Global Monitoring Database Harmonization Guidelines

Data for Goals (D4G) Team

Poverty Global Practice

The World Bank

November 2019

Table of Contents

1 Introduction

1.1 Global Micro Database (GMD) Project

Global Micro Database (GMD) constructs globally comparable microdata across countries, regions and across years for global poverty monitoring and welfare measurement. GMD is based on the best available multi-purpose surveys[1]: surveys that cater for many applications and users, have wide coverage of years, and are available, accessible, and sharable across the World Bank.

Household surveys are among three major sources of social and demographic statistics in many countries.[2] They are used for the collection of detailed and varied socio demo-graphic data pertaining to the conditions under which people live, their well-being, the activities in which they engage, and demographic characteristics and cultural factors that influence behavior, as well as social and economic change.

Despite the common focus on the study of patterns of consumption expenditures and income of private households, their characteristics, etc., household surveys represent a diversity of structures and designs, with their own coding standards for variable. As a result, merging or comparing across multiple surveys can be a very costly and time-consuming exercise. Additionally, the lack of standardized survey structures negatively impacts the comparability of survey indicators and, as a result, limits significantly their analytical applications and usefulness. This presents a serious limitation to cross-national or even national research.

In this context, the GMD project and the Guidelines developed by the Poverty Global Practice team are the first steps towards filling the critical knowledge gap on harmonization and comparability of survey indicators for measuring inequality, poverty and living conditions on a global scale.

Consistency and comparability among the indicators resulting from the survey can only be assured by using a common framework for the harmonization of survey indicators, which envisages standardizing indicators and indicator definitions across time and space.

The primary goal of these Guidelines is to ensure that relevant statistical and research teams within and outside the Bank follow accepted standards when creating harmonized data and documentation files and apply a common strategy that best fits the original source materials. It also aims to provide country, regional and global teams, including the user community and data harmonizers, with the tools and material resources they need in order to build an effective information system to collect, process, harmonize and analyze survey data.

Moreover, there is an increased need for data on the SDG indicators. The GMD will support the development of products in line with the corporate goals of the bank in addition to the products developed by the Poverty and Equity Global Practice.

[1] In most of the countries, Household Budget Survey is the best available type of multi-purpose household level information. But Living Standard Measurement Studies (LSMS), EU Statistics on Income and Living Conditions (EU-SILC), and other types of surveys are also used.

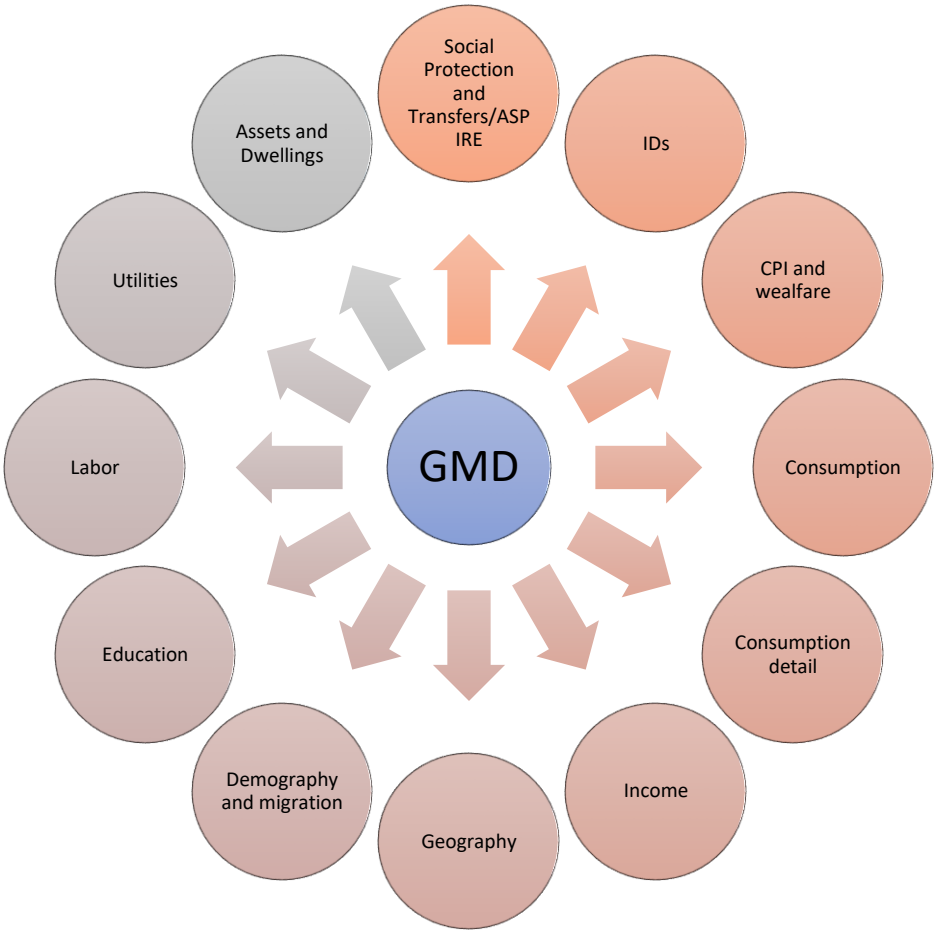
[2] The other two major sources are population and housing census and Administrative record systems.

While the Guidelines mainly target regional statistical teams that carry out routine survey harmonization, they also provide useful information for analysts, policymakers, researchers and other users of survey data.

1.2 File Structure and Content

The Global Monitoring Databases consists of 10 modules, as presented in Figure 1. The modules covered comprise 1. IDs, 2. CPI and welfare, 3. Consumption, 3.a Consumption detailed, 4. Income, 5. Geography, 6. Demography and migration, 7. Education, 8. Labor, 9. Utilities, 10. Assets and Dwellings, and 11. Social Protection.

Figure 1. Modules of the Global Monitoring Database.



Source: Based on initial GSG proposal.

Each module file contains data for all module subject areas available in a survey. The files also contain selected identification and demographic information from the core dataset, making it possible to conduct the analysis independently from the core file and full panel files. If more detailed social or demographic information is required for an analysis, users can obtain that information by merging module files with each other or with any other statistical information.

To facilitate documentation consistency, all topical module files have been assigned with their unique identification numbers. For example, the "Geography" module can be identified under the code 5. Furthermore, some modules may be broken down further into sub-modules. For example, the "Assets and Dwellings" module is comprised of two sub-sections – Assets and Dwellings.

The next section provides a brief overview of the GMD Data Dictionary and its variables.

1.3 Data Dictionary

This section discusses the names, definitions, and attributes of data elements within the GMD data system. The goal of the GMD Data Dictionary is to provide a comprehensive information catalog of data definitions, relationships, collection groupings, and sources of data.

The inventory revealed about 300 variables available in the latest version of the GMD Data Dictionary. It should be noted that as the GMD team continues updating and revising the Data Dictionary, the number of variables may differ from the number of variables covered in the current version of the Guidelines.

All topical GMD modules consistently use the same format and naming convention. As noted, all GMD modules contain a set of common identification and demographic variables, such as a country code, year variable, household identifier, individual identifier (if applicable), and household weights. The number of module specific variables typically varies across the subject areas depending on the type of module under consideration.

To facilitate common understanding, each topical GMD module contains a table summarizing all the variables covered in the module, including variable name, variable label, variable description, acceptable variable type after harmonization, as well as sources of variable information. Each variable element is also mapped to its tier, GMD module, as well as sub-module.

1.4 File organization

Since GMD is a collaborative effort where different individuals work on the harmonization of different surveys, it is essential that everyone follows the same file structure and do-file organization to allow for easy access and understanding.

Where to find the original raw data files

The first step for each harmonization is to obtain the original raw data files. The original data files used for the GMD harmonization should be cataloged and stored in the regional shared drive and the central Microdata Library catalog (http://microdatalib/). We assume that the depositing of the original microdata in the Microdata Library is a collective responsibility shared by all World Bank staff working in all the World Bank's regions, irrespective of global practice, since the dialogue and data acquisition through the NSO and line ministries often take place in a decentralized manner. Those original data files are stored in a secured server, to which access is limited only to the regional admins (see Azevedo and Cancho, 2013 for further details).

To avoid having multiple data files stored and saved in different location, it is recommended that the users should use and query the original files within the datalibweb system.

Do-file and data file naming guidelines

Data sets and do-files have identical file naming conventions. Do-files used to create a data file must have the same file name and follow the naming guidelines below.

Example: The following files are the do file and STATA file for the GMD consumption module harmonization using the Kazakhstan 2011 HBS.

KAZ_2011_HBS_v01_M_v01_A_GMD_CON.do

KAZ_2011_HBS_v01_M_v01_A_GMD_CON.dta

Table 1. GMD File Naming Convention

Generic Structure: CCC_YYYY_SurveyName_vnn_M_vmm_A_GMD_mod		
Component	**Description**	**Values**
CCC	Country Code	Example: KAZ, GEO, TJK
YYYY	Year of the Survey	The starting year of the survey
SurveyName	Abbreviation of the survey	Example: HBS, KIHS, LFS, EUSILC Character length can vary.
v*nn*	stands for the version of the master file	*nn*=01, 02, …
v*mm*	stands for the version of the harmonization, as there can be revisions after the first release	*mm*=01, 02, …
mod	denotes the specific GMD dataset (module)	*mod*= CON, EDU, DEM, LBR… Always 3 letters, can be found on each module heading

Where to save and store do files and GMD harmonized files

While the following instructions are specifically for data harmonizers at the World Bank, following a clear set of naming and file storing guidelines will save time and ensure accuracy for all researchers.

In the past, harmonized data files in the datalib folder used file paths that followed no particular formatting rules, often referring to folders that did not exist or were inaccessible. This limited the speed with which replication can be completed. Instead, a standardized approach was adopted, which should be followed for every harmonization exercise:

*Identify output file for harmonized data [CCC is 3 letter country code]

global out “$root\\datalib_review\CCC\”

The file names and versions would then be saved using the global links within the .do file, allowing others to run the .do files with few (or even without any) updates needed.

In the review process, versions will be zipped with time stamp for version control.

Saving .dta base files

In addition to the clean, harmonized files created in the process described below, an intermediate version of the .dta files should also be saved in the same location. This version should include the variables used in the process of creating the harmonized data. The file should be saved using the following convention:

CCC_YYYY_SurveyName_vnn_M_vmm_A_GMDBASE_mod

Working versions of the harmonized files

Over the course of harmonizing raw data into the appropriate format, files invariably go through several versions. To keep track of these versions, the last version of the harmonization that is “live” on datalib should be used as the root name, but with a slight modification to differentiate updates that occur between datalib versions. As an example, a first version of a harmonization would start with 00wrk1:

CCC_2012_Surveyname_v01_M_v00wrk1_A_GMD_mod.dta

Once the file is moved to datalib for the first time, the version changes to v01. Any updates from that point would be done using the new root version: v01wrk, and so forth. For clarity and ease of navigation, each time a new file is created it should be placed in its own folder, identically named to the harmonization .do file.

How to use the datalibweb system to load the raw/original data

Often the original/raw data was saved with Stata format in the datalib folder with a structured folder. This well-defined structure allows programs to query the data easily, making the sharing of the work much easier.

Harmonization from original/raw data should start with datalibweb, thus ensuring that the data is available for all to use.

For example: the code below is used to query the file “NHIES_2015_16_individual_level.dta” from the NAM 2015 survey.

```
datalibweb, country(NAM) year(2015) type(SSARAW) survey(NHIES) filename(NHIES_2015_16_individual_level.dta) clear files
```

For more examples on how to use datalibweb to query the data, please see the helpfile of datalibweb.

1.5 Do-File Organization and Guidelines

Do-files: header guidelines

Each module (*mod*) and each do-file must have the same file name structure.

All relevant information about the project should be included in the preamble of each do file (see

Box 1). The relevant information includes:

- Description of the original survey
- Name of the researcher who created the current do-file (and of researchers who worked on previous versions)
- Complete names of the input and output dataset
- Global paths where temporary output data will be stored
- Essential variables, such as country name, country code, country management unit, year and survey name

Box 1: Do file – Preamble

```
/* ------------------------------------------------------------------------
                          GMD Harmonization
 ------------------------------------------------------------------------
 <_Program name_>                    MNE_2013_HBS_v01_M_v02_A_GMD_CON.do </_Program name_>
 <_Application_>                     STATA 13 <_Application_>
 <_Author(s)_>                       ABC, XYZ </_Author(s)_>
 <_Date created_>                    2014-03-01 </_Date created_>
 <_Date modified>                    2015-03-01 </_Date modified_>
 ------------------------------------------------------------------------
 <_Country_>                         Montenegro (MNE) </_Country_>
 <_Survey Title_>                    Household Budget Survey (HBS) - 2013 </_Survey Title_>
 <_Survey Year_>                     2013 </_Survey Year_>
 <_Reference Year_>                  2013 </_Reference Year_>
 <_Study ID_>                        1234 </_Study ID_>
 <_Data collection from (M/Y)_>      MM/YYYY </_Data collection from (M/Y)_>
 <_Data collection to (M/Y)_>        MM/YYYY </_Data collection to (M/Y)_>
 <_Source of dataset_>               Poverty and Education teams </_Source of dataset_>
 <_Sample size (HH)_>                12240 </_Sample size (HH)_>
 <_Sample size (IND)_>               52240 </_Sample size (IND)_>
 <_Sampling method_>                 Stratified, multi-stage </_Sampling method_>
 <_Geographic coverage_>             National </_Geographic coverage_>
 <_Geo_1_>                           division </_Geo_1_>
 <_Geo_2_>                           region </_Geo_2_>
 <_PSU variable_>                    psu </_PSU variable_>
 <_Number of food items_>            133 </_Number of food items_>
 <_Food recall period(s)_>           Last 14 days </_Food recall period(s)_>
 <_Number of non-food items_>        221 </_Number of non-food items_>
 <_Non-food recall period(s)_>       One month/3 months/12 months </_Non-food recall period(s)_>
 <_Rent imputation_>                 Actual and owner's estimate rental values, taken from existing
data </_Rent imputation_>
 <_Durables_>                        Use value of durable goods could not be calculated as value
of the items when purchased are not available </_Durables_>
 <_Regional price deflators_> Paasche price index using food consumption items </_Regional price
deflators_>
 <_Consumer price index_>            1.56 value of base year 2005 </_Consumer price index_>
 <_Currency_>                        EURO </_Currency_>
 ------------------------------------------------------------------------

 <_Version Control_>
 Date: 2015-03-01
 File: MNE_2013_HBS_v01_M_v02_A_GMD_CON.do - Changes on education variables (5 level to 3 level)
                                                                                - Add more
```

```
durable items to consumption aggregates (items are ...)
 Date: 2014-03-01
 File: MNE_2013_HBS_v01_M_v01_A_GMD_CON.do - First version
 </_Version Control_>
 */
```

Do-files: variable and variable note tagging guidelines

The location of i) variable creation and ii) survey/headings information in the .do files will be tagged using a standardized approach. All harmonized variables in the in each module in the data will be tagged according to the following convention:

- The beginning of the code relating to a harmonized variable should be proceeded by **<_var_>*
- The end of the code relating to the variable creation should read **</_var_>* where "var" is the harmonized variable being created.
- Variables that are already named (such as in the case when "hhid" is already defined) should be noted when the file when opened, using the same convention as above. Between the "open" and "close" codes, a starred out line should read:"**'var' brought in from 'source'*"
- If a variable is created more than once (for example, hhid is created from several sources, then used to merge), it should be tagged only once.
- If possible, the DDI tags should include any value labels for the variable
- For survey information (such as spatial deflation) a similar convention should be used: **<_var_s_>*, the "s" signifies that the variable is "survey" information that is included in the general description of the data in the DDI.

