

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

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)

ORGANISATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT
(OECD)
STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 9-11 February 2004)

Topic (ii): Metadata interchange

ILO EXPERIENCE WITH GATHERING AND DISSEMINATING METADATA ON HOUSEHOLD INCOME AND EXPENDITURE STATISTICS

Contributed Paper

Submitted by the International Labour Organization¹

Summary: This paper summarises ILO experience in the gathering, processing and dissemination of metadata on household income and expenditure statistics.

Introduction

In late 2002 and early 2003, as part of its preparations for the 17th International Conference of Labour Statisticians, Geneva, 24 November to 3 December 2003, the ILO took action to update the information presented in the ILO's *Sources and Methods* publication that covers household income and expenditure statistics (HIES)². The results from the last systematic updating of this publication were published in 1994.

Previous practice

The ILO continually monitors the methods used by countries to produce labour statistics. These methodological descriptions are included in the on-line statistical database LABORSTA (laborsta.ilo.org) as well as in the ILO's *Sources and Methods* publications. The methodological descriptions also form the basis for footnotes to national statistics presented in ILO's *Yearbook of Labour Statistics* and other statistical publications.

In addition to this, efforts are made at infrequent intervals to ensure that the methodological descriptions on a particular topic are complete and up-to-date for all countries. This updating is generally carried out in rotation for different topics.

¹ Prepared by Bob Pember (pember@ilo.org).

² In the division of labour between statistical units within the UN system, ILO serves as the custodian for household income and expenditure statistics.

In the past, the practice has been:

- (a) for countries that have previously submitted a description: to send them the previous text and ask for a revision or update;
- (b) for countries that have not previously submitted a description: to ask for a description of the source.

This process takes time and places a burden on both the ILO Bureau of Statistics and the national statistical offices.

The resulting descriptions are translated into the other two of the three working languages of the ILO (English, French and Spanish) and converted into electronic form for dissemination, both in print and in electronic form, e.g. on the ILO's websites. These descriptions do not lend themselves to easy analysis of differences in concepts, definitions and techniques.

Latest method

In an effort to reduce the burdens in reporting, processing, translation and reformatting for dissemination, the ILO Bureau of Statistics decided to use a new approach for updating the meta information on national HIES, in which:

- (a) A questionnaire was designed with tick boxes offering a choice of answers for all main features to the extent possible. (Offering a choice of answers simplified and speeded country reporting. As the design had to allow for all possible alternatives it required sound knowledge of likely methods. "Other, specify" options were provided throughout.);
- (b) This questionnaire was tested in one country. This resulted in some revisions to the first draft;
- (c) The finalised questionnaire (amounting to over 80 questions plus annexes) was translated into French and Spanish;
- (d) By oversight, codes were not pre-printed alongside each tick box in the final version, and this slowed data entry, especially for those questions where there were many alternatives;
- (e) A set of complex computer programs were developed to:
 - i. Permit data entry into a database³ of codes representing country responses (and accompanying text where the tick box responses were not sufficient);
 - ii. Carry out automatic checks on data entry for valid codes, coherency between answers, etc to reduce errors during data entry;
 - iii. Allow extensive data analysis (cross-tabulations) of the country responses. These analyses were automatically updated as each new set of data was added to the database;
 - iv. Produce output in HTML format of the textual descriptions of each source in three languages. Again these were automatically updated as each new set of data was added to the database;
 - v. This HTML output could then be directly inserted into website descriptions or used in printed publications.

³ There were in fact six separate databases each covering a different part of the questionnaire, but for ease of presentation, this paper refers to one combined database. The six databases were linked by a common variable, which was the identification number of the methodological description.

- (f) For each question, different output descriptions were prepared in the three languages reflecting the variations allowed for by the tick box format. In order to create a full description of a particular national HIES, these separate individual sentence elements were then concatenated in such a way that any source could be described. As indicated in (e)(iv) above, these elements were used to generate text in all three languages as part of the computerised output;
- (g) The database that was created allowed extensive cross-classification and analysis. (Significantly more than previous methods had permitted.) This database will be made available to external users on request.

Resources required

Previous work by the Bureau of Statistics to generate publications of this type required many months of clerical preparation (and professional involvement) of a methodological description for each country/source separately and then the translation of each of these into the other two languages. The total cost of this work for each publication has not been fully estimated but it must have been close to \$US 100,000 because the translation process alone amounted to well over \$US 45,000 (external translators) for each publication. It was the extent of translation that the above computerised system was aiming to replace.

The total cost of the new system (producing a publication as well as a database for more detailed analysis) was just over \$40,000, comprising:

- (a) one work month of questionnaire design and computer specification (in-house, estimated at \$US 10,000);
- (b) one work month of translation (mainly in-house, estimated at \$US10,000);
- (c) three work months of computer effort (in-house, estimated at \$US 15,000); and
- (d) two work months of data entry (\$US8,500).

It should be noted that the investment in the computerized system will benefit the production of future similar outputs.

Experience gained

In practice, it was found that most countries added comments and qualifications to what had been expected to be adequate response alternatives. This complicated and lengthened the data entry and added a considerable burden to the translation process. The translation of these additional (unexpected) comments will take place shortly.

It is believed, nevertheless, that the approach facilitated and improved country response and that the total translation burden was reduced with consequent savings for the ILO. In addition, the use of a computerised database permits a much richer analysis of methods and facilitates electronic dissemination to external researchers.

However, the approach required more care in designing the questionnaire and its pre-coded responses than with the previous approach. For the future, it would be prudent to design the questionnaire and database to allow comments in respect of all responses and to permit multiple responses to most questions. It has also been proposed that the resulting questionnaire might be prepared and dispatched to countries in an electronic format in order to avoid delays due to postage⁴.

⁴ The use of electronic questionnaires and dispatch may not be possible or effective in all cases. Occasionally, the contact e-mail address is not available and/or more than one agency is involved in completing the questionnaire.

Conclusion

Would we do it again? We are already doing so in preparing methodological descriptions for statistics on the employment situation of persons with disabilities, which however, are based on a much shorter, simpler questionnaire that has received similar types of replies. In balance, we believe that this approach has risks and high computing costs but it is worth repeating for the next round of methodological description updates, with details modified in light of the experiences gained.