

WP. 8
ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

Joint UNECE/Eurostat work session on statistical data confidentiality
(Bilbao, Spain, 2-4 December 2009)

Topic (i): Harmonization of statistical data confidentiality – legal and methodological aspects

**HARMONISED LABOUR FORCE AND MIGRATION STATISTICS
BASED ON MICRODATA**

Supporting Paper

Prepared by David Brackfield and Nicolas Ruiz, OECD

Harmonised Labour Force and Migration Statistics Based on Microdata

David Brackfield¹ and Nicolas Ruiz²

1. Principal Administrator – Short-Term Economic Statistics, Statistics Directorate, OECD, 2 rue André Pascal, Paris 75016, France, david.brackfield@oecd.org

2. Statistician – Household Income and Microdata, Statistics Directorate, OECD, 2 rue André Pascal, Paris 75016, France, Nicolas.ruiz@oecd.org

Abstract: This document is based on a proposal to undertake a harmonisation project of labour force and migration statistics based on microdata datasets among the OECD member countries, and the set-up of a remote access interface. The project (like the proposal) is split into two phase: Phase 1 will investigate the feasibility of the project by studying the legal and technical issues that could potentially arise, while Phase 2 is the concrete implementation. Thereby, this document is a kick-off for Phase 1.

1 The Proposal

In early 2009 a well known Foundation in the United States (which supports work in about 60 countries around the globe) was approached by the OECD to see if there was any interest in supporting an investigation into building a Labour Force Survey microdata database to undertake serious economic and social analysis of migration. The Foundation indicated that it would be willing to support a first phase investigation into the feasibility of building such a database and asked that a formal proposal be put forward. What follows in this paper is a reduced version of the final proposal that was agreed on between the OECD Statistics Directorate and the Foundation. All things being normal, the grant should begin in December 2009.

2 Introduction

The project aims to provide an interface to access Labour Force Survey (LFS) microdata statistics and able to offer for interested users and labour economists detailed labour force data (including data on migrants in the labour force). This was one of the main recommendations contained in the recent report of the Commission on “Migration Statistics” organised by the Global Centre for Development (CGD). The project is split into two phases, the first phase was described as a general stocktaking of what is available; Phase II is implementation.

This document deals with Phase I of the project. Clearly Phase I needs to provide a solid base for the full implementation (Phase II) and in this respect a number of tasks associated with Phase I have been determined directly as feeding from the requirements of Phase II. It should also be noted that the field of microdata (access, collection, storage and analysis) is a new one for International Organisations and currently there are a number of initiatives underway within and outside the OECD. The present project will draw, to the extent possible, on the lessons learned from these various initiatives. Internally the project will be undertaken by two Directorates within the OECD: the Division for Structural Economic Statistics, of the Statistics Directorate (STD); and the Division for Employment Analysis and Policy, of the Directorate for Employment, Labour and Social Affairs (ELS).

In a broad sense STD will provide the knowhow and direction for the statistical aspects of the project such as the microdata classification, database construction etc. ELS will provide input into the labour statistics themselves, for example theoretical understanding of LFSs, current data coverage, and access to the relevant community of labour force statisticians and economists. As deadlines are tight for this project (one year for Phase 1), the focus will be maintained on access to LFS microdata for countries outside the European Union as the EU countries are harmonised via

statistical regulations passed by the European Council and Commission. These OECD member countries outside the EU include Canada, Mexico, the United States, Australia, Japan, Korea, New Zealand, Switzerland and Turkey.

3 What is a Labour Force Survey?

Labour force surveys are rotating random sample surveys of persons living in private households. They are designed to collect information on labour market issues by personal and telephone interviews. The samples of households are drawn from Census addresses databases and are stratified geographically. The results, while robust for most of the labour market and individual characteristics of the population covered, can be subject to greater sampling variability at detailed level of disaggregation due to small sampling base at those levels.

The main objective of a labour force survey is to provide descriptive and explanatory data on the labour market status of persons aged 15 (or 16) years and above in order to develop, manage, evaluate and report on labour market policies. The survey collects information on a broad range of personal, labour market and related characteristics of persons in the labour force, whether employed or unemployed, and those not in the labour force; also called not economically active. The following topics are usually covered in labour force surveys: demographic background, migrant status, education and training, labour status, professional status, type of employment contract, industry and occupation, job tenure, hours worked, underemployment, unemployment, unemployment duration, attachment to the labour force, previous work experience, reasons for leaving last job, union status, wages, etc.

