

IPUMS International: A review and future prospects of a unique global statistical cooperation programme

Alphonse L. MacDonald

Former Senior United Nations Official, United Nations, New York, USA and Geneva, Switzerland
Tel.: +356 2156 1009; E-mail: fonzhan@gmail.com

Abstract. At the invitation of the University of Minnesota Population Center (MPC) the author carried out an assessment of the IPUMS International integrated census microdata programme during January–March 2016. The terms of reference included the assessment of the measures taken by the MPC to safe guard the security of the microdata, the quality and adequacy of services provided, characteristics of users and satisfaction with IPUMS, use of available microdata, support to participating developing country National Statistical Offices (NSOs) and adequacy of a proposed Remote Data Center (RDC).

The conclusions of the review are that IPUMS International is a unique, flexible, successful and secure programme for managing access to anonymized, harmonised and integrated microdata to academic users and policy makers. While currently the user base is predominantly in developed countries, steps are being taken to expand usage by researchers world-wide. The physical, methodological and technical arrangements for safeguarding the security and confidentiality of the data files are excellent; the possibilities of breaches are minimal. Data users have very positive opinions of the quality of the data, scope of services and expertise of staff but desire more detailed, up-to-date microdata. NSOs rate IPUMS International and its services positively but request advanced methodological training for staff and regular information on the use of their country's data. IPUMS International planned activities are presented and their contributions to census methodology are highlighted.

Keywords: IPUMS International, population census, microdata, metadata, sample data extracts, remote data enclave

1. Introduction

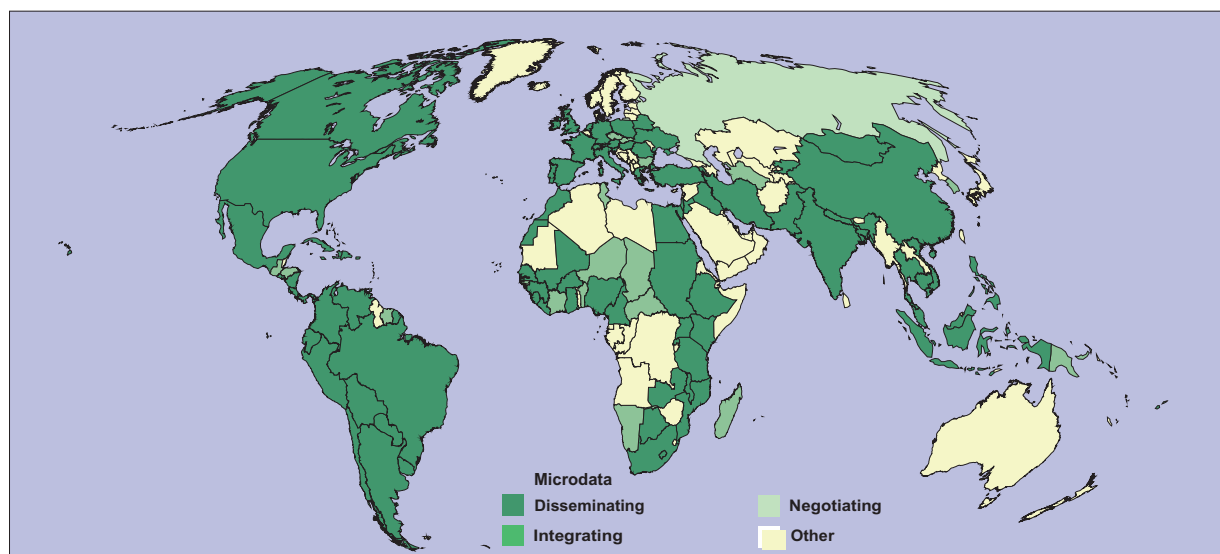
In early January 2016 the author was invited by Prof. Robert McCaa on behalf of the senior management of the Minnesota Population Center (MPC), University of Minnesota, Minneapolis, USA to carry out an evaluation of the Integrated Public Use Microdata Series – International (IPUMS International) project following a suggestion by Mr. Iwan Sno, the Director of the General Bureau of Statistics (ABS) of Suriname.

The terms of reference included the assessment of the security of both original microdata received from National Statistical Offices (NSOs) and integrated microdata,¹ produced by IPUMS International

staff, based on legal, administrative and technical measures taken; the quality of the service provided by IPUMS International, the use of available microdata of censuses, the characteristics of users, satisfaction with IPUMS and compliance with commit-

dealing with data files and microdata. The author uses the term data file to describe a data matrix containing individual information of M variables of N (total population censuses) and uses the terms sample data file when information is provided for a sample of the N individuals. The term data set is used if information is provided for m variables ($m < M$) for all N individuals, and sample data set when information is provided for m variables ($m < M$) for the n individuals (or a sub-sample) contained in a sample data file. IPUMS International uses the term “extracts” for pooled microdata of m variables, n censuses, p sample percent density, s subpopulations and x countries as requested by the individual researcher. When referring to literature from other authors their terms will be used.

¹There is no standard internationally accepted terminology when



Source: Robert McCaa email communication May 14, 2016.

Fig. 1. Countries participating in the IPUMS International partnership in 2016 (includes countries entering in 2016).

ments, the type services and/or support provided to data providers, especially developing countries NSOs; and the relative merits of a Remote Access Facility compared to Microlabs.

The review was carried out from January to end of March 2016 and included a field mission to the Minnesota Population Center, 28 February–3 March, participation and observation of IPUMS International organised workshops in New York 4–6 March and attendance at the 47th Meeting of the United Nations Statistical Commission in New York, USA, 7–11 March. A draft report on the findings of the assessment was submitted to the Minnesota Population Center on 27 March. The overall conclusions of the review are that IPUMS International is a UNIQUE, flexible, successful and secure programme for providing anonymized, harmonised and integrated data sets to researchers from across the globe.

In this paper the author will briefly review the history and objectives of IPUMS International, its current mode of operations, its partners and their opinions, and its future plans. He will provide an overview of their contributions to the enhancement of research and census methodology.

