

Implementing IPUMS-International Confidentiality Protocols using CPro/IMPS: 1991 Census Microdata of Saint Lucia

Michael J. Levin* and Robert McCaa**

* Harvard University School of Public Health, jlevin00@yahoo.com

** Minnesota Population Center, Minneapolis, MN 55455, USA, rmccaa@umn.edu

Research funded in part by the National Science Foundation of the United States, grant SES-0433654 ‘International Integrated Microdata Series (IPUMS-International II)’

“Inadequate use of microdata has high costs”

—Len Cook (2003)

Abstract. Confidentiality protections for census microdata depend not only on the sensitivity and heterogeneity of the data, but also on the potential users. It is widely recognized that statistical agencies expend substantial effort to protect microdata from misuse by academics, their most trust-worthy users. The IPUMS-International project, by disseminating only integrated, anonymized microdata and restricting access to licensed academic users, shifts the risk-utility curve sharply rightward—substantially increasing utility with only marginal increments in risk. This paper discusses legal, administrative and technical protocols of the IPUMS-International project. In addition, we discuss an experiment to implement IPUMSi technical protocols using , first CPro, then IMPS. The 1991 census microdata of Saint Lucia is used as an example. For statistical offices that wish to implement technical anonymization measures, this experiment may be of interest.

1 Introduction: IPUMS-International.

The IPUMS-International is a global initiative led by the University of Minnesota Population Center to confidentialize, harmonize and disseminate high-density census microdata samples on a restricted access basis to academic users. Begun in 1999 with funding provided by the National Institutes of Health and the National Science Foundation of the United States, to date the initiative enjoys the endorsement of official statistical institutes of more than ninety countries. Marginal costs of constructing and maintaining the database are born by the MPC, its funding agencies, the University of Minnesota and academic partners—not by the statistical institute partners. On the contrary each is paid a modest fee per census to license microdata and documentation to the project. As of June 2009, anonymized samples for 130 censuses (44 countries) are integrated into the database. Work is continuing on an additional 90 censuses (35 countries, see Appendix A), thanks to funding through 2014 by NSF and NIH. More than 3,000 users representing more than 76 countries are currently licensed to obtain custom-tailored extracts free of charge from the project website:

<https://www.ipums.org/international>

The IPUMS project has developed a widely accepted series of protocols for anonymizing microdata. Many statistical offices, such as Statistics Austria, the Federal Statistical Office of Switzerland, and others, rely upon IPUMS to anonymize microdata entrusted to the project. Nonetheless, for those statistical offices that prefer to anonymize microdata prior to entrusting to IPUMS and lack the capacity to do so, we offer this experiment using CPro/IMPS.

2 Confidentiality.

In practice, disclosure of confidential information from census microdata samples is highly improbable. Moreover, researchers have no interest or incentive to even attempt to identify individuals. There are compelling reasons for jealously guarding confidentiality, both for individual users and the academic community as a whole. Any partially successful effort, such as that by a rogue intruder, will require an enormous investment of resources to obtain rather trivial details invariably with a high degree of uncertainty about whether any one record truly corresponds to a targeted individual or entity (Dale and Elliot 2001). Indeed, over the past forty years of disseminating census microdata in the United States and elsewhere there are few allegations of misuse or breach of statistical confidentiality by an academic researcher.

Len Cook (2003) notes that increased access is not a threat to statistical systems. On the contrary he observes that increasingly there is an expectation that analysis of microdata will inform research and evaluation of policy. Increased access builds trust in statistical systems, while lack of access leads to suspicion. He advocates that different forms of access be granted for different degrees of trust. Moreover academic researchers possess a range and depth of expertise that national statistical institutes cannot replicate.

Julia Lane (2003) highlights five classes of benefits which accrue from broader access to microdata: address more complex questions, calculate marginal effects, replicate findings, assess data quality and build new constituencies or stakeholders. Replication is extremely important because there is an overwhelming temptation for scientists to misrepresent results when the data are unlikely to be available to others. The IPUMS system facilitates replication by providing access to microdata to all approved academic users on an equal basis.

