

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)