



**IPUMS-EurAsia, 2009-2014:  
Changing Patterns of Microdata Use**

**[www.ipums.org/international](http://www.ipums.org/international)**

\* \* \*

**Robert McCaa, Professor of Population History  
University of Minnesota**

**[rmccaa@umn.edu](mailto:rmccaa@umn.edu)**

**for additional details, please see:  
[www.hist.umn.edu/~rmccaa](http://www.hist.umn.edu/~rmccaa)**



# Outline

**A. Why Census Microdata?**

**B. Why IPUMS?**

**1. Preserving microdata and metadata**

**2. Integrated microdata**

**3. Integrated metadata**

**4. Statistical confidentiality**

**5. Microdata access**

**C. Whither IPUMS-EurAsia? ..an invitation**



# Introduction

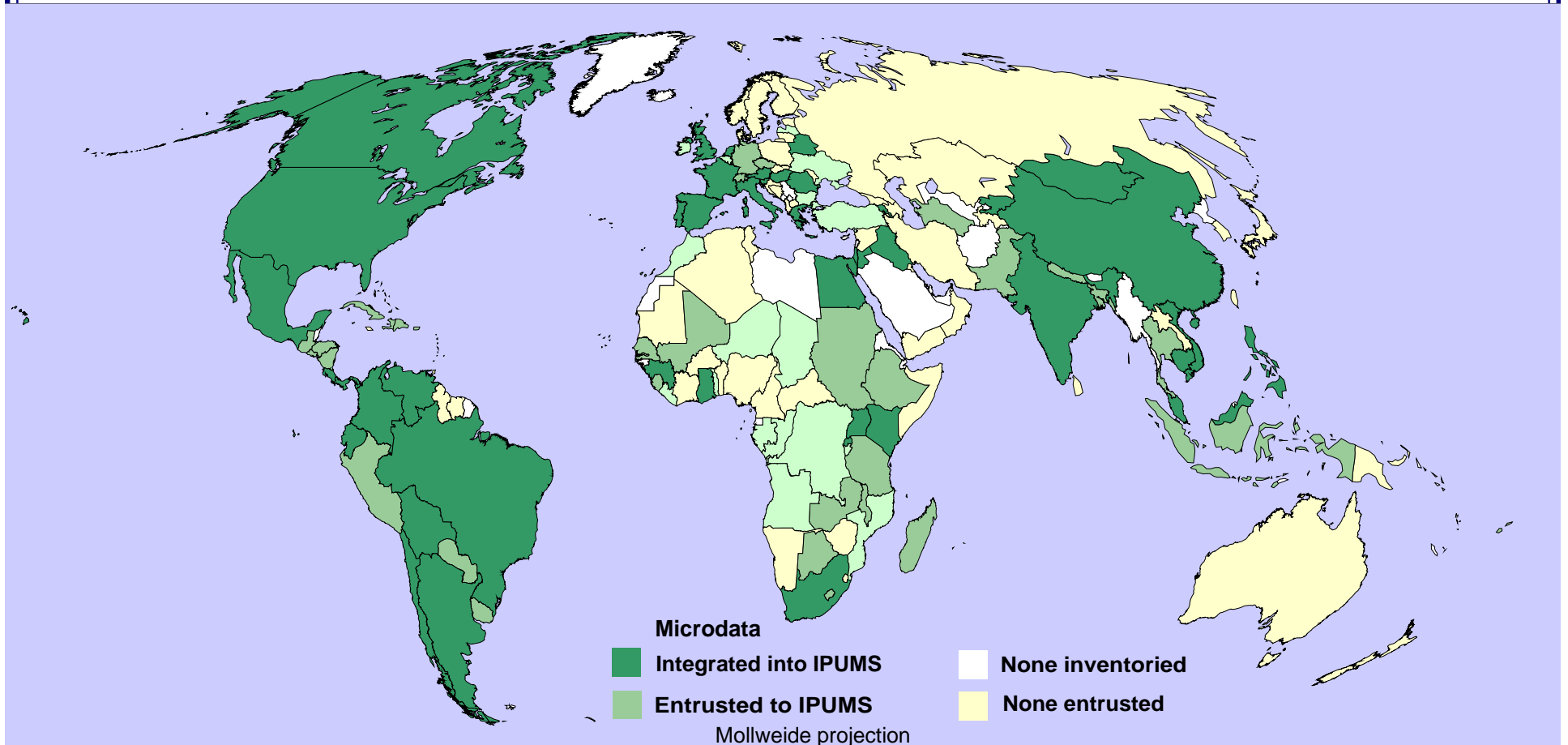
- 1. 1960s: dissemination of census microdata began**
- 2. 1999: IPUMS-International; 2009, 83 countries:**
  - a. Preserve census microdata and documentation world-wide**
  - b. Integrate microdata and metadata**
  - c. Disseminate to researchers world-wide, without cost;  
2009: 130 samples, 43 countries, 279 million person records sustained, major funding:**
    - » National Science Foundation (USA)– renewed through 2014
    - » National Institutes of Health (USA)– eager to fund IPUMS-Africa
- 3. Asia-Pacific region: Vietnam 2001, China 2002 ...  
2009: 12 countries**

