UNIVERSITY OF MINNESOTA

Twin Cities Campus

RobertMcCaa
rmccaa@umn.edu
Minnesota Population Center
College of Liberal Arts

537 Heller Hall
271 19th Avenue South
Minneapolis, MN 55455
612-624-3507
Fax: 612-626-8375
http://www.ipums.org

IPUMS-Europe Census Microdata Harmonization Project
Overview (draft: July 12, 2004)
http://www.hist.umn.edu/~rmccaa/ipums-europe
dissemination site: www.ipums.org/international (harmonized data for France & other countries)

Welcome! Thanks to the support of official statistical agencies in 18 European countries (see list of partners below) and major funding by the National Institutes of Health, one of the world's largest, integrated scientific instruments for the study of human populations is now under construction. The database will contain anonymized microdata samples, encompassing as many as 50 censuses and totalling more than 70 million person records. Researchers already are using, free of cost, integrated microdata for 7 countries: China, Colombia, France, Kenya, Mexico, USA and Vietnam from the IPUMS web-site. Brazilian microdata (5 censuses) are undergoing final testing and will be placed on the main site in the near future (please see http://beta.ipums.org/international for a preview).

Constructing the European database will be easier than what one might imagine. From the work completed so far, we have learned that many country projects can be underway simultaneously without the need for a great deal of travel or consultation between them. Many problems are readily resolved by email. The Latin America project, which began last year and serves as a prototype, may be viewed at: www.hist.umn.edu/~rmccaa/ipumsla Computer work will be performed at the Minnesota Population Center (MPC) using software tools already developed from working on the first eight countries.

The project has 4 phases: collection, preparation, harmonization and dissemination. In each country, the transition from one phase to the other is determined by when the work is completed:

1. **Collection Phase** ("year 1"): See below: "Billing Instructions", "Documentation Needed" & "Microdata Specifications". Please send these materials via courier service.
    - NSI: Provide copies of original documentation and microdata for each census
    - MPC: Pay license fees to NSI; recover endangered microdata
    - Centre d'Estudis Demogràfics (CED, Barcelona): Workshop, Sept. 2005
2. **Preparation Phase** ("year 2"): Country-specific details will be provided once we have examined the country's census documentation and microdata.
    - MPC & CED: Scan and translate documentation as necessary; validate and clean microdata, confirming record structure and contents.
    - NSI or in-country expert consultants: Provide methodological assistance with anonymization issues and difficult country-specific variables, such as administrative geography, economic and educational variables, etc.
3. **Harmonization Phase** ("year 3"):
    - MPC & CED: Design integration tables for each variable (original code, harmonized code); write metadata for users explaining comparability strengths and weaknesses
    - NSI, or in-country consultants: Test alpha site for methodological coherence and consistency
4. **Dissemination Phase** ("year 4" and beyond):
    - MPC & CED: Provide harmonized extracts to researchers; conduct workshops at regional or international meetings to promote use by national researchers; provide NSI partners with usage data and publications resulting from research using integrated microdata

UNIVERSITY OF MINNESOTA

Twin Cities Campus

RobertMcCaa
rmccaa@umn.edu
Minnesota Population Center
College of Liberal Arts

537 Heller Hall
271 19th Avenue South
Minneapolis, MN 55455
612-624-3507
Fax: 612-626-8375
http://www.ipums.org

IPUMS-Europe Census Microdata Harmonization Project
Billing Instructions for Microdata & Dissemination Licenses

Upon presentation of invoice, the project pays a fee to the corresponding national statistical authority, of approximately US$5,000 per census microdataset supplied. This fee is intended to cover marginal costs of preparing copies of the microdata and documentation as well as incidentals involved in providing a modest amount of consulting and translation assistance with unusual terms or concepts.

Payment is made upon provision of microdata and receipt of official, signed invoice. To reduce administrative costs as well as transaction charges, please bill for all microdata on a single invoice, unless a long delay is expected in providing one or more datasets (such as the most recent census or a historical census requiring recovery). Please note that while all microdata licenses are budgeted to be paid in the first year of the project, payment will be made upon receipt of microdata. Payment should be received within two months of submission of invoice.

The invoice should provide the following information:

- Name of Statistical Agency
- Address
- Telephone, fax, email contact
- Amount (as negotiated prior to submission of grant)
- Description of datasets (i.e., census years)
- Signature and date

Payment will be made by check. If payment by direct wire deposit into the Statistical Agency's bank account is preferred, please provide the following additional information:

- bank name
- branch name and street address
- bank swift code and routing number
- account holder's name (Statistical Agency or related official account)
- account number

Please send the signed invoice (and banking information for direct deposit) along with the microdatasets by courier mail, at project expense (FEDEX account #: 2221-6454-0) to:

Robert McCaa, Minnesota Population Center
271 19th Ave. S. 537 Heller Hall
Minneapolis, MN 55455 USA
Tel. 1+612-624-5818
Reference: IPUMS: [Country]

UNIVERSITY OF MINNESOTA

Twin Cities Campus

RobertMcCaa
rmccaa@umn.edu
Minnesota Population Center
College of Liberal Arts

537 Heller Hall
271 19th Avenue South
Minneapolis, MN 55455
612-624-3507
Fax: 612-626-8375
http://www.ipums.org

IPUMS-Europe Census Microdata Harmonization Project

Documentation Needed

If this pan-European initiative is to succeed in a mere five years, the essential first step is to obtain without undue delay copies of all relevant census documentation. Preliminary study of the documentation of each participating country will aid greatly in designing the over-all integration: country-by-country, census-by-census, question-by-question, concept-by-concept, and code-by-code.

