IPUMS-International Statistical Disclosure Controls: 159 Census Microdata Samples in Dissemination, 100+ in Preparation

Robert McCaa*, Steven Ruggles, and Matt Sobek

Minnesota Population Center, 50 Willey Hall
Minneapolis MN 55455 USA
rmccaa@umn.edu

Abstract. In the last decade, a revolution has occurred in access to census microdata for social and behavioral research. More than 325 million person records (55 countries, 159 samples) representing two-thirds of the world's population are now readily available to bona fide researchers from the IPUMS-International website: www.ipums.org/international hosted by the Minnesota Population Center. Confidentialized extracts are disseminated on a restricted access basis at no cost to bona fide researchers. Over the next five years, from the microdata already entrusted by National Statistical Office-owners, the database will encompass more than 80 percent of the world's population (85 countries, ~100 additional datasets) with priority given to samples from the 2010 round of censuses. A profile of the most frequently used samples and variables is described from 64,248 requests for microdata extracts. The development of privacy protection standards by National Statistical Offices, international organizations and academic experts is fundamental to eliciting world-wide cooperation and, thus, to the success of the IPUMS initiative. This paper summarizes the legal, administrative and technical underpinnings of the project, including statistical disclosure controls, as well as the conclusions of a lengthy on-site review by the former Australian Statistician, Mr. Dennis Trewin.

Keywords: Census microdata samples, data privacy, data dissemination, IPUMS-International.

1 Introduction

A revolution occurred in access to population census microdata for social and behavioral research in the first decade of the twenty-first century. The most successful initiative, with the cooperation of some 85 National Statistical Agencies world-wide, is the IPUMS-International project led by the Minnesota Population Center (MPC, Figure 1).

At this writing, datasets for 55 countries—159 anonymized, integrated samples totaling 325,430,447 person records—are available to registered researchers at no cost via the IPUMS-International web-site (Table 1). From the 250-odd datasets already

* Corresponding author.

J. Domingo-Ferrer and E. Magkos (Eds.): PSD 2010, LNCS 6344, pp. 74–84, 2010.

entrusted to the project, the number of countries represented is likely to increase to 85 or more over the next five years, and the number of datasets to some 250. Twenty to thirty samples are integrated into the database each year. 2010 round census data will be assigned the highest priority for integration, as they become available. For each country, an effort is made to construct a series of samples for all censuses for which microdata survive. Of the 159 samples currently in the database, 37 are from the 2000 round compared with 44 for the 1990s, 37 for the 1980s, 27 from the 1970s and only 13 from the 1960s. High precision household samples with a density of five percent or more number 128. Of the 30 lower precision samples, many consist of all the surviving microdata for the respective census. Notable exceptions are the samples for four censuses of Canada, and two each for China, the Netherlands, and the United Kingdom. The Chinese household samples, with a density of only one percent, number over ten million person records each.

