Ex-post harmonization – Global Monitoring Database guideline in responding to SDG

The World Bank

The Poverty and Equity Global Practice

Global Solutions Group on Welfare Measurement and Statistical Capacity

Minh Cong Nguyen

Paul Andres Corral Rodas

João Pedro Wagner De Azevedo

SDG indicators

- 17 Goals

- 169 targets

- About 250 indicators

- Data are from many sources such as multi-topic household surveys, administrative database, and more.

WORLD BANK GROUP

Examples of relevant SDG indicators

- Goal 1/Indicator 1.1.1: Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural).

- Goal 1/Indicator 1.1.2: Proportion of population living below the national poverty line, by sex and age

- Goal 10/Indicator 10.1.1: Growth rates of household expenditure or income per capita among the bottom 40 per cent of the population and the total population.

- Goal 10/Indicator 10.1.2: Proportion of people living below 50 per cent of median income, by age, sex and persons with disabilities

WORLD BANK GROUP

Indicators with statistics from sub-groups/disaggregation

- Same source of data should be used for those indicators disaggregated by sub-groups such as gender, age, employment.

- We acknowledge that countries have different surveys/sources for such indicators (such as Admin records and LFS), thus indicators produced from different sources might not be necessarily aligned.

WORLD BANK GROUP

Global Monitoring Database

- The Global Monitoring Database (GMD) is an ex-post harmonization effort based on available multi-topic household surveys, including household budget surveys and Living Standards Measurement Survey.

- The main purpose of the GMD is to create globally comparable harmonized micro database across countries, regions, and survey years for the purpose of creating, monitoring, and tracking of international poverty and shared prosperity indicators as well as by groups such as gender, age, employment status and urban/rural.

WORLD BANK GROUP

Ex-post harmonization process

- Survey data was received through country engagement with the country team, National Statistical Office or relevant line ministries.

- The data was securely stored in confidential servers for each geographical regions and their teams.

- Each statistical team in its region will start the harmonization process based on the agreed global GMD guidelines and timeline, as well as regional or country specific context.

- The data quality checks will be implemented for the harmonized data for consistency within countries, across countries in regions.

- Harmonized database with passed quality checks will be deposited in the GMD for usage.

WORLD BANK GROUP

Some principles

- One data source for the harmonization to ensure replicability and scalability

- Ensuring comparability data and statistics within countries and years

- Enabling greater efficiency in the use of statistics for regional and global research and policy dialogue

- Dofile/scripts for harmonization process are stored and shared for replicability and consistency and no duplication of work across WB teams.

- Original survey and harmonized survey data can be retrieved through the **datalibweb** system (Stata-based package)

WORLD BANK GROUP

What are the compromises?

Prioritize country–time comparison more than cross-country comparison

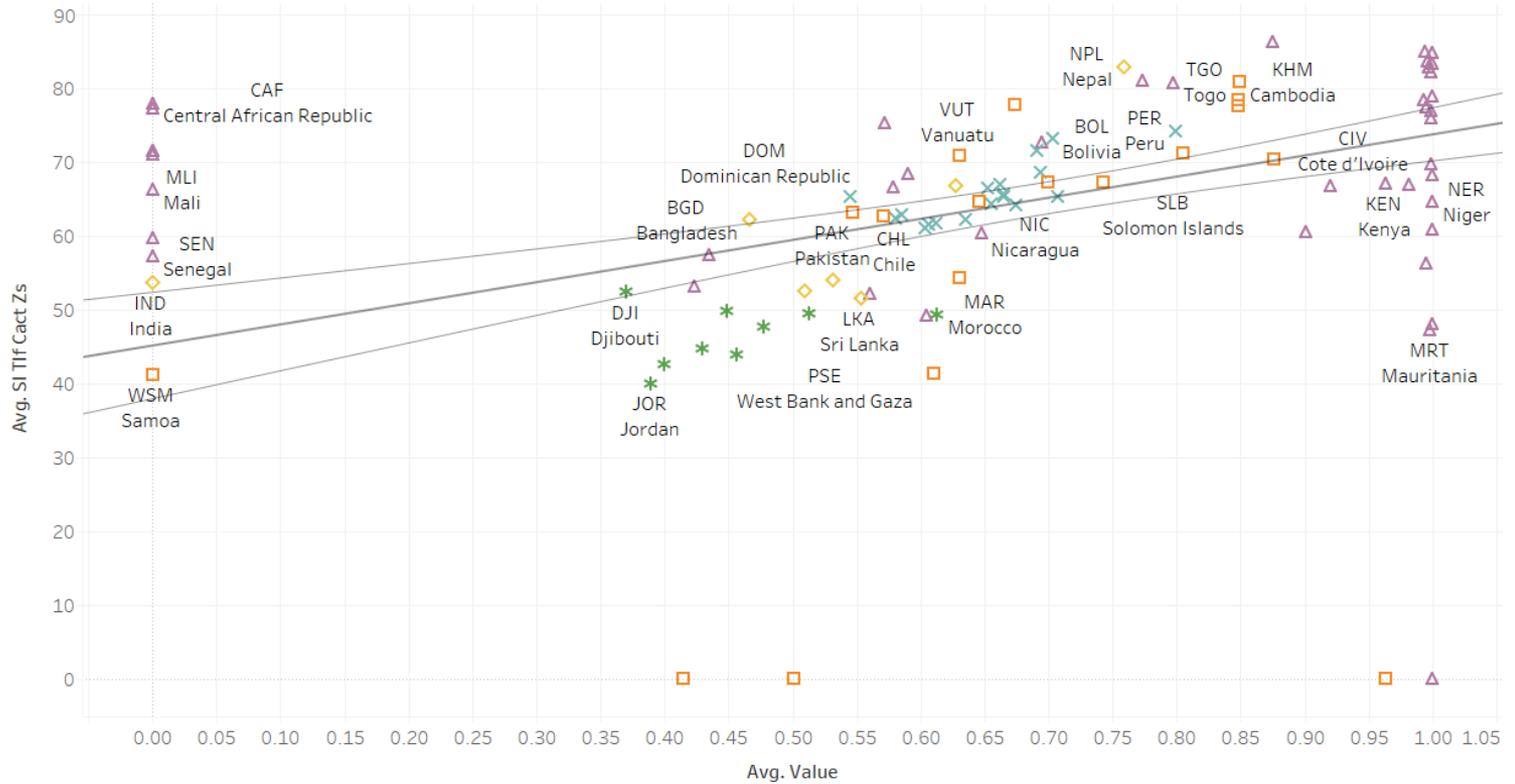The first priority is comparability within all years of available data for each country.