It would also be important to note the comments if necessary when creating the harmonized variables. For any variables with notes or comments on how the variable is created, that information is also needed to tag so one can pull out those variable-specific notes by using similar taggings **<_var_note_> and *</_var_note_>*.

- For example, the variable "lfstatus" is created only for individuals with age of 15 and above, then one can put the variable-specific note as follow: **<_lfstatus_note_> Only for individuals with age of 15 and above *</_lfstatus_note_>.*

These tags will then be located automatically using the *strpos* command in Stata. It is incorporated in the ado called **ddi2dta** to either generate DDI or to extract information from the dofile.

There are two useful purposes of tagging the harmonization variables: (1) tagging is useful when cross checking the definitions of harmonized variables overtime, and when comparing the comparability of such variables with different countries; (2) tagging will improve the automated updating of the DDI by adding the block of codes used for generating the harmonized variables in the variable description of the DDI. Tagging will also improve the transparency of the metadata DDI for basic users in the Microdata Library.

Some rules/comments in tagging (to be updated when testing):

- No double quotes: " in the tagging block, especially with * comment. If double quotes are included in the label, the ado will extract the double quote and put the rest in the DDI.

- Take the size of the block into account of (not too long, focus on the main idea). For example, for the spatial deflator variable, a short version of the code can be included.
- Tagging should be done for one variable at a time, not a group of variables.

Updates tracking/vintages

If there are any changes resulting into the new harmonization .do files, a summary of the changes in a text (.txt) document should be included. The name of the file should be the same as the harmonized Survey ID. Duplicate do-file and data file, put all changes in the .txt (same study ID and in the do-file). Cumulative and start with the latest one and goes backwards. This readme file should be stored together with the do-file (same folder). For example, the name could be:

```
"MNE_2013_HBS_v01_M_v03_A_GMD_readme.txt"
*<_Version Control_>
 Date: 2015-12-09
 File: MNE_2013_HBS_v01_M_v03_A_GMD_CON.do - File created
 Harmonization - outcome: MNE_2013_HBS_v01_M_v03_A_GMD_CON.dta
 Notes: Missing consumption in COICOP 9 and 10 added. GMD Harmonization
3.0 conducted.
*</_Version Control_>
```

If there are changes in content, we should make the new version of the data and the dofile all together, and the readme should explain the changes. For other modules where the data/dofiles are not affected, the names should also be updated to match the updated module.

For example, changes made to the labor module (LBR) only, we have "ARM_2012_ILCS_v01_M_v05_A_GMD_LBR.dta", then version v05 should be in other modules of this harmonization.

It is also important that the notes on the changes are added to the data cumulatively, so we keep the historical records of all the changes. In Stata, one can do that by:

```
note: ARM_2012_ILCS_v01_M_v05_A_GMD_LBR.dta; 11/02/2015; age is fixed, and labor
information is added.

For more detail, please see the template header and labeling dofiles for each module.
```

2 ID module (IDN)

2.1 Framework for Harmonization

The ID module covers identification variables that are commonly derived from survey data sets or that can be constructed using existing variables. The primary objective of the indicator harmonization is to generate a unified data source of globally comparable identification statistical indicators to support routine analytical and corporate business functions of the World Bank.

The ID variables are essential for necessary for keeping data well-organized and attributing the findings to a specific country or year. When the World Bank receives survey data from NSOs, those data sets can be in one file or spread between different files. When the data is spread between different files, household IDs are necessary to merge the data. Household IDs mat contain several identifying variables such as the region and/or PSU they come from. In the case that households don't have clear hhid, harmonizers can reconstruct new hhid to ensure that data is correctly attributed to the correct household. In this module, the primary units of analysis are the level of individuals within a household and the household.

2.2 Creating IDs

The ID module is a file at individual level with basic household and individual identifiers. The household identifier is *hhid*, and the individual identifier is *pid*. Those variables must be present and must be the same across the different modules of the harmonized database. For efficiency of the data and merging across different files, those *hhid* and *pid* variables must be stored in the same format (numeric preferable) throughout all data files within each survey. In addition, the ID module must also have the original variables that were used to construct the *hhid* and *pid* variables. Those variables are useful when users want to merge the harmonized data with the original data files. Those original variables must be in the same format and type with that variables in the original data files. In practice, the identifier variables should be common in all the raw/original data files so that they can be merged between them. Those variables are often the good candidates for the ID variables.

Example 1: the below codes are used in creating the *hhid* and *pid* variables for Albania's LSMS 2012:

```
datalibweb, country(ALB) year(2012) filen(Modul_2B_education.dta)
type(ECARAW) surveyid(ALB_2012_LSMS)

*<_pid_>
rename idcode pid
*</_pid_>

*<_hhid_>
gen hhid= psu * 100 + hh
*</hhid_>
```

In this example, those original variables are *idcode*, *psu*, and *hh*. Therefore, in the ID module we have the following variables: *countrycode, year, hhid, pid, idcode, psu,* and *hh*.

Example 2: For the case of *hhid* and *pid* are already in the raw/original database, either in the same numeric format or in different format, it would be still important to have the ID variables standardized and also keep the original variables. Those variables can be renamed with the suffix _orig. For example, in a country where *hhid* and *pid* are used as ID variables, then in the ID module, we should have the following variables: *countrycode, year, hhid, pid, hhid_orig,* and *pid_orig*; where *hhid_orig* and *pid_orig* are the same in content as *hhid* and *pid*, respectively.

Example 3: Create the *hhid* variable from the *-group-* function in Stata from two variables. Note that *-sort-* must be used before the *-group-* function, otherwise the *hhid* variable will point to different households in each data files. One good practice is to do that *group()* function only one time and use the variables in the *group()* to merge across the data files.

```
*<_hhid_>
sort folio e10
egen hhid =group(folio e10)
*</hhid_>

*<_pid_>
clonevar pid = eglin
*</_pid_>
```

2.3 Mapping and Description of Variables

The aim of this section is to provide the readers with basic information about the main sources of information and then some details about how GMD ID are produced, and the issues surrounding them.

countrycode
countrycode is a string variable that specifies the 3-character country code used by the World Bank to identify each country. Although there are different naming conventions, it is necessary to use those specified to ensure that the data for each country is appropriately labeled.

year
year is a numeric variable that denotes the year in which the implementation of the household survey was begun. For example, if a survey was implemented during October 2018 and September 2019, the *year* would be 2018.

int_year
int_year is a numeric variable that specifies the year when the survey questionnaire was administered to the household.

int_month
int_month is a numeric variable that specifies the month when the survey questionnaire was administered to the household.

hhid
hhid specifies the unique household identification number in the data file. The original format, string or numeric, of original data should be kept. If there is Household ID in the original data, hhid and hhid_orig should be the same. If hhid_orig is missing, it is constructed by "variable names in raw data" variables.

hhid_orig
hhid_orig is the household identifier available in the original data.

pid
This variable allows identification of individuals. Variable will vary in length depending on how the identification code was constructed in each country. Depending on individual countries, this variable may be a concatenation of several variables in the raw data file. Keep format (string or numeric) of original data. If there is Personal ID in the original data, pid and pid_orig should be the same. If pid_orig is missing, it is constructed by "variable names in raw data" variables.

pid_orig
pid_orig is the individual identifier within the household available in the original data set.

variable names in raw data
These are the variables in the raw data used to construct hhid and/or pid, when hhid_orig and pid_orig are not available.

2.4 Lessons Learned/Challenges

Creating and checking IDs

Avoid using the sequential index of the observation as the ID (i.e. gen hhid = _n). This is dangerous as the order of each observation may be different, even across vintages of the same file sorted by to different variables.

Create `hhid` like this:

✓ `gen hhid= psu * 100 + hh`

Not like this:

✘ `gen hhid = _n`

When creating *hhid* and *pid*, especially from string variables or from *group(varlist)* or *concat(varlist)* functions, users should try to create them from roster data files first where all information or observations are available. In addition, the order of the variables in the *varlist* option above must be the same across the files. Across the data files, the order and the sort on the variables in the varlist must be done in the same way across files.

When the *hhid* and *pid* are in numeric format but less precision, it is recommended to bring them the accurate precision level so it can be used in the merging correctly. For example, the value of the hhid for an observation might be 100021210121 (a long number), users should format the variable by “format %15.0g hh”.

In case a household survey is conducted more than once per year – e.g. quarterly HH surveys – you may want to use this as panel data, in which case the household ID can remain as is. However, if you want to use the data as cross-sectional, then new HHIDs can be constructed for each HH for each quarter.

Quarter	Quarter 1	Quarter 2	Quarter 3	Quarter 4
hhid_orig	hhid=1	hhid=1	hhid=1	hhid=1
hhid	hhid=1Q1	hhid=1Q2	hhid=1Q3	hhid=1Q4

hhid should never be missing and if there is any missing this variable should be checked.

```
assert missing(hhid)
```

It is recommended to check the uniqueness level of the data files with identifier variables at the corresponding level of the data (i.e. household vs individual level data).

hhid and pid need to be unique in the database.

```
isid hhid pid
cap destring pid, replace
duplicates report hhid pid
local n=r(unique_value)
`N'!= `n'
```

Harmonizers should also ensure that there is a perfect match between the hhid and the pid of each household and individual, respectively, across modules. Even if the some variables are missing for some individuals/households, these observations should be included in the dataset with missing values for the related information to ensure that the datasets merge.

Checks on the country code and year

Ensure that country is a three-letter country code.

```
cap confirm str3 var country _rc!=0
```

Harmonizers should also ensure that country codes are updated according to the ISO country codes (can be found in Appendix A.) Some common adjustments include the following:

```
cap replace countrycode="XKX" if countrycode=="KSV"
cap replace countrycode="TLS" if countrycode=="TMP"
cap replace countrycode="PSE" if countrycode=="WBG"
cap replace countrycode="COD" if countrycode=="ZAR"
```

Furthermore, harmonizers should check that the years used are in an appropriate range.

The year needs to be a four-digit number in the range of 1980 to the current year (assumed here to be 2020).

```
(year<1980 | year>2020) & mod(year, 1) == 0
```

Table 1 GMD ID Module – key variables and descriptions

Module Code	Variable label	Variable name	Allowed codes after standardization
ID	country code	**countrycode**	string
ID	Year	**year**	numeric
ID	int_year	**int_year**	numeric
ID	int_month	**int_month**	numeric
ID	Household identifier	**hhid**	string or numeric
ID	Household identifier in the raw data	**hhid_orig**	numeric or string
ID	Personal identifier	**pid**	string or numeric
ID	Personal identifier in the raw data	**pid_orig**	numeric or string
ID	Household weights	**weight**	Numeric
ID	Variables used to construct Household identifier	***variable name in raw data***	numeric or string

3 Geography (GEO)

3.1 Framework for Harmonization

The GMD's Geography module presents the geographic context in which households are situated. This chapter describes geographical dimensions that are commonly derived from survey data sets or that can be constructed using existing variables. The overall objective of this indicator harmonization is to generate a unified data source of globally comparable geographical indicators to support routine analytical and corporate business functions of the World Bank.

The international guidelines presented in this section contribute to the development of internationally comparable geographical statistics. Geographical indicators are instrumental in building an evidence base of the drivers and consequences of regional and local inequities and in informing policies for fostering regional convergence and other development outcomes. Conversely, the lack of adequate and internationally comparable geographical data is a major impediment to informed and effective policies. Regional breakdowns of information are necessary for several SDGs, such as SDG 1.1, which requires information for populations in different geographic areas. Moreover, it is crucial to understand uneven progress on SDGs within countries wherever possible, which requires information at the lowest subnational level available.

In this module, the primary unit of analysis is the household. Geography module variables typically cover the entire country with all its subdivisions, unless indicated otherwise. The harmonization framework consists of fifteen variables in the sub-national administrative structure of the country. For clarity and convenience, the largest administrative subdivision of a country is referred to as the "first-level administrative division" or "first administrative level". The next level is called "second-level administrative division" or "second administrative level." Each variable can then be used to identify a country subdivision in a global context uniquely. The geographical mapping is also compatible with the classification of territorial units for statistics. In ECA, the NUTS classification is used for regional classification.[3] In other regions, the remaining sub-national indicators are country-specific and follow existing naming conventions. To the extent possible, however, the administrative codes used in the survey should match an existing shapefile used by the national statistics office that describes the location of the administrative units.

3.2 Mapping and Description of Variables

This section describes all variables available in the GMD Geography module.

urban
urban is a dummy variable that specifies the location type – urban or rural - of the household. This variable is country specific as each country uses its own criterion to distinguish urban from rural areas. In many

[3] NUTS areas aim to provide a single and coherent territorial breakdown for the compilation of EU regional statistics.

cases there is no clear division between urban and rural areas, and areas are classified as “semi-urban” or “mixed”. Harmonizers are advised to classify such categories as “urban.”