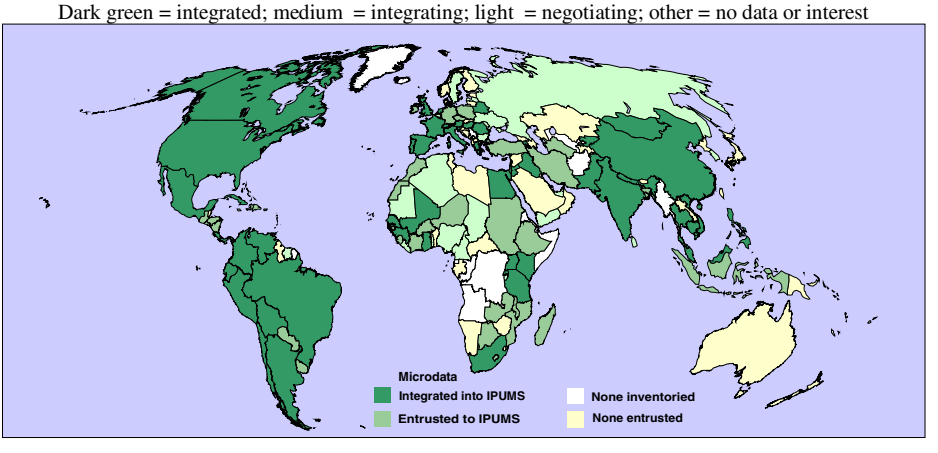

Fig. 1. The IPUMS-International World Map

3,750 researchers, representing 84 countries, are approved for access to the microdata from the IPUMS-International web-site (usage statistics as of July 1, 2010). 64,248 extracts, excluding those by MPC personnel, have been made, totaling 731,531 integrated variables extracted. Ten variables account for one-quarter of the usage: educational attainment, employment status, age, marital status, person weight, relationship to head (or reference person), sex, person number, sample identifier, and class of worker. The microdata of a mere seven countries account for one-half of the extracts: Mexico, USA, Brazil, Colombia, France, Chile and Argentina. The striking preference for American microdata is due to the fact that samples for these countries were among the first integrated into the database. Moreover, these countries have long series of censuses with extant microdata stretching back to the 1960s. Finally, the samples are rich, with at least 50 person variables and 10 household or dwelling variables.

At the PSD2006, we laid out the statistical disclosure controls to protect the privacy of persons, households and other entities developed on the first 47 samples integrated into the IPUMS-International database [1]. In this paper we describe how

the legal, administrative and technical procedures are being implemented to protect privacy and statistical confidentiality and, thus, to facilitate access to this massive trove of data. Restricting access to trusted users is the key to our success. To date, there have been no allegations of misuse of IPUMS-International census microdata extracts. On the contrary, what is most remarkable is the substantial usage by researchers, given the fact that the usability of "public use" microdata is sometimes deemed "limited" [2]. Despite the "PU" in the IPUMS acronym, "RA," "TU," or "SA" might be more appropriate because the data are disseminated as "restricted access," "trusted user," or "scientific access" files [3, 1].

2 Thwarting Intruders

The casual intruder (and casual user) is readily thwarted by the IPUMS-International registration form and policy statements. At 1,100 words, the IPUMS form is considerably shorter than the 5,830 words that constitute the FACEBOOK privacy policy, but, unlike FACEBOOK (where most registrants click "I agree" without reading the small print), the IPUMS registration requires the applicant to agree to each of eight detailed conditions of use. Failure to agree to even one condition results in an automatic rejection of the registration. The successful applicant provides not only personal details, but also must identify institutional affiliation, including name, official email address and phone number, web-link identifying affiliation, name and email of supervisor, the name, title and other pertinent information of any grant used to conduct the research and, most importantly, the name of "an Institutional Review Board (IRB), or Office for Human Subject Protections, Professional Conduct or similar committee." Applications that omit this information are reviewed, but a positive decision is delayed until bona-fides are explained and verified. Perhaps the biggest obstacle for a successful application is the research project description, which is carefully scrutinized to confirm that access to the database is needed for the proposed research. A researcher may possess adequate technical and professional qualifications, but if there is no research need for the microdata, access will be denied. Approximately one-third of completed applications are denied. In complete registrations—those that are begun and but never submitted—go uncounted, but it is likely that their number is not inconsiderable.

The rogue intruder—armed with the appropriate bona-fides but with malevolent intent—faces legal and institutional sanctions as well as substantial technical obstacles. If the violation occurs in the United States, the intruder risks civil prosecution with a maximum fine of US$250,000 and/or three years imprisonment. Elsewhere, since the laws of the country in which the violation occurs would apply, the discretion to prosecute would rest with the National Statistical Authority. The legal counsel of the University of Minnesota is committed to providing vigorous legal assistance. This threat of legal action is probably less a deterrent than institutional and professional sanctions. The IPUMS Case Study in [4, Annex 1.23] describes the sanctions as follows:

1. "sanctions against both the individual and the institution with which the individual is associated (e.g., University, international organization) [would be imposed];

2. "denial of access would immediately be invoked against the individual and his/her institution and would continue until corrective measures were deemed to be sufficient by the University of Minnesota and the National Statistical Office whose data were violated. If the institution where the breach occurred was the recipient of a grant from the National Institutes of Health of the United States, each researcher at the institution could be required to undergo Human Subjects Protection training and re-certification before access was re-instituted for individuals at that institution."