2. IPUMS International: A short historical overview

The Integrated Public Use Microdata Series – International (IPUMS International) is a project initi-

ated and managed by the MPC. It is an international statistical collaboration initiative of the University of Minnesota, National Statistical Offices, International Organisations, international and national academic and research institutions. Following the success of the IPUMS-USA project of the MPC, initiated by Prof. Steven Ruggles in the early 1990s, in the late 1990s, Prof. Robert McCaa started lobbying to establish an international IPUMS project. With support of the National Institutes of Health and the National Science Foundation (USA) the project started in 2000 with an inventory of world microdata and the acquisition of census microdata and documentation with the authorization of the owners, the National Statistical Offices. Much documentation and some microdata were initially preserved at CELADE (Santiago de Chile, Chile), the East-West Center (Honolulu, Hawaii), and elsewhere. In 2002 the first 21 anonymised, harmonised and integrated sample data files became available free of charge for scientific and policy research through the web site www.ipums.org/international from a MPC portal in Minneapolis. Data users were and are subjected to detailed screening procedures and are licensed, annually, to access the IPUMS International system. In 2005 the Integrated European Census Microdata (IECM) web portal (<http://www.iecm-project.org/>) was established in Barcelona, Spain providing access to harmonised data files of participating European countries. At the end of 2007 the cumulative number of harmonised

Table 1
Participating countries by region and number of datasets entrusted to IPUMS-International (May 2016)

Africa		Americas		Asia & Pacific		Europe	
Country	N	Country	N	Country	N	Country	N
Benin	3	Argentina	5	Armenia	2	Austria	5
Botswana	4	Bolivia	3	Bangladesh	4	Belarus	2
Burkina Faso	4	Brazil	6	Cambodia	2	Bulgaria	0
Cameroon	3	Canada	5	China	3	Czech Republic	3
Cape Verde	0	Chile	5	Fiji	6	France	8
Central African Republic	0	Colombia	5	India (NSSO)	5	Germany	4
Chad	1	Costa Rica	5	Indonesia	9	Greece	5
Cote d'Ivoire	0	Cuba	1	Iran	2	Hungary	5
Egypt	3	Dominican Republic	6	Iraq	1	Ireland	9
Ethiopia	3	Ecuador	6	Israel	5	Italy	3
Ghana	3	El Salvador	2	Jordan	1	Netherlands	3
Guinea-Bissau	0	Guatemala	4	Korea, RO	0	Poland	5
Guinea, Conakry	2	Haiti	3	Kyrgyzstan	2	Portugal	4
Kenya	5	Honduras	5	Malaysia	4	Romania	4
Lesotho	3	Jamaica	3	Mongolia	3	Russia	0
Liberia	2	Mexico	8	Nepal	1	Slovak Republic	3
Madagascar	1	Nicaragua	3	Pakistan	3	Slovenia	3
Malawi	3	Panama	6	Palestine	2	Spain	4
Mali	4	Paraguay	5	Papua New Guinea	3	Switzerland	4
Mauritius	2	Peru	3	Philippines	3	Turkey	3
Morocco	3	Puerto Rico	6	Thailand	4	Ukraine	1
Mozambique	2	Saint Lucia	2	Turkmenistan	1	United Kingdom	2
Namibia	3	Suriname	0	Vietnam	3		
Niger	3	Trinidad and Tobago	5	Yemen (LFS)	3		
Nigeria (GHS)	5	United States	7				
Nigeria (PES)	2	Uruguay	6				
Rwanda	3	Venezuela	4				
Senegal	3						
Sierra Leone	1						
South Africa	4						
South Sudan	1						
Sudan	4						
Tanzania	3						
Tunisia	1						
Uganda	2						
Zambia	3						
Total (360)	89		119		72		80
Countries (109)	36		27		24		22

Source: Robert McCaa email communication May 14, 2016. Note: Excludes household surveys, except as noted for India, Nigeria, and Yemen.

integrated sample data files had increased to 80. In 2010 the number had further increased to 160, and the African Integrated Census Microdata (AICMD) web portal (<http://ecastats.uneca.org/aicmd/>) was established in Addis Ababa, Ethiopia providing access to the harmonised sample data files of participating African countries. Over time a number of technological improvements in data access as well as methodological and data quality modifications were introduced which resulted in the current data access system through the IPUMS International general website and the African and European portals.

As of mid-March 2016 the cumulative number of

available harmonised data files had increased to 277. Figure 1 provides an overview of the participating countries by the nature of their involvement.

In Table 1 (above) the details of the number of integrated sample files of participating countries by major geographical regions are provided. The continent with the highest rate of participation is the Americas. All major countries are participating in IPUMS International, except for Guyana in South America, Belize in Central America and the smaller Caribbean countries. The region with the lowest participation rate is Oceania, with only Fiji and Papua New Guinea participating.

3. IPUMS International: Its goals and objectives

At its inception in 1999 IPUMS International had “two principal goals: first, preservation of the World’s census microdata resources, and second, democratization of access to these resources.” [1]. Democratization of access would be achieved by making extracts of the integrated microdata available to researchers worldwide free of charge. A project was designed with major funding from the National Science Foundation to demonstrate the feasibility of the idea. The project had four principal components which still form the core of IPUMS International activities. These are:

- Inventory and Preservation: Identify surviving machine readable census microdata and convert them in current standardised IT formats. In case the data are in non-machine readable form, IPUMS International offers modest support to finance the creation of microdata files. As of mid-March 2016 a total of 461 data files of 109 countries were archived by IPUMS International. Approximately 100 data files were either rescued through IPUMS International with the cooperation of the NSOs or were discovered at academic or research institutions. Not all of the latter are suitable for harmonisation and integration because of incomplete or defective data structures and lack of relevant documentation, but will be preserved for posterity.
- Processing: Convert the sample data files received in anonymized, harmonised and integrated microdata suitable for scientific research and teaching. As of mid-May 2016 101 participating countries had provided 360 sample data files to IPUMS International of which 277 were fully processed and were accessible.
- Documentation: Develop comprehensive metadata on census methodology, questionnaires, manuals and guidelines for data collection and processing, variables and coding schemes to establish comparability across space and time for the benefit of researchers.
- Dissemination: Each approved data user may request microdata specific to his needs which can be analysed on-line, or down-loaded for further detailed analysis. IPUMS International supports ASCII, CSV, SAS, SPSS and STATA formats and is considering R. [2]

4. Services provided by IPUMS International

In addition to the activities carried out to achieve the objectives described above, at present IPUMS International provides the following services to the academic community, the data providers and data users.

