Data Collection, Production and Dissemination

General Report Sections 4.1-4.5
ILO Department of Statistics
15 October 2018

Agenda

1. ILOSTAT
2. ILOSTAT questionnaire
3. Microdata
4. ILO estimates and projections
ILOSTAT

Steven Kapsos
General Report Sections 4.1 and 4.3

Uses of ILOSTAT data

World Employment & Social Outlook
Global Wage Report
World Social Protection Report

Sustainable Development Goals (SDGs)
Minimum Set of Gender Indicators
Monthly Bulletin of Statistics Online

World Development Indicators (WDI)

World Indicators of Skills for Employment

Human Development Index (HDI)
Gender Development Index (GDI)
Gender Inequality Index (GII)
Life-course Gender Gap Dashboard

ILO

UNSD

OECD

UNDP

WIPO

Global Innovation Index (GII)

ILOSTAT
ILOSTAT
THE WORLD'S LEADING SOURCE OF LABOUR STATISTICS

User-friendly portal for timely, comprehensive and internationally-comparable labour statistics

Challenges

- Labour statistics are complex
- Large data gaps
- Lag in data dissemination
- Lack of comparability
- Reporting burden
- Reporting mistakes
- Limited resources nationally and at the ILO

possible
Website
IMPROVING THE USER INTERFACE

- Simple navigation
- Customized & bulk download
- Multi-indicator reports with 1-click access
- Section on SDGs
- Quick access to data tools

Data tools
PROVIDING MORE OPTIONS FOR DATA USERS

- Excel Add-In
- WESO Data Finder
- SDMX API
- Country Profiles
- R-package

Excel Add-In
WESO Data Finder
SDMX API
Country Profiles
R-package
**Methods**

MORE RESOURCES FOR BETTER LABOUR STATISTICS

**METADATA**
- Glossary
- Standards and guidelines
- Indicator descriptions
- National sources and methods

**PUBLICATIONS > Statistical guides and manuals**
- ILO Guidebook on SDG Labour Market Indicators
- Quick Guide on Sources and Uses of Labour Statistics
- Labour Force Estimates and Projections: Methodological Description
- Key Labor Market Indicators: Analysis with Household Survey Data
- Sampling elusive populations: Applications of child labour
- Decent Work Indicators: Concepts and Definitions (revised)
- Measuring informality: A statistical manual on the informal sector and informal employment

**New indicators in ILOSTAT**

<table>
<thead>
<tr>
<th>Nationally-reported series</th>
<th>Harmonized series</th>
<th>ILO modelled estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum wages</td>
<td>Informal employment and informal sector</td>
<td>Global, regional, and country-level estimates on:</td>
</tr>
<tr>
<td>Gender pay gap</td>
<td>Earnings (minimum wage, hourly earnings, monthly earnings)</td>
<td>Labour force (by sex and age)</td>
</tr>
<tr>
<td>Temporary employment</td>
<td>Labour underutilization LU2-LU4</td>
<td>Employment (by sex and age, status, sector, occupation)</td>
</tr>
<tr>
<td>Unemployment levels and rates by disability status</td>
<td>International labour migration statistics (22 indicators)</td>
<td>Unemployment (by sex and age)</td>
</tr>
<tr>
<td></td>
<td>Labour cost</td>
<td>Labour productivity</td>
</tr>
</tbody>
</table>
Data sources and channels

Data channel comparison

**AUTOMATED**

**Bulk**
- Official NSO data available online
- Efficient
- Timely
- Reduces reporting burden
- Limited to LFS data (focus on short-term indicators)
- May not follow international standards

**Microdata**
- Increasingly available online
- Extensive cross-tabulations
- Reduces reporting burden
- Harmonized (consistently apply int’l standards)
- Limited to household surveys (usually LFS)

**Data entry**
- Official NSO data available online
- Reduces reporting burden
- Manual process
- Prone to human error
- May not follow international standards

**Questionnaire**
- Main channel to obtain non-LFS data
- Main channel for countries lacking online database
- Manual process
- Significant reporting burden
- Prone to human error
- May not follow international standards

**MANUAL**

**Bulk**
- Reduces reporting burden
- Limited to LFS data
- May not follow international standards

**Microdata**
- Extensive cross-tabulations
- Harmonized (consistently apply int’l standards)
- Limited to household surveys (usually LFS)