The IPUMS-International procedures are designed to extend this nearly perfect record of protecting census microdata by means of three types of confidentiality protections:

1. legal: dissemination agreements between the University of Minnesota and each participating Official Statistical Institute
2. administrative: licenses between the University of Minnesota and each user, specifying conditions and restrictions of use
3. technical: perturbations of the data (swapping, recoding, etc.) to make exceedingly unlikely the identification of individuals, families or other entities in the data. Technical measures have the additional benefit that any assertion of absolute certainty in identifying anyone in the data is false.

While much of the literature on statistical confidentiality ignores the legal and administrative environment (and in doing so exaggerates the risk of improper use), we remain firmly persuaded that the strongest system of protections must take into account all three types of guarantees (Thorogood 1999). IPUMS-International confidentiality standards seek to comply with EC Regulation 831/2002, although this regulation encompasses only four datasets at present: European Community Household Panel, Labor Force Survey, Community Innovation Survey, and Continuing Vocational Training Survey (King 2003).

2.1 Legal protections.

First, with regard to legal protections, IPUMS-International projects are undertaken only in countries where explicit authorization is forthcoming, usually in the form of a memorandum of understanding endorsed by the official statistical institute and the legal authority of the University of Minnesota (see Appendix B). No work is begun with the microdata of a

country without prior signed authorization from the corresponding OSI. The agreement is highly general and uniform across countries. Details specific to each country such as fees and sample densities are negotiated separately with each official agency and do not form part of the agreement. Under a carefully worded legal arrangement, the Regents of the University of Minnesota are responsible for enforcing the terms of the accords. The ten clauses spell out: 1) rights of ownership, 2) rights of use, 3) conditions of access (in which statistical institutes cede their gate-keeping authority to grant individual licenses to the IPUMS-International project), 4) restrictions of use, 5) the protection of confidentiality, 6) security of data, 7) citation of publications, 8) enforcement of violations, 9) sharing of integrated data, 10) and arbitration procedures for resolving disagreements. There are no secret clauses or special considerations. Although minor rewording of clauses is permissible, all members of the consortium are treated equally.

Nonetheless, the protocols are revised, indeed expanded, as OSIs suggest, or request, modifications. Any request for modification is reviewed by the legal cabinet of the University of Minnesota. Compare for example the violations clause in Appendix B (as signed by Statistics Austria in January 2002) with the current text (additions in italics), as follows:

Violations. Violation of the user license may lead to professional censure, loss of employment, and/or civil prosecution. *The University of Minnesota, national and international scientific organizations, and the [the Statistical Agency of Country X] will assist in the enforcement of provisions of this accord.*

In 2005, the tenth clause, which establishes jurisdiction for the settlement of a dispute between the University and any signatories to the memorandum, was amended, substituting the International Court of Arbitration for the Chamber of Commerce of Paris. At the same time, an eleventh clause, regarding order of precedence, was added, specifying that the clauses in the letter of understanding supersede any contract, purchase order or other document signed between the parties. Under the agreement, the Minnesota Population Center and its authorized partners are obliged to share the integrated data and documentation with the official statistical institutes and to police compliance by users.

2.2 Administrative measures.

Second, researchers must apply for a license to gain access to the microdata extraction system. Grounds for approval are based upon three considerations:

1. whether the data are appropriate for the proposed project as stated in the applicant's project description
2. whether the applicant is an academic, non-commercial user
3. whether the applicant agrees to abide by the restrictions on conditions of use.

The vetting of applications is performed by the Principal Investigators of the IPUMS-International project. It is noteworthy that approximately one-third of applications are denied because of a failure to adequately satisfy one or more of the specified conditions. It is gratifying to report that few users appeal denial of access.

Administrative measures limit access to the extract system to users, who:

1. sign the electronic non-disclosure license;

2. endorse prohibitions against a) attempting to identify individuals or the making of any claim to that effect, b) reporting statistics that might reveal an identity and c) redistributing data to third parties;
3. agree to use the data solely for non-commercial ends and to provide copies of publications to ensure compliance;
4. place themselves under the authority of educational institutions, employers, institutional review boards, professional associations, and other enforcement agencies to deal with any alleged violations of the license.