Urban categories:

1 = Urban
0 = Rural

subnatid1
subnatid1 refers to a subnational identifier at the highest level within the country’s administrative structure. This is typically a province or state. The variable is string and country-specific categorical. Numeric entries are coded in string format using the following naming convention: “1 – Hatay”.

subnatid2
subnatid2 refers to a subnational identifier at which survey is representative at the second highest level within the country’s administrative structure. This is typically a district. The variable is string and country-specific categorical. Numeric entries are coded in string format using the following naming convention: “1 – Hatay”.

subnatid3
subnatid3 refers to a sub-national identifier at which survey is representative at the third level within the country’s administrative structure. This is typically a sub-district. The variable is string and country-specific categorical. Numeric entries are coded in string format using the following naming convention: “1 – Hatay”.

subnatid4
subnatid4 refers to a sub-national identifier at which survey is representative at the lowest level within the country’s administrative structure. In some countries, this is effectively a village. The variable is string and country-specific categorical. Numeric entries are coded in string format using the following naming convention: “1 – Hatay”.

subnatidsurvey
subnatidsurvey is a string variable that refers to the lowest level of the administrative level at which the survey is representative. In most cases this will be equal to “subnatid1” or “subnatid2”. However, in some cases the lowest level is classified in terms of urban, rural or any other regional categorization cannot be mapped to subnatids. The variable would contain survey representation at lowest level irrespective of its mapping to subnatids.

subnatid1_prev
subnatid1_prev is coded as missing unless the classification used for subnatid1 has changed since the previous survey. In that case, it refers to the subnatid1 code used in the previous survey. This provides a way of tracking splits. For example, if province “32 – West Java” split into province “32 – West Java” and “36 – Banten” since the most recent survey, this variable would contain "32 – West Java” for both provinces 32 and 36.

subnatid2_prev

subnatid2_prev is coded as missing unless the classification used for subnatid2 has changed since the previous survey. In that case, it refers to the subnatid2 code used in the previous survey.

subnatid3_prev
subnatid3_prev is coded as missing unless the classification used for subnatid3 has changed since the previous survey. In that case, it refers to the subnatid3 code used in the previous survey.

subnatid4_prev
subnatid4_prev is coded as missing unless the classification used for subnatid4 has changed since the previous survey. In that case, it refers to the subnatid4 code used in the previous survey.

strata
strata refer to the division of the target population – typically the census sample frame -- into subpopulations based on auxiliary information that is known about the full population. Sampling is conducted separately for each strata. The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded. Sampling strata need to be considered when constructing the variance (or confidence intervals) of population estimates. strata is needed for the correct calculation of standard deviation for each sample design. A unique identifier is created for each stratum. In Stata, users are advised to specify strata through the svyset command. The variable is in string format with the following naming convention "code of stratum – stratum name", for example: "1 – Dar-es-salaam"

psu
primary sampling unit (psu) refers to sampling units that are selected in the first (primary) stage of multi-stage sample design. These sampling units typically correspond to a number of large aggregate units (clusters), each of which contains sub-units. For example, a primary sampling unit can represent the set of all housing units contained in a well-defined geographic area, such as a municipality or a group of contiguous municipalities. Primary sampling units are numeric and country-specific. A unique identifier is created for each primary sampling unit. In Stata, users are advised to specify the primary sampling unit through the svyset command.

gaul_adm1_code
gaul_adm1_code is numeric and country-specific based on the GAUL database. It should be taken from the same data in the GAUL database (a copy of those codes is available at the D4G team, contact Minh Nguyen at mnguyen3@worldbank.org or David Newhouse at dnewhouse@worldbank.org) where the geographical area can be identified in the survey based on the name of the location/area. *The number of unique values from the subnatid1 and the gaul_adm1_code could be different or the same.* For example, in the case of a fictional country, if the highest-level representation is the state level (53 states) and Gaul also has 53 states, it is the same in this case. In a different example, the survey is representative at the level of statistical regions (7) while the identifiable GAUL code is at state level (53 states); with this setup, one can know how the seven statistical regions are constructed.

gaul_adm2_code
gaul_adm2_code is numeric and country-specific based on the GAUL database. It should be taken from the same data in the GAUL database where the geographical area can be identified in the survey based on the name of the location/area.

3.3 Missing Value Codes

Harmonizers need to clearly differentiate missing values of variables from variables that were present in the survey but could not be harmonized due to reasons such as time unavailability. This will help the future harmonizers to focus on the unharmonized variables. The missing value code for these two scenarios are:

- For variables unavailable in survey = .
- For variables available in the survey but not harmonized = .a . To do so use this Stata command:
  gen *varname*=".a"

3.4 Challenges / common mistakes (Module specific)

- subnatid codes should reflect the most recent codes that pertain to that survey. subnatid_prev codes can be used to track splits and new administrative units that have been introduced since the previous survey. It is important to ensure there is consistency in geographic variables across time. Sub-nationally representative units may be added in later additions of surveys, so names of subnational units must be consistent across time. This will allow analysts to make the current administrative units "backwards-compatible" with little additional effort.
- Harmonizers should ensure the subnatid1 through subnatid4 are string variables NOT categorical.
- The urban variable cannot be different from zero or one.

```
urban!= 1 & urban!= 0
```

Table 2 GMD Geography Module – key variables and descriptions.

Module Code	Variable label	Allowed codes after standardization
ID	country code	string
ID	Year	numeric
ID	Household identifier	string or numeric
ID	Household weights	Numeric
Geography	Urban	1=Urban 0=Rural
Geography	Subnational ID - highest level	string, country-specific categorical variable; numeric entries in string format using the following naming convention: “1 – Hatay” (as string)
Geography	Subnational ID - second highest level	string, country-specific categorical variable; numeric entries in string format using the following naming convention: “1 – Hatay” (as string)
Geography	Subnational ID - third highest level	string, country-specific categorical variable; numeric entries in string format using the following naming convention: “1 – Hatay” (as string)
Geography	Subnational ID - fourth highest level	string, country-specific categorical variable; numeric entries in string format using the following naming convention: “1 – Hatay” (as string)
Geography	Lowest level of Subnational ID	string
Geography	Strata	string
Geography	PSU	numeric, country-specific
Geography	Subnatid previous - highest level	string, country-specific categorical variable;
Geography	Subnatid previous – second highest level	string, country-specific categorical variable;
Geography	Subnatid previous - third highest level	string, country-specific categorical variable;
Geography	Subnatid previous - lowest level	string, country-specific categorical variable;
Geography	Gaul Code	numeric
Geography	Gaul Code	numeric

4 Demography (DEM)

4.1 Framework for Harmonization

Multiple-topic household surveys collect data on the characteristics of both households and individuals within those households. The GMD Demography module covers demographic indicators that are commonly derived from survey data or that can be constructed using existing variables. The primary objective of the indicator harmonization is to generate a unified data source of globally comparable demographic characteristics to support routine analytical and corporate business functions of the World Bank.

In this module, the primary unit of analysis is the level of individuals within a household.

Since there is no commonly agreed-upon framework for demographic and migration indicators, these guidelines draw on various major projects and internationally accepted principles for the collection and production of demographic and migration statistical indicators based on household surveys. The mentioned projects include The World Bank Living Standards Measurement Study- Integrated Surveys on Agriculture (LSMS-ISA), the Demographic Health Surveys (DHS), International Income Distribution Database (I2D2) and other WB global and regional harmonization referenced in line with UN Sustainable Development Goal indicators.

Building upon the conceptual and operational foundations of this work, and in collaboration with a wide range of various relevant partners within and outside the World Bank, this chapters proposes a set of commonly collected and used indicators to assess the demographic and migration characteristics of individuals and households.

Demographic and migration statistics are essential for planning and monitoring socio-economic development programs. They are instrumental in building an evidence base of the drivers and consequences of poverty and inequality and in informing the necessary policy approaches for fostering poverty reduction and boosting shared prosperity, as well as other relevant development outcomes. Statistics on population composition by age and sex are among the most basic data necessary to describe a population and/or a subgroup of a population. The need for data decomposed by urban/rural, gender and age is clear from the indicators for the SDGs which call for breakdowns on these lines in indicators on poverty (SDG 1), education (SDG 4), gender issues (SDG 5), decent work (SDG 8), inequality (SDG 10), and urban issues (SDG 11). Conversely, the lack of adequate and internationally comparable demographic and migration statistics is a major impediment to informed and effective policies.

4.2 Mapping and Description of Variables

GMD Demography module contain a large amount of metadata that provides a wealth of information about the variables, including their type, description, sources, etc. To improve readability, only the most significant information has been included in this section. For a complete list of all variables captured in the module please consult Table at the end of the chapter. This section aims to provide the readers with basic information about the main sources of information and then some details about how GMD demographic indicators are produced, and the issues surrounding them.

Characteristics such as race, gender, age, and marital status, are all typical examples of demographics that are used in surveys. These are fundamental building blocks for most statistical analysis.

weight

weight contains household weights, typically inversely proportional to the probability of the household being selected for the sample, that should be applied to all analysis to make the results representative of the population.

language

language is a string variable that refers either to the one the respondent normally speaks in his or her present home (usual language) or the language usually spoken in the individual's home in his or her early childhood (mother tongue), or the language that the person commands best (main language). Its classification is country specific. Information on language (including any sign language) should be harmonized for all persons. In the tabulated results, the criterion for determining the language for children not yet able to speak should be clearly indicated. Numeric entries are coded in string format using the following naming convention: "2 – language".

age

age refers to the interval of time between the date of birth and the date of the survey. Every effort should be made to determine the precise and accurate age of each person, particularly of children[4] and older persons. Information on age may be secured either by obtaining the date (year, month, and day) of birth or by asking directly for age at the person's last birthday. In addition, in the case of children aged less than or equal to 60 months, variable age should be expressed in the number of completed ***years and months in decimals. For example, If the interview of a 4 years old was in December and he was born in June, his age should be recorded as 4.5***. Lastly, if the information on age is not available, it should be coded as missing rather than some other value such as "99" or "999".

agecat

agecat is a string variable that refers to age groups defined in the survey if information on age is only available in age categories rather than in years. For example:

"15 years or younger"
"15-24 years old"
"25-54 years old"
"55-64 years old"
"65 years or older"

male

male is a dummy variable that specifies the sex – male or female – of an individual within a household. While constructing this variable, it is important to make sure that all relevant values are included. Variable values coded as '98' or other numeric characters should be excluded from the values of the `male' variable. Sex of household member, two categories after harmonization:

1 = male

[4] Especially in the case of children aged less than 5 years old, age is used to interpret Anthropometrics data.

0 = female

relationharm
relationharm is a string variable that indicates a relationship to the reference person of household (usually the head of household). Variable values coded as '98' or other numeric characters should be excluded from the values of relationharm variable.

Relationship to head of household, six categories after harmonization:

1=head;
2=spouse;
3=children;
4=parents;
5=other relatives;
6=non-relatives

Note: In cases where head is missing or a migrant, we assign spouse as the head of the household. If spouse is also not available, then we will use oldest member of the household as the head and recode all the relations to head accordingly.

relationcs
relationcs is a country-specific categorical variable that indicates the relationship to the head of the household. The categories for relationship to the head of the household are defined according to the region or country requirements.

marital
Marital is a categorical variable that refers to the personal status of each individual in relation to the marriage laws or customs of the country. This variable should include at least the following: (a) married; (b) never married; (c) living together; (d) divorced/separated; (e) widowed. In some countries, category (a) may require a subcategory of persons who are contractually married but not yet living as man and wife. In all countries, category (d) should comprise both the legally and the de facto separated, who may be shown as separate subcategories if desired. The marital variable should not be imputed but rather calculated only for those to whom the question was asked (in other words, the youngest age at which information is collected may differ depending on the survey). The consistency between age and marital needs to be cross-checked. In most countries, there are also likely to be persons who were permitted to marry below the legal minimum age because of special circumstances. To permit international comparisons of data on marital status, however, any tabulations of marital status not cross-classified by exact age should at least distinguish between persons under 15 years of age and over. If it is not possible to distinguish between married and living together, then it should be assumed that the individual is married. Variable values coded as '98' or other numeric characters should be excluded from the values of the 'marital' variable.

Marital status, five categories after harmonization:

1=married
2=never married
3=living together

4=divorced/separated
5=widowed

eye_dsablty
eye_dsablty is a numerical variable that indicates whether an individual has any difficulty in seeing, even when wearing glasses. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

hear_dsablty
hear_dsablty is a numerical variable that indicates whether an individual has any difficulty in hearing even when using a hearing aid. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

walk_dsablty
walk_dsablty is a numerical variable that indicates whether an individual has any difficulty in walking or climbing steps. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

conc_dsord
conc_dsord is a numerical variable that indicates whether an individual has any difficulty concentrating or remembering. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

slfcre_dsablty
slfcre_dsablty is a numerical variable that indicates whether an individual has any difficulty with self-care such as washing all over or dressing. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

comm_dsablty

comm_dsablty is a numerical variable that indicates whether an individual has any difficulty communicating or understanding usual (customary) language. Categories after harmonization:

1 = No – no difficulty
2 = Yes – some difficulty
3 = Yes – a lot of difficulty
4 = Cannot do at all

4.3 Missing Value Codes

Harmonizers need to clearly differentiate missing values of variables from variables that were present in the survey but could not be harmonized due to reasons such as time unavailability. This will help the future harmonizers to focus on the unharmonized variables. The missing value code for these two scenarios are:

- For variables unavailable in survey = .
- For variables available in the survey but not harmonized = .a

4.4 Challenges/common mistakes (Module specific)

Data sets that are harmonized incorrectly can lead to skewed and/or incorrect data analysis. Harmonizers should run a series of checks to ensure data is harmonized properly, including the following:

Check to make sure that age is an integer since 5 years old.

```
age/int(age)!= 1 & age!= . & age > 5
```

age cannot have negative or extreme values (>120)

```
(age < 0 | age>120) & age<.
```

Male variable can only take one of two values 0 or 1 (or missing).

```
male!=. & male!= 1 & male!= 0
```

Check if male is missing.

```
male==.
```

Check to make sure that there is variation in male

```
egen sdmale = sd(male)      sdmale==0
```

relationharm must be an integer in the range [1,6].

```
relationharm<1 & relationharm>6 & mod(relationharm, 1) == 1
```

marital must be an integer in the range [1,5].

```
marital<0 & marital>5 & mod(marital, 1) == 1
```

Children are “Never married” and should be coded as so even though it may be perceived as obvious. The marital status of individuals should be harmonized for all individuals. Harmonizers should check to make sure children are not systematically left with missing values for marital.

```
tab age marital, missing
```

weight cannot be missing

```
weight==.
```

Additionally, harmonizers should ensure that the household size variable is calculated correctly. Not all the individuals reported in a household that form the raw data are current household members. For example, for the EU-SILC survey, a household contains the current member, but also the members of the previous survey who have left the household for reasons such as death or migration.

