



**Disclosure Controls for One Stop, Trans-Border
Access to Census Microdata via a Single License
and Dissemination Point
The IPUMS-IECM partnership**

*** * ***

Robert McCaa and Albert Esteve Palos

Minnesota Population Center and Centre d'Estudis Demografics--Barcelona

www.ipums.org/international

**“You have to do due diligence, something to assure yourself
that the people you’re giving your data to can be trusted.”**

--<http://www.nytimes.com/2011/09/09/us/09breach.html?hp>



IPUMS at the 59th ISI (Hong Kong, Aug 24-30, 2013)

<http://www.isi2013.hk/>

ENGLISH | 繁體 | 简体

The 59th World Statistics Congress of the International Statistical Institute
25-30 August 2013
Hong Kong, China

Welcome Message | The 59th World Statistics Congress | Host City | Meeting Venue | International Statistical Institute | Census and Statistics Department | Contact Us

See You in Hong Kong in 2013 WSC

To support browsing of this website, you need to use the Adobe Flash Player 10.

Census and Statistics Department
Hong Kong Special Administrative Region
People's Republic of China

INTERNATIONAL STATISTICAL INSTITUTE

- » **IPUMS Workshop**
- » **Microdata session**
- » **IPUMS Funding for delegates from developing countries**
- » **IPUMS booth**



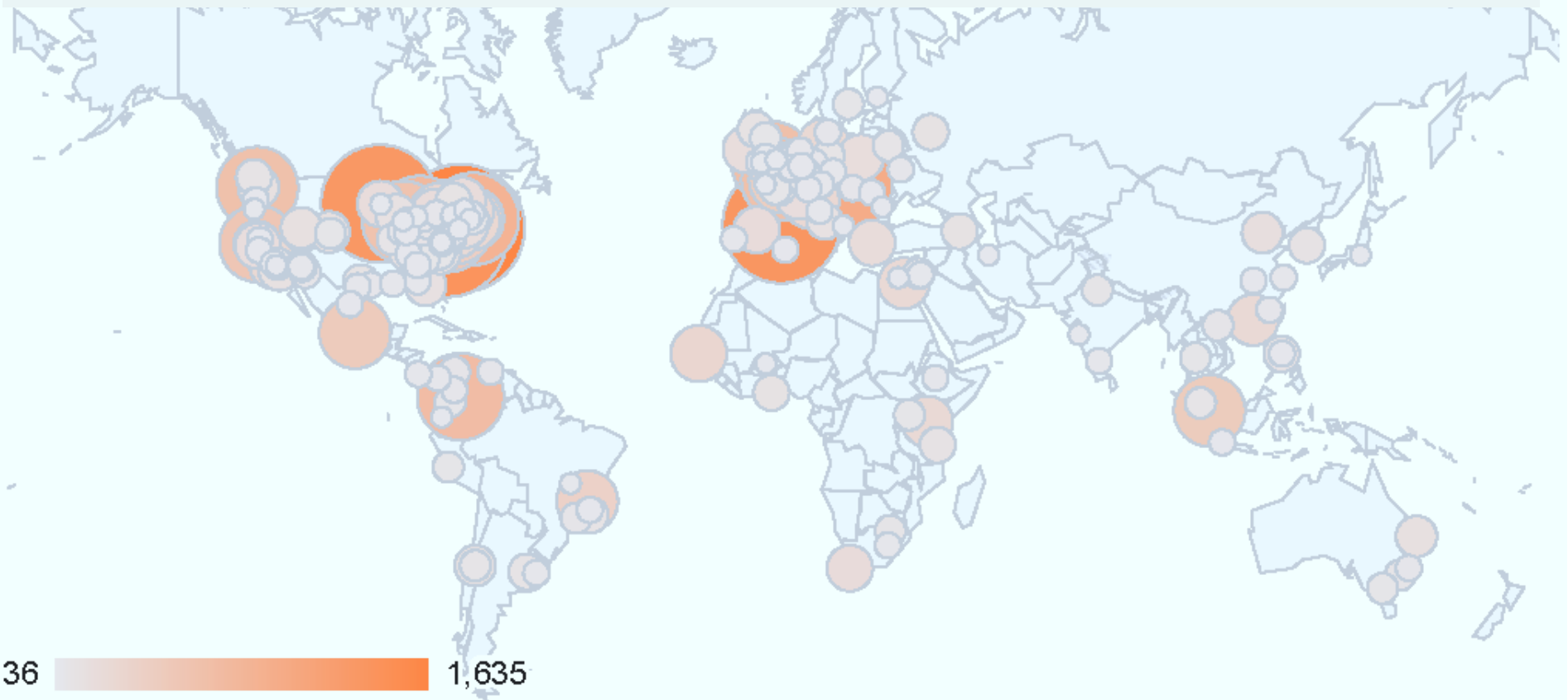
**Trans-Border access is essential in 21st Century.
Many demographers (membership of the IUSSP)
reside outside their country of birth**