The license is granted to users, individually, not to research groups, classes, or institutions. The license application instructs the applicant regarding conditions of use. The license is not transferable. Should the individual change institutions or employment, the license must be updated. Data can be reassigned within an institution, but the person responsible for the microdata must apply for access. Once licensed, the user is permitted to download data extracts of samples and variables according to need. Licensees import the extracts into their statistical software of choice to analyze at the convenience in their own institutional setting.

Since its adoption in 2002, the basic application procedure remains unchanged. Few suggestions for enhancing the application form or approval process have been forthcoming, even though advice is solicited from users, statistical institutes, funding agency review boards, and outside experts,. Nevertheless in 2006, we plan to strengthen application and vetting procedures, primarily to guard against fraudulent applications. In addition to requesting additional details about the applicant and institutional affiliation, the form will contain the following statement as a heading:

Legal Notice: Submission of this application constitutes a legally binding agreement between the applicant, the applicant's institution, the University of Minnesota, and the relevant official statistical authorities. Submitting false, misleading or fraudulent information constitutes a violation of this agreement. Misusing the data by violating any of the conditions detailed below also constitutes a violation. Violation of this agreement may lead to professional censure, loss of employment, civil prosecution under relevant national and international laws, and to sanctions against your institution, at the discretion of the University of Minnesota and the official statistical authorities.

In the United State, an Institutional Review Board for the protection of human subjects is required of any academic research institution applying for funding from the National Institutes of Health. IRBs provide a strong mechanism for enforcing of the IPUMS-International license agreement in the United States. Most developed and developing countries have similar mechanisms. Oversight boards are nearly universal. It is these boards that provide a strong shield for insuring the highest standards of scientific conduct.

Finally, once these revisions to the application are in place on the website, licenses will be valid for one year and is renewable. A license may be suspended at any time.

2.3 Technical protections.

Third are the technical measures taken to ensure statistical confidentiality. Sampling of datasets alone “provides the additional uncertainty needed to protect many data releases...” (Anderson and Fienberg 2001). Census errors and non-response error also provide their own confidentiality protections. As Fienberg (2005) has noted the principal threats are geographic

detail and extreme values. Many statistical institute partners anonymize the microdata and implement technical measures of confidentiality protection before the data are entrusted to the project. When the OSI provides a sample that is also made available to others no additional protections are implemented by the project. Usually the project is not informed of the precise technical measures imposed on the data. Where the samples are unique, we impose the following technical protections (based on Thorogood 1999):

1. adopt sample density according to official norms or conventions (see Appendix A);
2. limit geographical detail by means of global recoding to administrative units with a minimum number of inhabitants. For some countries, this limit is as high as 100,000 and for others as low as 20,000. For the European project, NUTS3 is likely to be the lowest level of identifiable administrative geography, with the minimum threshold varying from 20,000 to 100,000 inhabitants according to the most recent census
3. top and bottom code unique categories of sensitive variables (identified by the OSI);
4. round, group, or band age as necessary;
5. suppress exact date of birth (age is always provided; for some year and even month of birth is also provided);
6. suppress detailed place of birth (<20/100,000 population);
7. suppress detailed place of residence, work, study, and migration (<20/100,000 population);
8. systematically “swap” (recode) place of enumeration for a fraction of households, inversely proportional to population size at the NUTS3 level; Data swapping protects confidentiality by introducing uncertainty about sensitive data values, yet maintains the strength of statistical inferences by preserving summary statistics (see Fienberg and McIntyre, 2004).
9. randomly order households within administrative units (NUTS3);
10. and, conduct a sensitivity analysis once these measures are imposed to determine what additional measures may be required.

These technical protections are implemented in the IPUMS-International database using software tools developed at the Minnesota Population Center as one of many steps followed in the processing of microdata integration. The tools were not designed for use by third parties. The remainder of this paper describes an experiment to implement IPUMS protocols using CPro/IMPS.

3 Implementing IPUMS-International technical measures of confidentiality control using CPro/IMPS.

3.1 The 1991 census microdata of Saint Lucia.

The Department of Statistics of Saint Lucia endorsed the IPUMS protocol in 2007 and entrusted microdata for the 1981 and 1991 censuses. The 1991 census of Saint Lucia enumerated 33,079 households and 133,308 persons. Integration has been delayed because of incomplete source documentation. It is expected that the integrated microdata will be

launched in 2010, although the launch may be postponed if adequate documentation is not forthcoming soon.

