

# Service Delivery Indicators: Data Harmonization Protocol (February 2021)

## Objectives

In order to render the SDI dataset more usable, it is necessary to ensure the comparability of variables across surveys and time, particularly given the country-specific adaptations that SDI instruments usually undergo. To this end, and with the intention of making modules and variables across SDI surveys comparable, the following steps are taken:

1. Identification of unit of analysis that makes the most sense for an easy use and merge of data as needed: modules for health (i.e. facility, provider, and knowledge) and survey levels for education (i.e. school, teacher, student, and time-on-task)
2. Break down of multi-level data by unit of analysis
3. Identification of variables that contain the same information or information that can be recoded into comparable values.
4. Mapping of such variables, usually with different variable names, across SDI surveys.
5. Identification and creation of the most useful value label categories with a common denominator across countries<sup>1</sup>. For instance, some countries might differentiate between teacher or medical provider types. In such instances, only fully comparable categories will be harmonized to a common value and categories that provide further disaggregation might either be collapsed or left as they are depending on how comparable and useful they are expected to be.
6. Creation, combination, recoding, and/or modification of variables to align with the selected standard.
7. Inclusion and renaming of country-specific variables for completeness.

## Results

The resulting datasets retain the great majority of the data while improving comparability across surveys. Most variables (~90%) have been harmonized across countries. In a harmonization process, some information is inevitably lost. For instance, variables that were measured by two separate questions in one survey but only by one in other surveys will be combined, with some loss in specificity. Similarly, multiple country-specific answers to the same question would be consolidated into fewer categories to be comparable across countries. An example is information about teacher training, which in most countries is reported in terms of highest education level achieved, but in some is recorded instead according to country-specific diploma and systems. Harmonized consolidates it all in terms of “highest level of education completed”. Other variables lost in this process include country-specific information collected for data checks but with little relevance for research purposes (e.g. timestamps throughout different sections of the survey, multiple enumerators’ data, field supervisor checks, etc.).

In order to compare two or more country surveys, a simple data append would be enough to have clean data ready to be analyzed. An Excel variable map with all variables recodings could be provided for each harmonized dataset (i.e. each module or unit level) in case researchers want to refer to them or observe discrepancies with the questionnaires.

---

<sup>1</sup> This process requires some arbitrary decisions that might collapse information into a more concise set but greatly improves comparability across countries.