Among those topics, migrant status is characterised by either one or all three questions on: country of birth, nationality, number of years of residence in the host country. Moreover, the survey coverage is a key source to describe and analyse the labour market situation of migrants and other of household's members (sometimes including the place of birth or nationality of their parents).

OECD activities on international migration comprehensively rely on labour force survey results for comparing labour market performances of native and non-native persons by analysing their labour market outcomes, their exposure and vulnerability to unemployment and joblessness, etc. Labour force survey results are also thoroughly used for labour market analyses in various OECD policy and analytical reports. The value added of this project would be the creation of a unique “virtual repository” of microdata derived from labour force survey, which will dramatically enhance the possibility to compile meaningful and internationally comparable statistics on migration and the status of migrants.

4 The Harmonisation of Labour Force Surveys

Over the past three decades, the content of labour force surveys in OECD countries has become more uniform. Almost all now apply the International Labour Organisation (ILO) guidelines on the classification of labour force status and collect information on activity during a reference week, on job search or impediments to it and on job characteristics (hours of work, sector, occupation, job tenure, etc.). LFS follow the Recommendation of the 13th International Conference of Labour Statisticians, convened in 1982 by the International Labour Organisation.

At the EU level, the Statistical Office of the European Communities (EUROSTAT) is responsible for collecting quarterly results of continuous national labour force surveys. National statistical offices (NSO) are responsible for selecting the sample, preparing the questionnaires, conducting the direct interviews among households. The results are then reported to Eurostat in accordance with a harmonised structure of variables using a common coding scheme – so called European Labour Force Survey (EULFS) – as stipulated by a European Council Regulation (577/98) and by several other Commission Regulations defining the list of variables and the technical aspects for data collection. Eurostat is responsible for processing and disseminating the information sent by the national statistical offices. The EULFS includes variables to characterise migrant population and their work status namely country of birth, nationality, number of years of residence in the host country.

Completing the harmonisation procedure to extend the coverage to all OECD countries requires access to labour force survey micro-datasets (microdata) for the remaining eight non-European OECD countries – Australia, Canada, Japan, Korea, Mexico, New Zealand, Turkey and the United States. The OECD has direct or indirect access only to some versions of LFS micro-datasets of Canada, Mexico and the United States. These datasets are currently used for various labour market analyses of policy relevance. However, the relevant variables for characterising immigrants may be not always be available in national labour force surveys. Moreover, even in the case they are, those variables are not necessarily reported in the public use file (PUF) versions of the LFS micro-datasets available to ordinary users. Further investigation is therefore required.

4.1 Different forms of Microdata Access

What follows is a description of the various ways a NSO can make microdata available to interested users. In virtually all cases the NSO is required to undertake at least some work (transformation) on the core microdata so that it is understandable (usable) to users outside their organisation and it should not be forgotten that in many cases there is little incentive to the NSO for this.

4.2 Statistical Tables and Data Cubes

This can include both standard tables and special tables (or special analyses for that matter) generated at the request of the researcher. Some NSOs now release very detailed matrices, known as data cubes, which researchers can manipulate to support their own needs.

4.3 Anonymised Microdata Files (AMFs) - Public Use Files (PUFs)

These are microdata files that are disseminated for general public use outside the NSO. They have been anonymised and are often released on a medium such as CD-ROM, sometimes through a data archive. Note: The term anonymised implies that not only are names and addresses removed, but other steps are taken (e.g. collapsing of geographic details) to ensure that identification of individuals is highly unlikely.)

4.4 Anonymised Microdata Files- Licensed Files

Licensed files are also anonymised but are distinct from Public Use Files in that their use is restricted to approved researchers and an undertaking or contract is signed before files are provided to the researchers. Even if advertised as generally available to the public, they are not released before an undertaking or contract is provided by the researcher. Even though anonymisation and other steps are taken to ensure that identification of individuals is highly unlikely when used in isolation, they may contain potentially identifiable data if linked with other data sources; this is one reason why a preventive undertaking or contract is required. There may be other conditions of use that the NSO may impose on researchers.

4.5 Remote Access Facilities (RAFs)

Arrangements are now being made in many countries that allow researchers to produce statistical outputs from microdata files through computer networks, without the researchers actually 'seeing' the microdata. Because of the additional controls that are available through RAF, and the fact that microdata do not actually leave the NSO, access to more detailed microdata can be provided this way.