3.2 Drawing a systematic random sample

The first step in anonymizing the dataset is to draw a systematic random sample. Systematic random samples capitalize on low-level geographic sorting. By ensuring a representative geographic distribution of sampled cases, they are equivalent to extremely fine geographic stratification with proportional weighting. Since many economic and demographic characteristics are highly correlated with geographic location, this implicit stratification yields substantially greater precision than would a simple random sample of households. To the extent the strata used to draw a high precision sample are associated with the variables of interest (e.g., orphanhood, poverty, unemployment, etc.), the resulting estimates of these variables will have lower standard errors than what would have resulted had a simple random sample of records been drawn (Davern, et. al., 2009).

We draw a ten percent sample of households by computing a household serial number for each household in the dataset using CSPro. Then, a single digit random number was computed (3), and each household serial number ending in that digit was selected for inclusion in the sample (e.g., 3, 13, 23, ... 33,073). Only households, and individuals enumerated in them, with serial numbers ending in “3” are retained. The resulting sample contains 3,308 households and 13,405 persons.

3.3 Suppression of low-level geography regarding place of enumeration

The Saint Lucia microdata contain 5 geographical identifiers of place of residence: 8 regions, 12 districts, 28 towns, 99 settlements and 344 enumeration districts. Only 1 region, 1 district and 1 town were enumerated with more than 20,000 inhabitants. For this reason, all geography was suppressed. Researchers using this sample must analyze the population of Saint Lucia, as a geographic whole. Due to a technicality, apparently, CSPro does not permit the suppression of geographical identifiers. At this point the experiment was continued with IMPS.

3.4 Suppression of categorical attributes with fewer than 250 individuals.

The frequencies of every variable in the dataset were computed. Any count of less than 25 for individuals and 6 for households was identified and recoded as “other,” as follows:

1. Type of dwelling: suppress townhouse, barracks
2. Land occupation: suppress sharecrop
3. Type of ownership: suppress squatted, leased
4. Type of roof: suppress 5 categories
5. Wall material: suppress 5 categories
6. Water supply: suppress pubwell
7. Type of lighting: suppress gas
8. Ethnic origin: suppress Chinese, Portuguese, Syrian-Lebanese
9. Religion: suppress 6 categories
10. School, work mode of transport: bicycle
11. Type of school: technical institute, university
12. Number of hours worked last week: 5 hour groups; aggregate 70+
13. Pay period: suppress quarterly, annually
14. Occupation, industry, training code: reduce from 4 digits to 1/2/3

For continuous attributes, such as number of rooms, age, and the like, top and bottom coding was applied, as follows:

1. Number of rooms: 10+
2. Number of bedrooms: 7+
3. Number of radios: 4+
4. Number of tvs: 3+
5. Number of videos: 2+
6. Number of emigrants in dwelling: 2+
7. Age: 81+
8. Age at first child: ≤ 14
9. Age at first union: $\leq 14, 41+$
10. Age at last child: $\leq 14, 45+$
11. Number of school subjects: $\leq 3, \geq 7$
12. Income categories: 8+
13. Year of immigration: < 1948

Additional measures were applied, as follows:

1. Suppress: date of birth, precise place of birth, type of work wanted
2. Suppress detail: timing/place of migration
3. Suppress: 37 categories of country last lived

3.5 Assign a random household number and re-sort.

Once these steps were completed, a random serial number was applied to each household and the dataset re-sorted to insure that any implicit ordering of households is suppressed.

4 Discussion.

The experiment was a success in the sense that it is possible to confidentialize microdata using CSPro/IMPS. Nonetheless, what we learned is that these measures are not trivial to implement and the possibility of introducing error is considerable. We favour that confidentialization be performed by the IPUMS team for two reasons:

1. if errors occur, it is the responsibility of IPUMS, not the national statistical office
2. if the anonymization is unsuccessful, this too is the responsibility of IPUMS.

The IPUMS team has now anonymized more than 100 censuses to the complete satisfaction, to date, of every statistical agency partner.