Commercial researchers are prohibited from accessing the data. Some petition for access, but are denied because of the restriction to non-commercial users. None seek access to identify individuals. Instead, commercial users often require population statistics that are not readily obtainable elsewhere, such as to compute weights or expansion factors for specific population sub-groups. There is no interest by commercial or other entities in linking confidentialized population census samples to other sources because much more valuable data are readily available elsewhere. Then too, leaving aside the difficulties of constructing successful links, sample microdata are too ordinary to excite the slightest interest for the purpose of linking.

3 Statistical Disclosure Controls

Threats to privacy and statistical confidentiality by intruders have long provided the rationale for National Statistical Offices to simply deny access to census microdata, regardless of the professional qualifications and scientific needs of would-be researchers. IPUMS International is successful in overcoming these objections because our procedures are designed to thwart intruders, first, by screening to permit trusted researchers to use the data while denying access to potential intruders; second, by erecting strong sanctions against misuse; and, third, by imposing stringent statistical disclosure controls. We endorse the standard of the Office of National Statistics (UK) [5] that statistical controls should be such that it "would take a disproportionate amount of time, effort and expertise for an intruder to identify a statistical unit to others, or to reveal information about that unit not already in the public domain." Population census variables are mundane. Census attributes are relatively crude in comparison to the details available in employment or health surveys. IPUMS-International suppresses variables considered to be sensitive by the official statistical agency, but to date, there has been only one such request: "tribe" for a census from an African country, where ethnic violence is a grave concern.

Census operations produce a considerable amount of data that is less than perfect. Editing and imputation are necessary to produce coherent datasets. Few statistical agencies report the details. The rich samples of the 2001 population census microdata of the United Kingdom make it possible to assess the degree of imputation and of perturbation—the introduction of intentional error to protect confidentiality in the data. The ONS relied upon the Post Randomization Method (PRAM) to produce the 2001 Licenses Individual SAR. De Kort and Wathan [6] compared the Individual Controlled Access Microdata Sample (not perturbed) with the Licensed Individual SAR (perturbed—note that this is the sample integrated into IPUMS-International)

and discovered that the relative frequency of imputation was several times greater than perturbation. Of the twelve variables analyzed, the authors found that for "Social Grade of Reference Person" 15% of attributes were imputed, versus 2% perturbed. For "Age" imputation and perturbation were roughly the same at 1%. The frequency of perturbations was typically less than one percent, whereas imputations for five variables were 5% or more. For researchers inclined to ignore the imputed data, de Kort and Wathan warn that "raw data are not necessarily to be preferred." The ONS-UK is to be lauded for producing flags to indicate imputation for every variable in its samples. Flags empower researchers to gauge the effects of imputation and editing and take appropriate action.

Purdam and Elliott [7] carried the analysis a step farther to assess the effects of perturbation on published analyses. Thanks to the ability to replicate certified samples such as the SARS and the CAMS, replication of research results can be accomplished with a degree of confidence. Their findings are disconcerting for researchers: "disclosure control measures had a significant impact on the usability of the data (analytical completeness) and on the accuracy of the data in relation to the findings reached when the data were used in analyses (analytical validity)."

As in the case of the 2001 SARS, a few statistical agencies entrust samples that have already been subjected to privacy protections. Sometimes these go seriously awry, as in the case of the United States PUMS [8]. Beginning with datasets from 2000 through 2008, serious errors were introduced into the public use files for males and females aged 65 years and over. Due to a programming error, statistical disclosure controls corrupted age attributes so that published distributions differed from those computed from microdata samples by as much as 15%. Three series of microdata samples were corrupted: the 2000 census, the American Community Survey (2003-6), and the Current Population Survey (2004-9) [9]. Despite the uproar in the media only one dataset was corrected, and some researchers fear that the correction may actually make matters worse.

Most statistical offices entrust "raw" microdata to the IPUMS-International project (not truly raw because names and addresses are stripped out thereby anonymizing the data before shipment). In Table 1, these instances are identified by "IPUMS" in the column headed "Confidentiality Protocols". In such cases, we apply a series of straight-forward SDC measures. First, the data are anonymized by suppressing any names, addresses, or precise geographic identifiers. Second, a sample is drawn so that researchers have access to only a minor fraction of the complete dataset. Third, additional disclosure protections are imposed on the sample, variable-by-variable and code-by-code. Finally, a small fraction of households is swapped across geographic boundaries.

