The 2019 LIS Template: why a new template?
May 1ˢᵗ 2019

The LIS template consists of the set of variables of the *LIS* and *LWS Databases*, their definitions and their harmonisation rules. The 2019 Template is the fifth template revision over the last three decades. Survey methods have changed, theoretical foundations of key concept have been revised, and data from new countries have become available. While this explains the need for periodical revision, it does not fully clarify what the drivers of the current revision are.

The motivation of this template revision stemmed from our determination to provide easier to use variables (clearer categories, definitions and aggregation rules), higher temporal and cross-country variables' coverage, and more data points - possibly annual data series - and expansion to new geographical areas.

We started the project of the new template revision, by assessing the country coverage and the usage of the provided variables. We observed that 65% of all the variables available in the previous template were filled in less than one third of all the datasets, with a large fraction of those variables – especially among the income variables - filled for a very small subset of the datasets. In addition, we observed that a large number of the variables available in the previous template was hardly used, while the additional resources needed to both harmonize and document those variables were significant.

Reassured by the feedback received from a group of LIS "power" users that we consulted about this revision, we decided to implement a significant reduction in the number of LIS variables to be included in the new template - from 737 to 191 - in conjunction with a reorganisation of all of the included variables. The reorganisation entailed adjusted definitions, refined categories and easier aggregation rules, as well as the introduction of few new crucial variables. For more information about the new variable list and user guide for LIS, see [here](#).

With this revision, we believe we have achieved at least four major improvements:

1. *Higher coverage of the income variables*: by regrouping the incomes variables, their availability across all datasets increased from 29% to 50% on average.

2. *Higher degree of data usability and comparability*: this is achieved through the following changes:
   - Restructuring of the income variables, resulting in a disaggregation that is more suitable for cross-national comparison; where: i) the five income sub-components (*hilabour*, *hicapital*, *hipension*, *hipubsoc*, and *hiprivate*) perfectly sum up to total income; ii) systematic distinction between monetary and non-monetary amounts has been suppressed; and iii) the treatment of pensions has been improved.

- Adjustment of the consumption concept: due to data availability issues, the imputed rent is no longer included in the consumption expenditure variable (*hcexp*), but, similarly to the incomes, provided separately (*hrenti*).
- Conversion of income and consumption amounts expressed in old currencies into the currency actually in force (this concerns at87-at00, be85-be00, de73-de01, es80-es00, fi87-fi00, fr78-fr99, ie87-ie00, it86-it00, lu85- lu00, nl83-nl99, si97-si04, sk92-sk07, ee00-ee10, lt10- lt13, il79, pl86-pl92, ro95-ro97).
- Simplification of the missing values policy of income and consumption variables: when an income source was not collected for the whole sample by the data provider, the corresponding variable is set at zero instead of missing, allowing for an easier aggregation of variables; as a result, only item unit non-response is coded with the missing.
- Restructuring of some categorical variables, with the creation of more easy-to-use and comparable dummy variables, among which *hpartner* and *farming* at the household level and *partner*, *enrol*, *illiterate* (previously *literate*), *informal* (previously included in *oddjob_c* and *notoff_c*), *parleave* (previously included in *leave*), *public1* (previously *sector1*) and *temp1* (previously *perm1*), as well as the renaming of *hourstot* (previously *hours*).
- Rationalisation of the labour force status, with the provision of one clear variable for labour force status (*lfs*, replacing *clfs*, *cmas* and *umas*), containing ideally the main status as reflected in the *emp* dummy; in addition, there is a new dummy for ILO employment (*emp_ilo*).
- All of the LIS variables are now included in the LWS datasets as well; the variables present in LWS and not in LIS are only the ones that concern wealth and households' behaviour.
- Suppression of non-household members from the micro data files, so that the aggregation of individual level information into household level (whether for incomes or demographic variables) is now straightforward.

3. *Addition of new variables*:
   - Education: two new standardised education variables 1) *educlev*: for the highest education level achieved (with more details than the low/medium/high *educ* variable) 2) *edyrs*: for the number of years of schooling.
   - Consumption aggregates: one new variable (*hhouscost*) that summarises the total amount of housing costs.
   - Wealth aggregates: one new aggregate for net worth (integrated net worth, *inw*).
   - Behavioural variables: three new variables providing information on household financial expectations (*boef_c* and *boee1/2_c*), two new variables on constraints in debt repayment (*bocd1/2_c*) and a few additional country-specific variables on attitudes towards household finance (*bafr2/3_c*, *bafp2_c* and *bafl4_c*).
   - Technical variables: inclusion of the 3-letter country abbreviation identifier (*iso3*).

4. *Extended documentation*: in order to provide our users with better understanding of the income and consumption variables, a field has been created in our METIS documentation system to show the contents of such variables in each dataset.

We are confident that this template revision will also help us achieve new milestones in the near future. The flexibility offered by the new template will enable us to extend the dataset coverage both in the temporal and geographical dimensions: we envision to start providing annual data whenever available, and to add new countries (especially among the middle-low income ones). Moreover, the new template is accompanied by an overhaul of the technical tools underlying the harmonisation process that were designed in such a way to allow for a much easier introduction of new blocks of variables, which could become interesting in future research.

Overall we are proud of this further step towards increasing the quality and easing the use of the *LIS* and *LWS Databases*. We continue in our policy of serving the research community with the best quality for the largest number of countries.