Whether a sample dataset with “only” 13,000 individuals and a large number of suppressions will prove to be a success from the researcher’s point of view will be determined by use. Our reaction is that the technical confidentializing measures are so “heavy” that few researchers will wish to make use of this sample.

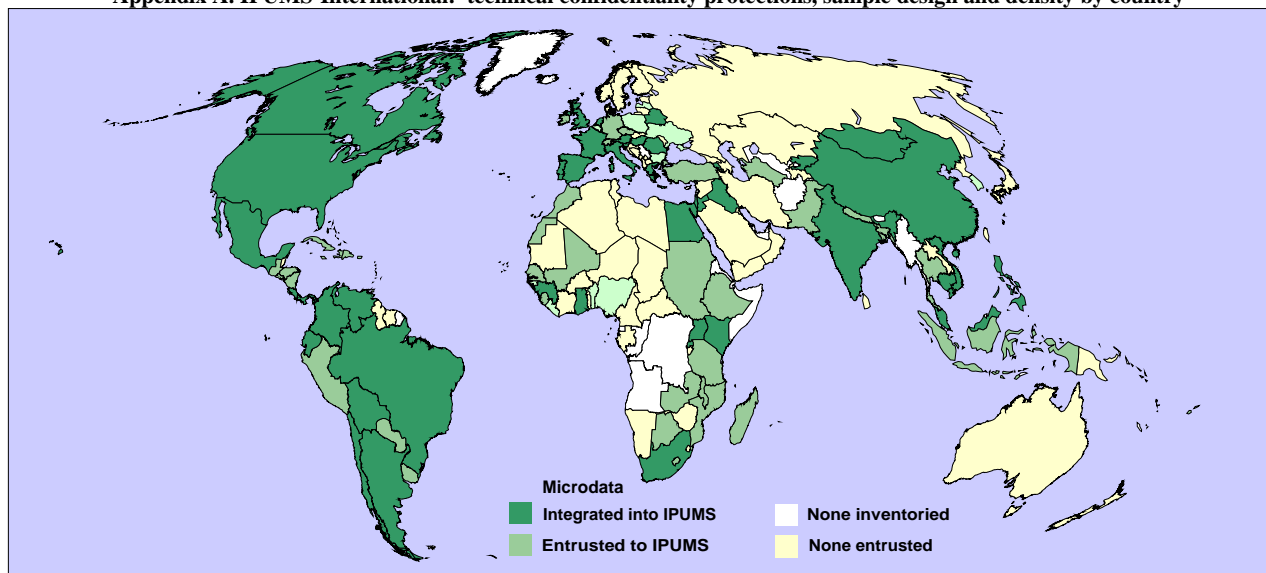
Indeed, we would propose that a second experiment be conducted in which the complete, but lightly anonymized, series of integrated datasets for the 2001, 1991 and 1981 censuses are made available in a secure data enclave perhaps at the Department of Statistics, the Minnesota Population Center, ECLAC and/or CARICOM for use by researchers. Then after a period of say five years, the experiment should be evaluated in terms of the number of users, the kinds of uses, and the knowledge acquired. At that point a determination could be

made on whether to continue with an anonymized, restricted access version disseminated over the web or only a lightly anonymized series available via a secure enclave.