Our procedures are summarized in a contract with one of our statistical agency partners, as follows:

> (1) Detailed geographic codes will be suppressed.
> (2) Any geographical unit with fewer than 20,000 individuals will be aggregated to the next highest geographical unit.

Table 1. Microdatasets entrusted, confidentiality protocols and sample densities

Sample density			Country	Confidentiality protocols	Census decade				
10%+	~5%	<=4%			2000s	1990s	1980s	1970s	1960s
Integrated and Disseminating 2002-2010: 55 countries, 159 censuses, 87 million households and 325 million person records									
4			**Argentina**	INDEC	**2001**	**1991**	**1980**	**1970**	1960
1			**Armenia**	SCS	**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**
2			**Cambodia**	IPUMS	**2008§**	**1998**			1962
		4	**Canada**	STATSCAN	**2001p**	**1991p**-6	**1981p**-6	**1971p**	1961
4		1	***Chile**	IPUMS	**2002**	**1992**	**1982**	**1970**	**1960p**
		2	**China**	NBS	**2000**	**1990**	**1982**		1964
3		2	***Colombia**	IPUMS	**2005**	**1993**	**1985**	**1973**	**1964p**
3	1		***Costa Rica**	IPUMS	**2000**		**1984**	**1973**	**1963**
1			**Cuba**	IPUMS	**2002**		1981	1970	
4		1	***Ecuador**	IPUMS	**2001**	**1990**	**1982**	**1974**	**1962p**
3			**Egypt**	IPUMS	**2006§**	**1996**	**1986**	1976	1964
1	6		**France**	INSEE	**2006§**	**1990,9**	**1982**	**1975**	**1968,2**
2			***Ghana**	IPUMS	**2000**		**1984**	1970	
4			**Greece**	IPUMS	**2001**	**1991**	**1981**	**1971**	1961
2			***Guinea, C.**	IPUMS		**1996**	**1983**		1960
	4		**Hungary**	CSO	**2001**	**1990**	**1980**	**1970**	
		5	**India**	NSSO	**2005m**	**1993,9m**	**1983,7m**		
1			***Iraq**	IPUMS		**1997**	1987	1977	1967
5			**Israel**	CBS	**2008**	**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 Rep.**	IPUMS	**2009**	**1999**	1989		
		4	**Malaysia**	IPUMS	**2000**	**1991**	**1980**	**1970**	1960
3			***Mali**	IPUMS	**2008**	**1998**	**1987**	**1976**	
4		3	**Mexico**	INEGI	**2000,5**	**1990,5**	**1980**	**1970**	**1960p**
2			***Mongolia**	IPUMS	**2000**		**1989**	1979	**1956**
1			**Nepal**	CBS	**2001**	1991?	1981	1971	1961
		3	**Netherlands**	CBS	**2001pm**			**1971p**	**1960p**
2			**Palestine**	CBS	**2007§**	**1997**			
3			***Pakistan**	IPUMS		**1998**	**1981**	**1973**	1961
5			***Panama**	IPUMS	**2000**	**1990**	**1980**	**1970**	**1960**

Table 1. (*continued*)