4.1. Archives

The MPC maintains a special archival service which collects and preserves documentation, hard copy and electronic documents, pertaining to the different historical data archives, including IPUMS International.

4.2. Secure storage of microdata

IPUMS International has taken extra-ordinary measures to protect the physical and electronic security of the data. The original data files, the intermediate products and the anonymized, harmonised and integrated data files are stored in a separate section of the University’s IT Core. The original microdata files received from NSOs are archived in a separate storage facility independent of the rest of IPUMS International data. Access to this area is restricted to a selected number of IPUMS International project staff, who work with the data within this restricted area until it has been anonymized. Adequate provisions for backup and additional archiving of the data files and other documents have been made.

The IT Core complies with the provisions of the security principals and codes of practice of the National Institute of Standards and Technology (NIST) and the International Organization for Standardisation (ISO). The University IT system has taken the necessary measures to ensure storage security and data replication at alternative university data centres and has taken precautions against power and cooling systems failures. Access to the facility where the IPUMS International data are stored is limited to select IT personnel (3 MPC IT staff) and authorized vendors, who obtain security clearances following background checks. All others persons must be escorted by authorised personnel. Among the authorised MPC IT staff are two persons who are holders of the Certified Information Systems Security Professional (CISSP) certificates which are issued by the International Information System Security Certification Consortium (ISC2). These certificates need to be, and are renewed annually!

4.3. *The IPUMS International website (<https://international.ipums.org/international/>)*

IPUMS International maintains a comprehensive website to assist researchers to use the integrated harmonised sample data files. The website contains detailed information about its objectives, a brief history, conditions of use, registration procedures, tutorials for users on how to register and make use of the IPUMS International data and facilities, a bibliography, with the facility to upload new publications, a Help facility, a User's Forum, a list of the international partners and options to use IPUMS data for teaching including the establishment of a classroom account. The website allows registered users to select the countries and census years they would like to analyse, either by downloading the selected variables or by using the on-line tabulation facility.

4.4. *Classroom accounts*

IPUMS International offers lecturers and instructors the possibility to create (https://international.ipums.org/international/classroom_accounts.shtml) classroom accounts which simplifies the registration for bona fide students. Instructors are requested to remove students who drop out from the classroom list, but are not obliged to report back to IPUMS International which students have successfully completed their assignments. A number of institutions in the USA and abroad (Belgium, Hong Kong, Scotland, Singapore, and Uruguay) covering a variety of subjects, such as demography, economics, population, and sociology have used this facility.

4.5. *The promotion of IPUMS International*

To enhance the participation of countries, IPUMS International management led by Prof Robert McCaa meets with senior managers of National Statistical Offices, either through specific missions to countries or at relevant international conferences such as the sessions of the United Nations Statistical Commission in New York or meetings of regional and sub-regional statistical organisations. In March 2015 Dr. Patricia Kelly-Hall undertook a mission to New Zealand. In 2015 representatives of national statistical offices and other population experts were contacted at the 46th session of the United Nations Statistical Commission meeting in New York and regional meetings in Kampala,

Uganda, Kuala Lumpur, Malaysia, Geneva, Switzerland, Santiago, Chile, and Pretoria, South Africa.

To inform statisticians and demographers about IPUMS International and its activities information booths are staffed in exhibition halls of relevant national and international conferences. In 2015 information booths were installed at professional associations in the USA such as the American Economic Association (AEA), the Population Association of America (PAA) and the International Association for Social Science Information Services & Technology (IASSIST). IPUMS International was also present at regional and international meetings abroad, such as the African Symposium on Statistical Development (ASSD) in Kampala, Uganda., the Third Asian Population Association (APA) Conference in Kuala Lumpur, Malaysia, the 7th African Population Conference of the Union of African Population Studies (UAPS) in Pretoria, South Africa, the International Association for Statistical Education (IASE) in Rio de Janeiro, Brazil, and the 60th World Statistics Congress of the International Statistical Institute (ISI) also in Rio de Janeiro. Additionally in many of these meetings demonstration workshops for current and potential users were also organised.

In collaboration with academic partners IPUMS International co-organised and participated in the following technical workshops: the Mekong Region Development Research Group (MRDRG), Workshop on Demographic Data and Techniques, with Resources and Methods from IPUMS & REVES, in May 2015 in Ho Chi Minh City, Vietnam, and Integrating expertise in inclusive growth (INGRID) Call 22: Expert workshop 'Research uses of high-precision census samples' in June 2015 in Barcelona, Spain.

4.6. *Training for developing country NSOs*

IPUMS International, as an initiative funded by the National Institutes of Health and the National Science Foundation (USA), does not have regular funds to provide training or technical assistance to NSO staff of developing countries. However, in 2015 IPUMS International obtained support of the Inter-American Development Bank (IDB) and the Public Capacity Building Korea Fund for Economic Development (KPC) of the Republic of Korea to organise a Workshop on Digitizing Historical Census Maps of Latin America and the Caribbean Part 1, in Minneapolis, 19–22 May 2015, for cartography technicians of 17 Latin American and Caribbean countries. In 2016 a follow up workshop was organised in New York City, NY 5–

7 March: attended by cartography technicians of 15 Latin American and Caribbean countries. Also, an intensive Workshop on digitizing historical census maps for participants from Bolivia, Costa Rica, Guatemala, Honduras and Suriname was held on Sunday 5 March in New York City. IPUMS-International, with appropriate funding from regional organizations, is eager to undertake similar initiatives in other regions of the world.