References

- Anderson, Margo and Stephen E. Fienberg. (2001). "U.S. Census Confidentiality: Perception and Reality," International Statistical Institute Biennial Meeting (Seoul).
- Cook, Len. (2003). "Summary of Discussants' Main Points," in *Statistical Confidentiality and Access to Microdata: Proceedings of the Seminar Session of the 2003 Conference of European Statisticians*. Geneva, pp. 7-10.
- Dale, A. and Elliot, M. (2001). 'Proposals for 2001 SARS: An assessment of disclosure risk.' *Journal of the Royal Statistical Society, Series A*, 164, part 3, pp.427-447.
- Davern, Michael, Steven Ruggles, Tami Swenson, J. Trent Alexander and J. Michael Oakes. (2009) "Drawing statistical inferences from historical census data, 1850-1950," *Demography*, 46(3):589-603.
- King, John (2003). "Recent European Union Legislation for Research Access to Confidential Data: Implementation and Implications," in *Statistical Confidentiality and Access to Microdata: Proceedings of the Seminar Session of the 2003 Conference of European Statisticians*. Geneva, pp. 97-116.
- Lane, Julia (2003). "Uses of Microdata: Keynote Speech," in *Statistical Confidentiality and Access to Microdata: Proceedings of the Seminar Session of the 2003 Conference of European Statisticians*. Geneva, pp. 11- 20.
- McCaa, Robert and Albert Esteve. (2006). "IPUMS-Europe: Confidentiality measures for licensing and disseminating restricted access census microdata extracts to academic users," Monographs of official statistics: Work session on statistical data confidentiality. Luxembourg: Office for Official Publications of the European Communities, pp. 37-46.
- McCaa, Robert; Steven Ruggles, Michael Davern, Tami Swenson, and Krishna Mohan, Palipudi. (2006) "IPUMS-International high precision population census microdata samples: Balancing the privacy-quality tradeoff by means of restricted access extracts," Privacy in Statistical Databases (New York: Springer), 375-382.
- Thorogood, D. (1999). 'Statistical Confidentiality at the European Level.' Paper presented at: Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, Thessaloniki, March.

Appendix A. IPUMS-International: technical confidentiality protections, sample design and density by country



Note: **bold country** = Memorandum of Understanding signed with Regents of the University of Minnesota;
 IPUMS = systematic household sample(s) drawn to IPUMS specifications: every nth household stratified by enumeration district.
 Year = census conducted; **Bold year** = microdata survive; * = 100% microdata entrusted, where extant; m = microcensus; p = person sample

Sample density			Country	Technical Confidentiality Protections	Census decade				
10%	~5%	<=4%			2000s	1990s	1980s	1970s	1960s
<i>Integrated and Disseminating 2002-2009 (44 countries, 130 censuses, 76 million households and 279 million person records)</i>									
4			Argentina	IPUMS	2001	1991	1980	1970	1960
1			Armenia	IPUMS	2001		1989	1979	1970
4			Austria	IPUMS	2001	1991	1981	1971	1961
1			Belarus	IPUMS		1999	1989	1979	1970
3			*Bolivia	IPUMS	2001	1992		1976	
5			Brazil	IBGE	2001	1991	1980	1970	1960p
1			Cambodia	NIS		1998			1962
		4	Canada	STATSCAN	2001	1991-6	1981-6	1971-6	1961, 6
4		1	*Chile	IPUMS	2002	1992	1982	1970	1960p
		2	China	NBS	2000	1990	1982		1964
3		2	*Colombia	DANE/IPUMS	2005,6	1993	1985	1973	1964p
3	1		*Costa Rica	INE	2000		1984	1973	1963
4		1	*Ecuador	IPUMS	2001	1990	1982	1974	1962p
2			Egypt	IPUMS	2006	1996	1986	1976	1964
	6		France	INSEE	1999	1990	1982	1975	1968, 2
2			*Ghana	IPUMS	2000		1984	1970	
4			Greece	IPUMS	2001	1991	1981	1971	1961
2			*Guinea, Conakry	IPUMS		1996	1983		1960
	4		Hungary	IPUMS	2001	1990	1980	1970	
		5	India (microcensuses)	NSSO	2005m	1993,9m	1983,7m		
1			*Iraq	IPUMS		1997	1987	1977	1967
3			Israel	CBS		1995	1983	1972	1961,7
	1		Italy	ISTAT	2001	1991	1981	1971	1961
1			Jordan	IPUMS	2004	1994	1979		
	3		Kenya	IPUMS	1999	1989	1979	1969	
1			Kyrgyz Republic	IPUMS		1999			
		4	Malaysia	IPUMS	2000	1991	1980	1970	1960
3		3	Mexico	INEGI	2000,5	1990,5	1980	1970	1960p