2			**Peru**	IPUMS	**2007**	**1993**	**1981**	1972	1961
3			***Philippines**	IPUMS	**2000**	**1990**	**1980**	**1970**	**1960p**
	3		**Portugal**	INE	**2001**	**1991**	**1981**	1970	1960
	4		**Puerto Rico**	USCB	**2000**	**1990**	**1980**	**1970**	1960
3			**Romania**	IPUMS	**2001**	**1992**		**1977**	1965
2			***Rwanda**	IPUMS	**2002**	**1991**			
2			***Saint Lucia**	IPUMS	**2001**	**1991**	**1980**	1970	1960
3			***Senegal**	IPUMS	**2002**		**1988**	**1976**	
1			**Slovenia**	SORS	**2001**	1991	1981		
6		1	**South Africa**	StatsSA	**2001,7**	**1996-1**	**1985-0**	**1970**	1960
	3		**Spain**	INE	**2001**	**1991**	**1981**	1970	1960
	4		**Switzerland**	IPUMS	**2000**	**1990**	**1980**	**1970**	1960
2			***Tanzania**	IPUMS	**2002**		**1988**	1978	1967
		4	**Thailand**	NSO	**2000**	**1990**	**1980**	**1970**	1960
2			***Uganda**	IPUMS	**2002**	**1991**	1980		1969
		2	**United King.**	ONS	**2001p**	**1991**	**1981**	**1971**	**1966,1**
	6		**USA**	USCB	**2000,5**	**1990**	**1980**	**1970**	**1960**
4			***Venezuela**	IPUMS	**2001**	**1990**	**1981**	**1971**	1961
	2		**Vietnam**	IPUMS	**2009**	**1999**	**1989**	1979	
Europe									
			Albania	-	**2001**	1989	1979	1969	1960
			Bulgaria	-	**2001**	**1992**	**1985**	1975	1965
			Belgium	-	**2001**	**1991**	**1981**	**1970**	1961
	2		**Czech Rep.**	IPUMS	**2001**	**1991**	**1980**	**1970**	1961
			Estonia	-	**2000**	**1989**	1979	1970	1959
4			**Germany §**	FSO	**2001m**	**1991m**	**1981-7**	**1970,1**	1961
8			**Ireland §**	CSO	**2002, 6**	**1991, 6**	**1981, 6**	**1971,9**	
			Latvia	-	**2000**		**1989**	1979	
			Poland	-	**2001**	**1995**	**1988**	**1970,8**	1960
			Russia	-	**2002**		**1989**	1979	1970
			Turkey	TurkSTAT	**2000**	**1990**	**1985, 0**	1975,0	1960
			Ukraine	IPUMS	**2001**		1989	1979	1970
North America and the Caribbean									
1	1	2	***DominicanR.**	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
South America									

Table 1. (*continued*)

4		1	***Paraguay**	IPUMS	**2002**	**1992**	**1982**	**1972**	**1962**
4			***Uruguay**	IPUMS		**1996**	**1985**	**1975**	**1963**
Africa									
			Benin		**2002**	**1990**		**1979**	
3			***Botswana**	IPUMS	**2001**	**1991**	**1981**	1971	1964
			Burkina Faso		**2006**	**1996**	**1985**	1975	
			Burundi		**2008**	1990?	1979?	1970?	
			Cameroon		**2005**		**`1987**	**1976**	
			Cape Verde	IPUMS	**2000**	**1990**	1980	1970	1960
			Central Afr. R.		**2003**		**1988**	1974	
			Chad		**2008**	**1993**	**1989**		1969
			Côte d'Ivoire		**2009**	**1998**	**1988**	1975	
2			***Ethiopia**	IPUMS	**2007**	**1994**	**1984**		
			Gabon		**2003**	**1993**	1980		1969
			Guinea-Bis.	IPUMS	**2009**	**1991**		1979	
2			**Lesotho**	IPUMS	**2006**	**1996**	**1986**	**1976**	1966
			Liberia		**2008**		1984	**1974**	
1			***Madagascar**	IPUMS		**1993**			
2			***Malawi**	IPUMS	**2008**	**1997**	**1987**	1977	1967
			Mauritania		2001		1988	1977	
2			***Mauritius**	IPUMS	**2000**	**1990**	**1983**	1972	1962
	3		**Morocco**	IPUMS	**2004**	**1994**	**1982**	1971	1960
1			**Mozambique**	IPUMS	**2007**	**1997**	1980		
2			***Niger**	IPUMS	**2001**		**1987**	**1977**	
			Nigeria	NatPopCom	**2006**	**1991**		1973	1963
1			***Sierra L.§**	IPUMS	**2004**		1985	1974	1963
3			***Sudan**	IPUMS	**2008**	**1993**	**1983**	**1973**	
			Togo		2010		**1981**	**1970**	**1958**
2			***Zambia**	IPUMS	**2000**	**1990**	1980	1969	1963
Asia and Oceania									
1		1	***Bangladesh**	IPUMS	**2001**	**1991**	**1981**	1974	1961
5			***Fiji Islands**	IPUMS	**2007**	**1996**	**1986**	**1976**	**1966**
8			**Indonesia §**	BP/IPUMS	**2000, 5**	**1990, 5**	**1980, 5**	**1971,6**	1961
1			Iran §	SCI	**2006**	**1996**	**1986**	**1976**	**1966**
			Korea, Rep.	KOSTAT	**2005, 0**	**1995, 0**	**1985, 0**	**1975**	**1960,6**
			Sri Lanka	DCS	**2001**		**1981**	**1971**	1960
1			**Turkmenistan**	IPUMS		**1995**	1989	1979	1970
			United A. E.		**2005**	1995	1985, 0	1975	1968

