

IPUMS-International Harmonized Census Microdata Extract System: users and uses

*** * ***

**Robert McCaa, Steven Ruggles, & Matt Sobek
University of Minnesota Population Center
rmccaa@umn.edu**

5 steps to using www.ipums.org/international



Variable Name	Variable Description	Case Selection	KE	KE	99	99
COUNTRY	Country of residence		X			
YEAR	Year		X			
<input type="checkbox"/> DATANUM	Dataset number		X	X		
<input type="checkbox"/> SERIAL	Serial number		X	X	X	

1 slide
4 slides

- Goals: inventory, preserve, anonymize, organize, disseminate (free of cost: www.ipums.org/international)
- First release, May 2002: 6, now 8 countries

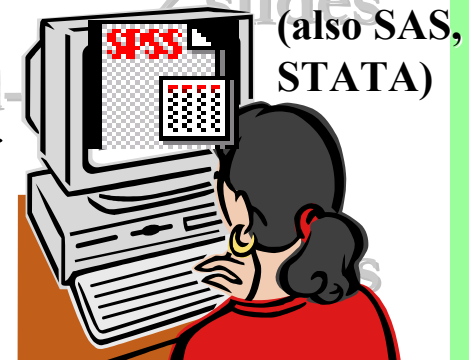


- 3. How do we use the data?
- 4. Who uses the data internationally?

2 slides
2 slides

- Major research centers in the US & world-
- Policy research centers: World Bank, WHO, etc.

- 5. What do they use IPUMS extracts for?



Outline

- » **1. What are census microdata?** **1 slide**
- » **2. What is IPUMS-International?** **4 slides**
 - » **Partners: > 59 official statistical agencies**
 - » **Goals: inventory, preserve, anonymize, harmonize, disseminate (free of cost: www.ipums.org/international)**
 - » **First release, May 2002: 6, now 8 countries**
- » **3. How do we integrate?** **2 slides**
- » **4. Who uses IPUMS-International?** **2 slides**
 - » **Major research universities in the US & world-wide**
 - » **Policy researchers: World Bank, WHO, etc.**
- » **5. What do they use IPUMS extracts for?** **4 slides**

1. What are “census microdata”?:
anonymized, computerized census records
of individuals, households & dwellings

PERSON NUMBER

SEX AGE

1	2	1	0	0	1	0	2	6	0	0	7	0	0	7	2	0	0	0	0	1	1	2	1	0	0	0	0	1	0	4	
2	2	2	0	0	2	0	2	6	0	0	7	0	0	7	2	0	0	0	0	0	1	1	2	1	0	0	0	0	1	0	4
3	2	3	0	0	1	0	0	0	0	0	7	0	0	7	2	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
4	2	3	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	U	
5	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Study any desired set of characteristics.

Easier to harmonize than tables.

Facilitates comparative research.

2. What is IPUMS-International?

a global collaboratory of Official Statistical Institutes & Universities to:

- » 1. Inventory the world's census microdata
- » 2. Preserve endangered microdata and documentation
* * *
- » 3. Integrate census microdata
 - » a. use standards of UNSD, Eurostat, ISCO, ISCED, etc.
 - » b. facilitate comparative research in time and space
- » 4. Anonymize census microdata to preserve statistical confidentiality, using highest standards
- » 5. Disseminate *restricted access, custom extracts* to approved researchers/research projects at no cost

8 countries available now (Appendix table 1)

1998 – Colombia, National Institutes of Health, 1 country

1999 – International, National Science Foundation, 7 countries

Brazil, China, France, Kenya, Mexico, USA, Vietnam

30 country-projects underway

2003 –Latin America, NIH, 14 countries

Argentina, Bolivia, Chile, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Venezuela

2004 –Europe, NIH, 16 countries + 3 (in 2005)

Austria, Belarus, Bulgaria, Czech Republic, Germany, Greece, Hungary, Netherlands, *Poland, Portugal, Romania, *Russia, Slovenia, Spain, *Turkey, UK plus Armenia, Georgia, Ireland

