Session 3: Tools and policies for collecting, managing and disseminating data and metadata (1)

ILO’s Labour Market Indicators Library Network (LMIL) and Key Indicators of the Labour Market (KILM) as tools for collecting, managing and disseminating data and metadata

Prepared for the Conference on Data Quality for International Organizations, Wiesbaden, Germany, 27-28 May 2004

Sara Elder and Isabelle Guillet
Employment Trends Unit
Employment Strategy Department
International Labor Office, Geneva
1. Introduction

Since 1999 the International Labour Organization in Geneva and the field have developed and refined the process of assembling, transmitting and disseminating labour market indicators via ICT tools such as the Labour Market Indicators Library Network (LMIL) system, the Key Indicators of the Labour Market (KILM) as well as the database of the Bureau of Statistics (Laborsta). The overall objective of all projects is to contribute to an increase in the use of labour market information to formulate economic policies and to monitor employment while also increasing the availability and timeliness of labour market information.

The LMIL system works as an internal Internet-based database system developed to allow the electronic transfer of information from the field to headquarters, and vice versa, in a manner that complies with ILO methodological standards. Data transferred, including the latest updates, via the LMIL system feeds into ILO products such as the KILM which relies on the availability of the most up-to-date labour market information. While the LMIL is an internal LMI gathering, validation and management tool, the KILM database and software are tools whereby the end-user can query and manage data according to specific metadata criteria.

The paper will be organized according to the functionality of the ICT tools, first outlining the updating and management capability of the LMIL system and then the disseminating advantages of the KILM.

2. The Labour Market Indicators Library Network (LMIL) as LMI collection and management tool

What is it?

The overall objective of the LMIL Network is set out in the introduction. The strategy of the Network consists of consolidating a knowledge-based network by providing appropriate products and technical expertise, while at the same time increasing the availability and timeliness of labour market indicators. In this respect, the LMIL Network assists ILO sub regional offices (SRO) in strengthening national and regional capacities to produce, analyze and disseminate LMI with a view to better meet policy makers needs to assess employment trends and poverty reduction. The PARIS21 Trust Fund on Statistical Capacity Building, managed by the World Bank, supports the development of the LMIL Network through 2004, including conducting a series of training workshops in the field of LMI and assisting countries to strengthen or develop dissemination tools. At present, the project is implemented in five regions (Western Africa, Central Africa, Central and Eastern Europe, East Asia, and Central America).

An Internet-based database system has been developed to support the LMIL Network. It allows the electronic transfer of indicators and methodological information from the field to headquarters, and vice versa, in a manner that complies with ILO
methodological standards. The system is currently designed and housed within the ILO Employment Trends (TREND) unit but functions in close collaboration with the Bureau of Statistics and the ILO field offices.

The LMIL is a multi-user and multi-functional ICT tool that manages the internal transfer of labour market information within the ILO. Firstly, it serves as an organization tool by housing all ILO-based sources of labour market data in one database thus making it easier to streamline and identify where gaps exist, to increase the consistency of LMI within the house and to avoid duplications of efforts. The shared database will serve as a one-stop-shop for both data and metadata for ILO officials who will extract and disseminate the data for their own purposes.

Secondly, the LMIL has a data entry and validation component that will insure quicker, more accurate, more transparent, and less burdensome processing of data. This feature allows having a “live” shared dataset between several users. Thirdly, beyond the ICT tool functionality there is the added benefit of setting up an in-house communication network on the topic of labour market information. Participants of the LMIL project interact to discuss topics such as data quality, standardization of definitions, selection of indicators, etc.

LMIL currently offers the following functions (on the local/field level) to facilitate the in- (display) and out- (updating) flow of information:

Retrieval and display features
• downloading via the Internet the current data tables into a local LMIL database (that is, a replication of the LMIL database housed on the headquarters sever);
• querying – defining the criteria and retrieving the targeted indicator data;
• presenting data – generating graphs and reports;
• transferring data files or graphs to, or copying and pasting data and graphs into other software applications;

Updating data feature (SRO level)
• entering regional data into the local LMIL database to fill in the gaps for countries that are under the mandate of the concerned SRO;
• sending new inputs data to ILO headquarters for validation;
• editing and deleting records that had been entered in the local LMIL database by the concerned SRO;
• adding repositories.

