Committee for the Coordination of Statistical Activities

Conference on Data Quality for International Organizations
Rome, Italy, 7 and 8 July 2008

Session 1: Quality assessment - use of checklist; user surveys; collection of quality metadata; disseminating quality information; training in quality management

Assessment of statistical data quality: The example of the Wages part in the ILO *Yearbook of Labour Statistics* [1] (chapter V)

Lê Anh Hua, ILO

The aim of this paper is to provide information to persons interested in using ILO statistics of *wages*, currently published in the ILO *Yearbook of Labour Statistics* and disseminated online in the Laborsta website at http://laborsta.ilo.org. This paper describes the main characteristics of these data as well as their main limitations regarding comparability over time and between countries. These are examined in the context of the ILO responses to the questionnaire on "Self-Assessment Checklist for Quality Assessment of Statistics compiled by International and Supranational Organisations" developed for the CCSA by Eurostat. Many users of the wages statistics and the associated metadata express their interest and satisfaction when consulting and making use of the wage statistics disseminated by the ILO. However, some users report that they find it difficult to compare these statistics between some countries because of differences in (i) currency units, (ii) time units, (iii) the coverage in terms of jobs, and most importantly, (iv) concepts (i.e. the components which are covered in the reported 'wages'). It should be noted that while the ILO Bureau of Statistics takes steps to improve the quality and comparability of the data, the main problem with the statistics of wages disseminated by the ILO relates to the *gaps* in the available statistics. There has been some improvement over time thanks to the close cooperation with the countries, ILO field offices, regional efforts to harmonize statistics (e.g. by Eurostat), and extensive use of data disseminated on national statistical websites. However, still many countries do not produce statistics on wages or when they do, they do not always communicate them to the ILO either because of concerns with quality or for another reason. Some users' requests that are not yet satisfied by the ILO relate to 'indices of wages',

[1] also available on-line in Laborsta, http://laborsta.ilo.org and can be downloaded free-of-charge

'real wages' 'wages by age group', 'wages converted into a common currency', or 'wages based on a yearly time unit'.

1. International standards and guidelines

In order to assist countries in compiling their statistics on labour topics, including wages, the ILO has established many standards and guidelines. These take the form of ILO Resolutions which are adopted in the International Conference of Labour Statisticians (ICLS), the Labour Statistics Convention No. 160 (1985) (which supersedes the Convention concerning Statistics of Wages and Hours of Work, 1938), and the Labour Statistics Recommendation No. 170 (1985), which supplements the guidelines in Convention 160. The aim is to provide the best guidance, given the current state of knowledge and experience, to help countries produce the types of statistics they require for their own purposes, at the highest possible level of quality. Additionally, by following the international guidelines, countries will improve the comparability of their data with those of other countries.

These international standards and guidelines are the framework for the ILO's activities in each field of statistics, including data gathering and dissemination, and technical assistance.

Countries that have ratified Convention 160 are required to report periodically to the ILO their application of the various articles, and the ILO in turn provides comments to these countries concerning the data compiled under the terms of the Convention. This channel of on-going dialogue represents an important component in ILO's effort to assist countries to improve the quality of their statistics.

2. Questionnaires and Data providers

The national statistics and the associated meta-data are collected by means of questionnaires prepared in the three official languages of the ILO (English, French and Spanish). They are either printed or computer readable Excel files and are addressed to National Statistical Offices (NSOs) and/or relevant Ministries, mostly Ministries of Labour (MoLs). The questionnaires contain statistics for previous years when available, with the request that respondents insert new/updated statistics. When the completed questionnaires reach the ILO Bureau of Statistics, they are processed as quickly as possible to allow time for any follow-up. Automatic reminders are sent out to countries which have not returned the completed questionnaires by the requested deadline.