17 country-projects late 2004, early 2005

2005 – Global, NSF, 17 countries

Bangladesh, Cambodia, Canada, Egypt, *Iran, Israel, *Indonesia, Madagascar, Malawi, Malaysia, Mongolia, Pakistan, Palestinian Authority, Philippines, Tajikistan, Turkmenistan, Uganda

Appendix Table 2. IPUMS-International Integrated Census Microdata Sample Characteristics
120 million person records

Source: www.ipums.org/international/sample_descriptions.html

<u>Country census</u>	<u>Sample %</u>	<u>No. of Person records</u>	<u>Additional details</u>
Brazil 1960	5.0	3,001,000	Long-form, cluster sample
1970	5.0	4,954,000	Same
1980	5.0	5,870,000	Same
1990	5.0	8,523,000	Same
2000	6.0	10,136,000	Same
China 1982	0.1	1,003,000	Every thousandth household
Colombia 1964	2.0	350,000	Every fiftieth person
1972	10.0	1,989,000	Every tenth household
1985	10.0	2,643,000	Long-form, cluster sample
1993	10.0	3,247,000	Every tenth household
France 1962	5.0	2,321,000	Every twentieth household
1968	5.0	2,488,000	Same
1975	5.0	2,629,000	Same
1982	5.0	2,714,000	Same
1990	4.2	2,361,000	Every twenty-fourth household
Kenya 1989	5.0	1,074,000	Every twentieth household
1999	5.0	1,410,000	Same
Mexico 1960	1.5	503,000	Every 67th individual
1970	1.0	483,000	Every hundredth household
1990	10.0	8,028,000	Every tenth household
2000	10.6	10,099,000	Long-form, cluster sample
USA 1960	1.0	1,800,000	Stratified, random sample
1970	1.0	2,030,000	Same
1980	5.0	11,337,000	Same
1990	5.0	12,500,000	Stratified, cluster sample
2000	5.0	14,082,000	Same
Vietnam 1989	5.0	2,627,000	Long-form, cluster sample
1999	3.0	2,368,000	Same

3. I
N
T
E
G
R
A
T
E
S
I
P
U
M
S
i

- » Integrate (harmonize), not standardize
 1. retain all original detail
 2. harmonize every digit

- » How is this possible?
Composite codes (multiple digits, 111)
Not serial (1, 2, 3,)
(example: next slide)

- » Why? Researcher confidently
understands the logic
uses as much detail as needed

Composite coding scheme: employment status

Harmonized Codes and Labels		Source Data Codes (selected samples)									
IPUMSI	IPUMSI	Col	Col	Fra	Fra	Ken	Mex	Mex	US	Viet	Viet
Code	Label	1964	1993	1962	1975	1999	1970	2000	1960	1989	1999
0000	N/A	*,5	B	*	B	BB	0	BB	00	B	B,1
	1000 Employed, not specified	1								1	
1100	At work		4	1	1	01	1	10	10		
1101	At work, and 'student'							14			
1102	At work, and 'housework'							15			
1103	At work, and 'seeking work'							13			
1104	At work, and 'retired'							16			
1105	At work, and 'no work'							18			
1108	At work, family holding, not agricultural					03					
1109	At work, family holding, agricultural					04					
1200	Have job, not at work last week		3			02		20	12		
1300	Armed forces								13		
1303	Military trainee (France)			8	6						
2000	UNEMPLOYED, not specified	2			3	05	2	30	20		
2001	Unemployed (Vietnam)									4	5
2002	Worked less than 6 months, permanent job									2	
2003	Worked less than 6 months, temporary job									6	
2100	Unemployed, experience worker		1						21		
2200	Unemployed, new worker		2	7					22		
3000	INACTIVE (Not in Labor Force)								30		
3100	Housework	3	6			10	3	50	31	6	2
3200	Unable to work/disabled	7	7			09		70	32	7	4
3300	In school	4	5	9	5	07		40	33	5	3
3400	Retirees and living on rent	8						60			
3402	Retirees/pensioners		8		4	08					
3500	Elderly	6									
3600	No work available/discouraged					06					
3700	Inactive, other reasons	9	0	0	0	11	4	80	34		6
9000	UNKNOWN/MISSING		9			00	9	99			9