bold country = Memorandum of Understanding with Regents of the University of Minnesota;
IPUMS = systematic household sample: every n^{th} household stratified by enumeration district; confidentiality specifications (see text).
Year = census conducted; **Bold year** = microdata survive; §= samples for launch in 2011
*** =** 100% microdata entrusted, where extant; m = microcensus; p = person sample

(3) Any social characteristic (categorical variables such as place of birth, occupation, etc.) with fewer than 250 individuals in the population will be re-coded as missing, suppressed or aggregated.
(4) Continuous variables (such as income, size of rooms, etc.) will be top/bottom coded to prevent identification of individuals or other entities with unique characteristics.

(5) The geographical identifiers of a fraction of households will be recoded to a different geographical unit so that any allegation that an individual or other entity is positively identified is false. Swapping of individuals and households across geographical boundaries (that is, editing the geographical identifiers of a small fraction of individuals and households to one that is false) introduces uncertainty into any attempts at identification.

The thresholds in this contract are those usually authorized by most statistical agencies that entrust "raw" microdata to the IPUMS-International project. Nonetheless, the thresholds may be adjusted at the request of the National Statistical Office-owner. For example, in the case of France, place of residence is limited to 22 regions. The smallest region has a population exceeding 80,000 in the 1990 census (sample n > 4,000). The population count for any identifiable single year of age is >2000. For any identifiable country of citizenship the threshold is also >2000. Recently, INSEE, the French national statistical authority, began a reconsideration of these thresholds, particularly for the historical datasets that are now more than a decade old. A comprehensive assessment is being prepared to develop lower thresholds so that detailed attributes may be made available for several key variables, such as place of residence, country of birth, occupation, and industry.

During the process of confidentializing international microdata at the Minnesota Population Center, all work is performed by senior staff who have taken the appropriate training and signed official statements to protect the data. Once the statistical disclosure controls are in place, junior staff may begin integrating the microdata. Original source microdata, whether "raw" or confidentialized, are encrypted and archived off-line and thus are preserved in case there are questions about errors introduced by the SDCs. To date there have been no queries about the validity of any IPUMS samples. Errors have been discovered—some due to the integration and others in the source microdata, but none attributed to the process of confidentializing samples.

4 An Evaluation of Security

Statistical data privacy is more than simply SDC. All procedures and processes associated with the microdata must be secure and must be perceived as such by the public. With the large stock of international microdata archived at the Minnesota Population Center, protecting these assets is a major concern of the Center, the University, and official statistical agencies, international as well as national—whether associated with the project or not.

The first author of this paper invited Mr. Dennis Trewin to conduct an on-site inspection of the IPUMS-International facilities and procedures [10]. Mr. Trewin is well qualified for the undertaking. As Australian Statistician one of his achievements was the extension of microdata services to researchers while maintaining public trust and abiding by the conditions outlined in the legislation of Australia governing microdata access. He also chaired the Conference of European Statisticians Task Force on Guidelines and Core Principles of Confidentiality and Microdata Access. The Guidelines were adopted by the CES plenary session in June 2006 and published

as Managing Statistical Confidentiality & Microdata Access: Principles and Guidelines of Good Practice [4]. Not surprisingly, Mr. Trewin is noted for his critical acumen and professional probity. The terms of reference for his review was to identify weaknesses and lapses so that IPUMS-International could improve its procedures and thereby provide an additional layer of protection for official statisticians as well as trust for the public. Mr. Trewin's review included attendance at a side-meeting of the IPUMS-International at the 2007 International Statistical Institute in Lisbon as well as bilateral interviews with official statisticians. His report concludes:

> Without question IPUMS International meets the four Core Principles outlined in CES (2007). It is cited in CES (2007) as a Case Study of good practice. This review confirms its status as good practice for Data Repositories. Indeed it is likely to provide the best practice for a Data Repository of international statistical data sets.
> …
> The security of the computing environment used by IPUMS-International is first class and appears to be of the standard of the best statistical offices.
> …
> IPUMS-International is a valuable and trustworthy microdata service. It meets the fundamental principles of good practice with respect to confidentiality and microdata. Consequently, my recommendations are limited.

