

IPUMS/CED Harmonization of European Labour Force Survey Metadata and Microdata

IPUMS Minneapolis and CED Barcelona have begun to harmonize Labour Force Surveys of cooperating European National Statistics Offices in a global, scientific use, extract management system.

The IPUMS/CED initiative complements conventional, whole-file dissemination with a twenty-first century system offering:

1. open access interactive metadata
2. restricted access harmonized microdata via bespoke, pooled extracts—for selected countries, years, variables, and even subpopulations—tailored precisely to the needs of each individual researcher or research team.

The new IPUMS/CED initiative extends the highly successfully Census Microdata Series, currently encompassing 98 countries, with a world-wide Labour Force/Household Series piloted for 5 countries: India, Italy, Nigeria, Spain, and the USA.

Cooperating National Statistical Office entrust data directly to IPUMS/CED under the terms of a uniform Memorandum of Understanding endorsed by more than 100 NSOs worldwide (see example below).

The IPUMS/CED initiative adds value in many ways. For example:

- Pooled, custom datasets. Users can create extracts that pool data into a single file containing only their desired variables and including as many samples and countries as they wish. They need not download potentially hundreds of files to conduct largescale comparative research.
- Easy exploration of metadata. Integrated variable documentation provide descriptions, comparability information, codes, and frequencies in one place for all samples. Users may easily assess whether particular categories have sufficient cases for their research or to identify major complicating factors due to some peculiarity of a particular sample or country. Questionnaire wording is associated via metadata to the relevant variables, so users can easily see how different wordings might affect the data in specific time periods or countries.
- User-defined variables. The extract system allows users to create new variables using the sequence number variables for parents and spouses. Thus, a researcher can make a variable for co-resident mother's occupation or spouse's educational attainment as part of their data extract.

- Exploratory on-line data analysis. The website offers a full-featured online data analysis system to registered users using the SDA software that has been applied to numerous IPUMS collections. The system is more than a tabulator. It allows multiple cross-tabulations; multivariate regression analysis; has many statistical features such as calculating confidence intervals; and it allows users to leverage the full detail of the microdata by recoding variables and applying case selection.

Other benefits. There are additional meta-level advantages for the IPUMS/CED initiative. The system is designed for the easy incorporation of new samples in the future and can accommodate changing codes via metadata-driven harmonization. We will create regional labour force data collections for other world regions using IPUMS principles, which would make it easy for researchers to do truly world-spanning research. From the start, European researchers may take advantage of the IPUMS-CPS database that harmonizes the U.S. employment survey across time. New variables will make those data even more easily interoperable with the European collection.