- | | | |
|--|-------------|--|
| • New Zealanders | 60% | reside outside country of birth |
| • Dutch | 40% | |
| • Germans | 38% | |
| • Danes | 34% | |
| • Chinese | 30% | |
| • Belgians | 31% | |
| • British | 25% | |
| • Australians | 22% | |
| • Canadians, Finns, French, Japanese, Swiss | ~20% | |

Requiring on-site access to microdata is old-fashioned, inefficient, costly, & unfair. Encourages violations.



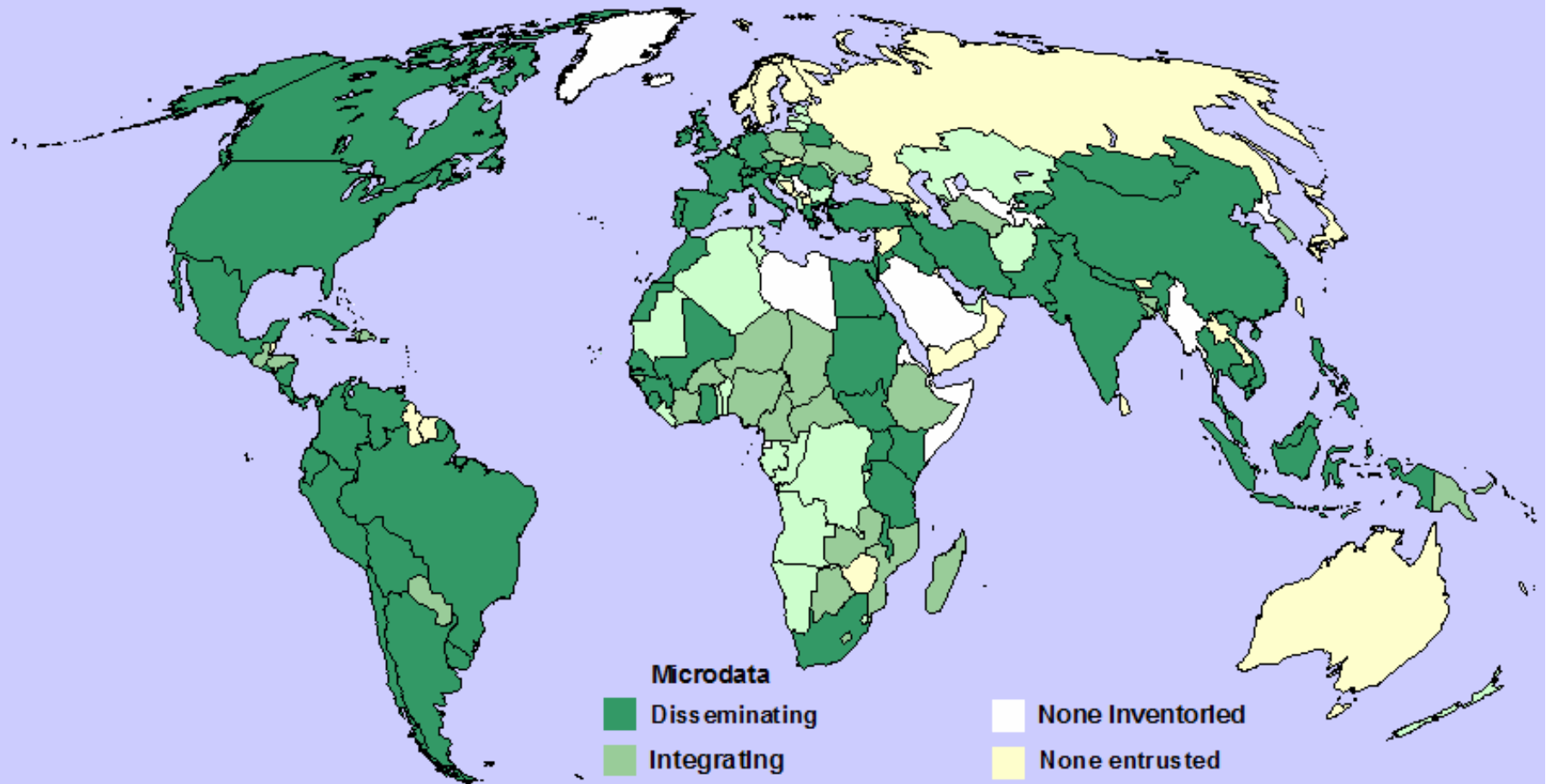
**IPUMS-International Google Analytics: 2011
Trans-Border Access: 169 countries/territories
3,033 cities, 45,000 page views. Up 4X from 2010**