**Data entry**
- Reduces reporting burden
- Manual process
- Prone to human error
- May not follow international standards

**Questionnaire**
- Main channel to obtain non-LFS data
- Main channel for countries lacking online database
- Manual process
- Significant reporting burden
- Prone to human error
- May not follow international standards
More data
EVLoving Processes for More Efficiencies

Cumulative number of values

More data (cont.)
Number of values by data channel

- Microdata
- Automated
- Questionnaire

LABORSTA 2008 annual
ILOSTAT 2016 annual

LABORSTA 2008 annual
ILOSTAT 2016 annual + short-term
More comparable FEWER CAVEATS SINCE THE 19TH ICLS

Percent of values with footnotes, selected concepts

Percent of values with footnotes, selected classifications

Datasets and source types
DISTRIBUTION OF 62 MILLION VALUES BY DATASET & SOURCE TYPE
ILOSTAT Questionnaire

Marie-Claire Sodergren
General Report Section 4.2

ILOSTAT Questionnaire

TAILORED TO YOUR COUNTRY

- Email in May
  - Due 29 June
- Country-specific page
  1. Instructions with detailed guide and Excel files
  2. Contacts
  3. Data availability
Questionnaire split by source type

- 4 Excel files
- 18 topics/sheets
- 73 indicators

**HOUSEHOLD SURVEYS & POPULATION CENSUSES**
- Working-age population, Labour force, Employment, Informality, Time-related underemployment, Unemployment, Persons outside the labour force and NEET, Working poor

**ADMINISTRATIVE SOURCES**
- Minimum wage, Occupational injuries, Labour inspection, Strikes and lockouts, Industrial relations

**ESTABLISHMENT SURVEYS & NATIONAL ACCOUNTS**
- Employees, Earnings, Working time, Labour cost, Public sector employment, Labour income share

**OFFICIAL ESTIMATES & OTHER SOURCES**
- All of the above

---

Breakdowns

**Recurring**
- Sex
- Age
- Economic activity (ISIC)
- Education (ISCED)
- Occupation (ISCO)

**Topic-specific**
- For LFS
  - Rural/urban
- For employment
  - Status in employment (ICSE)
  - Hour bands
  - Private/public sector
- For unemployment
  - Duration
  - Disability status
- For injuries
  - Migrant status
# Questionnaire and the SDGs

<table>
<thead>
<tr>
<th>SDG</th>
<th>Indicator</th>
<th>Quest</th>
<th>Most used Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1</td>
<td>Working poverty rate</td>
<td></td>
<td>ILO</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Population covered by social protection floors/systems</td>
<td>socpro</td>
<td>ADM</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Female share of employment in managerial positions</td>
<td>✓</td>
<td>LFS</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Labour productivity</td>
<td></td>
<td>ILO</td>
</tr>
<tr>
<td>8.3.1</td>
<td>Informal employment</td>
<td>✓</td>
<td>LFS</td>
</tr>
<tr>
<td>8.5.1</td>
<td>Average hourly earnings by sex (by occupation)</td>
<td>✓</td>
<td>ES</td>
</tr>
<tr>
<td>8.5.2</td>
<td>Unemployment rate (by sex, age, disability status)</td>
<td>✓</td>
<td>LFS</td>
</tr>
<tr>
<td>8.6.1</td>
<td>NEET rate</td>
<td>✓</td>
<td>LFS</td>
</tr>
<tr>
<td>8.7.1</td>
<td>Number and proportion of children in child labour</td>
<td></td>
<td>CLS</td>
</tr>
<tr>
<td>8.8.1</td>
<td>Fatal and non-fatal occupational injuries (by sex, migrant status)</td>
<td>✓</td>
<td>ADM</td>
</tr>
<tr>
<td>9.2.2</td>
<td>Manufacturing employment</td>
<td>✓</td>
<td>LFS</td>
</tr>
<tr>
<td>10.4.1</td>
<td>Labour income share as a percent of GDP</td>
<td>✓</td>
<td>SNA</td>
</tr>
</tbody>
</table>

## Questionnaire length

**BASED ON OTHER DATA CHANNELS & DATA AVAILABILITY**

- **Full length**
  - 72 countries
  - No automated procedures in place. Must complete entire questionnaire.