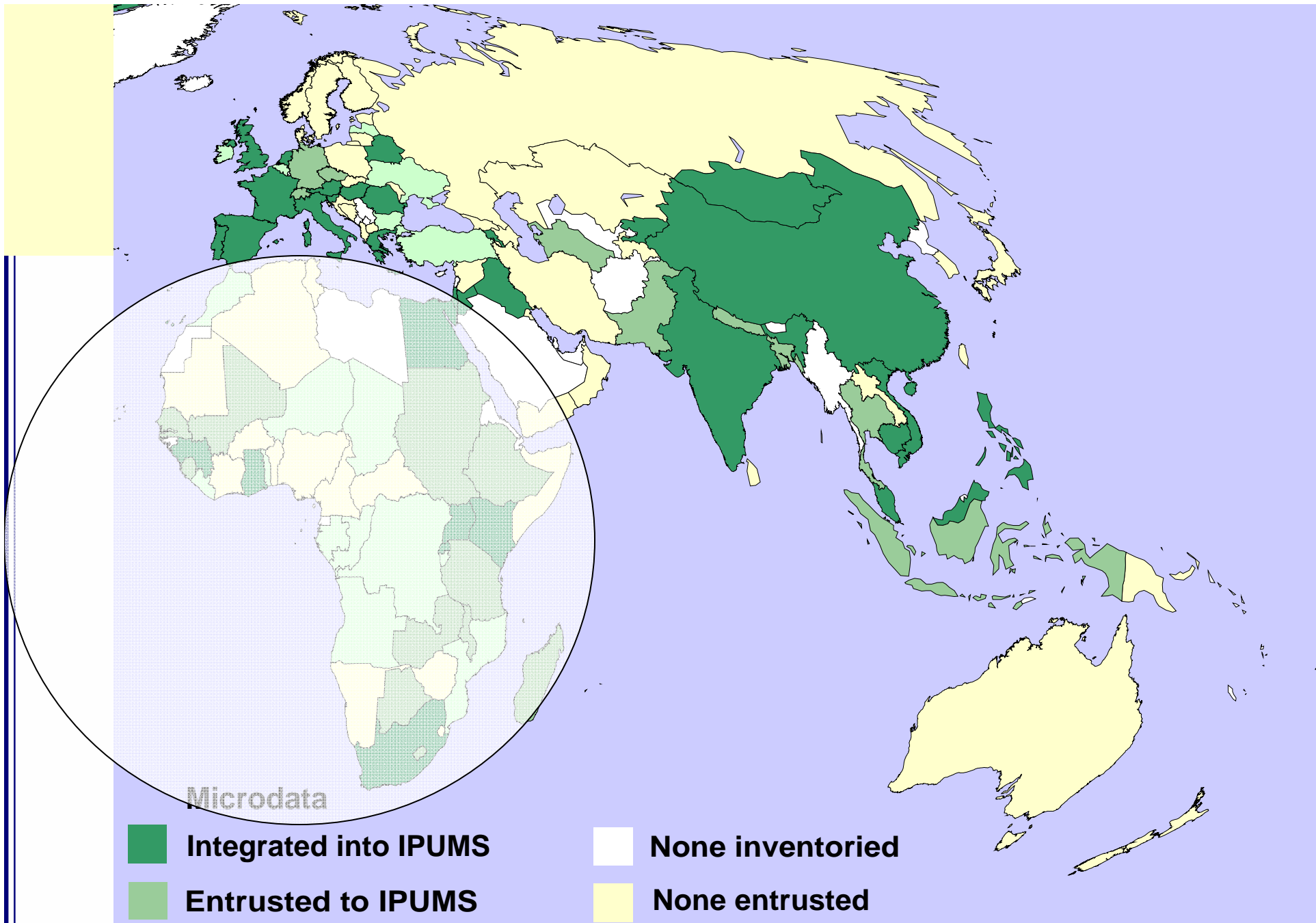
## *IPUMS-EurAsia in global context*

*dark green = integrated 2002-2009*

*(44 countries, 130 censuses, 279 million person records)*

*green = to be integrated (40 countries, 120 censuses, ~200 mpr)*







## Why census microdata?

- 1. Detailed information on individuals, households, and dwellings**
- 2. Researchers may analyze any desired characteristics**
- 3. Essential for comparative research over time and across countries**
- 4. Most widely used source for demographic and social research**



## Why IPUMS?

- 1. Preserve microdata and metadata**
- 2. Integrate microdata**
- 3. Integrate metadata**
- 4. Protect statistical confidentiality**
- 5. Disseminate microdata**



# 1. Data recovery. Example: Bangladesh Bureau of Statistics--1981 census, 276 tapes, recovered Sep. '08)



**>3,000 tapes  
recovered: 1971 Germany  
1980 Mexico,  
Mali 1976, Sudan 73  
and many more**

## 2. Microdata integration

**composite codes (multiple digits)**  
 retain not only significant distinctions  
 but also integrate comparable concepts

Code	Label	Chile		México	
		1992	2002	1990	2000
0	NIU	X	X	X	X
	<b>ACTIVE (In Labor Force)</b>				
100	EMPLOYED, not specified	.	.	.	.
110	At work	X	X	X	X