IPUMS-International: 2012

*dark green = anonymized, harmonized and disseminating
(69 countries, 212 censuses, 480 million person records)*

medium green = to be integrated (27 countries, 71 censuses, ~100 mpr)

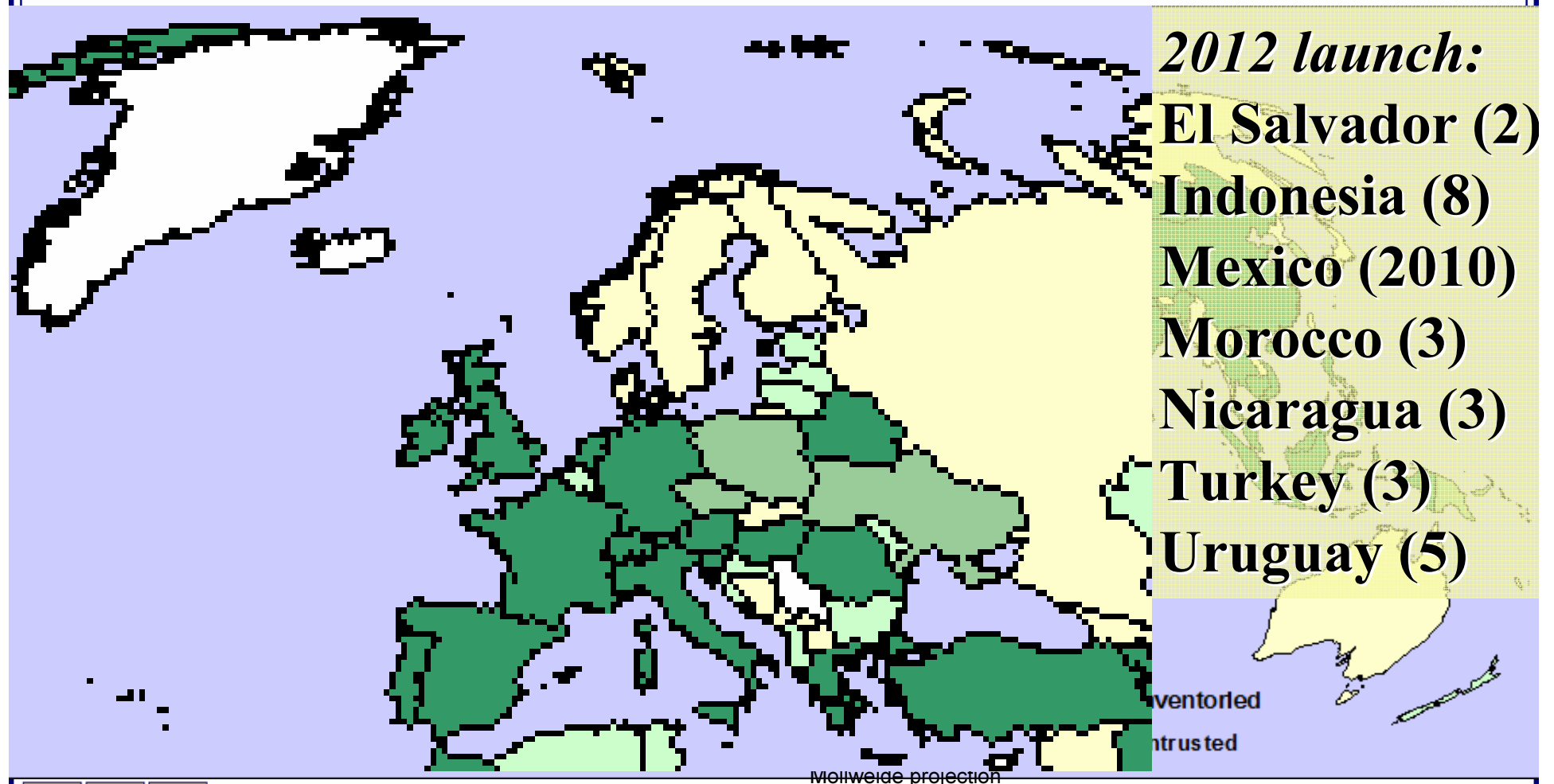


Mollweide projection

IPUMS-International: 2012

*dark green = anonymized, harmonized and disseminating
(69 countries, 212 censuses, 480 million person records)*

medium green = to be integrated (27 countries, 71 censuses, ~100 mpr)