5. The IPUMS International data providers, legal and administrative arrangements

To establish the legal and administrative conditions under which participating NSOs entrust microdata to IPUMS International an agreement (Memorandum of Understanding) between the national statistical office and the University of Minnesota is signed. The ownership of data remains with the national statistical office. The data are to be used exclusively for teaching, scientific and policy research and publishing and cannot be used for any commercial or income-generating venture. IPUMS International is committed to ensuring that data users will abide by a number of conditions to safeguard the confidentiality and security of the data and privacy of respondents. Failure to honour conditions of use will lead to cancellation of the license and professional censure and may lead to civil prosecution. IPUMS International commits itself to provide the national statistical office with electronic copies of the integrated sample data files, and metadata, as well as reports on their use. The National Statistical Office and the University of Minnesota agree that in cases of conflict these will be resolved by friendly negotiations, or arbitration according to international law and that in case of conflicting legal instruments this Agreement will have precedence over any other agreements. In some agreements an article is included in which IPUMS International provides a monetary contribution to the national statistical office as compensation for costs incurred in the preparations of the microdata and supporting documentation.

6. The sample data files, processing and quality assessment

Participating NSOs generally provide IPUMS International with a sample, usually ten percent, of the data file of the censuses. In some cases complete

data files are provided from which 10% samples are drawn. Data files from participating NSOs are of varying quality, and some may already have been (partially) anonymized. In all cases, the sample data files are cleaned by erasing duplicate entries, the file structure adjusted, if needed, to create a hierarchical data file of households and household members, and edited for inconsistent information. The data are anonymized beyond what the data provider already may have done, by removing any identifiers, such as name and address of the respondents. Reference to geographical areas is restricted by specific instructions from the NSO-owner. The most common cut-off size is 20,000 inhabitants. In the case of the USA it is as high as 100,000 while the Netherlands does not allow any reference to sub-national geographical location. In general the IPUMS International integrated harmonised microdata offer second-level administrative geographies. Indirect identifiers may be removed, or are top and/or bottom coded while quantitative variables are recoded in grouped data. Certain sensitive variables may be excluded or treated as indirect identifiers. Swapping of households across lower level geographical units may also be applied. Variables are harmonised across time and countries ensuring that the same code refers to the same concept for all samples [3,4].

The initial data processing is performed by a small number of IPUMS International senior staff who have access to the original files. Once the files are anonymized graduate and under-graduate research assistants work on limited portions of the files to carry out specific tasks. They are not permitted to access the original files. Both staff and research assistants are vetted by IPUMS International management, trained in handling sensitive data and sign agreements to respect the strict conditions of access. The process of preparing the data files is labour intensive but IPUMS International is currently able to process a number of datasets simultaneously, as many as 20–25 in twelve month cycles.

The IPUMS International coding structure is rather unique as it provides both general and detailed harmonised coding schemes to retain all significant concepts in the original national codes. This may be illustrated by the harmonised variable “marital status” based on information from the population census samples of Bangladesh 2011, Mexico 2010 and Kenya 2009. In Table 2, (below) it is apparent that the terminology and codes used in the national coding schemes for the variable marital status vary widely.

By harmonising the terminology and the codes a unique coding system is developed consisting of three

Table 2

Example of a correspondence table of the harmonised variable marital status based on information from population census samples of Bangladesh 2011, Mexico 2010 and Kenya 2009

Correspondence table marital status				
Harmonised		Input census data codes and labels		
Code	Label	Bangladesh 2011	Mexico 2010	Kenya 2009
100	Single	1 = Unmarried	8 = Single	1 = Never married
200	Married or in union	2 = Married		
210	Married formally			
211	Civil		5 = Married, civil	
212	Religious		6 = Married, religious	
213	Civil and religious		7 = Married, civil and religious	
214	Monogamous			2 = Monogamous
215	Polygamous			3 = Polygamous
220	Consensual union		1 = Consensual union	
300	Divorced/separated	4 = Divorced/separated		
310	Separated		2 = Separated	6 = Separated
320	Divorced		3 = Divorced	5 = Divorced
400	Widowed	3 = Widowed	4 = Widowed	4 = Widowed

Source: https://international.ipums.org/international-action/variables/KE2009A_0408#codes_section, etc.

digits. The first digit is the harmonised general code consisting of four categories: 1 = Single, 2 = Married or in union, 3 = Divorced or separated and 4 = Widowed. The second and third digits allow the retention of additional details as in the original census material. The IPUMS system of composite codes is applied to all but the simplest variables.

To assess the quality of the original source files a number of tests are carried out. As far as possible completeness of the census as well as the quality of age reporting is assessed using the standard indices such as Whipple's Index, Myers Index and the United Nations age/sex accuracy index. [5] In addition a cohort based coherence test is applied, measuring the consistency of the distribution of an invariant characteristic in two successive censuses. The IPUMS International staff has mainly analysed and verified the education variable. In 2014 IPUMS International and its European counterpart, the Centre d'Estudis Demogràfics, of the Autonomous University of Barcelona, Spain, presented the results of a comparative study of the consistency between the coherence of completed secondary education of two European countries that participated in IPUMS International and the data archived in the EUROSTAT's Census Hub at the meeting of Group of Experts on Population and Housing Censuses organized jointly with Eurostat and Conference of European Statisticians in Geneva [6]. The IPUMS International staff has verified the coherence of completed primary education for a number of African [7], American [8], and Asian countries [9]

The staff members of IPUMS International also prepare a detailed and consistent set of documentation,

metadata, about the census methodology, the variables, their coding instructions and harmonisation. Both the original questionnaire and instructions to field workers in the official language as well as English translations are readily accessible.

7. Harmonised census maps (boundary files)

Among the documentation submitted by the NSOs are boundary files, maps and sketch maps of the geographical divisions used in the census. These maps reflect the situation at the time of the census date and can vary over time. How much geographical detail is included in the integrated data files available to users is determined by the data providers. In general, countries allow geographical details up to the second level to be included. The maps are digitized and processed by IPUMS International staff in such a way that the harmonised geographical variables are consistent over the years covered by the data. In harmonising the data some detailed information may be suppressed to enhance confidentiality and security, by merging small areas with larger ones. In the harmonised geographic variables the names, labels and codes of the sub-divisions are as much as possible harmonised over time. However, the year specific maps are retained and archived.