- **Reduced length**
  - 75 countries
  - Automated procedures in place, usually for household survey-based data. Complete partial questionnaire.

- **Mini version**
  - 48 countries
  - No or very low response rate usually due to lack of data available. Complete only 10 key tables.
**Questionnaire length (cont.)**

**DISTRIBUTION BY WB INCOME GROUP AND BY ILO REGION**

<table>
<thead>
<tr>
<th>Income Group</th>
<th>Full</th>
<th>Reduced</th>
<th>Mini</th>
</tr>
</thead>
<tbody>
<tr>
<td>High income</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Middle income</td>
<td>25%</td>
<td>75%</td>
<td>0%</td>
</tr>
<tr>
<td>Low income</td>
<td>50%</td>
<td>50%</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Region</th>
<th>Full</th>
<th>Reduced</th>
<th>Mini</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe and Central Asia</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>Americas</td>
<td>25%</td>
<td>75%</td>
<td>0%</td>
</tr>
<tr>
<td>Arab States</td>
<td>50%</td>
<td>50%</td>
<td>0%</td>
</tr>
<tr>
<td>Asia and the Pacific</td>
<td>75%</td>
<td>25%</td>
<td>0%</td>
</tr>
<tr>
<td>Africa</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

**Self-help to avoid inconsistencies**

**IMPROVING THE QUESTIONNAIRE PROCESS**

- **Data problems**
  - Sheet «Summary for review» flags typical data problems. Please carefully review and address issues to extent possible.

- **File problems**
  - Workbook protected to avoid structural problems. Also cannot upload data when survey and metadata are missing.

- **Revisions**
  - Do not forget to enter revisions!

- **Tips and Targets**
  - Provide quick overview. Refer to the Guide for details on concepts and definitions.
Data reporting
RESPONSE TO THE 2018 QUESTIONNAIRE
AND INCOMING 2017 DATA FROM OTHER CHANNELS

Reporting rate
(percent of countries sending back the questionnaire)

50% in 2018

- No incoming data
- Questionnaire only
- Web only
- Questionnaire + Web
- Microdata only
- Microdata + Questionnaire
- Microdata + Web + Questionnaire

Reducing the Questionnaire Burden

Provide microdata
- Preferably make available online
- Substitutes the household survey portion of the questionnaire

Online static files
- Allows for automation
- Release data from all source types, not just LFS

Send CSV file
- Any format, but prefer similar to ILOSTAT csv files
- Allows for some automation
- Less prone to human error than questionnaire

Online database
- Allows for automation
- Release data from all source types, not just LFS
Microdata
Mabelin Villarreal-Fuentes & Yves Perardel
General Report Section 4.5

Benefits of microdata processing

The challenges
- Data gaps
- Lack of comparability
- Reporting burden
- Reporting mistakes
- Non-response

The Solution
- More data
- Comparable
- Significantly reduced reporting burden
- Reduced reporting errors
- Highlights areas for technical assistance
The process

Documentation
• Document deviations if international concepts cannot be applied
• Metadata documentation for ILO microdata repository

Acquisition
• Obtain files online or establish agreement with NSO
• Store microdata files in ILO safe drive

Processing
• Map national to international concepts and classifications
• Run routine to derive variables
• Apply reliability compliance test
• Review/assess resulting figures

Dissemination
• Publish results on ILOSTAT, including any caveats
• Publish metadata on ILO microdata repository; no files for external users

Number of LFS/household micro datasets

Large acquisition in recent years

Most datasets from Europe and the Americas
Challenges

01 Missing necessary documentation

02 Unaware of updated population benchmarks

03 Undesirable skip patterns

04 No oversampling for vulnerable groups

Acknowledgements

To National Statistical Systems
That widely and freely disseminate HHS or LFS’s anonymized datasets.

To Countries
Which despite not disseminating files online, engaged themselves to sending them to us regularly.

We encourage countries to share anonymized microdata sets
Action Point

Participants in the Conference may wish to reflect on the importance of providing access to anonymized national microdata sets to promote more and better data globally and to discuss ways in which the ILO can leverage its microdata processing efforts to enhance technical assistance to constituents.