2			*Mongolia	IPUMS	2000		1989	1979	1970
		3	Netherlands	CBS	2001m			1971	1960
1			Palestine	CBS		1997			
5			*Panama	IPUMS	2000	1990	1980	1970	1960
3			*Philippines	IPUMS	2000	1990	1980	1970	1960p
	3		Portugal	IPUMS	2001	1991	1981	1970	1960
3			Romania	IPUMS	2001	1992		1977	1965
2			*Rwanda	IPUMS	2002	1991			
1			Slovenia	IPUMS	2001	1991	1981		
3			South Africa	StatsSA	2001,7	1996-1	1985-0	1970	1960
	3		Spain	INE	2001	1991	1981	1970	1960
2			*Uganda	IPUMS	2002	1991	1980		1969
		2	United Kingdom	ONS	2001p	1991	1981	1971	1966,1
	6		United States	USCB	2000,5	1990	1980	1970	1960
4			*Venezuela	IPUMS	2001	1990	1981	1971	1961
	2		Vietnam	IPUMS		1999	1989	1979	
Europe (21 countries, 79 censuses—including samples for 11 countries released above)									
			Bulgaria	-	2001	1992	1985	1975	1965
			Belgium (negotiating)	-	2001	1991	1981	1970	1961
	2		Czech Republic	IPUMS	2001	1991	1980	1970	1961
1			Germany	FSO	2001m	1991m	1981-7	1970-1	1961
			Ireland (in process)	IPUMS/CSO	2002, 6	1991, 6	1981, 6	1971, 9	
			Latvia (negotiating)	-	2000		1989	1979	
			Poland (negotiating)	-	2001		1988	1970-8	1960
			Russia (negotiating)	-	2002		1989	1979	1970
	4		Switzerland (2010 launch)	IPUMS	2000	1990	1980	1970	1960
3			Turkey (in process)	TurkSTAT	2000	1990	1980-5	1970-5	1960, 5
1			Ukraine (in process)	IPUMS	2001		1989	1979	1970
North America and the Caribbean (15 countries, 48 censuses—including samples for 5 countries released above)									
1			Cuba (2010 launch)	IPUMS	2002		1981	1970	
1	1	2	*Dominican Republic	IPUMS	2003	1993	1981	1970	1960p
1			*El Salvador	IPUMS	2007	1992		1971	1961
2		3	*Guatemala	IPUMS	2002	1994	1981	1973	1964
3			*Jamaica	IPUMS	2001	1991	1982	1970	1960
2			*Haiti	IPUMS	2003		1982	1971	
3		1	*Honduras	IPUMS	2000		1988	1974	1961
2		1	*Nicaragua	IPUMS	2005	1995		1971	1963
	4		Puerto Rico	USCB	2000	1990	1980	1970	1960
2			*Saint Lucia (2010 launch)	IPUMS	2001	1991	1980	1970	1960
			Trinidad and Tobago (negotiating)		2000	1990	1980	1970	1960
South America (9 countries, 40 censuses—including samples for 6 countries released above)									
4		1	*Paraguay	IPUMS	2002	1992	1982	1972	1962
2			Peru (2010 launch)	IPUMS	2007	1993	1981?	1972	1961
4			*Uruguay	IPUMS		1996	1985	1975	1963
Africa (22 countries, 51 censuses—including samples for 7 countries released above)									
			Benin (negotiating)		2002	1990		1979?	
3			*Botswana	IPUMS	2001	1991	1981	1971	1964
			Cameroon (negotiating)		2002		1987	1976?	
			Central African Rep. (negotiating)		2003		1988	1974	
			Chad (negotiating)		2008	1993	1989		1969
2			*Ethiopia	IPUMS	2007	1994	1984		
			Guinea-Bissau (negotiating)	IPUMS	2009	1991		1979	

