H. 'Malaria eradication in the Americas: a retrospective analysis of childhood exposure. *Am Econ J* 2010;**2**:1–45; Barofsy J, Chase C, Anekwe T, Farshad F. The economic effects of malaria eradication: Evidence from an intervention in Uganda. Working Paper No. 70. Harvard University Program on the Global Demography of Aging (PGDA), 2011.
3. Ruggles S, Sobek M, Esteve A, McCaa R. Using integrated census microdata for evidence-based policy making: the IPUMS-International global initiative. *Afr Stat J* 2006;**2**:83–100.
4. Ruggles S, McCaa R, Sobek M. Using census microdata disseminated by ipums-international to assess millennium development goals of literacy, education and gender equity in the Ugandan censuses of 1991 and 2002. Scientific Statistics Conference; 11–13 June 2007. Kampala, 2007.
5. Cuesta A, Lovaton R. Millennium Development Goals (MDGs): measuring within-country inequalities for selected indicators for South America using IPUMS-International Data (1990–2010). VI

-
- Congress of the Latin American Population Association, 12–15 August. Lima, 2014.
6. McCaa R, Sobek M, Cleveland L, Ruggles S. 2013. The IPUMS big data revolution: liberating, integrating and disseminating the globe's census microdata free of cost. Chaire Quetelet 2013. Demography revisited. The past 50 years, the coming 50 years, 12–15 November. Louvain-la-Neuve, France, 2013.
 7. Kelly Hall P, Cleveland L, Sobek M. IPUMS International: a data resource for statistics education. ICOTS: 9th International Conference on Teaching Statistics, 13–18 July 2014. Flagstaff, AZ, 2014.