8. The data users, country of residence and other characteristics

Potential data users are required to register with IPUMS International through an internet based reg-

istration process (see: https://international.ipums.org/international-action/users/request_access). This initial registration process requests a number of key professional details that must be correctly completed before the request for registrations is transmitted to IPUMS International staff. These requests are further subjected to a strict verification process to ensure that only bona fide researchers with a need to access the microdata are approved as users. Table 3 below shows a variable rate of approval of nearly 60% for the years 2013 to 2015, which indicates a denial rate of 43.8%, 40.6% and 42.9% respectively.

The high rate of denial is mainly due to the large number of incomplete applications. For 2015 the detailed information is as follows: 3,541 applications were received, 1,303 were incomplete and were not considered, 211 were not accepted, 2,021 were approved and 6 were still under review.

The number of active IPUMS International users increased monotonically from about 300 in 2005 to over 2,000 in 2014 tapering off slightly in 2015; the annual growth rate from 2011 onward is higher than in the period 2005–2010.

The majority of users are individuals with links to academic and research institutions but staff and consultants of the World Bank, the Population Division of the United Nations, the Inter-American Development Bank and the World Health Organisation also use the data. In 2015 out of the 2,021 approved users about 5.0% were non-academics, and 5.7% had no institutional affiliations; 66% were students; 11% faculty members; 14% academic researchers and the rest other academic personnel. This is only slightly at variance with the overall finding that cumulatively 50% are students, 15% are faculty members, 15% are researchers, and 5% are other academics and non-academics respectively.

In terms of academic disciplines just over 50% were from the department of economics, 16% of demography, 8% of sociology and roughly 5% each of departments of public policies and statistics; 36% claimed to do research for a scientific article and 1% for a book, but 28% stated that this was for a class assignment; 13% for a PhD thesis and 11% for another thesis, masters or similar.

For the period 2012 to the end of September 2015 the cumulative percentage of active users was 77% of the total number of approved users, but the cumulative percentage of publications was only 10% of the cumulative number of users. This is rather low given that over 50% of the users claimed that the output of

Table 3
The number of applications, approved and current users and publications by years

Year	Applied	Approved		Active	Number of publications
		N	%		
2013	3,178	1,787	56.2	1,189	125
2014	3,467	2,058	59.4	1,578	158
2015	3,541	2,021	57.1	2,402	109
To Sep 2015	*	12,399	*	9,552	1,251

Source: Robert McCaa email communication February 22, 2016.
Note*: This information is not available due to changes in the administration.

their research would be either a thesis or a scientific article or book! However, each author of a multi-author publication is required to register to access the data, but the bibliographical entries are per publication, not per author. Also, there is no information of the number of users that access the IPUMS International data for class assignments, as these normally do not result in publications.

For 2014 and 2015 detailed information is available for country of residence of users. The country of residence is not necessarily country of nationality. Especially for developed countries it is likely that some of the users are not nationals of those countries and even could be nationals of developing countries. In 2014 and 2015 registered users were located in 78 and 84 countries respectively. In both years there were 27 countries in which there were more than ten users, representing 90.6% and 91.6% of the data users of those years. Among the 27 countries there were 24 that were represented in both years, namely in order of number of users in 2014: the United States, the United Kingdom, Germany, Canada, France, Spain, Brazil, China, Argentina, Japan, Colombia, South Africa, Italy, Mexico, Belgium, Chile, India, Netherlands, Australia, Sweden, Uruguay, Singapore, Thailand, and Austria. Of the ten countries with the highest number of users in 2014, the first eight are also among the countries with the ten highest numbers of users in 2015.

In 2015 the ten institutions with the highest number of users were all located in the United States; nine university institutes and the World Bank. The universities are: the University of Minnesota, Harvard University, University of California, Berkeley, University of Michigan, The World Bank, University of Chicago, Columbia University, Dartmouth College, University of California, Davis and Wellesley College.

The number of institutions of developing countries involved with IPUMS International is steadily increasing. At present academic institutions from the following regions and countries participate: in Africa, Egypt,

Kenya, Nigeria, South Africa (2) and Uganda; in Latin America, Argentina, Brazil, Chile, Colombia and Mexico; and in Asia, China, Hong Kong S.A.R., Japan, Korea, Malaysia, and Singapore.

For 2015 out of the 27 countries with at least ten users five countries do not participate in IPUMS International, namely: Singapore (34 users), Australia (32 users), Belgium (18 users), Japan (15 users) and Sweden (14 users). In the countries with less than ten users the following do not provide sample data files to IPUMS International: Hong Kong S.A.R. (9 users), Denmark (8 users), Norway (4 users), United Arab Emirates, Laos, Lebanon, Luxembourg, New Zealand and Taiwan, each with 2 users, and Albania, and Cyprus, each with one user.

9. The data sets, characteristics

Registered users do not have access to the original or complete integrated harmonised sample data files. Instead they select the countries, censuses, variables and even portions of samples or subpopulations (“SELECT IF” command) they wish to analyse. The IPUMS system then constructs the requested dataset (extract) to which the data users have access, using their own statistical package for analysis. They may also access an on-line facility to create statistical tables which can be downloaded. Hence there is no way in which data users can access or modify the data stored at the IPUMS data storage facility.

As of February 2016 the cumulative number of extracts requested was over 80,000. This number increased steadily from less than 100 in 2005 to over 11,000 in 2014 and 2015. On average each user will access data from three countries, from six censuses and 35 variables. The ten most used samples are from: Mexico, Brazil, the United States, Colombia, Argentina, Indonesia, Chile, South Africa, Ecuador and France; six are from countries from the Americas. As has been observed above most countries from the Americas are actively participating in IPUMS International and have entrusted microdata for many censuses. The quality of the original data files and of the IPUMS integrated documentation (metadata) are good and these are undoubtedly positive factors of the attractiveness of these data for researchers.

10. The IPUMS International future plans

For the immediate future IPUMS International intends to expand beyond censuses to include data from

household surveys, specifically Labour Force surveys. Such surveys are already integrated and accessible for India (NSSO Schedule 10) and Nigeria (GHS).