			Lesotho (in process)	IPUMS	2006	1996	1986	1976	1966
1			*Madagascar	IPUMS		1993			
2			*Malawi	IPUMS	2008	1997	1987	1977	1967
3			*Mali (2010 launch)	IPUMS		1998	1987	1976	
2			*Mauritius	IPUMS	2000	1990	1983	1972	1962
			Morocco (in process)	IPUMS	2004	1994	1982	1971?	1960?
1			Mozambique	IPUMS	2007	1997	1980		
2			Niger		2001		1987	1977	
			Nigeria (negotiating)	NatPopCom	2006	1991		1973	1963
3			*Senegal (2010 launch)	IPUMS	2002		1988	1976	
1			*Sierra Leone	IPUMS	2004		1985?	1974	1963
3			*Sudan	IPUMS	2008	1993	1983	1973	
2			*Tanzania (2010 launch)	IPUMS	2002		1988	1978	1967
			Togo (negotiating)		2009	1993	1981	1970	1961
2			*Zambia	IPUMS	2000	1990	1980	1969	1963
<i>Asia and Oceania (22 countries, 52 censuses—including samples for 13 countries released above)</i>									
1		1	*Bangladesh	IPUMS	2001	1991	1981	1974	1961
3			*Fiji Islands	IPUMS	2007	1996	1986	1976	1966
7			Indonesia	BP/IPUMS	2000	1990	1980	1971	1961
			Korea, Republic of (negotiating)	KNSO	2005, 0	1995, 0	1985, 0	1975, 0	1960, 6
1			Nepal	CBS	2001	1991?	1981?	1971	1961
3			*Pakistan (2010 launch)	IPUMS		1998	1981	1973	1961
		4	Thailand (2010 launch)	NSO	2000	1990	1980	1970	1960
1			Turkmenistan	IPUMS		1995	1989	1979	1970

Appendix B

Letter of Understanding
**Integrated Public Use Microdata Series International
and Statistics Austria**

Purpose. The purpose of this letter is to specify the terms and conditions under which metadata and microdata provided by **Statistics Austria** shall be distributed by **Integrated Public Use Microdata Series International** of the University of Minnesota.

1. Ownership. **Statistics Austria** is the owner and licensee of the intellectual property rights (including copyright) in the metadata and microdata supplied to the University of Minnesota to be distributed by **Integrated Public Use Microdata Series International**.
2. Use. These data are provided for the exclusive purposes of teaching, academic research and publishing, and may not be used for any other purposes without the explicit written approval, in advance, of **Statistics Austria**.
3. Authorization. To access or obtain copies of integrated microdata of Austria from **Integrated Public Use Microdata Series International**, a prospective user must first submit an electronic authorization form identifying the user (i.e., principal investigator) by name, electronic address, and institution. The principal investigator must state the purpose of the proposed project and agree to abide by the regulations contained herein. Once a project is approved, a password will be issued and data may be acquired from servers or other electronic dissemination media maintained by **Integrated Public Use Microdata Series International**, **Statistics Austria**, or other authorized distributors. Once approved, the user is licensed to acquire integrated metadata and microdata of Austria from **Integrated Public Use Microdata Series International** or other authorized distributors. No titles or other rights are conveyed to the user.
4. Restriction Users are prohibited from using data acquired from the **Integrated Public Use Microdata Series International** or other authorized distributors in the pursuit of any commercial or income-generating venture either privately, or otherwise.
5. Confidentiality. Users will maintain the absolute confidentiality of persons and households. Any attempt to ascertain the identity of a person, family, household, dwelling, organization, business or other entity from the

microdata is strictly prohibited. Alleging that a person or any other entity has been identified in these data is also prohibited.

6. Security. Users will implement security measures to prevent unauthorized access to microdata acquired from **Integrated Public Use Microdata Series International** or its partners.
7. Publication. The publishing of data and analysis resulting from research using metadata or microdata of Austria is permitted in communications such as scholarly papers, journals and the like. The authors of these communications are required to cite **Statistics Austria and Integrated Public Use Microdata Series International** as the sources of the data of Austria, and to indicate that the results and views expressed are those of the author/user.
8. Sharing. **Integrated Public Use Microdata Series International** will provide electronic copies to **Statistics Austria** of documentation and data related to its integrated microdata as well as timely reports of authorized users.
9. Violations. Violation of this agreement may lead to professional censure and/or civil prosecution.
10. Jurisdiction. Disagreements which may arise shall be settled by means of conciliation, transaction and friendly composition. Should a settlement by these means prove impossible, a Tribunal of Settlement shall be convened which will rule upon the matter under law. This Tribunal shall be composed of an (1) arbitrator, which shall be elected by lot from the list of Arbitrators of the Chamber of Commerce of Paris. This agreement shall be governed by, and construed in accordance with, generally accepted principles of International Law.

Date: December 20, 2001

Signed: *K. M. Y.*
Regents of the University of Minnesota
By: Kevin McKoskey, Grants Manager, Sponsored Projects Administration

Date: January 28, 2002

Signed: *[Signature]*