Outline:

Trans-Border access to Census Microdata: IPUMS-International protects confidentiality 3 ways: Legal, Administrative and Statistical

- 1. Legal. uniform agreement between NSOs and University of Minnesota**
- 2. Administrative. Intimidating application form: self-selection
Over 5,000 views of the registration page
1,057 applications. A mere 46 had to be denied access.**
- 3. Statistical Disclosure controls**
- 4. Much usage: 8,000 extracts in 2011; top 20 institutions.**
- 5. Invitation to participate: update law, where necessary**
- 6. Conclusion: 10 years, 5000 users, hundreds of publications, no problems.**



1. Legal: Uniform Memorandum of Understanding with 98 National Statistical Office Partners (Appendix A)

- 1. Ownership** **NSO**
- 2. Use** **Teaching, research and publishing only**
- 3. Authorization** **Submit electronic application**
- 4. Restriction** **No commercial use**
- 5. Confidentiality** **Strictly prohibited: attempt or allege I.D.**
- 6. Security** **Prevent unauthorized access**
- 7. Publication** **Publication permitted—cite IPUMS & NSO**
- 8. Violations** **Sanction individual and his/her institution**
- 9. Sharing** **Share integrated microdata and metadata**
- 10. Jurisdiction** **ICC Court of Arbitration**
- 11. Order of precedence** **First, MOU; then, other...**



2. Uniform License agreement with researchers and their institutions (Appendix B)

- **Establish individual bona-fides**
- **Institutional affiliation**
- **Describe research proposal and need for access to microdata**
- **Agree to conditions of use**
 - 1. Scholarly research and educational purposes**
 - 2. Commercial use and re-distribution is prohibited**
 - 3. Obey strict rules of confidentiality**
 - 4. Secure microdata from 3rd parties**
 - 5. Publications permitted; cite IPUMS and NSO; provide copies**
 - 6. Violation: recall of microdata and sanction institution**



3. 7 Statistical Disclosure Controls (Thorogood, 1999)

Example: Switzerland 1970-2000

- » 1. Sub-sampling; 95% of records suppressed**
- » 2. Suppression of names, addresses, and precise place of birth/residence. Floor: 20,000 pop (others 65k, 100k, 250k, etc.)**
- » 3. Recode unique categories--top and bottom**
- » 4. Suppression of digits for hierarchical codes (e.g., 4 digit ISCO becomes 3)**
- » 5. Suppression of sensitive variables (e.g., tribe)**
- » 6. Swapping of residence codes for a tiny fraction**
- » 7. Randomly re-order households.**

Table 2. Rank of the Top Five and all European Countries plus Canada and the USA by Number of Extracts for the 2000 round census (statistics for calendar year 2011)

Rank	Country	Sample %*	Variables (n)*	Years of census samples	Extracts
1	Brazil	5	106	1960, 70, 80, 91, 2000	712
2	Mexico	10	120	1960p, 70, 90, 95, 2000, 05	626
3	United States	5	92	1960, 70, 80, 90, 2000, 05	554
4	Colombia	10	120	1964p, 72, 85, 93, 2005	516
5	South Africa	10	108	1996, 2001, 2007	428
7	Canada	2.5	59	1971p, 81p, 91p, 2001p	409
9	France	33	94	1962, 68, 75, 82, 90, 99, 06	380
10	Spain	5	99	1981, 91, 2001	366
13	Greece	10	89	1971, 81, 91, 2001	327
18	Austria	10	75	1971, 81, 91, 2001	310
25	Italy	5	81	2001	285
26	Portugal	5	96	1981, 91, 2001	283
29	Romania	10	97	1976, 92, 2002	272
30	Switzerland	5	79	1970, 80, 90, 2000	266
32	United Kingdom	3	47	1991, 2001p	263
38	Hungary	5	74	1970, 80, 90, 2001	222
42	The Netherlands	1	33	1960p, 71p, 2001p	211
45	Slovenia	10	80	2002	185
48	Belarus	10	84	1999	179
Total samples extracted for 55 countries (162 samples) available from January 1, 2011.					8,048



*2000 round census; refers to all integrated variables, including IPUMS constructed variables

POPULATION CENTER

UNIVERSITY OF MINNESOTA



IPUMS 2011 Top 20 (Appendix C—excludes UMN)