For some time MPC management has considered the possibility of creating an IPUMS International Restrictive Data Enclave (RDE) or Remote Data Center (RDC) to provide users the opportunity to access, but not download, microdata with enhanced security of the highest precision and densities possible (even 100%) with the permission of the owner-NSO. Such a facility would be set up in a special section of the IPUMS International data storage area with additional security arrangements. It would be managed by IPUMS International and would be accessible only by specially vetted and certified users. Strict security measures would be implemented over and above the standard IPUMS International requirements: no options to download data sets, special online data processing facilities, and outputs would be reviewed and vetted before they would be downloadable by the users. The system could be operated through specially configured workstations in participating national statistical offices, certified research centres and universities. Full details are yet to be determined but the IPUMS International IT staff has already carried out some preliminary studies and it is expected that a pilot could be launched in the second half of 2016. The costs for setting up and managing the system will be borne by IPUMS International without cost to participating national statistical offices.

IPUMS International is also planning to expand its efforts to include more detailed geographical variables in the integrated data files and further promote the linkage between statistical population data and environmental and geographical variables as is currently being done by the Terra Pop programme.

In view of the success of the Workshops on Digitizing Historical Census Maps of Latin America and the Caribbean, IPUMS International plans to organise similar workshops for NSOs in Africa and Asia, as funding permits.

11. Conclusions and future developments

Mr. Dennis Trewin, the former Australian Statistician and ex-President of the International Statistical Institute, in 2007, after a week-long on-site inspection at the MPC, concluded: “IPUMS International provides a range of deeply appreciated services with rapidly increasing demand particularly in the United States. It

also has the potential for even greater use internationally especially by the international agencies. It could become one of the most important global statistical assets... Indeed it is likely to provide the best practice for a Data Repository of international statistical data sets" [10]. His expectations have become reality.

IPUMS International is a UNIQUE, flexible, secure and successful programme for providing access to high quality, anonymized, harmonised, and integrated census microdata to researchers across the globe.

Currently, IPUMS disseminates microdata free of cost for 82 countries, totalling more than 600 million person records, to researchers of 132 nationalities. Each year, microdata for a half dozen countries are added to the database. The methodological approach of IPUMS is comprehensive and fully compliant with internationally recognized scientific principles and technical requirements. The legal, administrative, physical, methodological and technical arrangements for safeguarding the security and confidentiality of the data files entrusted to IPUMS are excellent and possibilities of breaches are minimal. IPUMS has no direct means of verifying compliance with the terms of the license with researchers, but to date no case of breach of the terms of the license has come to the attention of IPUMS staff, the University of Minnesota, any National Statistical Office, nor any media outlet. The quality of the staff, permanent and temporary, involved in processing of the data is excellent and their dedication exemplary. IPUMS International complies fully with the provisions of the Fundamental Principles of Official Statistics (see <http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>).

The benefits that IPUMS International bestows on data users are universally recognised and appreciated. What is less well recognised are the benefits to the data providers, the NSOs. First, the country's census microdata files will be processed to high quality standards and be transformed into anonymised, harmonised, and integrated microdata, metadata, and geographic boundary files. Second, the microdata will be accessed both nationally and internationally, relieving the NSO from the burden of managing a facility to make its data available for international research. Third, availability of a country's microdata for international use invariably has a positive effect on the image of the national statistical system. Fourth, NSOs may also use the IPUMS integrated microdata for internal, national and international purposes. Fifth, the methodology to produce high quality integrated samples is available to NSOs, which could take advantage of IPUMS methods to up-

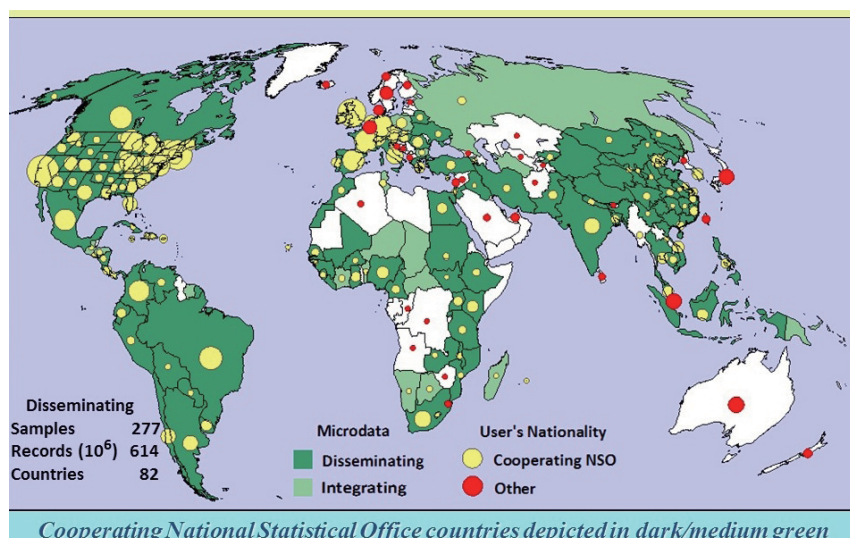
grade full-count census microdata files to a higher standard. Sixth, IPUMS provides a recovery service to restore "lost" and damaged microdata files, which are often reconstructed to high quality standards. Finally, IPUMS also offers a secure archival service for original source microdata. Unfortunately, IPUMS offers only limited possibilities of training of staff of developing country NSOs. Due to the limited, scientific nature of funding, the project actively seeks resources from international development banks and similar agencies to support capacity development activities.

11.1. Opinions of data providers

Practically since its inception IPUMS International has been well regarded by the statistical community. The participation of NSOs has been impressive. As of May 2016, 109 countries have endorsed the IPUMS Memorandum of Understanding, and 101 have entrusted microdata (see Table 1). In Africa, Asia, Europe and Oceania there remain scope for a larger number of countries to cooperate. It was difficult for the reviewer to contact individuals who had taken part in the original decisions to provide data of their country to IPUMS. Discussions with senior officials of NSOs who were aware of their country's participation were generally positive in the sense that IPUMS was considered a trusted partner, and that it was generally a good idea to share anonymized samples for international comparison. Few were aware of interactions between staff of their offices with the IPUMS team. Some indicated that it would be beneficial for their staff to be trained in innovative techniques. Few were aware of the on-line bibliographical tool (<http://bibliography.ipums.org>) to inform themselves regularly on the use of the data of their country. The efforts of IPUMS to conserve and preserve historical census data were greatly appreciated.