There are two basic types of RAF:

- Remote execution, where a researcher submits a program and receives the output later over the Internet. The researcher doesn't see the any data, they submit a program code that will then be performed by the NSO on the data; output is checked and made available.
- Remote facilities, where the researcher performs the analysis and can immediately see the answer on the screen. In effect the researcher can 'play' around with the data on screen. However the output is checked by the NSO before it can be downloaded.

These facilities are becoming increasingly important but the way RAFs are implemented varies considerably from country to country. The key characteristic is that researchers do not have access to the microdata themselves: rather tasks are submitted remotely over the Internet. Often there is a contractual arrangement between the NSO and the researcher or the institution of the researcher.

By way of illustration, a NSO could provide researchers with dummy microdata files and allow them to submit a program against the full file via the computer network. The NSO then run the requests offline and sends the results back via computer networks after checking for confidentiality.

4.6 Example of what has been done internationally already

The Luxembourg Income Study, known as LIS, is a non-profit microdata archive and research institute. LIS, located in Luxembourg since 1983, serves a global community of researchers, educators and policy makers. LIS acquires datasets on household income and demographic characteristics from a large number of countries, harmonizes them to enable cross-national comparisons, and makes them available for public use by providing registered users with remote access. The LIS archive includes two primary databases, the Luxembourg Income Study Database, which focuses on income data, and the newer, smaller Luxembourg Wealth Study Database, which focuses on wealth data.

5 Task List

Clearly in undertaking this project a number of tasks will need to be carried out to set the groundwork and give a clear direction of what is feasible and under which conditions.

5.1 LFS Availability

The first stage of the project would need to cover what LFS microdata is required to undertake the detailed analysis envisaged e.g. migration and the labour force; and what data is actually available for each country. Evidently the first requirement would heavily influence the second. The project would need to consult interested users and analysts of labour force data (including those inside the OECD) to understand what data fields would be required to meet their needs. This would involve in-depth communication and possibly an international workshop to fully gauge and cover the range of possibilities.

Once what data variables that are required to fulfil the needs of the analysts are well understood, and then the project would need to line up this requirement with what NSOs collect via their LFS questionnaires. This would require investigating LFS questionnaires of OECD member countries, the way the LFS questionnaire data is stored, compiled and disseminated.

5.2 LFS Current Situation in the OECD

The next major task that would need to be completed, to start implementation, is a full assessment of what has already been done in the field of LFS statistics and microdata at the OECD. The OECD has a long association with collecting tabulated LFS data through country questionnaires and direct country access and this data. However in recent years more focus has been given to accessing microdata to undertake analysis at a more detailed level.

The project will need to focus on fully understanding what has already been done in this field, what LFS microdata is currently available and how they could be used for this project. It will also need to investigate what agreements with countries have been put in place and how ongoing microdata delivery (if any) is taking place. There will also need to be some coverage of what tabulated LFS data already exists at the OECD and how this possibly could be used to aid the project.

5.3 Microdata

The project will need to look into and fully understand the concepts and operational aspects of microdata and in particular the issue of remote access. While not a new idea, it would be fair to say that in recent years the power of using microdata at an international level has gained ground and the OECD is currently involved in a number of initiatives at this level. A theme to emerge from this investigation into microdata is the issue of remote access, in other words the ability to access microdata held in National Statistics Office's (NSO) without having to physically receive (or download) the datasets. The project would need to look at the operation aspects of remote access e.g. legal, administrative; and the technical issues such as information technology requirements, for example software and hardware.

The project would also need to gauge how access to the microdata could be undertaken either directly or remotely, the technical issues that would need to be resolved and tools required to perform such access. The issue of designing a 'microdata interface' would also need to be investigated.

5.4 LFS Microdata Matrix and Report

The final stage of the project would require the construction of a matrix linking all the aspects of the investigation: data requirements, data availability, current situation in the OECD, remote access possibilities and the building of a generic interface.

The final report, as well as building on the matrix, would clearly describe how a harmonised labour force and migration database based on microdata from labour force surveys could be built. It would need to provide a clear understanding of what

could realistically be achieved in the current environment within a reasonable time period.

It would also need to provide a detailed estimate of the costs involved and the resources that would need to be sourced to ensure a successful outcome in Phase II of this project.

6 Timeline

6.1 December-March 2010

Complete investigation into variable requirements for analysis
Complete investigation into country questionnaires and data practices
Complete investigation into current situation regarding LFS data in the OECD

6.2 April- June 2010

Complete investigation into LFS microdata, remote access and the possibility of a generic interface

6.3 June 2010

Workshop

6.4 June-September 2010

Completion of LFS matrix
Completion of final report