3. Statistics and Metadata

The ILO collects statistics on *gross* average earnings or, when these are not available in the country, wage rates,[2] per worker or per job, by type of economic activity (using tabulation categories (alpha codes) of the International Standard Industrial Classification, in its second or third revision, ISIC rev.2 or ISIC rev. 3) and sex. In addition, for manufacturing, statistics on average earnings or wage rates are also requested by division (two digit codes).

The first pages of the questionnaire contain instructions on the concepts and definitions of "earnings" that should form the basis of the statistics reported. These correspond to the recommendations in the latest ICLS resolution on the subject. If it is not possible to provide figures on **earnings**, then the country may provide statistics on "**wage rates**". The various components of remuneration to be included in these concepts are listed. Some explanations are also given relating to the possible data sources as well as the preferred time unit (ideally earnings per hour) and coverage (ideally, the whole country, all employees and all industries).

In the questionnaire, information is requested in the first table concerning the department or agency completing the questionnaire and coordinates of the person who may be contacted for further information. When a country has provided statistics in previous years, the questionnaire presents these together with information provided about the concepts, the time units and the coverage, as well as the sources (name of survey, type of administrative record, etc), the reference periods and the national publications or website, if any, in which the statistics appear. The respondent is requested to modify this information if it is incorrect or update it if it has changed.

Each questionnaire contains columns covering statistics for ten years: the current year (e.g. 2007 collected in 2008) and pre-filled data (when available) for the previous years. Thus reporting agencies can check the statistics previously provided, and if necessary amend or revise them.

The statistics supplied may be drawn from a number of sources, such as establishment surveys, household surveys or administrative records.

Only the most essential m*etadata* are requested in the questionnaire, i.e. methodological information is not collected beyond what is needed to distinguish between series. This information is introduced

[2] For the definitions of these concepts, see http://www.ilo.org/global/What_we_do/Statistics/topics/Income/lang--en/index.htm ; for the full text of the Resolution concerning an integrated system of wages statistics adopted by the Twelfth International Conference of Labour Statisticians (Geneva, 1973), see http://www.ilo.org/wcmsp5/groups/public/---dgreports/---integration/---stat/documents/normativeinstrument/wcms_087496.pdf.

in the database either as variables or as footnotes explaining e.g. a change in currency or major methodological changes. More comprehensive metadata are collected separately from this exercise, for the preparation/updates of ILO's "Sources and Methods" publications. The information in these publications is also made available through Laborsta.

The number of countries responding to the ILO questionnaire on wages statistics varies from year to year, and, for the reasons given in the introduction, the rate of response - the response rate for these statistics has been calculated as the proportion of the dispatched relevant part of the questionnaires that were returned, regardless of how many series were submitted by the country - tends to be 35 per cent or less (see below):

2007 34%
2006 32%
2005 31%
2004 29%
2003 34%
2002 32%
2001 35%
2000 32%
1999 31%

The introduction of e-mailed Excel questionnaires has led to improvements in response in recent years, as illustrated by these numbers. In 2007, about 85% of all responses were returned by e-mail.

4. Analysis of the statistics received and steps taken to ensure reasonable comparability over time

The data received by the ILO Bureau of Statistics (STAT) are examined with regard to the (i) consistent use of **sources** (some countries provide many wage series from different sources and within these, for different time units); (ii) **trends** over time and (iii) whether or not a change in **currency** has occurred.

Some wages statistics may not be used for the *Yearbook* for reasons including the following: they relate to *yearly* wages (i.e., if the time unit is a year rather than a period such as an hour, day, week or month); they represent *total* wage sums for the group of jobs covered (with no indication of the number of jobs) instead of average wages per worker or job; they represent *forecasts* instead of

actual averages; they are presented in *constant* values instead of current ones; the statistics are *not classified* according to ISIC Rev.2 or ISIC Rev.3; it is not possible to translate the national categories to corresponding ISIC categories.

It should also be noted that the quality of reporting by some countries leaves much to be desired. For example, some have provided average wage data for aggregate industry groups which clearly correspond to the simple averages of the average wage data for the component industry groups. This may be because the respondents did not fully understand the instructions in the questionnaire, or were not trained in statistical methods.