111	At work, and 'student'	.	.	.	X
112	At work, and 'housework'	.	.	.	X
113	At work, and 'seeking work'	.	.	.	X
114	At work, and 'retired'	.	.	.	X
115	At work, and 'no work'	.	.	.	X
116	At work, and 'other'	.	.	.	X
117		.	.	.	.
118	At work, family holding, not agricultural	.	.	.	.
119	At work, family holding, agricultural	.	.	.	.
120	Have job, not at work last week	X	X	X	X

**INDEC-Argentina evaluated IPUMS integration:  
 A couple of minor errors and misinterpretations**



## 3. Metadata integration

- » **Comprehensive source documentation:**
  - » **Data dictionaries and codebooks**
  - » **Questionnaires, manuals, etc.**
  - » **All translated to English and converted into metadatabase for each census**
- » **New metadata for each census and sample**
  - » **Census title, year, universe, de-jure/de-facto, census day, forms, field work period, etc.**
  - » **Sample: source, design, density, unit, weights, etc.**



### 3. Metadata integration

- » **New, systematic metadata for each variable**
  - » **Codes**
  - » **Universes**
  - » **Definitions**
  - » **Comparability**
  - » **Dynamic System—facilitates comparing the wording of questionnaires and instructions for any combination of countries and censuses**