Table 3 GMD Demography Module – key variables and descriptions.

Module Code	Variable label	Variable name	Allowed codes after standardization
ID	country code	countrycode	string
ID	Year	year	numeric
ID	Household identifier	hhid	string or numeric
ID	Personal identifier	pid	string or numeric
Survey Characteristics & Welfare	Weight	weight	numeric
Demography & Migration	Language	language	Codes vary by country.
Demography & Migration	Age of individual (continuous)	age	numeric, continuous
Demography & Migration	Age of individual (categorical)	agecat	String, categorical
Demography & Migration	Sex of household member (male=1)	male	1=male 0=female
Demography & Migration	Relationship to head of household harmonized across all regions	relationharm	1=head; 2=spouse; 3=children; 4=parents; 5=other relatives; 6=non-relatives
Demography & Migration	Relationship to head of household country/region specific	relationcs	string, country-specific categorical variable
Demography & Migration	Marital status	marital	1=married; 2=never married; 3=living together; 4=divorced/separated; 5=widowed
Demography & Migration	Eye Disability	eye_dsablty	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all
Demography & Migration	Hearing Disability	hear_dsablty	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all

Demography & Migration	Walk Disability	walk_dsablty	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all
Demography & Migration	Concentration Disorder	conc_dsord	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all
Demography & Migration	Selfcare Disability	slfcre_dsablty	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all
Demography & Migration	Communication Disability	comm_dsablty	1 = No – no difficulty 2 = Yes – some difficulty 3 = Yes – a lot of difficulty 4 = Cannot do at all

5 Labor Module (LBR)

5.1 Framework of Harmonization

The GMD labor module contains a variety of variables relating to individuals' labor status and jobs. This includes their primary activity (employed, unemployed, or out of the labor force), as well as types, sectors of employment and wages of workers. This set of variables is identical to the labor variables coded in the International Income Distribution Database (I2D2). The primary objective of this module is to generate a unified data source of globally comparable labor indicators that can be easily linked to poverty and welfare indicators and support routine analytical and corporate business functions of the World Bank.

In this module, the primary unit of analysis is at the individual level. The age at which the labor module starts being applied, or the legal working age, is different for each country. Hence, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The data harmonized in this module is critical for understanding progress on SDG indicators. SDG 8 calls for data on the proportion of informal employment in non-agriculture sectors by sex (8.3), the average hourly earnings of female and male employees by occupation and age (8.5, partially), the unemployment rate by sex and age (8.5, partially), and the proportion and number of children aged 5-17 years engaged in child labor by sex and age (8.7). The data for these indicators can be linked to poverty, welfare, and other GMD variables through this module.

The module contains variables that encompass two different time units: the last seven days and the past twelve months. Considering that labor market dynamics differ among regions and countries, the idea of having two different time units is to be able to get a better characterization of the labor market at a country level. Developing countries, for example, usually have a substantial share of their labor force dedicated to agricultural activities. This type of labor activity is characterized by an important seasonal component, meaning that labor participation and the economic status of individuals varies significantly throughout the year. Therefore, it is probable that labor variables regarding last seven days will be affected by this seasonality, in contrast with annual variables, which are less prone to be sensitive to the seasonal component but may be more prone to recall error.

The last seven days variables include the employment status (paid employee, employer, self-employed, etc.), sector (public or private), industry and occupation classification as well as other important job's characteristics. GMD also contains variables for contract, social security, health insurance, and union membership, which are proxy measures of the extent to which a worker's job is formal. The last 12 months variables include employment status and number of jobs.

Consistent with the above, in many developing regions individuals usually have more than one job. The GMD labor module also contains variables regarding the second job both in the last seven days and during the past year. The second job variables include employment status, sector, industry and occupation classification.

ILO Resolution I

Labor variables are provided only for working-age population, which might be different for different countries. Furthermore, even though most countries have an established retiring age, GMD files do not consider an upper bounding age for the working population, which is line with ILO recommendations.[5] In this way, working age population is defined as those individuals whose age is greater than or equal to the minimum legal age to work, defined by each country. Besides this, many surveys (particularly in developing regions) also include a child labor module; this information is not provided in GMD files, which only include labor outcomes for those deemed old enough to be included in each country's labor module.

In 2013, The ILO adopted a new definition of employment, which reclassifies own-production and many forms of unpaid labor as not working.[6] As of 2019, these guidelines have yet to be widely adopted. Because most labor modules do not contain the required information to follow these guidelines, and to maintain comparability with older surveys, this module maintains the traditional pre-2013 definitions of employment.

In the next section, a more detailed description of labor module variables is provided.

5.2 Mapping and Description of Variables

The GMD Labour module contains a large amount of metadata that provides a wealth of information about variables, including their types, descriptions, sources, etc. There are four sections of variables: labour screening questions (last seven days), primary employment (last 7 days), secondary employment (last 7 days) and employment (last 12 months). To, improve readability, only the most significant information has been included in this section.

5.2.1 Labor status, 7-day reference period

All variables are numeric unless specified.

minlaborage

This is the lowest age for which the labor module is implemented in the survey or the minimum working age in the country. For this reason, the lower age cutoff at which information is collected will vary from country to country.

lstatus

[5] Resolution Concerning statistics of work, employment and labor underutilization http://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/normativeinstrument/wcms_230304.pdf

[6] ILO. 2013. Statistics and Databases. See here: https://www.ilo.org/global/statistics-and-databases/statistics-overview-and-topics/WCMS_470295/lang--en/index.htm

lstatus is an individual's ***labor status in the last 7 days***. The value must be missing for individuals less than the required age (minlaborage).

Three categories are used after harmonization:

1 = Employed
2 = Unemployed
3 = Not-in-labor force

All persons are considered active in the labor force if they presently have a job (formal or informal, i.e., employed) or do not have a job but are actively seeking work (i.e., unemployed).

1 = Employed

> Employed is defined as anyone who worked during the last 7 days or reference week, regardless of whether the employment was formal or informal, paid or unpaid, for a minimum of 1 hour. Individuals who had a job, but for any reason did not work in the last 7 days are considered employed.

2 = Unemployed

> A person is defined as unemployed if he or she is, presently not working but is actively seeking a job. The formal definition of unemployed usually includes being 'able to accept a job.' This last question was asked in a minority of surveys and is, thus, not incorporated in the present definition. A person presently not working but waiting for the start of a new job is considered unemployed.

3 = Not-in-labor force

> A person is defined as not-in-labor force if he or she is, presently not working and it is not actively seeking a job during the last 7 days or reference week.

nlfreason

nlfreason is the reason an individual was not in the labor force ***in the last 7 days***. This variable is constructed for all those who are not presently employed and are not looking for work (lstatus=3) and missing otherwise.

Five categories after harmonization:

1= Student (a person currently studying.)
2= Housewife (a person who takes care of the house, older people, or children)
3= Retired
4 = Disabled (a person who cannot work due to physical conditions)
5 = Other (a person does not work for any other reason)

Fill this information for all people interviewed in the labor section of the questionnaire regardless of their age.

unempldur_l
unempldur_l is a continuous variable specifying the ***duration of unemployment in months*** (***lower bracket***).

The variable is constructed for all unemployed persons (lstatus=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical it records the lower boundary of the bracket.

Missing values are allowed for everyone who is not unemployed.

unempldur_u
unempldur_u is a continuous variable specifying the ***duration of unemployment in months (upper bracket)***.

The variable is constructed for all unemployed persons (lstatus=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical it records the upper boundary of the bracket. If the right bracket is open a missing value should be inputted.

Missing values are allowed for everyone who is not unemployed.

If the duration of unemployment is not reported as a range, but as continuous variables, the **unempldur_l** and **unempldur_u** variables will have the same value. If the high range is open-ended the **unempldur_u** variable will be missing.

5.2.2 Primary Employment, 7-day reference period

All variables are numeric unless specified.

empstat
empstat is a categorical variable that specifies the ***main employment status in the last 7 days*** of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to take into account the data available.

Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer

4 = Self-employed
5 = Other, workers not classifiable by status

1 = Paid Employee

Paid employee includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid for piece work or in-kind. The 'continuous' criteria used in the ILO definition is not used here as data are often absent and due to country specificity.

2 = Non-Paid Employee

Non-paid employee includes contributing family workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment.

All apprentices should be mapped as 'non-paid employee'

3 = Employer

An employer is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and contributing family workers, the person is not considered an employer (as has no employees) and is, instead classified as self-employed.

4 = Self-employed

Own account or self-employment includes jobs where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and where one has not engaged any permanent employees to work for them on a continuous basis during the reference period.

Members of producers' cooperatives are workers who hold a self-employment job in a cooperative producing goods and services, in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.

5 = Other, workers not classifiable by status

Other, workers not classifiable by status include those for whom insufficient relevant information is available and/or who cannot be included in any of the above categories.

ocusec

ocusec is a categorical variable that specifies the ***sector of activity in the last 7 days***. It classifies the main job's sector of activity of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire.

Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish

1. Public Sector, Central Government, Army (including armed forces)

   Public sector is the part of economy run by the government.

2 = Private, NGO

   Private sector is that part of the economy which is both run for private profit and is not controlled by the state, it also includes non-governmental organizations

3 = State-owned enterprises

   State-owned includes para-state firms and all others in which the government has control (participation over 50%).

4 = Public or State-owned, but cannot distinguish

   Select this option is the questionnaire does not ask for State-owned enterprises, and only for Public sector.

Notes:

- Do not code basis of occupation (ISCO) or industry (ISIC) codes.

industry_orig
industry_orig is a string variable that specifies the ***original industry codes in the last 7 days for the main job*** provided in the survey (the actual question) and should correspond to whatever is in the original file with no recoding. The variable is constructed for all individuals that respond to this question, even if they are below the working age. It classifies the main job of any individual with a job (lstatus=1) and is missing otherwise

industrycat10
industrycat10 is a categorical variable that specifies the 1-digit industry classification ***in the last 7 days for the main job*** of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. The codes for the main job are given here based on the UN International Standard Industrial Classification. It classifies the main job of any individual with a job (lstatus=1) and is missing otherwise

Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining
3 = Manufacturing
4 = Public Utility Services
5 = Construction
6 = Commerce
7 = Transport and Communications
8 = Financial and Business Services
9 = Public Administration
10 = Other Services, Unspecified

Notes:

- In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes.
- Category 10 is also assigned for unspecified categories or items.
- If all 10 categories cannot be identified in the questionnaire create this variable as missing and proceed to create industrycat4.

industrycat4

industrycat4 is a categorical variable that specifies the 1-digit ***industry classification in the last 7 days for the main job*** for Broad Economic Activities. This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10.

Four categories after harmonization:

1 = Agriculture
2= Industry
3 = Services
4 = Other

This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10.

occup_orig

occup_orig is a string variable that specifies the ***original occupation code in the last 7 days for the main job***. This variable corresponds to whatever is in the original file with no recoding.

occup

occup is a categorical variable that specifies the 1-digit ***occupational classification for the main job in the last 7 days*** of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason,

the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Most surveys collect detailed information and then code it, without keeping the original data, no attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO). It classifies the main job of any individual with a job (lstatus=1) and is missing otherwise.

Eleven categories after harmonization:

1 = Managers
2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

wage_nc

wage_nc is a continuous variable that specifies the ***last wage payment in local currency*** of any individual (lstatus=1 & empstat<=4) in its primary occupation at the reference period reported in the survey and it is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- For all those with self-employment or owners of own businesses, this should be *net revenues* (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.
- By definition, non-paid employees (empstat=2) should have wage=0.
- The reference period of the wage_nc will be recorded in the **unitwage** variable.

unitwage

unitwage is a categorical variable that specifies the time reference for the *wage_nc* variable. It specifies the time unit measurement for the wages of any individual (lstatus=1 & empstat<=4) and it is missing otherwise. Acceptable values include:

1 = Daily
2 = Weekly
3 = Every two weeks
4 = Every two months

5 = Monthly
6 = Quarterly
7 = Every six months
8 = Annually
9 = Hourly
10 = Other

whours

whours is a continuous variable that specifies the hours of work last week for the main job of any individual with a job (lstatus=1) and is missing otherwise. The main job defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire. Notes:

- If the respondent was absent from the job in the week preceeding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.
- Sometimes the questions are phrased as, "on average, how many hours a week do you work?".
- For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.
- In the case of a question that has hours worked per month, divide by 4.3 to get weekly hours.

wmonths

wmonths is a continuous variable that specifies the number of months worked in the last 12 months for the main job of any individual with a job (lstatus=1) and is missing otherwise. The main job is defined as that occupation that the person dedicated more time to over the past week. The variable is constructed for all persons administered this module in each questionnaire.

wage_total

wage_total is a continuous variable that specifies the ***annualized wage payment*** (regular wage plus bonuses, in-kind, compensation, etc.) for the primary occupation in local currency of any individual (lstatus=1 & empstat<=4) and is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the week preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_total should be equal to wage_nc in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire. The annualization of the *wage_total* should consider the number of months/weeks the persons have been working and receiving this income. Harmonizer should not assume the person has been working the whole year.

Example: Creation of wage_total when there are no bonuses nor other compensations

```
gen double wage_total=.
replace wage_total=(wage_nc*5*4.3)*wmonths      if unitwage==1 //Wage in daily unit
replace wage_total=(wage_nc*4.3)*wmonths        if unitwage==2 //Wage in weekly unit
```

```
replace wage_total=(wage_nc*2.15)*wmonths          if unitwage==3 //Wage in every two weeks unit
replace wage_total=(wage_nc)/2*wmonths if unitwage==4 //Wage in every two months unit
replace wage_total=( wage_nc)*wmonths              if unitwage==5 //Wage in monthly unit
replace wage_total=( wage_nc)/3*wmonthsif unitwage==6 //Wage in every quarterly unit
replace wage_total=( wage_nc)/6*wmonthsif unitwage==7 //Wage in every six months unit
replace wage_total= wage_nc/12*wmonths if unitwage==8 //Wage in annual unit
replace wage_total=(wage_nc*whours*4.3)*wmonths if unitwage==9 //Wage in hourly unit
```

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

contract

contract is a dummy variable that classifies the contract status (yes/no) of any individual with a job (lstatus=1) and is missing otherwise. It indicates whether a person has a signed (formal) contract, regardless of duration. The variable is constructed for all persons administered this module in each questionnaire. Two categories after harmonization:

0 = No
1 = Yes

healthins

healthins is a dummy variable that classifies the health insurance status (yes/no) of any individual with a job (lstatus=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. However, this variable is only constructed if there is an explicit question about health insurance provided by the job. Two categories after harmonization:

0 = No
1 = Yes

socialsec

socialsec is a dummy variable that classifies the social security status (yes/no) of any individual with a job (lstatus=1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. However, this variable is only constructed if there is an explicit question about pension plans or social security. Two categories after harmonization:

0 = No
1 = Yes

union

union is a dummy variable that classifies the union membership status (yes/no) of any individual with a job (lstatus=1) and is missing otherwise. Variable is constructed for all persons administered this module

in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. However, this variable is only constructed if there is an explicit question about trade unions. Two categories after harmonization:

0 = No
1 = Yes

firmsize_l
firmsize_l specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 7 days for the main job***. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

firmsize_u
firmsize_u specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 7 days for the main job***. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the upper boundary of the bracket. If the right bracket is open, this variable should be missing.