11.2. Opinions of data users

In general comments received from data users on IPUMS International, its products, services and staff members are very positive. Frequently mentioned are the wide choices of data sets over time and space, the comprehensive documentation and the ease of access. The users have some suggestions for improvements, mostly requesting more frequent and more detailed data from more countries, and more detailed documentation especially about changes in the period between censuses. There are very few negative com-



Source: Robert McCaa email communication May 14, 2016. Note: Although concentrated in developed countries IPUMS International data users are scattered across the globe, even in non-cooperating countries indicated by white background with a red dot.

Fig. 2. IPUMS-International registered users May 2016: 132 Nationalities.

ments, mostly related to the lengthy and complicated registration procedure, delays in reactions to negative comments, the restrictive nature of access and delays in the release of data.

The opinions of researchers of the World Bank and the Population Division of the United Nations are very positive. A consultant of the World Bank considered the data available through IPUMS International to be more useful than the Bank's own data collected through the Living Standards Measurement Surveys (LSMS). Colleagues of the Population Division of the United Nations published a paper in 2013 in which they highlighted why the microdata provided through IPUMS were of particular benefit to their activities, namely their wide geographic and historical coverage, their degree of integration and harmonisation, ease of use, the quality of the documentation and ready access to household microdata. They also indicated that IPUMS disseminated microdata had been crucial in completing tabulations for a number of standard publications of the Population Division. Their overall conclusion was that "...IPUMS International data are of great relevance to the analytical work of the Population Division (...), and all the features and functionalities provided by the IPUMS web site make the access to existing microdata and documentation as convenient as possible." [11]. In March 2016 a senior official of the Population Division of the United Nations confirmed that IPUMS was and remains invaluable to the activi-

ties of the Division, especially since timely access to data for some countries was problematic.

The overwhelming majority of the data users have academic connections and are mainly located in the developed countries, although there is an increasing number of institutions in the developing countries that make use of IPUMS (see Fig. 2). With some minor additional effort and initiatives on the part of IPUMS and its Associates more institutions and researchers from developing countries could be making greater use of the opportunities offered. Future developments planned by IPUMS will undoubtedly result in an increase of users in those countries.

11.3. Opinions of workshop participants

Although IPUMS has limited options for training and technical assistance to NSOs of developing countries the experience of the Workshops on Digitizing Historical Census Maps of Latin America and the Caribbean held in New York on 5 and 6 March 2016 were very highly rated by the participating technical staff. The fact that the instruments (camera and tripod) to apply what they had learned had been provided free of charge, at a prior training workshop at the MPC in 2015 was much appreciated. One of the participants stated: "We had no excuse for not digitizing maps available in the office".

11.4. *Opinions of non-data providers*

It was possible to obtain some indications to what the reasons were for the reluctance to participate in IPUMS International of countries not yet doing so. Some indicated that they were not fully aware of IPUMS activities and others mentioned that they did not feel comfortable entrusting data to a third party that was not a UN agency where their data would be accessible without their control. Some had the incorrect impression that IPUMS provided full copies of the original sample to users. Explaining IPUMS policies on data access may reduce misunderstanding and distrust by NSOs not yet cooperating. Among some small nations there were reservations about maintaining confidentiality of their data, but they accepted that the procedures used by IPUMS made breaches of confidentiality and data integrity difficult, if not impossible.

11.5. *Prospects for the IPUMS International Remote Data Center*

The IPUMS International Remote Data Center is an option that will allow countries, especially small countries, to fully participate in the IPUMS International programme benefitting from the advantages that such participation bestows on the country. Moreover, it would allow NSOs of countries with restricted IT infrastructure and facilities to allow its staff and national researchers, including students, access to a wide range of high quality census data, including their own, if they partner with IPUMS International. This would reduce if not eliminate the need to maintain a national Micro laboratory. However, depending on the access to national data which are not part of the IPUMS International would justify the operation of a Micro laboratory in addition to the IPUMS International Remote Data Center. Finally, the Remote Data Center offers a secure facility for small countries, the Caribbean and Pacific island nations, to participate in the IPUMS programme.

11.6. *International recognition of IPUMS for policy monitoring*

Recognition of the high quality of IPUMS International integrated data and the services it provides is not limited to just academic and scientific research as was manifest in the paper of the staff the Population Division of the United Nations and the opinion of the World Bank consultant. The usefulness of IPUMS

and integrated census microdata for assessing the 2030 Agenda for Sustainable Development, adopted by the General Assembly in September 2015, has recently been acknowledged. In a report of the Secretary General of the United Nations on Strengthening the demographic evidence base for the post-2015 development agenda to the Forty-ninth session of the Commission on Population and Development reference is made that IPUMS “facilitates statistical analysis of integrated, high-precision census samples for 80 countries in roughly equal proportions by continent” (para 12). Paragraph 54 is dedicated to a summary overview of IPUMS International as “A notable source of demographic microdata...” In paragraph 67 reference is made to the TerraPop project in which population census data are linked to geographical variables and other geo-referenced information including environment and climate change [12].

11.7. *The challenge of more detailed geography*

Further processing maps beyond the second administrative level may be required especially since in the MPC’s TerraPop project population characteristics will be linked to environmental and climatological variables within specific geographical units. Depending on the subject of the research project the information needed may be based on geographical units below the second administrative level. The contributions of IPUMS to TerraPop are crucial for providing census information for standardised and harmonised geo-referenced units.