Who enters the data and how are they guided to ensure data and metadata quality?
Data is entered by LMIL focal points that were designated in each participating ILO sub regional office. Each focal point, under the supervision of the SRO senior employment specialist or senior statistician, has or will have been briefed on how the system works and how data should be located and entered. Moreover, a database developer and a LMI specialist from the TREND team are available to provide technical and methodological assistance to LMIL focal points whenever required.

Gathering sources of information
The focal point will review which indicators are available for a country – the view table will include all currently ILO accepted LMI – in order to identify data gaps in the time series for the countries covered by their field office. They will then turn to available local sources with an eye toward locating the missing information. The missing information can likely be found in statistical national publications which may not have made their way to headquarters, local websites, etc., and whenever necessary the LMIL focal point may obtain further information via national statistical offices. LMIL focal points are not authorized to alter any records that currently exist in the LMIL database. When they have alternate sources of data at their disposal for the same indicator and year, they may wish to discuss the quality and validity of the alternative source with other network members. If all groups, SRO, STAT and TRENDS, assess the new data or sources as of better quality, they will replace the existing records.

**Updating the database**

Having gathered national reports and located sources of data and metadata so that he/she is ready to begin the data entry process, the LMIL focal point would make use of the “Add record” feature of the system (eg: ADD RECORD button). The process of adding data is fairly simple and guided by the electronic tools.

There are numerous benefits of using an electronic interface for data entry in terms of insuring better quality of data entry including completeness, accuracy, comparability, compliance with international standards and assessment of the source:

1. The layout of the input grid in itself prompts the data entrant to collect and note the necessary metadata. An entry cannot be submitted until the pre-defined minimally acceptable metadata have been included. The following metadata are required for most tables:
   a. age
   b. type of source (census, labour force survey, administrative record, etc.)
   c. scope of population coverage (e.g. civilian or total)
   d. coverage limitation (any variation of total population coverage is noted – e.g. “excluding persons temporarily laid off”),
   e. reference period (annual coverage is the default)
   f. geographical limitation (entire country coverage is the default; any variation of this – urban areas, etc. – is noted)
   g. remarks (generally explanations of breaks in series – e.g. “Sample design revised.”)
   h. repository (including bibliographic reference)
2. Use of drop-down lists so that the entrant has to choose from a pre-defined list of standard choices. (See figure 1.) The use of predefined dropping lists limits the entrant to make use of “reliable” sources such as aggregates from labour force survey. It also helps the entrant to figure out whether there is any discrepancy between national definitions and international standards. If so, the entrant is

---

1 This constraint is ensured via the use of a primary key composed of the following three variables: country, year and sex.
expected to write appropriate notes to document the definition used at national level.

3. There are programme “tricks” that guide the data entrant, including mouse-overs, links to standard definitions, pop-up boxes that inform an entrant when something has been entered incorrectly.

4. The program includes validation routines that check for basic arithmetic errors, logical inconsistencies - e.g. Do the entries for males and females sum to the entry for both sexes? Is total employment less than total population? Do sub-elements add up to the total? The program’s checks are individualized to each indicator table. (See figure 2.)

5. The program can be set to automatically calculate rates once the necessary variables are entered.

6. When a new repository is entered, it will automatically be added in the drop-down list of repositories used by a particular SRO so that the entrant can select the referenced document and avoid retyping the information (thus avoiding the potential appearance of typos which would result in multiple entries for the same documents).

Figure 1. Adding metadata – drop-down lists
As part of the updating process, SROs also have the possibility of editing and deleting records that are under their ownership. This means that teams can edit and delete records which they had entered themselves even though the entry had been approved by STAT and TRENDS. When they do so, records are resubmitted to the validation process (see below). Indeed, it might occur for example that an input was entered as “provisional data”, which may need to be modified when final aggregates are released.

**Help files**

Detailed help files are available within the LMIL programme. (See figure 3.) Methodological manuals will also be incorporated into the system which will include web links to relevant statistical resolutions and classifications.
**Validation process**

As long as new data have not been approved by the headquarters’ STAT and TRENDS units, the data is visible only to the ‘owner’ SRO. Newly entered data can then be uploaded (sent) to ILO headquarters as soon as the entrant and the senior employment specialist or the senior statistician is satisfied with the data content.