5.2.3 Secondary Employment, 7-day reference period

empstat_2
empstat_2 is a categorical variable that specifies employment status of the secondary job ***with reference period of last 7 days*** of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to take into account the data available.

Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer
4 = Self-employed
5 = Other, workers not classifiable by status

1 = Paid Employee

Paid employee includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid

for piece work or in-kind. The 'continuous' criteria used in the ILO definition is not used here as data are often absent and due to country specificity.

2 = Non-Paid Employee

Non-paid employee includes contributing family workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment.

All apprentices should be mapped as non-paid employee.

3 = Employer

Employer is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and 'contributing family workers, the person is not considered an employer (as has no employees) and is, instead classified as own account.

4 = Self-employed

Own account or self-employment includes jobs are those where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and have not engaged any permanent employees to work for them on a continuous basis during the reference period.

Members of producers' cooperatives are workers who hold a self-employment job in a cooperative producing goods and services in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.

5 = Other, workers not classifiable by status

Other, workers not classifiable by status include those for whom insufficient relevant information is available and/or who cannot be included in any of the above categories.

ocusec_2

ocusec_2 is a categorical variable that specifies the ***sector of activity in the last 7 days***. It classifies the secondary job's sector of activity of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire.

Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)

2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish

2. Public Sector, Central Government, Army (including armed forces)

Public sector is the part of economy run by the government.

2 = Private, NGO

Private sector is that part of the economy which is both run for private profit and is not controlled by the state, it also includes non-governmental organizations

3 = State-owned enterprises

State-owned includes para-state firms and all others in which the government has control (participation over 50%).

4 = Public or State-owned, but cannot distinguish

Select this option is the questionnaire does not ask for State-owned enterprises, and only for Public sector.

Notes:

- Do not code basis of occupation (ISCO) or industry (ISIC) codes.

industry_orig_2

industry_orig_2 is a string variable that specifies the original industry codes for the second job ***with reference period of the last 7 days*** and should correspond to whatever is in the original file with no recoding. The variable is constructed for all individuals that respond to this question, even if they are below the working age. It classifies the main job of any individual with a job (lstatus=1) and is missing otherwise

industrycat10_2

industrycat10_2 is a categorical variable that specifies the 1-digit industry classification that classifies the second job ***with reference period of the last 7 days*** of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. The codes for the second job are given here based on the UN International Standard Industrial Classification.

Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining

3 = Manufacturing
4 = Public Utility Services
5 = Construction
6 = Commerce
7 = Transport and Communications
8 = Financial and Business Services
9 = Public Administration
10 = Other Services, Unspecified

Notes:

- In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes.
- Category 10 is also assigned for unspecified categories or items.

industrycat4_2

industrycat4_2 is a categorical variable that specifies the 1-digit industry classification for Broad Economic Activities for the second job ***with reference period of the last 7 days***. This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from industrycat10_2.

Four categories after harmonization:

1 = Agriculture
2= Industry
3 = Services
4 = Other

occup_orig_2

occup_orig_2 is a string variable that specifies the ***original occupation code in the last 7 days for the secondary job***. This variable corresponds to whatever is in the original file with no recoding.

occup_2

occup_2 is a categorical variable that specifies the 1-digit occupation classification. It classifies the second job of any individual with a job (lstatus=1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. Most surveys collect detailed information and then code it, without keeping the original data. No attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO). In the case of different classifications, re-coding has been done to best match the ISCO.

Eleven categories after harmonization:

1 = Managers

2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

wage_nc_2

wage_nc_2 is a continuous variable that specifies ***the last wage payment in local currency of any individual (lstatus=1 & empstat_2<=4) in its secondary occupation*** and is missing otherwise. The wage should come from the second job, in other words, the job that the person dedicated the second most amount of time in the week preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.
- By definition, non-paid employees (empstat_2=2) should have wage=0.
- The reference period of the wage_nc_2 will be recorded in the unitwage_2 variable

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

unitwage_2

unitwage_2 is a categorical variable that specifies the time reference for the wage_nc_2 variable. It specifies the time unit measurement for the wages for the secondary job of any individual (lstatus=1 & empstat<=4) and is missing otherwise.

Ten categories after harmonization:

1 = Daily
2 = Weekly
3 = Every two weeks
4 = Every two months
5 = Monthly
6 = Quarterly
7 = Every six months
8 = Annually

9 = Hourly
10 = Other

whours_2

whours_2 is a continuous variable that specifies the hours of work in ***last week for the second job with reference period of the last 7 days*** of any individual with a job (lstatus=1) and is missing otherwise. The second job defined as that occupation that the person dedicated the second most amount of time to over the past week. The variable is constructed for all persons administered this module in each questionnaire. The lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Notes:

- If the respondent was absent from the job in the week preceeding the survey due to holidays, vacation, or sick leave, then record the time worked in the previous 7 days that the person worked.
- Sometimes the questions are phrased as, "on average, how many hours a week do you work?".
- For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.
- In the case of a question that has hours worked per month, divide by 4.3 to get weekly hours.

wmonths_2

wmonths_2 is a continuous variable that specifies the number of months worked in the last 12 months for the secondary job of any individual with a job (lstatus=1) and is missing otherwise. The secondary job is defined as that occupation in which the person dedicated less time than the primary job over the past week. The variable is constructed for all persons administered this module in each questionnaire.

wage_total_2

wage_total_2 is a continuous variable that specifies the *annualized wage payment* (regular wage plus bonuses, in-kind, compensation, etc.) in local currency of any individual (lstatus=1 & empstat<=4) in its secondary occupation and is missing otherwise. The wage should come from the secondary job, in other words, the job that the person dedicated the second most amount of time in the week preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_total_2 should be equal to wage_nc_2 in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- The annualization of the *wage_total_2* should consider the number of months/weeks the persons have been working and receiving this income. Harmonizer should not assume the person has been working the whole year.
- For an example on how to annualize wage see 5.2.2 Primary Employment last 7-days variable wage_total (Example: Creation of wage_total when there are no bonuses nor other compensations).

firmsize_l_2

firmsize_l_2 specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

firmsize_u_2

firmsize_u_2 specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the upper boundary of the bracket. If the right bracket is open, a missing value should be inputted.

5.2.4 Other Employment, 7-day reference period

t_hours_others

t_hours_others is a continuous variable that specifies the hours of work in last 12 months **in all jobs excluding the primary and secondary ones**.

t_wage_nc_others

t_wage_nc_others is a continuous variable that specifies the annualized wage **in all jobs excluding the primary and secondary ones** . This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

t_wage_others

t_wage_others is a continuous variable that specifies the annualized wage **in all jobs excluding the primary and secondary ones**. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_others should be equal to wage_nc_ others in case there are no bonuses, tips etc. offered as part of any of the jobs.

5.2.5 Total Employment Earnings, 7-day reference period

t_hours_total

t_hours_total is a continuous variable that specifies the hours of work in last 12 months in **all jobs including primary, secondary and others**.

t_wage_nc_total

t_wage_nc_total is a continuous variable that specifies the total annualized wage income in **all jobs including primary, secondary and others**. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

t_wage_total
t_wage_total is a continuous variable that specifies the total annualized wage income in **all jobs including primary, secondary and others**. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. t_wage_total should be equal to t_wage_nc_total in case there are no bonuses, tips etc. offered as part of any of the jobs. If the number of months worked in this job is missing the harmonizer could assumed that the person worked the whole year in this job.

5.2.6 Labor status, 12-month reference period

> This section must be filled only for those individuals who responded to labor questions with a reference period of 12 months, regardless of whether they responded to questions with a reference period of the last 7 days.

All variables are numeric unless specified.

minlaborage_year
This is the lowest age for which the labor module is implemented in the survey or the minimum working age in the country. For this reason, the lower age cutoff at which information is collected will vary from country to country.

lstatus_year
lstatus_year is an individual's ***labor status in the last 12 months***. The value must be missing for individuals less than the required age (minlaborage).

Three categories are used after harmonization:

1 = Employed
2 = Unemployed
3 = Not-in-labor force

All persons are considered active in the labor force if they presently have a job (formal or informal, i.e., employed) or do not have a job but are actively seeking work (i.e., unemployed).

1 = Employed

> Employed is defined as anyone who worked during the last 12 months or reference week, regardless of whether the employment was formal or informal, paid or unpaid, for a minimum of 1 hour. Individuals who had a job, but for any reason did not work in the last 7 days are considered employed.

2 = Unemployed

> A person is defined as unemployed if he or she is, presently not working but is actively seeking a job. The formal definition of unemployed usually includes being 'able to accept a job.' This last question was asked in a minority of surveys and is, thus, not incorporated in the present definition. A person presently not working but waiting for the start of a new job is considered unemployed.

3 = Not-in-labor force

> A person is defined as not-in-labor force if he or she is, presently not working and it is not actively seeking a job during the last 12 months or reference week.

nlfreason_year

nlfreason_year is the reason an individual was not in the labor force ***in the last 12 months***. This variable is constructed for all those who are not presently employed and are not looking for work (lstatus_year=3) and missing otherwise.

Five categories after harmonization:

1= Student (a person currently studying.)
2= Housewife (a person who takes care of the house, older people, or children)
3= Retired
4 = Disabled (a person who cannot work due to physical conditions)
5 = Other (a person does not work for any other reason)

Fill this information for all people interviewed in the labor section of the questionnaire regardless of their age.

unempldur_l_year

unempldur_l_year is a continuous variable specifying the ***duration of unemployment in months*** (***lower bracket***). The variable is constructed for all unemployed persons (lstatus_year=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical it records the lower boundary of the bracket. Missing values are allowed for everyone who is not unemployed.

unempldur_u_year
unempldur_u_year is a continuous variable specifying the ***duration of unemployment in months (upper bracket)***.

The variable is constructed for all unemployed persons (lstatus_year=2, otherwise missing). If it is specified as continuous in the survey, it records the numbers of months in unemployment. If the variable is categorical it records the upper boundary of the bracket. If the right bracket is open a missing value should be inputted. Missing values are allowed for everyone who is not unemployed. If the duration of unemployment is not reported as a range, but as continuous variables, the **unempldur_l_year** and **unempldur_u_year** variables will have the same value. If the high range is open-ended the **unempldur_u_year** variable will be missing.

5.2.7 Primary Employment, 12-month reference period

.

> This section must be filled only for those individuals who responded to labor questions with a reference period of 12 months, regardless of whether they responded to questions with a reference period of the last 7 days.

All variables are numeric unless specified.

empstat_year
empstat is a categorical variable that specifies the ***main employment status in the last 12 months*** of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to take into account the data available.

Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer
4 = Self-employed
5 = Other, workers not classifiable by status

1 = Paid Employee

Paid employee includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid for piece work or in-kind. The 'continuous' criteria used in the ILO definition is not used here as data are often absent and due to country specificity.

2 = Non-Paid Employee

Non-paid employee includes contributing family workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment.

All apprentices should be mapped as 'non-paid employee'

3 = Employer

An employer is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and contributing family workers, the person is not considered an employer (as has no employees) and is, instead classified as self-employed.

4 = Self-employed

Own account or self-employment includes jobs where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and where one has not engaged any permanent employees to work for them on a continuous basis during the reference period.

Members of producers' cooperatives are workers who hold a self-employment job in a cooperative producing goods and services, in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.

5 = Other, workers not classifiable by status

Other, workers not classifiable by status include those for whom insufficient relevant information is available and/or who cannot be included in any of the above categories.

ocusec_year

ocusec_year is a categorical variable that specifies the ***sector of activity in the last 12 months***. It classifies the main job's sector of activity of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all persons administered this module in each questionnaire.

Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned

4 = Public or State-owned, but cannot distinguish

3. Public Sector, Central Government, Army (including armed forces)

   Public sector is the part of economy run by the government.

2 = Private, NGO

   Private sector is that part of the economy which is both run for private profit and is not controlled by the state, it also includes non-governmental organizations

3 = State-owned enterprises

   State-owned includes para-state firms and all others in which the government has control (participation over 50%).

4 = Public or State-owned, but cannot distinguish

   Select this option is the questionnaire does not ask for State-owned enterprises, and only for Public sector.

Notes: Do not code this variable on the basis of occupation (ISCO) or industry (ISIC) codes.

industry_orig_year

industry_orig_year is a string variable that specifies the ***original industry codes in the last 12 months for the main job*** provided in the survey (the actual question) and should correspond to whatever is in the original file with no recoding. The variable is constructed for all individuals that respond to this question, even if they are below the working age. It classifies the main job of any individual with a job (lstatus_year =1) and is missing otherwise

industrycat10_year

industrycat10_year is a categorical variable that specifies the 1-digit industry classification ***in the last 12 months for the main job*** of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. The codes for the main job are given here based on the UN International Standard Industrial Classification. It classifies the main job of any individual with a job (lstatus_year =1) and is missing otherwise

Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining

3 = Manufacturing
4 = Public Utility Services
5 = Construction
6 = Commerce
7 = Transport and Communications
8 = Financial and Business Services
9 = Public Administration
10 = Other Services, Unspecified

Notes:

- In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes.
- Category 10 is also assigned for unspecified categories or items.
- If all 10 categories cannot be identified in the questionnaire create this variable as missing and proceed to create industrycat4_year.

industrycat4_year

industrycat4_year is a categorical variable that specifies the 1-digit ***industry classification in the last 12 months for the main job*** for Broad Economic Activities. This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10_year.