Variable Description: Please see web-page

Restricted Access web-based extraction system

Legally-binding license agreement

- »protects privacy and confidentiality
- »assures proper use

Access limited to:

- »Bona-fide researchers (credentials)
- »With a demonstrated scientific need
- »who agree to abide by license restrictions
 - »Confidentiality
 - »Educational, policy uses only (no commercial use)
 - »No redistribution
 - »Safely secured
 - »Proper citation

Restricted Access web-based extraction system facilitates chronological &/or cross-national research

Researcher selects

- »Countries,
- »Censuses,
- »Cases/sub-populations,
- »Variables, and
- »Sample densities

»Users (May 2002 – Jan 2005):

- »Of 766 applicants, 469 accepted; denial of access rate = 39%

Table 1. Country of residence and Countries of research interest (since August 2002)

Country of residence	%	Country/ies of interest	%
USA	72	Brazil (since Sept. 2004)	4
Canada	4	China (since May 2003)	11
Switzerland	3	Colombia	13
Brazil, Colombia, Kenya (total)	8	France	12
France, Italy, Mexico, Spain, UK, Vietnam (total)	6	Kenya	12
China (includes Hong Kong, etc.)	1	Mexico	20
Australia, Germany	1	USA (excludes IPUMS-USA)	17
19 other countries (total)	5	Vietnam	11

Affiliation, position, academic discipline and expected outcome

Table 2. User Profile: Institutional affiliation and Position

Institutional affiliation	%	Position	%
University	88	Student	48
Regional/International organization	8	Researcher	26
National policy institute	2	Professor	21
National statistical agency	2	Other	6

Table 3. Academic discipline and Expected outcome

Academic discipline	%	Expected outcome	%
Economics	37	Teaching, B.A./M.A. thesis	16
Demography	26	Paper, article, policy report	10
Sociology	13	PhD dissertation	9
Public policy	6	Book	2
History	5	Enhance DHS/other survey	6
Other	13	Other, Not mentioned	57

Research Topics (26 categories), n =

**Table 4. 446 Research topics classified in 26 categories
ordered by frequency**

Migration	64	Marriage	12
Schooling	57	Aging	12
Gender	30	Equality/inequality	12
Data management/development	26	Mortality	12
Teaching	37	Development	10
Health	21	Statistics	9
Fertility	21	Sampling	9
Methods	17	Demography	7
Wages	17	Brain drain/gain	6
Urbanization	15	Religion	4
Family	15	Population projection	3
Children	13	Disability	3
Poverty	12	Vital statistics evaluation	2

Research Topics (examples)

- » Living arrangements of the aged
- » Female labor-force participation and educational attainment
- » Regional inequality differentials
- » International migration
- » Divorce and family composition
- » U.S. National Academy of Sciences, “Transitions to Adulthood in Developing Countries” uses data from Colombia, Kenya, Mexico, and Vietnam to analyze changing outcomes such as schooling, work, fertility, and marriage as a function of age, gender, etc.

**I
P
U
M
S**
**B
E
N
E
F
I
T
S**

» **Researchers & policy-makers:**

High-quality, harmonized microdata—free of charge
Comprehensive, integrated metadata

» **Official Statistics Institutes:**

**Fair, free, uniform solution to providing census
microdata**

Trust, transparency, more stakeholders

Increased usage, better science/policy

Enhanced cost-benefit ratio

Little or no marginal cost (project pays license fee)