ILO Estimates and Projections

Steven Kapsos & Roger Gomis
General Report Section 4.4
Key features

01 GLOBAL AND REGIONAL ESTIMATES
Insights into broad global and regional trends

02 INTERNATIONALLY COMPARABLE
Consistent definitions and methods across countries

03 STRICT DATA QUALITY STANDARDS
Extensive review of inputs and outputs

04 FREQUENT UPDATES
Next edition in October 2018 (released January 2019)

Existing models

Labour Force and Population (LFEP) → Unemployment - Employment

Labour underutilization LU3 → Employment distributions:
- By economic class (working poverty)
- By status in employment
- By economic activity
- By occupation

Other related models
- Social protection coverage
- Global wage growth
## Forthcoming models

- Women in managerial positions
- NEET
- Labour income share
- Informal employment
- LU2 & LU4 (labour underutilisation measures)

## Use in SDG monitoring

<table>
<thead>
<tr>
<th>SDG</th>
<th>Indicator</th>
<th>Global and regional data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1</td>
<td>Working poverty rate</td>
<td>Yes</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Population covered by social protection floors/systems</td>
<td>Yes*</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Female share of employment in managerial positions</td>
<td>Planned</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Labour productivity</td>
<td>Yes</td>
</tr>
<tr>
<td>8.3.1</td>
<td>Informal employment</td>
<td>Yes*</td>
</tr>
<tr>
<td>8.5.1</td>
<td>Average hourly earnings by sex (by occupation)</td>
<td>No</td>
</tr>
<tr>
<td>8.5.2</td>
<td>Unemployment rate (by sex, age, disability status)</td>
<td>Yes</td>
</tr>
<tr>
<td>8.6.1</td>
<td>NEET rate</td>
<td>Planned</td>
</tr>
<tr>
<td>8.7.1</td>
<td>Number and proportion of children in child labour</td>
<td>Planned**</td>
</tr>
<tr>
<td>8.8.1</td>
<td>Fatal and non-fatal occupational injuries (by sex, migrant status)</td>
<td>No</td>
</tr>
<tr>
<td>9.2.2</td>
<td>Manufacturing employment</td>
<td>Yes</td>
</tr>
<tr>
<td>10.4.1</td>
<td>Labour income share as a percent of GDP</td>
<td>In progress</td>
</tr>
</tbody>
</table>

* Indicates single year estimates only; ** To be jointly developed with UNICEF
Key principles

1. **Input data review**
   - Extensive review of input data

2. **Non-response**
   - Identify and address non-response bias

3. **Data patterns**
   - Missing data patterns dictate methodology and scope of aggregation

4. **Modelling**
   - Use of econometric/statistical models developed by ILO experts

5. **Transparency**
   - Methods documented and publicly disseminated; imputed values clearly indicated

The process

- **Input preparation**
  - Data collection via various channels
  - Compile data & metadata from various datasets in ILOSTAT

- **Data cleaning**
  - Discard outliers and erroneous data
  - Source homogenization: LFS + very restrictive use of household surveys and population census
  - Coverage homogenization: nationally representative
  - Age-group homogenization: standard age bands

- **Harmonization**
  - Enforce international standards
  - Use the raw source (micro) data

- **Estimation procedures**
  - Country-level imputation
  - Model non-response and differential response
  - Model selection: pseudo out of sample performance
Strengths

Integration with ILOSTAT
- Consistency checks
- Updates and data revision
- Metadata management system
- Improvements in data production
- Open access

Country level imputation
- Flexible groupings
- Offsetting of non-systematic errors: more stable aggregates
- Valuable model byproduct: country data for international comparisons

Statistical approach
- Objective & transparent
- Maximizes performance (pseudo) out of sample
- Uncertainty management: estimates of confidence in results

Model integration
- Outputs are fully compatible across models
- The same classifications and standards
- Input review is done globally
- Methodological improvements shared for all models

Results: an example
- Data cleaning and harmonization avoid erroneous and non-comparable data entry
- Imputation deals with non-response and differential response
- Both reduce bias and volatility in global and regional aggregates
THANK YOU

www.ilo.org/ilostat
ilostat@ilo.org