# EMPSTAT, General Version, Case-Count View

## Example of IPUMS Metadata “Codes”

**IPUMS** International

[HOME](#) | [GET DATA](#) | [VARIABLES](#) | [FAQ](#) | [CONTACT US](#)

Include in extract

Select samples

**EMPSTAT** (general version, 1 digit )

Label: Employment status

- Category availability view       General codes  
 Case-count view                       Detailed codes

Code	Label	cambo 1998	china 1982	china 1990	malay 1970	malay 1980	malay 1991	malay 2000	philipp 1990	vietn 1989	vietn 1999
0	NIU	.	3,371,276	3,264,675	54,601	48,289	89,729	99,921	1,523,457	902,315	666,431
1	Employed	483,633	5,216,774	6,775,149	54,334	64,103	122,755	162,358	1,992,412	1,113,788	1,096,580
2	Unemployed	27,210	35,537	61,553	3,179	4,664	9,674	9,032	216,171	143,926	58,865
3	Inactive	630,411	1,415,604	1,734,570	61,854	63,559	120,885	163,022	2,281,873	463,654	545,853
9	Unknown/missing	.	.	.	2,029	1,986	4,849	967	.	3,302	438

[HOME](#) | [GET DATA](#) | [VARIABLES](#) | [FAQ](#) | [CONTACT US](#)

Copyright © Minnesota Population Center, University of Minnesota

# EMPSTAT, Variable Description

## Example of IPUMS Metadata

**IPUMS**

International

[HOME](#) | [GET DATA](#) | [VARIABLES](#) | [FAQ](#) | [CONTACT US](#)

Include in extract

Select samples

### EMPSTAT

**Label: Employment status**

---

Availability [[show](#)]

Universe [[show](#)]

[Codes](#)

[Enumeration text](#)

[Unharmonized variables](#)

### Description

EMPSTAT indicates whether or not the respondent was part of the labor force -- working or seeking work -- over a specified period of time. Depending on the sample, EMPSTAT can also convey further information.

THE FULL LIST OF EMPSTAT IS FULLY AVAILABLE AND CLASSIFIED THE VARIABLE INTO THREE GROUPS: EMPLOYED

# EMPSTAT, Variable Description

## Example of IPUMS Metadata



### Description

EMPSTAT indicates whether or not the respondent was part of the labor force -- working or seeking work -- over a specified period of time. Depending on the sample, EMPSTAT can also convey further information.

The first digit of EMPSTAT is fully comparable, and classifies the population into three groups: employed, unemployed, and inactive. The combination of employed and unemployed yields the total labor force. The second and third digits of EMPSTAT preserve additional information available for some countries and census years but not for others.

### Reference period

The reference period for the employment status question varies across censuses. For most samples, employment status was reported with respect to the day of the census or within a specified week prior to the census. In a few samples the period can be as month or as long as a year -- amounting to "usual employment status" over this period. Short reference periods can have an impact on seasonal labor measures, depending on when the census was taken. The variations among samples with respect to reference period are summarized below:

- 1) Current (at time of census):

# EMPSTAT, “Enumeration Text”

## = form and instructions

### Example of IPUMS Metadata

Cambodia 1998-- source variable [KH98A414](#) -- Economic activity [top](#)

#### Enumeration form

view entire document: [text](#) [image](#)

15 Main Activity (Enter Code from list below)

Main activity during last year

Codes for Column 15 Main Activity During Last Year

1 Employed (Fill in Cols. 16 to 20)

2 Unemployed (Employed any time before - Fill in Cols. 16 to 20 for last employment)

3 Unemployed (Never employed any time before)

Inactive

4 Homemaker

5 Student

6 Dependent

7 Rent-receiver, retired or other income recipient

8 \_\_\_ Other (Specify)

(For codes 3 to 8, put dash ( - ) in Cols. 16 to 20)]

***Click above for  
text or image in  
official language***

#### Enumeration instructions

view entire document: [text](#) [image](#)

#### Columns 15 to 20 on Economic Characteristics of Population

97. These columns relate to the economic characteristics of population both male and female. Information proposed to be collected on economic characteristics of population includes information on economically active and economically not active population.