» **Citizens, Society, and Governments:**

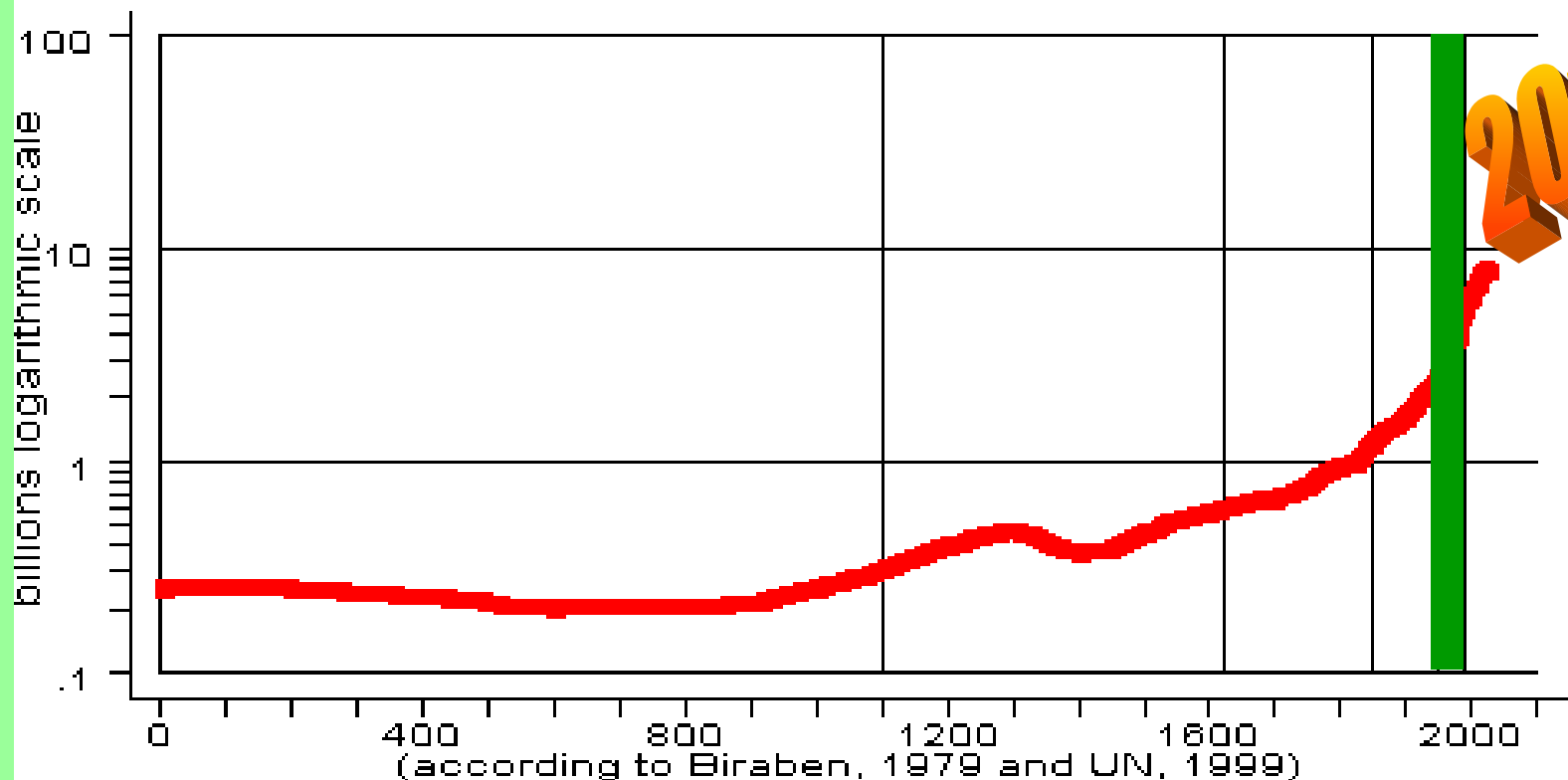
Who we are

What the future may bring

How policies might be improved

On a millennial scale, census microdata survive
for only a short, but significant period
May we help you preserve yours?
... And help promote good use?

Census microdata survive only for the period 1850 to the present
vertical lines show doubling intervals of world population



Please visit the web-site:

www.ipums.org/international

contact:

rmccaa@umn.edu

*** * * * ***

Thank you!!