11.8. *IPUMS contributions to census methodology*

The international statistical community is well advised to take note of, and perhaps adopt, two methodological developments pioneered by IPUMS. The first is the coding system of harmonised variables, which provides an option for a standard harmonised coding system while retaining specific details of each individual census or country. The second is the use of coherence tests for invariant variables. Up to now IPUMS has concentrated on the education variables but there are more invariant variables that could be used, such as sex/gender, ever-married, ever-employed, etc. Disaggregating these tests by sex/gender would provide an additional set of criteria for assessing data quality. Adopting these techniques into the global census methodology would be a fitting tribute to IPUMS as an outstanding international statistical programme.

Acknowledgements

The author wishes to express his gratitude to the senior management of the Minnesota Population Center (MPC), University of Minnesota, Minneapolis, USA for having invited him to assess the Integrated Public Use Microdata Series – International (IPUMS International) one of its current nine microdata projects. Special thanks are due to Mr. Iwan Sno, Director of the General Bureau of Statistics of Suriname and Prof. Robert McCaa for their confidence in his expertise and capabilities. He is very grateful to the staff of IPUMS International and the Associates of the Minnesota Population Center (MPC) for their willingness to discuss matters of relevance to the IPUMS programmes, for the preparations of special presentations on specific issues of the IPUMS International approach, their professional support and personal hospitality during his field mission to Minneapolis, 29 February–3 March 2016. The author appreciates the candid opinions provided by IPUMS International data users, and colleagues from participating and non-participating national statistical offices and international organisations. Special thanks are due to Prof. Robert McCaa for his copy editing skills. The author alone remains responsible for the content and opinions expressed.

References

- [1] S. Ruggles, The minnesota population center data integration projects: challenges of harmonising census microdata across time and place, University of Minnesota, 2006. Available from <http://www.hist.umn.edu/~ruggles/jsmx9.pdf>.
- [2] S. Ruggles, R. McCaa, M. Sobek and L. Cleveland, The IPUMS collaboration: Integrating and disseminating the world's population microdata, *Journal of Demographic Economics* **81**(2) (June 2015), 203–216. DOI: 10.1017/dem.2014.6, Published online: 20 May 2015. Link to this article: http://journals.cambridge.org/abstract_S2054089214000066, (p. 210).
- [3] R. McCaa, S. Ruggles, M. Davern, T. Swenson and K.M. Palipudi, IPUMS International high precision population census microdata samples: Balancing the privacy-quality trade-off by means of restricted access extracts, University of Minnesota, 2006. Available from: http://www.hist.umn.edu/~rmccaa/ipums_psd2006_0824.pdf.
- [4] L. Cleveland, R. McCaa, S. Ruggles and M. Sobek, When excessive perturbation goes wrong and why IPUMS International relies instead on sampling, suppression, swapping, and other minimally harmful methods to protect privacy of census microdata, University of Minnesota, 2012, Available from: [https://experts.umn.edu/en/publications/when-excessive-perturbation-goes-wrong-and-why-ipumsinternational-relies-instead-on-sampling-suppression-swapping-and-other-minimally-harmful-methods-to-protect-privacy-of-census-microdata\(685fa063-3fc2-49f5-ae2-ad29ab009eab\).html](https://experts.umn.edu/en/publications/when-excessive-perturbation-goes-wrong-and-why-ipumsinternational-relies-instead-on-sampling-suppression-swapping-and-other-minimally-harmful-methods-to-protect-privacy-of-census-microdata(685fa063-3fc2-49f5-ae2-ad29ab009eab).html).
- [5] U.S. Bureau of the Census, Evaluating Censuses of Population and Housing, Statistical Training Document, ISP-TR-S, Washington, D.C., 1985, pp. 133–140.
- [6] United Nations Economic Commission for Europe, Conference of European Statisticians Group of Experts on Population and Housing Censuses organized jointly with Eurostat Sixteenth Meeting, Geneva, 23–26 September 2014. Statistical Coherence in Secondary Education Completed: Census Hub Hyper Cubes and IPUMS/IECM integrated census sample results compared, Note by IPUMS/IECM, ECE/CES/GE.41/2014/27, Geneva, 10 September 2014.
- [7] R. McCaa, L. Cleveland, P. Kelly-Hall, S. Ruggles and M. Sobek, Statistical coherence of primary schooling in population census microdata: IPUMS International integrated samples compared for fifteen African countries, *African Population Studies* **29**(1) (2015), 1570–1586.
- [8] R. McCaa, L. Cleveland, R. Lovatón Dávila, S. Ruggles and M. Sobek, Census microdata as the historical demographer's time machine: The statistical coherence of primary schooling completed in Latin American census samples, 1960–2015, paper presented at the International Seminar on Registering and Counting the Population: The production and exploration of census information from early modern times to the end of 20th century, organized by the IUSSP Scientific Panel on Historical Demography and the Department of Demography, NEPO-UNICAMP, Campinas – São Paulo, Brazil, 19-21 May 2016.
- [9] R. McCaa, L. Cleveland, P. Kelly-Hall, S. Ruggles and M. Sobek, Statistical coherence of primary schooling in IPUMS International integrated population samples for China, India, Vietnam and ten other Asia-Pacific countries, *Chinese Journal of Sociology* **1**(3) (2015), 333–355.
- [10] D. Trewin, IPUMS International: “Best practice” A review. Minnesota, 2007. Available from http://www.hist.umn.edu/~rmccaa/IPUMSI/trewin_ipums_report.pdf.
- [11] P. Gerland, T. Spoorenberg, J. Bravo, P. Lattesc, C. Sawyer, V. Kantorovae and M.S. Lai, Uses of census microdata by the united nations population division, Paper prepared for the KOSTAT pre-event to XXVII IUSSP International Population Conference Busan, Korea on “Using Census Micro-data for Demographic Research” (23–25 Aug 2013), 4–14. Available: https://www.researchgate.net/publication/256375892_Uses_of_Census_Microdata_by_the_United_Nations_Population_Division.
- [12] United Nations Commission on Population and Development, Forty-ninth session, 11-15 April 2016. Strengthening the demographic evidence base for the post-2015 development agenda. Report of the Secretary-General. E/CN.9/2016/3, New York, 27 January 2016. Available: <http://www.un.org/en/development/desa/population/documents/cpd-report/index.shtml>.