Four categories after harmonization:

1 = Agriculture
2= Industry
3 = Services
4 = Other

This variable is either created directly from the data (if industry classification does not exist for ten categories) or created from industrycat10_year.

occup_orig_year

occup_orig_year is a string variable that specifies the ***original occupation code in the last 12 months for the main job***. This variable corresponds to whatever is in the original file with no recoding.

occup_year

occup_year is a categorical variable that specifies the 1-digit ***occupational classification for the main job in the last 12 months*** of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Most surveys collect detailed information and then code it, without keeping the original data, no attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO). It classifies the main job of any individual with a job (lstatus_year=1) and is missing otherwise.

Eleven categories after harmonization:

1 = Managers
2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

wage_nc_year

wage_nc_year is a continuous variable that specifies the ***last wage payment in local currency*** of any individual (lstatus_year =1 & empstat_year =1) in its primary occupation at the reference period reported in the survey and it is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the 12 months preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- For all those with self-employment or owners of own businesses, this should be *net revenues* (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.
- By definition, non-paid employees (empstat_year=2) should have wage=0.
- The reference period of the wage_nc_year will be recorded in the **unitwage_year** variable.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

unitwage_year

unitwage_year is a categorical variable that specifies the time reference for the *wage_nc_year* variable. It specifies the time unit measurement for the wages of any individual (lstatus_year =1 & empstat_year =1) and it is missing otherwise. Acceptable values include:

1 = Daily
2 = Weekly
3 = Every two weeks
4 = Every two months
5 = Monthly
6 = Quarterly
7 = Every six months
8 = Annually

9 = Hourly
10 = Other

whours_year
whours_year is a continuous variable that specifies the hours of work last week for the main job of any individual with a job (lstatus_year =1) and is missing otherwise. The main job defined as that occupation that the person dedicated more time to over the past 12 months. The variable is constructed for all persons administered this module in each questionnaire. Notes:

- Sometimes the questions are phrased as, “on average, how many hours a week do you work?”.
- For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.
- In the case of a question that has hours worked per month, divide by 4.3 to get weekly hours.

wmonths_year
wmonths_year is a continuous variable that specifies the number of months worked in the last 12 months for the main job of any individual with a job (lstatus_year =1) and is missing otherwise. The main job is defined as that occupation that the person dedicated more time to over the past 12 months. The variable is constructed for all persons administered this module in each questionnaire.

wage_total_year
wage_total_year is a continuous variable that specifies the ***annualized wage payment*** (regular wage plus bonuses, in-kind, compensation, etc.) for the primary occupation in local currency of any individual (lstatus_year =1 & empstat_year =1) and is missing otherwise. The wage should come from the main job, in other words, the job that the person dedicated most time in the year preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_total_year should be equal to wage_nc_year in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire. The annualization of the *wage_total_year* should consider the number of months/weeks the persons have been working and receiving this income. Harmonizer should not assume the person has been working the whole year.

- For an example on how to annualize wage see 5.2.2 Primary Employment last 7-days variable wage_total (Example: Creation of wage_total when there are no bonuses nor other compensations).

contract_year
contract_year is a dummy variable that classifies the contract status (yes/no) of any individual with a job (lstatus_year =1) and is missing otherwise. It indicates whether a person has a signed (formal) contract, regardless of duration. The variable is constructed for all persons administered this module in each questionnaire. Two categories after harmonization:

0 = No
1 = Yes

healthins_year

healthins_year is a dummy variable that classifies the health insurance status (yes/no) of any individual with a job (lstatus_year =1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. However, this variable is only constructed if there is an explicit question about health insurance provided by the job. Two categories after harmonization:

0 = No
1 = Yes

socialsec_year

socialsec_year is a dummy variable that classifies the social security status (yes/no) of any individual with a job (lstatus_year =1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. However, this variable is only constructed if there is an explicit question about pension plans or social security. Two categories after harmonization:

0 = No
1 = Yes

union_year

union_year is a dummy variable that classifies the union membership status (yes/no) of any individual with a job (lstatus_year =1) and is missing otherwise. Variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. However, this variable is only constructed if there is an explicit question about trade unions. Two categories after harmonization:

0 = No
1 = Yes

firmsize_l_year

firmsize_l_year specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 12 months for the main job***. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

firmsize_u_year

firmsize_u_year specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed ***in the last 12 months for the main job***. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the upper boundary of the bracket. If the right bracket is open, this variable should be missing.

5.2.8 Secondary Employment, 12-month reference period

empstat_2_year

empstat_2_year is a categorical variable that specifies employment status of the secondary job ***with reference period of last 12 months*** of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. For this reason, the lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country.

The definitions are taken from the International Labor Organization's Classification of Status in Employment with some revisions to take into account the data available.

Five categories after harmonization:

1 = Paid Employee
2 = Non-Paid Employee
3 = Employer
4 = Self-employed
5 = Other, workers not classifiable by status

1 = Paid Employee

> Paid employee includes anyone whose basic remuneration is not directly dependent on the revenue of the unit they work for, typically remunerated by wages and salaries but may be paid for piece work or in-kind. The 'continuous' criteria used in the ILO definition is not used here as data are often absent and due to country specificity.

2 = Non-Paid Employee

> Non-paid employee includes contributing family workers who hold a self-employment job in a market-oriented establishment operated by a related person living in the same households who cannot be regarded as a partner because of their degree of commitment to the operation of the establishment, in terms of working time or other factors, is not at a level comparable to that of the head of the establishment.
>
> All apprentices should be mapped as non-paid employee.

3 = Employer

> Employer is a business owner (whether alone or in partnership) with employees. If the only people working in the business are the owner and 'contributing family workers, the person is not considered an employer (as has no employees) and is, instead classified as own account.

4 = Self-employed

Own account or self-employment includes jobs are those where remuneration is directly dependent from the goods and service produced (where home consumption is considered to be part of the profits) and have not engaged any permanent employees to work for them on a continuous basis during the reference period.

Members of producers' cooperatives are workers who hold a self-employment job in a cooperative producing goods and services in which each member takes part on an equal footing with other members in determining the organization of production, sales and/or other work of the establishment, the investments and the distribution of the proceeds of the establishment amongst the members.

5 = Other, workers not classifiable by status

Other, workers not classifiable by status include those for whom insufficient relevant information is available and/or who cannot be included in any of the above categories.

ocusec_2_year

ocusec_2_year is a categorical variable that specifies the ***sector of activity in the last 12 months***. It classifies the secondary job's sector of activity of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age.

Four categories after harmonization:

1 = Public sector, Central Government, Army (including armed forces)
2 = Private, NGO
3 = State-owned
4 = Public or State-owned, but cannot distinguish

4. Public Sector, Central Government, Army (including armed forces)

Public sector is the part of economy run by the government.

2 = Private, NGO

Private sector is that part of the economy which is both run for private profit and is not controlled by the state, it also includes non-governmental organizations

3 = State-owned enterprises

State-owned includes para-state firms and all others in which the government has control (participation over 50%).

4 = Public or State-owned, but cannot distinguish

Select this option is the questionnaire does not ask for State-owned enterprises, and only for Public sector.

Notes: Do not code this variable on the basis of occupation (ISCO) or industry (ISIC) codes.

industry_orig_2_year

industry_orig_2_year is a string variable that specifies the original industry codes for the second job ***with reference period of the last 12 months*** and should correspond to whatever is in the original file with no recoding. The variable is constructed for all individuals that respond to this question, even if they are below the working age. It classifies the main job of any individual with a job (lstatus_year=1) and is missing otherwise

industrycat10_2_year

industrycat10_2_year is a categorical variable that specifies the 1-digit industry classification that classifies the second job ***with reference period of the last 12 months*** of any individual with a job (lstatus_year =1) and is missing otherwise. The variable is constructed for all individuals that respond to this question, even if they are below the working age. The codes for the second job are given here based on the UN International Standard Industrial Classification.

Ten categories after harmonization:

1 = Agriculture, Hunting, Fishing, etc.
2 = Mining
3 = Manufacturing
4 = Public Utility Services
5 = Construction
6 = Commerce
7 = Transport and Communications
8 = Financial and Business Services
9 = Public Administration
10 = Other Services, Unspecified

Notes:

- In the case of different classifications (former Soviet Union republics, for example), recoding has been done to best match the ISIC codes.
- Category 10 is also assigned for unspecified categories or items.

industrycat4_2_year

industrycat4_2_year is a categorical variable that specifies the 1-digit industry classification for Broad Economic Activities for the second job ***with reference period of the last 12 months***. This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from industrycat10_year.

Four categories after harmonization:

1 = Agriculture
2= Industry
3 = Services
4 = Other

This variable is either created directly from the data (if industry classification does not exist for 10 categories) or created from industrycat10_2_year.

occup_orig_2_year
occup_orig_2_year is a string variable that specifies the ***original occupation code in the last 12 months for the secondary job***. This variable corresponds to whatever is in the original file with no recoding.

occup_2_year
occup_2_year is a categorical variable that specifies the 1-digit occupation classification. It classifies the second job of any individual with a job (lstatus_year =1) and is missing otherwise. _Most surveys collect detailed information and then code it, without keeping the original data. No attempt has been made to correct or check the original coding. The classification is based on the International Standard Classification of Occupations (ISCO). In the case of different classifications, re-coding has been done to best match the ISCO.

Eleven categories after harmonization:

1 = Managers
2 = Professionals
3 = Technicians and associate professionals
4 = Clerical support workers
5 = Service and sales workers
6 = Skilled agricultural, forestry and fishery workers
7 = Craft and related trades workers
8 = Plant and machine operators, and assemblers
9 = Elementary occupations
10 = Armed forces occupations
99 = Other/unspecified

wage_nc_2_year
wage_nc_2_year is a continuous variable that specifies ***the last wage payment in local currency of any individual (lstatus_year =1 & empstat_2_year =1) in its secondary occupation*** and is missing otherwise. The wage should come from the second job, in other words, the job that the person dedicated the second most amount of time in the week preceding the survey. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- For all those with self-employment or owners of own businesses, this should be net revenues (net of all costs EXCEPT for tax payments) or the amount of salary taken from the business. Due to the almost complete lack of information on taxes, the wage from main job is NOT net of taxes.
- By definition, non-paid employees (empstat_year_2 =2) should have wage=0.
- The reference period of the wage_nc_year_2 will be recorded in the unitwage_2_year variable

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

unitwage_2_year
unitwage_2_year is a categorical variable that specifies the time reference for the wage_nc_2_year variable. It specifies the time unit measurement for the wages for the secondary job of any individual (lstatus_year =1 & empstat_2_year =1) and is missing otherwise.

Ten categories after harmonization:

1 = Daily
2 = Weekly
3 = Every two weeks
4 = Every two months
5 = Monthly
6 = Quarterly
7 = Every six months
8 = Annually
9 = Hourly
10 = Other

whours_2_year
whours_2_year is a continuous variable that specifies the hours of work in ***last week for the second job with reference period of the last 12 months*** of any individual with a job (lstatus_year =1) and is missing otherwise. The second job defined as that occupation that the person dedicated the second most amount of time to over the past year. The variable is constructed for all persons administered this module in each questionnaire. The lower age cutoff (and perhaps upper age cutoff) at which information is collected will vary from country to country. Notes:

- Sometimes the questions are phrased as, “on average, how many hours a week do you work?”.
- For individuals who only give information on how many hours they work per day and no information on number of days worked a week, multiply the hours by 5 days.
- In the case of a question that has hours worked per month, divide by 4.3 to get weekly hours.

wmonths_2_year
wmonths_2_year is a continuous variable that specifies the number of months worked in the last 12 months for the secondary job of any individual with a job (lstatus_year =1) and is missing otherwise. The secondary job is defined as that occupation in which the person dedicated less time than the primary job

over the past year. The variable is constructed for all persons administered this module in each questionnaire.

wage_total_2_year

wage_total_2_year is a continuous variable that specifies the *annualized wage payment* (regular wage plus bonuses, in-kind, compensation, etc.) in local currency of any individual (lstatus_year =1 & empstat_2_year =1) in its secondary occupation and is missing otherwise. The wage should come from the secondary job, in other words, the job that the person dedicated the second most amount of time in the year preceding the survey. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. wage_total_2_year should be equal to wage_nc_2_year in case there are no bonuses, tips etc. offered as part of the job. The variable is constructed for all persons administered this module in each questionnaire. For this reason, the lower age cutoff (and perhaps upper age cutoff) will vary from country to country. Notes:

- The annualization of the *wage_total_2_year* should consider the number of months/weeks the persons have been working and receiving this income. Harmonizer should not assume the person has been working the whole year.
- For an example on how to annualize wage see 5.2.2 Primary Employment last 7-days variable wage_total (Example: Creation of wage_total when there are no bonuses nor other compensations).

firmsize_l_2_year

firmsize_l_2_year specifies the lower bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the lower boundary of the bracket.

firmsize_u_2_year

firmsize_u_2_year specifies the upper bracket of the firm size. The variable is constructed for all persons who are employed. If it is continuous, it records the number of people working for the same employer. If the variable is categorical, it records the upper boundary of the bracket. If the right bracket is open, a missing value should be inputted.

5.2.9 Other Employment, 12-month reference period

t_hours_others_year

t_hours_others_year is a continuous variable that specifies the hours of work in last 12 months **in all jobs excluding the primary and secondary ones**.

t_wage_nc_others_year

t_wage_nc_others_year is a continuous variable that specifies annual wage **in all jobs excluding the primary and secondary ones**. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

t_wage_others_year

t_wage_others_year is a continuous variable that specifies the annual wage **in all jobs excluding the primary and secondary ones**. This wage includes tips, compensations such as bonuses, dwellings or clothes, and other payments. t_wage_others should be equal to t_wage_nc_ others in case there are no bonuses, tips etc. offered as part of any of the jobs.

5.2.10 Total Employment Earnings, 12-month reference period

t_hours_total_year

t_hours_total_year is a continuous variable that specifies the hours of work in last 12 months in **all jobs including primary, secondary and others**. Note:

t_wage_nc_total_year

t_wage_nc_total_year is a continuous variable that specifies the total annualized wage income in **all jobs including primary, secondary and others**. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

t_wage_total_year

t_wage_total_year is a continuous variable that specifies the total annualized wage income in **all jobs including primary, secondary and others**. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. t_wage_total should be equal to t_wage_nc_total in case there are no bonuses, tips etc. offered as part of any of the jobs.