Specifically we are requesting, from each national statistical authority, copies of documentation in the national language as well as in English, French, German, or Russian (where available), for each census for which microdata are to be included in the project. Three types of documentation are needed:

1. Census enumeration forms (Nearly complete; we are lacking: Austria 71; Bulgaria 92, 75; Czech Republic 91; Greece 91, Hungary 80; Poland 88, 70; Slovenia 81; Spain 81; Turkey 1960-2000)
2. Census enumerator instructions (all needed)
3. "Codebooks" for each dataset (definitions of record structures, column location of variables and labels for codes, such as the U.S. Census Bureau "IMPS" data dictionary files), including administrative geography, occupations, etc.—all needed

Also welcome are copies of technical reports on census operations, sampling methods, comparability, data quality, phrasing of questionnaires, etc.—any extant documentation that would be of assistance in attaining the highest standards of integration.

Documentation may be supplied in electronic, published, or photocopied form. Electronic is the preferred means, where available, and may be sent as emailed attachments or on CDs. To avoid loss of materials and to economize effort, please assemble the entire collection, and send by courier mail, at project expense (FEDEX account #: 2221-6454-0) to:

Robert McCaa
Minnesota Population Center
271 19th Ave. S. 537 Heller Hall
Minneapolis, MN 55455 USA
Tel. 1+612-624-5818
Reference: IPUMS: [Country]

Please confirm shipment, indicating contents, date shipped and package tracking number, by email to: rmccaa@umn.edu

As an example of the original source documentation needed, please consult the Latin America web page: www.hist.umn.edu/~rmccaa/ipumsla (the IPUMS-LA project began a year ago).

There you will find a nearly complete collection of electronic images of each of these three types of documents for each country and census in the region. Statisticians have lauded this body of documentation as a significant scientific accomplishment in itself! Work on harmonizing the LA census microdata is already underway.

Thank you for your cooperation!

UNIVERSITY OF MINNESOTA

Twin Cities Campus

RobertMcCaa
rmccaa@umn.edu
Minnesota Population Center
College of Liberal Arts

537 Heller Hall
271 19th Avenue South
Minneapolis, MN 55455
612-624-3507
Fax: 612-626-8375
http://www.ipums.org

IPUMS-Europe Census Microdata Harmonization Project: **microdata specifications**

Please note:

a. No reformatting or other processing of microdata by the National Statistical Agency is required. We prefer to receive the "best copy" of the microdata—sometimes this may be the only copy.
b. The original files supplied to the project will never be copied or distributed to anyone. They are for the sole purpose of developing the anonymized, integrated database.
c. Before integration is undertaken on datasets that are not already anonymized, a detailed anonymization plan will be drawn up for each census and discussed with each statistical agency.
d. Microdata, including all census records (dwelling, household, family, person, fertility, migration, etc.), should be supplied on CD(s) and shipped by courier mail (see "billing instructions").

Specifications:

a. Where a sample is all that exists (e.g., "long-form" sample taken in the field at enumeration):
    i) We accept that sample for anonymization and integration.
    ii) Please provide documentation on sample design for any such datasets.
b. Where 100% microdata exist:
    i) We prefer that copies of these be made available to us in Minnesota so that the samples may be drawn consistently, efficiently, and with a minimum of burden on our partners. Moreover, should there be imperfect records in the sample, we resolve these easily by replacement. Please note that any such datasets are maintained under total security ("Icebox") and are never reproduced for any person or institution under any circumstances. For the Latin America project, we have been entrusted with 100% microdata for some 30 censuses from a dozen countries. We have an unblemished security record, and have procedures in place to extend this record. If additional assurances are required, please contact me to discuss this.
    ii) Where 100% microdata cannot be supplied, we prefer samples according to the following simple protocol:
        a) Where necessary, sort the microdata files by major and minor administrative divisions down to the census tract level, dwelling, household, family and person.
        b) After a random start, select every nth private dwelling (our preference is every tenth, yielding 10% samples).
        c) For institutional dwellings, after a random start, every nth person using the same density as for private dwellings. If families are identifiable, include the family of the sampled individual with the sampled person indicated.

Example: Colombia, the first country to be integrated in the IPUMS-International system, is a good example of the variety of samples harmonized by the project with the cooperation of the national statistical agency (DANE).

- 1964 - every 50th individual (supplied by DANE, as constructed in the late 1960s using punch-cards!—remember, we are historians; no data are too old or rudimentary to be ignored);
- 1973 - 10% of dwellings drawn by MPC in 2001 (from microdata recovered by the project);
- 1985 - 10% long-form drawn in the field by DANE (as constructed in the 1980s);
- 1993 - 10% dwellings drawn by MPC in 2001 (from 100% microdata supplied by DANE).

Please see: http://www.ipums.umn.edu/international/sample_designs.shtml for designs for China, Colombia, France, Kenya, Mexico, and Vietnam. Brazil will go on-line shortly (long forms, 1960-2000)