Harmonized variables might be different from the NSO definitions.

This means that indicators could be different and can presents a challenge for discussing policy implication, for example when the poverty rates are different. Transparency is our main remedy.

Harmonization is done with available surveys

Most available surveys in ECA region are HBS surveys design to gather information about expenditure, not consumption.

WORLD BANK GROUP

What are the compromises?
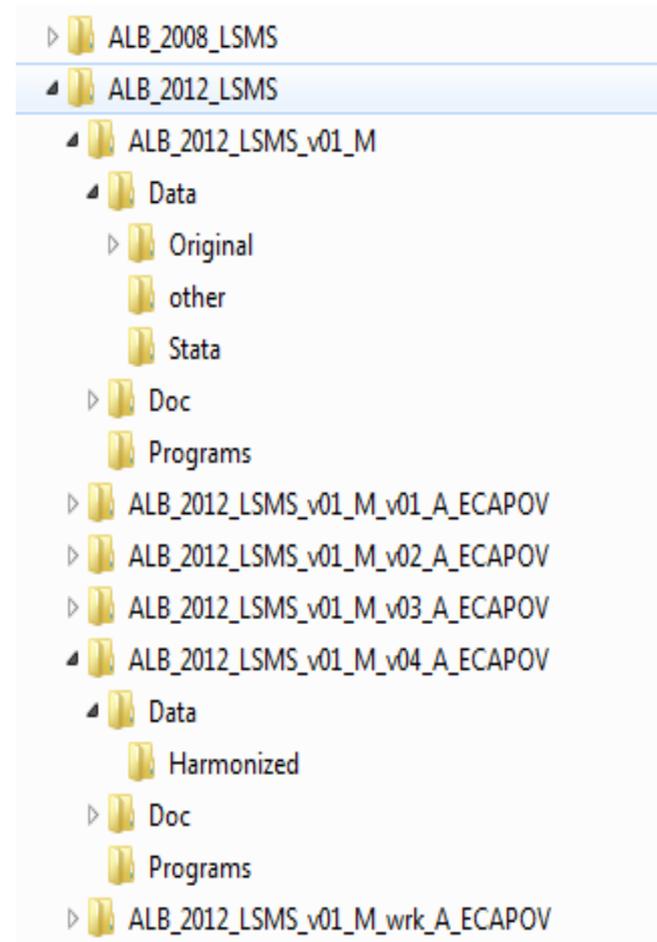
What are the compromises?

Global Monitoring Database guideline

- It is modular structure. Each module is a data file with a dofile/script to generate that data.

- Each module was consulted with relevant sectors and units within the WBG, and also with other international organizations such as ILO, UNESCO.