5.2.11 Total Labor Income

Total Labor income will be created based on either the 7 days or 12 months reference period variables or a combination of both. Harmonizers should make sure that all jobs are included and none of them are double counted.

njobs

njobs is a numeric variable that specifies the total number of jobs. Do not put missing value for people below working age, unemployed and people out of the labor force.

t_hours_annual

t_hours_annual is a continuous variable that specifies the annual numbers of hours worked in all the jobs including primary, secondary and others regardless of their period of reference.

linc_nc

linc_nc is a continuous variable that specifies the total annualized wage income in all the jobs including primary, secondary and others regardless of their period of reference. This excludes tips, bonuses, other compensation such as dwellings or clothes, and other payments.

Note: Use gross wages when available and net wages only when gross wages are not available. This is done to make it easy to compare earnings in formal and informal sectors.

laborincome

laborincome is a continuous variable that specifies the total annualized individual labor income in all jobs including primary, secondary and others regardless of their period of reference. This income includes tips, compensations such as bonuses, dwellings or clothes, and other payments. This variable should be used as the total annual labor income of an individual.

5.3 Missing Value Codes

Harmonizers need to clearly differentiate missing values of variables from variables that were present in the survey but could not be harmonized due to reasons such as time unavailability. This will help the future harmonizers to focus on the unharmonized variables. The missing value code for these two scenarios are:

- For variables unavailable in survey = .
- For variables available in the survey but not harmonized = .a

5.4 Main Challenges and Lessons Learned

Notes:

- For any variable not collected in a country, the variable should be created and left as missing (.) in the final harmonized file.
- Variables in the data files must follow the sequence in which they appear in the manual.
- Labor status during last 12 months only reflects if the person has worked during last year or not, as the vast majority of the surveys do not provide enough information to distinguish those individuals that are unemployed from those that are out of the labor force.
- Individuals working in cooperatives are considered as “paid employee” in the employment status variable.

Several checks should be conducted to ensure that the data is harmonized correctly.

lstatus should be an integer in the range [1,3].

```
lstatus<0 & lstatus>3 & mod(lstatus, 1) == 1
```

If lstatus is classified as employed then the employment type needs to be defined.

```
lstatus==1 & empstat>5
```

minlaborage should be an integer.

```
mod(minlaborage,1)==1
```

The minimum age for employment should not be higher than 20.

```
minlaborage >20 & minlaborage <.
```

empstat should be an integer in the range [1,5].

```
mod(empstat, 1) != 0 & empstat<1 & empstat>5
```

If employment type is defined then labor force status should be employed.

```
empstat<=5 & lstatus!=1
```

industrycat10 should be an integer in the range [1,10].

```
industrycat10<0 & industrycat10>10 & mod(industrycat10, 1) == 1
```

industrycat4 should be an integer in the range [1,4].

```
industrycat4<0 & industrycat4>4 & mod(industrycat4, 1) == 1
```

There should not be a mismatch between industry and industrycat4.

```
((industrycat4==1 & industrycat10!=1 ) | (industrycat4==2 &
(industrycat10 <2 | industrycat10 >5)) |  (industrycat4==3 &
(industrycat10 <6 | industrycat10 >9)) | (industrycat4==1 &
industrycat10 !=1 ) ) &  industrycat10 !=.
```

firmsize_u should not be lower than firmsize_l

```
firmsize_u<firmsize_l
```

Conversion of given wage to annual wage. Example: Creation of wage_total when there are no bonuses nor other compensations

```
gen double wage_total=.
replace wage_total=(wage_nc*5*4.3)*wmonths        if unitwage==1 //Wage in daily unit
replace wage_total=(wage_nc*4.3)*wmonths          if unitwage==2 //Wage in weekly unit
replace wage_total=(wage_nc*2.15)*wmonths         if unitwage==3 //Wage in every two weeks unit
replace wage_total=(wage_nc)/2*wmonths if unitwage==4 //Wage in every two months unit
replace wage_total=( wage_nc)*wmonths             if unitwage==5 //Wage in monthly unit
replace wage_total=( wage_nc)/3*wmonthsif unitwage==6 //Wage in every quarterly unit
replace wage_total=( wage_nc)/6*wmonths           if unitwage==7 //Wage in every six months unit
replace wage_total= wage_nc/12*wmonths            if unitwage==8 //Wage in annual unit
replace wage_total=(wage_nc*whours*4.3)*wmonths          if unitwage==9 //Wage in hourly unit
```

Table 4 GMD Labor Module – key variables and descriptions

Module Code	Variable label	Variable name	Allowed codes after standardization
ID	country code	countrycode	string
ID	Year	year	numeric
ID	Household identifier	hhid	string or numeric
ID	Personal identifier	pid	string or numeric
Survey Characteristics & Welfare	Weight	weight	numeric

Labor status, 7-day reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Labor module application age	minlaborage	numeric
Labor	Labor status (7-day ref period)	lstatus	1 = Employed 2 = Unemployed 3 = Not-in-labor force
Labor	Reason not in the labor force (7-day ref period)	nlfreason	1 = Student 2 = Housewife 3 = Retired 4 = Disabled 5 = Other
Labor	Unemployment duration (months) lower bracket (7-day ref period)	unempldur_l	Continuous variable

Labor	Unemployment duration (months) upper bracket (7-day ref period)	unempldur_u	Continuous variable

Primary Employment, 7-day reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Employment status, primary job (7-day ref period)	empstat	1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other, workers not classifiable by status
Labor	Sector of activity, primary job (7-day ref period)	ocusec	1 = Public sector, Central Government, Army 2 = Private, NGO 3 = State owned 4 = Public or State-owned, but cannot distinguish
Labor	Original industry code, primary job (7-day ref period)	industry_orig	Country specific
Labor	1 digit industry classification, primary job (7-day ref period)	industrycat10	1 = Agriculture, Hunting, Fishing, etc. 2 = Mining 3 = Manufacturing 4 = Public Utility Services 5 = Construction 6 = Commerce 7 = Transport and Communications 8 = Financial and Business Services 9 = Public Administration 10 = Other Services, Unspecified
Labor	4-category industry classification, primary job (7-day ref period)	industrycat4	1=Agriculture 2=Industry 3=Services 4=Other
Labor	Original occupational classification, primary job (7-day ref period)	occup_orig	Country specific
Labor	1 digit occupational classification, primary job (7-day ref period)	occup	1 = Managers 2 = Professionals 3 = Technicians and associate professionals 4 = Clerical support workers 5 = Service and sales workers 6 = Skilled agricultural, forestry and fishery workers 7 = Craft and related trades workers 8 = Plant and machine operators, and assemblers 9 = Elementary occupations 10 = Armed forces occupations 99 = Other/unspecified

Labor	Last wage payment, primary job, excl. bonuses, etc. (7-day ref period)	wage_nc	Continuous variable
Labor	Time unit of last wages payment, primary job (7-day ref period)	unitwage	1 = Daily 2 = Weekly 3 = Every two weeks 4 = Every two months 5 = Monthly 6 = Quarterly 7 = Every six months 8 = Annually 9 = Hourly 10 = Other
Labor	Hours of work in last week, primary job (7-day ref period)	whours	Continuous variable
Labor	Months worked in the last 12 months, primary job (7-day ref period)	wmonths	Continuous variable
Labor	Annualized total wage, primary job (7-day ref period)	wage_total	
Labor	Contract (7-day ref period)	contract	0 = No 1 = Yes
Labor	Health insurance (7-day ref period)	healthins	0 = No 1 = Yes
Labor	Social security (7-day ref period)	socialsec	0 = No 1 = Yes
Labor	Union membership (7-day ref period)	union	0 = No 1 = Yes
Labor	Firm size (lower bracket), primary job (7-day ref period)	firmsize_l	Continuous variable
Labor	Firm size (upper bracket), primary job (7-day ref period)	firmsize_u	Continuous variable

Secondary Employment, 7-day reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Employment status, secondary job (7-day ref period)	empstat_2	1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other, workers not classifiable by status
Labor	Sector of activity, secondary job (7-day ref period)	ocusec_2	1 = Public sector, Central Government, Army (including armed forces) 2 = Private, NGO 3 = State-owned 4 = Public or State-owned, but cannot distinguish
Labor	Original industry code, secondary job (7-day ref period)	industry_orig_2	Country specific

Labor	1 digit industry classification, secondary job (7-day ref period)	industrycat10_2	1 = Agriculture, Hunting, Fishing, etc. 2 = Mining 3 = Manufacturing 4 = Public Utility Services 5 = Construction 6 = Commerce 7 = Transport and Communications 8 = Financial and Business Services 9 = Public Administration 10 = Other Services, Unspecified
Labor	4-category industry classification, secondary job (7-day ref period)	industrycat4_2	1=Agriculture 2=Industry 3=Services 4=Other
Labor	Original occupational classification, secondary job (7-day ref period)	occup_orig_2	Country specific
Labor	1 digit occupational classification, secondary job (7-day ref period)	occup_2	1 = Managers 2 = Professionals 3 = Technicians and associate professionals 4 = Clerical support workers 5 = Service and sales workers 6 = Skilled agricultural, forestry and fishery workers 7 = Craft and related trades workers 8 = Plant and machine operators, and assemblers 9 = Elementary occupations 10 = Armed forces occupations 99 = Other/unspecified
Labor	wage payment, secondary job, excl. bonuses, etc. (7-day ref period)	wage_nc_2	Continuous variable
Labor	Time unit of last wages payment, secondary job (7-day ref period)	unitwage_2	1 = Daily 2 = Weekly 3 = Every two weeks 4 = Every two months 5 = Monthly 6 = Quarterly 7 = Every six months 8 = Annually 9 = Hourly 10 = Other
Labor	Hours of work in last week, secondary job (7-day ref period)	whours_2	Continuous variable
Labor	Months worked in the last 12 months, secondary job (7-day ref period)	wmonths_2	Continuous variable
Labor	Annualized total wage, secondary job (7-day ref period)	wage_total_2	Continuous variable
Labor	Firm size (lower bracket), secondary job (7-day ref period)	firmsize_l_2	Continuous variable

Labor	Firm size (upper bracket), secondary job (7-day ref period)	firmsize_u_2	Continuous variable

Other Employment, 7-day reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Total hours of work in the last 12 months in other jobs excluding the primary and secondary ones	t_hours_others	Continuous variable
Labor	Annualized wage in all jobs excluding the primary and secondary ones (excluding tips, bonuses, etc.).	t_wage_nc_others	Continuous variable
Labor	Annualized wage (including tips, bonuses, etc.) in all other jobs excluding the primary and secondary ones.	t_wage_others_year	Continuous variable

Total Employment Earnings, 7-day reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Annualized hours worked in all jobs (7-day ref period)	t_hours _total	Continuous variable
Labor	Annualized wage in all jobs excl. bonuses, etc. (7-day ref period)	t_wage_nc_total	Continuous variable
Labor	Annualized total wage for all jobs (7-day ref period)	t_wage_total	Continuous variable

Labor status, 12-month reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Labor module application age (12-mon ref period)	minlaborage_year	numeric
Labor	Labor status (12-mon ref period)	Lstatus_year	1 = Employed 2 = Unemployed 3 = Not-in-labor force
Labor	Reason not in the labor force (12-mon ref period)	nlfreason_year	1 = Student 2 = Housewife 3 = Retired 4 = Disabled 5 = Other
Labor	Unemployment duration (months) lower bracket (12-mon ref period)	unempldur_l_year	Continuous variable
Labor	Unemployment duration (months) upper bracket (7-day ref period)	unempldur_u_year	Continuous variable

Primary Employment, 12-month reference period

Module Code	Variable label	Variable name	Allowed codes after standardization

Labor	Employment status, primary job (12-mon ref period)	empstat_year	1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other, workers not classifiable by status
Labor	Sector of activity, primary job (12-mon ref period)	ocusec_year	1 = Public sector, Central Government, Army 2 = Private, NGO 3 = State owned 4 = Public or State-owned, but cannot distinguish
Labor	Original industry code, primary job (12-mon ref period)	industry_orig_year	Country specific
Labor	1 digit industry classification, primary job (12-mon ref period)	industrycat10_year	1 = Agriculture, Hunting, Fishing, etc. 2 = Mining 3 = Manufacturing 4 = Public Utility Services 5 = Construction 6 = Commerce 7 = Transport and Communications 8 = Financial and Business Services 9 = Public Administration 10 = Other Services, Unspecified
Labor	4-category industry classification primary job (12-mon ref period)	industrycat4_year	1=Agriculture 2=Industry 3=Services 4=Other
Labor	Original occupational classification, primary job (12-mon ref period)	occup_orig_year	Country specific
Labor	1 digit occupational classification, primary job (12-mon ref period)	occup_year	1 = Managers 2 = Professionals 3 = Technicians and associate professionals 4 = Clerical support workers 5 = Service and sales workers 6 = Skilled agricultural, forestry and fishery workers 7 = Craft and related trades workers 8 = Plant and machine operators, and assemblers 9 = Elementary occupations 10 = Armed forces occupations 99 = Other/unspecified
Labor	Last wage payment, primary job, excl. bonuses, etc. (12-mon ref period)	wage_nc_year	Continuous variable: wage in local currency.