Once transmitted, both STAT and TRENDS units will receive notification of updates and then be required to validate the new data within a one-week period, excepting special circumstances. Common criteria with which data will be assessed for approval or rejection include: completeness, accuracy, comparability, compliance with international standards, consistency of the series over time, assessment of the source and its reliability. Approved data is released - becomes visible - to all LMIL users. If headquarters reject a data entry - perhaps because they question the source, it is missing a necessary piece of methodological information, it does not “fit” within the rest of the time series, or other reason – notice of non-approval will be accompanied with comments and/or further guidance for SRO to improve the data entry and try again. The LMIL system will keep track of the “Not approved” data records and accompanying comments, so that they would not be included in the “Upload” list and re-submitted to ILO headquarters without evidence of modifications or improvements. Improved/corrected data should be uploaded again to ILO headquarters and the validation process repeated.

**How is data managed?**

As stated before, the LMIL system serves principally as a data management tool, housing existing ILO collections of statistics - the Key Indicators of the Labour Market (KILM), database of the ILO Yearbook of Labour Statistics (LABORSTA), collections from field offices) – under one “roof”. It is not an end-user product; rather, the intention is that the participating ILO teams extract the information from the LMIL database and disseminate it for their own purposes. The LMIL in the future, therefore, will be the core database from which end-user products such as the KILM will be developed.

**3. The Key Indicators of the Labour Market (KILM) as data dissemination tool**

**What is it?**

The KILM is an ILO tool for the dissemination of labour market information. The product consists of 20 labour market indicators - each of which is accompanied by text relating the use of the indicator, the sources and definitions the limitations to comparability as well as some basic analysis of the data highlights –, additional background indicators, annexes of employment-related international classifications, and finally a chapter on “key issues in the labour market” in which variable topics of current interest to the ILO are drawn out and analyzed in greater detail. The KILM, currently in its third edition, is available in both publication and CD-ROM format and released semi-annually in September.
Functionality of the electronic product

While we concede that it is necessary to have a printed publication of statistical information – for quick reference, for ease of use, for presentation of trends, for constituents without access to electronic products, among other reasons – there are numerous advantages to producing statistics in electronic format, amongst which are:

(1) Updating capability.

The KILM CD-ROM can be updated via the Internet. Once the programme is installed, users can click on the “upgrade” which will activate the programme to check the files installed on the user’s hard drive against the latest KILM tables on the ILO server and if discrepancy is found – that is, a more recent file on the KILM server than on the hard drive – the user will be notified that newer files exist and will be prompted to download the latest version. (See figure 4.) This means that users of the CD-ROM have access to the most recently available data as it is released from the KILM team, rather than having to wait two years until the next publication is printed. It also allows us to make corrections to records without having to go through a lengthy process of preparing errata sheets and distributing to user lists.

(2) Additional data.

Because an electronic product does not have the space/size limitation of a printed book (although file size has to be taken into consideration), more data records can be included. The KILM CD-ROM includes data from 1980 to the present, whereas the printed publication displays data tables with records for the years 1980, 1990, 1995 and the latest five years after 1995 (a maximum of eight records per country).

(3) User-defined querying.

In using the electronic product, a user can generate a data grid that shows exact what they want to see – stripping the extraneous details that are not crucial to their specific purposes – thus facilitating (and personalizing) data analysis. Upon entering the
KILM CD-ROM, the user is prompted to select the first three basic criteria: (1) Indicator; (2) Country/Region; and (3) Year. The user may select a maximum of seven tables at one time. The default for non-selection of country and year is all countries and all years. The selection will generate a data grid such as the example shown in figure 5 below.