Rank	Institution	Extracts	
1	Columbia University	558	
2	University of Hong Kong (Hong Kong, SAR)	309	
3	University of Michigan	296	
4	Arizona State University	256	
5	Institute for Health Metrics & Evaluation (Seattle)	245	
6	Universidad Nacional de La Plata (La Plata, Argentina)	205	
7	Harvard University	205	
8	National University of Singapore (Singapore)	198	
9	Centre of Demographic Studies (Barcelona, Spain)	185	
10	Dartmouth College	177	
11	Baruch College	152	
12	University of Virginia	141	
13	Vienna Institute of Demography (Vienna, Austria)	138	
14	United Nations-Habitat (Nairobi, Kenya)	134	
15	University of Stirling (Scotland, UK)	110	
16	University of Pennsylvania	108	
17	Institute of Political Studies (Paris, France)	107	
18	Vanderbilt University	105	
19	University of Colorado at Boulder	90	
20	University of Cape Town (South Africa)	88	12

Cooperation with new initiatives

1. Data without Boundaries: Service Center for Official Statistics



Search..

About DwB

- Work Packages
- Deliverables
- Participants
- Contact

Data Access

- Call for Research Proposals
- Research Data Centres and Datasets
- Application Process
- Research Outcomes
- FAQs

Events and Training

- European Data Access Forums
- Regional Workshops
- Users' Conferences
- Training Events and Material
- Visits to the Research Data Centres
- Additional events

Promotion and Publications

- Promotion
- Publications

Research Data Centres and Datasets

Below you will find links to the involved RDCs and to the available datasets.

Most RDCs require the successful applicant to visit the RDC for a face-to-face training session before the data can be accessed, whether onsite or remotely. For selected projects, DwB will refund travel expenses.

Research Data Centre	Contact	Datasets	Mode of access
GENES, France	Email	Datasets (in French only)	Onsite and remote 1
Federal Statistical Office of Germany	Email	Datasets (in German only)	Onsite only

1st European Data Access Forum (Update)

Final programme, slideshows and practical details are now available [More...]

1st European Data Access Forum

Indicative programme is now available [More...]

1st European Data Access Forum

DwB organises its first European Data Access Forum (EDAF) on 27 & 28 March 2012 in Luxembourg. Registration is now open. [More...]

Workshop on Accreditation Issues in Eastern Europe

DwB organised a joint workshop on Access to Official Microdata & Accreditation of Researchers in Eastern European Countries, on 23 January 2012. [More...]

Workshop on metadata standards

DwB organised a joint workshop on metadata standard issues, in collocation with EDDI 2011. [More...]

Cooperation with new initiatives:

2. Trans-border, on-line tabulator



Centre d'Estudis Demogràfics



Integrated European Census Microdata

MPC
Minnesota
Population Centre

IECM DATABASE

-Crosstabs of Variables

- Dwelling and Household
- Person Variables
- Combination

-Dictionary Information

- All Variables
- Household Variables
- Household Categories
- Person Variables
- Person Categories

-Documentation

- Help
- Questionnaires

OTHERS

-Full IECM Database

- All IECM Samples

-Samples Specific Queries

- Austria
- Belarus
- France
- Greece
- Hungary
- Netherlands
- Portugal
- Romania
- Spain
- United Kingdom

-Links

- IECM

Parameter Definition of All IECM Samples

Table Title: Educational Attainment by sex, country and s

Sample Selection: ALL [Edit]

Column: Sex

Row 1: Educational attainment, international recode

Row 2: (None)

Row 3: (None)

Sample Break: SAMPLE

Weight: Person weight

Percent Options: % Column

Filter: (None)

PERSON.AGE >= 30 AND PERSON.AGE < 40

[Build Filter]



Password protected tabulator: www.iecm-project.org



**Recap and Conclusion:
Trans-Border access to Census Microdata
IPUMS-International protects confidentiality 3 ways:
Legal, Administrative, and Statistical**

- 1. Legal. uniform agreement between NSOs and University of Minnesota**
- 2. Administrative. Intimidating application form: self-selection
Over 5,000 views of the registration page
1,057 applications. A mere 46 had to be denied access.**
- 3. Statistical Disclosure controls**
- 4. Much usage: 8,000 extracts in 2011; top 20 institutions.**
- 5. Invitation to participate: update law, where necessary**
- 6. Conclusion: 10 years, 5000 users, hundreds of publications, no problems.**



Thank you

**If your NSO is not participating yet, please
contact: rmccaa@umn.edu**

**If you, your colleagues or students are not
using integrated census microdata...**

Please visit: www.ipums.org/international