Economically active population includes those who are employed and unemployed. Economically inactive population includes home

1 Employed (Fill in Cols. 16 to 20)

2 Unemployed (Employed any time before - Fill in Cols. 16 to 20 for last employment)

3 Unemployed (Never employed any time before)



#### **4. Statistical confidentiality:**

**Conference of European Statisticians: “Good practice”  
Dennis Trewin on-site inspection:**

- » “The best practice for an international repository of microdata”**
- » “The security of IPUMS is first class...the standard of the best national statistical offices”**
- » “in full compliance with the principles and recommendations of the CES [Conference of European Statisticians]”**



## 5. Microdata access:

### IPUMS is a restricted-access, web-based system

- » Password protected: to make extracts and retrieve microdata
- » Licensed researcher selects:
  - » Countries,
  - » Censuses,
  - » Cases/sub-populations,
  - » Variables, and
  - » Sample densities
- » Extract engine queues request, generates extract
- » Researcher retrieves extract via web with SSL 128-bit encryption and analyzes using own wares (soft/hard/wet)
- » **NO** source files. **NO** complete datasets.



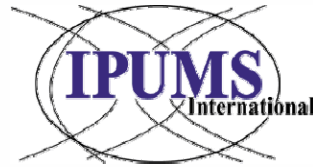
## C. Whiter IPUMS-EurAsia?

- » **Users: 2,482 researchers, 69 countries**
  - » **90% are University researchers; but also WHO, ILO, World Bank**
  - » **Economists: 45.7%**
  - » **Demographers: 19.0**
  - » **Sociologists: 10.1**
  - » **Public policy: 5.1**
  - » **Statisticians: 2.7**
  - » **Historians: 2.4**
- » **Asia and Pacific region—not so many users, due to few samples?**
  - » **China 37**
  - » **Japan 25**
  - » **Australia 24**
  - » **Singapore 10**
  - » **India 8**
- » **Looking Ahead**



# Looking ahead

- » **Countries:**
  - » **Soon: Bangladesh, Indonesia, Nepal, Pakistan, Thailand**
  - » **Later: ???**
- » **2010 census round**
- » **New methods: variance estimation**
  - » **Imputing pseudo-strata to simplify variance estimation for complex samples**
- » **2015: 200 censuses, 75 countries, 10,000 users??**



# IPUMS at the 57<sup>th</sup> ISI (Durban, Aug 16-21, 2009)

<http://www.statssa.gov.za/isi2009/index.aspx>

57th Session of the ISI  
**ISI 2009**

- ▶ Welcome
- ▶ Invitations
- ▶ About ISI
- ▶ About the 57th Session of the ISI
- ▶ About Host City
- ▶ About Durban ICC
- ▶ ISI and African Statistical Development
- ▶ The ISI 2009 Logo Defined
- ▶ Travel and Accommodation
- ▶ Conference Information
- ▶ Registration
- ▶ Scientific Programme
- ▶ Newsletters
- ▶ News Archive
- ▶ Image Gallery
- ▶ Contact Us

The 57<sup>th</sup> Session of the ISI  
International Statistical Institute

16-22 August 2009 • Durban, South Africa  
Statistics: Our Past, Present & Future

Registration now open. Click [here](#) to register or express interest.

## Welcome

Welcome to the website of the historic 57th Session of the International Statistical Institute (ISI), the first in Sub-Saharan Africa.

- » IPUMS-NSI Workshop (Aug 15-16)
- » STCPM session: cross-national microdata
- » IPUMS-Users Workshop
- » IPUMS Modest funding for delegates from developing countries



# IPUMS Global workshop, 56<sup>th</sup> ISI (Lisbon, Aug 2007)





**Thank you.**

---

**rmccaa@umn.edu**

---