Figure 5. KILM display data grid

<table>
<thead>
<tr>
<th>Country</th>
<th>Sex</th>
<th>Total employed (1000)</th>
<th>Part-time workers (1000)</th>
<th>Part-time employment rate (%)</th>
<th>Female share of part-time employment (%)</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>2001 MF</td>
<td>1.65</td>
<td>0.34</td>
<td>20.6</td>
<td>LML</td>
<td>HS</td>
</tr>
<tr>
<td>Aruba</td>
<td>1994 MF</td>
<td>35.54</td>
<td>2.41</td>
<td>6.7</td>
<td>71 CARIBBEAN HS</td>
<td></td>
</tr>
<tr>
<td>Aruba</td>
<td>1994 M</td>
<td>21</td>
<td>0.72</td>
<td>3.4</td>
<td>71 CARIBBEAN HS</td>
<td></td>
</tr>
<tr>
<td>Aruba</td>
<td>1994 F</td>
<td>14.54</td>
<td>1.71</td>
<td>11.4</td>
<td>71 CARIBBEAN HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1980 MF</td>
<td>6281.9</td>
<td>1376.1</td>
<td>21.7</td>
<td>70.7 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1980 M</td>
<td>3963</td>
<td>315.4</td>
<td>7.9</td>
<td>70.7 OECD/EMC</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1980 F</td>
<td>2295.5</td>
<td>760.6</td>
<td>33.1</td>
<td>70.7 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1981 MF</td>
<td>6333.5</td>
<td>1120.4</td>
<td>17.5</td>
<td>69.2 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1981 M</td>
<td>4057.8</td>
<td>345.6</td>
<td>9.5</td>
<td>69.2 OECD/EMC</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1981 F</td>
<td>2330.7</td>
<td>748.8</td>
<td>33.2</td>
<td>69.2 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1982 MF</td>
<td>6379.3</td>
<td>1157.2</td>
<td>18.1</td>
<td>69.3 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1982 M</td>
<td>4024.4</td>
<td>354.9</td>
<td>8.8</td>
<td>69.3 OECD/EMC</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1982 F</td>
<td>2354.9</td>
<td>602.3</td>
<td>34.1</td>
<td>69.3 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1983 MF</td>
<td>6240.8</td>
<td>1126.1</td>
<td>18.1</td>
<td>69.3 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1983 M</td>
<td>3903.4</td>
<td>330.2</td>
<td>8.7</td>
<td>69.3 OECD/EMC</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1983 F</td>
<td>2337.4</td>
<td>766.8</td>
<td>33.7</td>
<td>69.3 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1984 MF</td>
<td>6468.2</td>
<td>1189</td>
<td>18.1</td>
<td>70.3 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1984 M</td>
<td>4016</td>
<td>359.9</td>
<td>9.5</td>
<td>70.3 OECD/EMC HS</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1984 F</td>
<td>2443.2</td>
<td>628.1</td>
<td>33.9</td>
<td>70.3 OECD/EMC</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1985 MF</td>
<td>6575.6</td>
<td>1202.2</td>
<td>18.4</td>
<td>69.5 OECD/EMC HS</td>
<td></td>
</tr>
</tbody>
</table>

Coverage limitation: All jobs covered, excluding contributing family workers working less than 15 hours per week.

Control variables, data variables and methodology variables are displayed in columns. Specific record notes are shown for highlighted records in the box underneath the grid (in this case a note relating to a coverage limitation for the measurement of part-time workers in Australia).

(4) Presentation of methodological information and the ability to refine query according to the strictest comparability criteria.

A user can tailor the query even further in order to enhance the international comparability of the data displayed. This is done by selecting the “Parameters” tab, as seen in figure 6 below.
This screen allows users to refine their search criteria. A user might wish to see, for example, only part-time employment records for both sexes (MF), from the repository OECDEMO (a mouse over reveals the full title), based on household surveys (HS), covering the total population (TE) and where the part-time hourly cut-off is measured as 35. After making the selections and hitting the “Refresh data” parameter button, a new data grid would be generated according to the new criteria.

The “Definitions” and “Repository” tabs offer additional information about the methodological variables.

(5) Variety of output formats.

When the desired data is generated, the KILM user can cut and paste records into other software applications or alternatively export the data into a variety of applications (Excel, Access, comma-delimited file, clipboard.)
(6) Graphing and mapping functionality.

Alternatively, the KILM software allows users to generate their own graphs or maps.

Figure 8. KILM graphing and mapping functions

(7) Linking to methodological details (classifications posted on websites, working papers, etc.) via the Internet.
A final advantage in presentation of data in electronic format is the ability to include html links that connect users direct to websites of original data sources or websites where international classifications, resolutions, or definitions are displayed.