Mr. Trewin's recommendations to IPUMS-International for enhancing security and data confidentiality are, indeed, "limited". Nonetheless all have been or are being implemented, including his recommendation that "checks should be made of published outputs from time to time to provide some assessment of whether there has been any inappropriate use of microdata (e.g., reference to individual cases)."

5 Conclusion

The goals of IPUMS-International are, first, to recover census microdata that are at risk of loss; second, to archive microdata; and third, to disseminate confidentialized, custom-tailored, integrated extracts to researchers world-wide at no cost. In the first decade of operations, more than 3,700 researchers registered for access, a vast trove of microdata was entrusted to the Minnesota Population Center, and 159 datasets underwent the arduous process of confidentializing the microdata and integrating both data and documentation into the IPUMS-International system. Over the next five years, an additional hundred datasets are likely to be integrated into the IPUMS-International system. Academics and policy makers needing census microdata for research are invited to visit the project website: www.ipums.org/international.

Acknowledgements. Funded in part by the National Science Foundation of the United States, Grant Nos. SES-0433654 and 0851414; National Institutes of Health, Grant Nos. R01HD047283 and R01 HD044154.

References

1. McCaa, R., Ruggles, S., Davern, M., Swenson, T., Mohan Palipudi, K.: IPUMS-International High Precision Population Census Microdata Samples: Balancing the Privacy-Quality Tradeoff by Means of Restricted Access Extracts. In: Domingo-Ferrer, J., Franconi, L. (eds.) PSD 2006. LNCS, vol. 4302, pp. 375–382. Springer, Heidelberg (2006)
2. Museux, J.-M., Peeters, M., Santos, M.-J.: Legal, Political and Methodological Issues in Confidentiality in the European Statistical System. In: Domingo-Ferrer, J., Saygın, Y. (eds.) PSD 2008. LNCS, vol. 5262, pp. 324–334. Springer, Heidelberg (2008)
3. McCaa, R., Esteve, A.: IPUMS-Europe: Confidentiality measures for licensing and disseminating restricted-access census microdata extracts to academic users. In: Monographs of official statistics: Work session on statistical data confidentiality, pp. 37–46. Office for Official Publications of the European Communities, Luxembourg (2006)
4. United Nations Economic Commission for Europe. Conference of European Statisticians. Managing Statistical Confidentiality & Microdata Access: Principles and Guidelines of Good Practice. Geneva: United Nations (2007), http://www.unece.org/stats/publications/Managing.statistical.confidentialityandmicrodataaccess.pdf
5. United Kingdom. Office of National Statistics. National Statistics Code of Practice Protocol on Data Access and Confidentiality. HMSO, London (2004)
6. De Kort, S., Wathan, J.: Guide to Imputation and Perturbation in the Samples of Anonymised Records (2009) (unpublished)
7. Purdam, K., Elliot, M.: A case study of the impact of statistical disclosure control on data quality in the individual UK Samples of Anonymised Records. Environment and Planning A 39(5), 1101–1118 (2007)
8. Alexander, J.T., Davern, M., Stevenson, B.: Inaccurate Age and Sex Data in the (United States) Census PUMS Files: Evidence and Implications. NBER Working Paper No. 15703 (2010) (Forthcoming Public Opinion Quarterly)
9. Wolfers, J.: Can You Trust Census Data? Freakonomics blog. New York Times (February 2, 2010), http://freakonomics.blogs.nytimes.com/2010/02/02/can-you-trust-census-data
10. Trewin, D.: A Review of IPUMS-International (2007) (unpublished), http://www.hist.umn.edu/~rmccaa/IPUMSI/trewin_ipums_report.pdf