4.1 Use of assessment list prior to data entry

The following items are considered:

- Whether the statistics come from an official source: only official statistics are accepted; i.e. those produced or endorsed by an official government agency.

- Whether the figures provided in a table come from the same source, have the same coverage, and use the same time unit and same currency units.

- When new series have been provided, and the previously reported ones have not been updated, then it is necessary to enquire with the reporting country whether the old series has been discontinued and been superseded by the new one.

- When figures received show a significant upward/downward trend or shift, without any explanatory information, e.g. change of coverage or high inflation as shown by the available Consumer Price Index, then it is necessary to enquire with the reporting country about the reasons for that change.

- When a series is provided with some divergence from the main concept, then footnotes need to be added to provide the relevant explanations.

- As much as possible, national websites and publications are searched for missing or corroborating statistics .

- When a national classification is used by the reporting country which does not correspond to the ISIC, then correspondence tables between the two classifications need to be established or found.

Follow-up with respondents is needed in about 20% of countries in the case of the wages statistics provided for the *Yearbook*. This follow-up is carried out mainly by mail (traditional or e-mail) or by telephone. E-mail has improved correspondence with some countries considerably, because response is not always forthcoming. ILO country or area offices or ILO correspondents are asked to help in some cases, or ILO colleagues on mission or working on technical cooperation projects can provide valuable assistance.

4.2 Types of validation tests

After entering the statistics into the database, the following automatic validation and consistency tests are carried out:

- the average wage reported for economic activity D (manufacturing) in table 5A should be *identical* to the "total" in table 5B, which stands for "average wages in manufacturing as a whole";
- the average wages reported for *total* (men and women combined) should not exceed the maximum of men and women considered separately or should not be lower than the minimum of men and women considered separately
- the average wages reported for men and women separately should not be equal to the arithmetic mean of that related to men and women combined (except in the case of an identical wage rate for men and women separately)

When inconsistencies are important, queries are taken up with the reporting country.

4.3 Production of internal quality indicators

All the main correspondences, queries with countries, internal codes for dissemination, footnotes are electronically documented. Statistics on the number of series updated/modified/revised are produced on a monthly basis. Output tables showing the number of series (out of the total number of series) which still need to be verified, based on results of checks of coherence between the source of the statistics provided and the source described in volume "Sources and Methods" are also generated.

5. Publication/Dissemination

The statistics resulting from the above exercise are published/disseminated without adjustment, i.e. they are presented in the national currency and in nominal terms. No conversion into a common currency is carried out, even if this might facilitate comparability between countries for some purposes. It is also unavoidable that there are some differences between the concepts used, time units, reference periods, types of sources and methods of data collection used in the various countries and that these may influence the comparability of the statistics, both when comparing the level of wages between groups in different industries or in the same industries in different countries and when comparing the changes that have taken place between two years. For some countries there is a significant time-lag between the reference period for the statistics and their publication. For example, ideally, countries should send in 2008 data referring to 2007, but in practice, many countries are only able to send statistics for periods prior to 2007.

6. User consultations and satisfaction

The questionnaire for a *User Satisfaction Survey* was placed on the Laborsta website in November 2007. However, very few responses have been received, and it has not yet been possible to obtain valid results from them.

Users can consult "FAQ"s on Laborsta for responses to some of the questions.

7. Future work

Plans are underway to collect data for the *Yearbook* by means of *electronic* questionnaires, from which data will be uploading automatically into the ILO's database. This will reduce significantly the time devoted to processing the data and the risk of errors in manual data entry. However it should be noted that analysis of the replies and manual entry of codes and notes (metadata) will still be required.

The ILO is also working with the OECD and EUROSTAT on the use of a common computer language for the compilation and exchange of labour data and metadata. This exercise might address some of the problems relating to gaps in the series.

Another plan relates to updating methodological descriptions for the series published and disseminated.