- Modules:
  - 1. ID module
  - 2. Welfare module
    - 2.1: Consumption/expenditure welfare
    - 2.2: Income welfare
    - 2.3: National welfare and its poverty lines
  - 3. Labor module
  - 4. Education module
  - 5. Demography and geography module
  - 6. Utilities module
  - 7. Assets module
  - 8. Housing and dwelling module

WORLD BANK GROUP

How do we structure the data?

- The data and programs (scripts) were organized in the hierarchical structures, with the following structures:
- CCC – 3-letter country code
- YYYY – Survey year
- SSSS – acronym of the survey (i.e. HBS, LSMS, HIES)
- vNN_M – version of master/original data
- vMM_A – version of harmonized or adaptation data
- i – GMD modules (i=2,3…,10)

WORLD BANK GROUP

Global Monitoring Database guideline – ID module

- Single and unique IDs at each level of observation are important for merging/joining different harmonized data files, and importantly with the original data files.

- ID module:
  1. stores key information on the survey period, collectors and other sampling information if needed to be used for future calculations.
  2. stores the unique IDs and the variables that were used to construct that unique IDs. It is useful to join/merge with original data or other teams' harmonized database.

WORLD BANK GROUP

Global Monitoring Database guideline – Welfare module

- Metadata for the Prices framework: all information about the prices, currency, deflation across time and space, and survey period are stored as metadata – important for converting to PPP terms.

- For consumption/expenditure welfare:
  - Detailed mapping data to COICOP, when possible at 4 digit level with consumption/expenditure and unit prices when available.
  - Rent imputation process based on different modeling such as hedonic model, matching model, or user-cost model.

WORLD BANK GROUP

Global Monitoring Database guideline – Welfare module

- For income welfare:
  - Income items are grouped into 3 structured nested levels, allowing for regrouping and common item coverage across countries.
- Whenever possible, national welfare and national poverty lines will be also included in this module.
- Spatial adjustment across time and space within the survey data:
  - For consumption welfare, the Paasche price index is used to account for regional price differences.
  - There are some approaches on spatial adjustment for income welfare.

WORLD BANK GROUP

Labor module

- Labor module contains key information on labor related information encompassing two different times: last seven days and past twelve months, to get a better characterization of the labor market.
  1. Individual level data
  2. ILO definitions
  3. Variables: employment status, sector, industry and occupation classification, and other job's characteristics such as contract, social security, health insurance, union, etc.
  4. For main and secondary jobs.

WORLD BANK GROUP

Education module

- Education module contains all information about education for each member of the household.
  1. It is defined at the individual level.
  2. When available, use ISCED 2011 for country-specific educational levels mapping to one global standard mapping for both attainment and completion of education levels.

WORLD BANK GROUP

Education module – ISCED 2011

ISCED 0: Early childhood education ('less than primary' for educational attainment)

ISCED 1: Primary education

ISCED 2: Lower secondary education

ISCED 3: Upper secondary education

ISCED 4: Post-secondary non-tertiary education

ISCED 5: Short-cycle tertiary education

ISCED 6: Bachelor's or equivalent level

ISCED 7: Master's or equivalent level

ISCED 8: Doctoral or equivalent level

WORLD BANK GROUP

Demography and Geography modules

- Geography module contains all information about the coded geographic variables about the surveys, such as
  1. level of representative of the surveys such as region, province, or district, etc.
  2. Sampling information such as strata and psu, etc.
- Demography module contains all information about the demographic and migration variables for all members in the households.
  1. Individual level
  2. Some example variables such as age, gender, martial status, language, ethnicity, social group.

WORLD BANK GROUP

Utilities module

- Utilities module contains information about affordability and access to related services from the household surveys.
- Affordability section follows COICOP related energy expenditures (COICOP 4) such as water, gas, electricity, other fuels, etc.
- Access to services uses the frameworks of WASH (water, sanitation, and hygience) Access Plus Framework, and Multi-tier access Tracking Framework (MTF)
    1. Simple definition Yes/No for the access
    2. Exploring other tiers of the definitions such as access with quality, reliability, and affordability.

WORLD BANK GROUP

Assets module

- Assets module contains information about assets that households and its members have access to use such as TV, AC, bicycles, etc.
    1. It is household level data (unless the data allows for individual and access ownership).
    2. Ownership versus access and usage
    3. Yes/No and number of items
    4. Some usage/ownership items can be inferred from expenditure items.

WORLD BANK GROUP

Housing and dwelling module

- Housing and dwelling module includes information on housing, its characteristics and conditions, services and expenditures.
- The variables in this module can be used to impute housing rents when necessary.

WORLD BANK GROUP

Thank you!

Minh Cong Nguyen mnguyen3@worldbank.org
Paul Andres Corral Rodas pcorralrodas@worldbank.org
Joao Pedro Wagner De Azevedo jazevedo@worldbank.org

WORLD BANK GROUP