Labor	Time unit of last wages payment, primary job (12-mon ref period)	unitwage_year	1 = Daily 2 = Weekly 3 = Every two weeks 4 = Every two months 5 = Monthly 6 = Quarterly 7 = Every six months 8 = Annually 9 = Hourly 10 = Other
Labor	Hours of work in last week, primary job (12-mon ref period)	whours_year	Continuous variable: hours worked in the last 12 months
Labor	Months worked in the last 12 months, primary job (12-mon ref period)	wmonths_year	Continuous variable: months worked in the last 12 months
Labor	Annualized total wage, primary job (12-mon ref period)	wage_total_year	
Labor	Contract (12-mon ref period)	contract_year	0 = No 1 = Yes
Labor	Health insurance (12-mon ref period)	healthins_year	0 = No 1 = Yes
Labor	Social security (12-mon ref period)	socialsec_year	0 = No 1 = Yes
Labor	Union membership (12-mon ref period)	union_year	0 = No 1 = Yes
Labor	Firm size (lower bracket), primary job (12-mon ref period)	firmsize_l_year	Continuous variable
Labor	Firm size (upper bracket), primary job (12-mon ref period)	firmsize_u_year	Continuous variable

Secondary Employment, 12-month reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Employment status, secondary job (12-mon ref period)	empstat_2_year	1 = Paid Employee 2 = Non-Paid Employee 3 = Employer 4 = Self-employed 5 = Other, workers not classifiable by status
Labor	Sector of activity, secondary job (12-mon ref period)	ocusec_2_year	1 = Public sector, Central Government, Army (including armed forces) 2 = Private, NGO 3 = State-owned 4 = Public or State-owned, but cannot distinguish
Labor	Original industry code, secondary job (12-mon ref period)	industry_ orig_2_year	Country specific

Labor	1 digit industry classification, secondary job (12-mon ref period)	industrycat10_2_year	1 = Agriculture, Hunting, Fishing, etc. 2 = Mining 3 = Manufacturing 4 = Public Utility Services 5 = Construction 6 = Commerce 7 = Transport and Communications 8 = Financial and Business Services 9 = Public Administration 10 = Others Services, Unspecified
Labor	4-category industry classification, secondary job (12-mon ref period)	industrycat4_2_year	1=Agriculture 2=Industry 3=Services 4=Other
Labor	Original occupational classification, secondary job (12-mon ref period)	occup_orig_2_year	Country specific
Labor	1 digit occupational classification, secondary job (12-mon ref period)	occup_2_year	1 = Managers 2 = Professionals 3 = Technicians and associate professionals 4 = Clerical support workers 5 = Service and sales workers 6 = Skilled agricultural, forestry and fishery workers 7 = Craft and related trades workers 8 = Plant and machine operators, and assemblers 9 = Elementary occupations 10 = Armed forces occupations 99 = Other/unspecified
Labor	last wage payment, secondary job, excl. bonuses, etc. (12-mon ref period)	wage_nc_2_year	Continuous variable: wage in local currency
Labor	Time unit of last wages payment, secondary job (12-mon ref period)	unitwage_2_year	1 = Daily 2 = Weekly 3 = Every two weeks 4 = Every two months 5 = Monthly 6 = Quarterly 7 = Every six months 8 = Annually 9 = Hourly 10 = Other
Labor	Hours of work in last week, secondary job (12-mon ref period)	whours_2_year	Continuous variable: hours worked in the last 12 months
Labor	Months worked in the last 12 months, secondary job (12-mon ref period)	wmonths_2_year	Continuous variable: months worked in the last 12 months

Labor	Annualized total wage, secondary job (12-mon ref period)	wage_total_2_year	
Labor	Firm size (lower bracket), secondary job (12-mon ref period)	firmsize_l_2_year	Continuous variable
Labor	Firm size (upper bracket), secondary job (12-mon ref period)	firmsize_u_2_year	Continuous variable

Other Employment, 12-month reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Annualized hours worked in all but primary and secondary jobs (12-mon ref period)	t_hours_others_year	Continuous variable
Labor	Annualized wage in all but primary & secondary jobs excl. bonuses, etc. (12-mon ref period)	t_wage_nc_others_year	Continuous variable
Labor	Annualized wage in all but primary and secondary jobs (12-mon ref period)	t_wage_others_year_year	Continuous variable

Total Employment Earnings, 12-month reference period

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Annualized hours worked in all jobs (12-mon ref period)	t_hours_total_year	Continuous variable
Labor	Annualized wage in all jobs excl. bonuses, etc. (12-mon ref period)	t_wage_nc_total_year	Continuous variable
Labor	Annualized total wage for all jobs (12-mon ref period)	t_wage_total_year	Continuous variable

Total Labor Income

Module Code	Variable label	Variable name	Allowed codes after standardization
Labor	Total number of jobs	njobs	Continuous variable
Labor	Total hours worked in all jobs in the previous 12 months	t_hours_annual	Continuous variable
Labor	Total annual wage income in all jobs, excl. bonuses, etc.	linc_nc	Continuous variable
Labor	Total annual individual labor income in all jobs, incl. bonuses, etc.	laborincome	Continuous variable

6 Utilities (UTL)

6.1 Framework of Harmonization

Multiple-topic household surveys collect data on the characteristics of both households and individuals within those households. This module covers affordability and access related indicators that are commonly derived from survey data sets, or that can be constructed using existing variables. The primary objective of this indicator harmonization is to generate a unified data source of globally comparable indicators for utilities affordability and access to support routine analytical and corporate business functions of the World Bank.

In this module, the primary unit of analysis is the level of a household. The units of classification are consumption expenditures made by households for satisfying their needs or wants for various goods and services.

The typology of GMD utilities affordability variables follows the Classification of Individual Consumption According to Purpose (COICOP[7]) developed by the United Nations Statistics Division (UNSD). The objective of COICOP is to provide a framework for comparable classification of homogeneous categories of goods and services, which are considered a function or purpose of household consumption expenditure.

COICOP has 14 consumption categories, of which utility related expenditures are classified under the "4 Housing, water, electricity, gas and other fuels" category. While more detailed information on water and urban services consumption is provided in the "4.4 Water Supply and Miscellaneous Services related to the dwelling" category, expenditure on energy is sub-categorized into "4.5 Electricity, Gas, and other Fuels".

The "Water Supply and Miscellaneous Services (4.4)" group is further sub-divided into the "Water supply (4.4.1)", "Refuse collection (4.4.2)", "Sewage collection (4.4.3)" and "Other services relating to the dwelling (4.4.4)" categories. Similarly, the "Electricity, gas and other fuels (4.5)" group are sub-categorized into "Electricity (4.5.1)", 'Gas (4.5.2)", "Liquid fuels (4.5.3)", "Solid fuels (4.5.4)" and "Heat energy (4.5.5)". According to COICOP, energy-related household consumption is not sub-categories beyond the 3-digit level, but GMD divides them into commonly used categories except electricity expenditures.

Non-energy or non-WASH variables that are broadly classified as utilities are included as additional variables: Maintenance and Repair of the Dwelling (4.3); Part of Fuels and lubricants for personal transport equipment (transport fuels only, but not lubricants; 7.2.2); Telephone and telefax services (8.3.0); Part of Cultural Services (TV broadcasting services only; 9.4.2).

The harmonization framework for GMD utilities access to services variables is based on three independent frameworks: Human Opportunity Index (HOI)[8], developed by the World Bank's Latin America and Caribbean (LAC) department in 2008; WASH (Water, sanitation, and hygiene) Access Plus Framework developed by the Water Global Project in 2015 (draft); and the Multi-tier access Tracking Framework (MTF)[9] introduced by the

[7] The classification of individual consumption by purpose, abbreviated as COICOP, is a classification developed by the United Nations Statistics Division to classify and analyze individual consumption expenditures incurred by households, non-profit institutions serving households and general government according to their purpose. It includes categories such as clothing and footwear, housing, water, electricity, and gas and other fuels.

[8] For further reference, see http://web.worldbank.org/WBSITE/EXTERNAL/COUNTRIES/LACEXT/EXTLACREGTOPPOVANA/0,,contentMDK:21881102~pagePK:34004173~piPK:34003707~theSitePK:841175,00.html

[9] For further reference, see http://www.worldbank.org/en/topic/energy/publication/energy-access-redefined

World Bank and SE4ALL Knowledge Hub in 2015. According to the HOI typology, all WASH and energy access indicators are defined as binary indices, while the other two frameworks define WASH and energy indicators as multi-tier indices. Due to the limitation of ex-post harmonization, GMD variables may not match exactly multi-tier frameworks, but the tiers are embedded into the categories whenever possible.

Access to and affordability of WASH and energy are important nonmonetary dimensions of welfare. The SDGs call for universal and equitable access to safe and affordable drinking water and access to adequate and equitable sanitation and hygiene for all. The indicators require an understanding of the proportion of the population which has access to safely managed WASH services or facilities. In the same vein, understanding the proportion of the population with access to electricity, and more specifically, the types of fuels used is necessary to understand the progress made on SDG 7- Affordable and Clean Energy.

6.2 Mapping and description of variables

GMD Utilities module contains a large amount of metadata that provides a wealth of information about variables, including their types, descriptions, sources, etc. To improve readability, only the most significant information has been included in this section. For a complete list of all variables captured in the module, please consult table at the end of the module.

GMD Utilities module consists of affordability variables and access to services variables, which follow COICOP and MTF and Access Plus frameworks, respectively.

GMD utilities affordability variables are monetary variables expressed at current prices in the local currency unit (LCU) and non-deflated either temporal nor spatial. It should include not only monetary expenses, but also value of in-kind acquisitions.

6.2.1 Access to Services

GMD utilities access to services indicators include access to water, sanitation, and hygiene (WASH) and access to energy.

6.2.1.1 Access to Water, Sanitation and Hygiene (WASH)

water_source
water_source is a categorical variable that indicates the main source of drinking water for the household. If the main source of water differs between the wet and dry season, water source during dry season is referred. The best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect. Harmonizers should refer to the survey questionnaire to assess the best matches.

Main source of drinking water, fourteen categories after harmonization:

1 = Piped water into dwelling
2 = Piped water to yard/plot
3 = Public tap or standpipe
4 = Tubewell or borehole
5 = Protected dug well
6 = Protected spring
7 = Bottled water

8 = Rainwater
9 = Unprotected spring
10 = Unprotected dug well
11 = Cart with small tank/drum
12 = Tanker-truck
13 = Surface water
14 = Other

imp_wat_rec

imp_wat_rec is a categorical variable that estimates the "recommended" categorization for access to improved water sources in each country, or how evidence suggests that the expected error might be minimized. If the relevant survey was on file in the SDG calculations, this would be considered 1 if the majority of the problematic category was estimated therein to be of an improved type at the rural level, and otherwise considered 0. If the survey was not already in the SDG calculations, recommendations are based on the standard international classifications plus any relevant insights from other surveys on file for the specific country. In the few instances where there was no evidence, 0 is used. For an example of this, see the Main challenges/lessons learned section.

This variable is created with the help of the WASH team (Luis Alberto Andres landres@worldbank.org), the poverty GP is in charge of proposing the classification of the categories, the WASH team will be in charge of reviewing the proposed categorization, then either approve it or recommend any changes.

1. The Regional TSD will fill out the excel template including all categories for water access and the actual question in the questionnaire as it is shown below, each country year must be submitted in a separate worksheet.

Improved water

Harmonization

Category	Share of population (%)			Improved	
	National	Rural	Urban	Recommended estimate	WaterGP review
Centralize water supply	97.39	93.20	99.76	Yes	Y
Spring Water, wells	0.75	1.78	0.16	Yes	
Own system of water supply	1.40	3.81	0.04	Yes	
River, lake	0.00	0.00	0.00	No	
Delivered (imported) water	0.10	0.25	0.02	Yes	
Bought water	0.01	0.04	0.00	Yes	
Rainwater	0.00	0.00	0.00	Yes	
Other	0.35	0.93	0.02	No	
Missing	0.00	0.00	0.0	Missing	Missing

Share improved	Share of population (%)			JMP National	JMP Rural	JMP Urban
	National	Rural	Urban			
Recommended estimate	99.65	99.07	99.98			

Snippet from questionnaire

C.13. What are the sources of water you use?

1. centralized water supply
2. spring water, wells ⇨ C. 16.
3. own system of water supply
4. river, lake ⇨C.16.
5. delivered (imported) water ⇨ C. 16.
6. bought water (Noy, Byuregh, etc) ⇨ C. 16.
7. rainwater ⇨ C. 16.
8. other _____(specify) ⇨ C. 16

C.14. Where is your water tap located?
(respondents are points 1 and 3 of item C. 13)

1. in door
2. in the yard
3. in the street

2. The WASH team reviews and proposes a classification in the section "Water GP review"

Improved water

Harmonization

Category	Share of population (%)			Improved	
	National	Rural	Urban	Recommended estimate	WaterGP review
Centralize water supply	97.39	93.20	99.76	Yes	Yes
Spring Water, wells	0.75	1.78	0.16	Yes	Yes
Own system of water supply	1.40	3.81	0.04	Yes	Yes
River, lake	0.00	0.00	0.00	No	No
Delivered (imported) water	0.10	0.25	0.02	Yes	Yes
Bought water	0.01	0.04	0.00	Yes	Yes
Rainwater	0.00	0.00	0.00	Yes	Yes
Other	0.35	0.93	0.02	No	No
Missing	0.00	0.00	0.0	Missing	Missing

Share improved	Share of population (%)			JMP National	JMP Rural	JMP Urban
	National	Rural	Urban			
Recommended estimate	99.65	99.07	99.98	100%	100%	100%

Snippet from questionnaire

C.13. What are the sources of water you use?

1. centralized water supply
2. spring water, wells ⇨ C. 16.
3. own system of water supply
4. river, lake ⇨C.16.
5. delivered (imported) water ⇨ C. 16.
6. bought water (Noy, Byuregh, etc) ⇨ C. 16.
7. rainwater ⇨ C. 16.
8. other _____(specify) ⇨ C. 16

C.14. Where is your water tap located?

(respondents are points 1 and 3 of item C. 13)

1. in door
2. in the yard
3. in the street

3. Regional TSD will create the harmonized variables based on the recommendation of WASH team (based on "Water GP review").

The recommended access, two categories after harmonization:
0 = No
1 = Yes

water_original

water_original is a string variable that specifies the original survey response for the water_source variable. It is a country-specific variable. It must follow the naming convention: "1 – Piped Water " (as string).

watertype_quest

watertype_quest is a categorical variable that specifies the type of water questions used in the survey. This variable is to record the type of question(s) asked about access to water in the survey, for example,the survey had a specific question on the water source on drinking water, or on water source on general water or both.

Type of water question, four categories after harmonization:

1=drinking water
2=general water
3=both
4=others

piped

piped is a categorical variable that indicates whether the household has access to piped water. There are two major types of water supply – within premises and outside premises. 'Within premises' refers to water service piped connection to own tap. It includes both household connection (in-house plumbing) and yard connection (yard or plot outside the house plumbing). Conversely, outside premise refers to a public water point from which people can collect water, shared among houses. It includes public tap and standpipe or a public fountain.

Access to piped water, two categories after harmonization:

0 = No
1 = water_source 1,2 or 3 (Piped water into dwelling, piped water to yard/plot, or public tap or standpipe)

piped_to_prem
piped_to_prem is a categorical variable that specifies whether a household has access to piped water on premises. There are two major types of water supply – within premises and outside premises. 'Within premises' refers to water service piped connection to own tap. It includes both household connection (in-house plumbing) and yard connection (yard or plot outside the house plumbing). Conversely, outside premise refers to a public water point from which people can collect water, shared among houses. It includes public tap and standpipe or a public fountain.

Access to piped water on premises, two categories after harmonization:

0 = No
1 = water_source 1 or 2 (Piped water into dwelling or piped water to yard/plot)

w_30m
w_30m is a categorical variable that specifies whether a household has access to improved water within 30 minutes. This includes time taken for a *round trip and waiting time* in case of queues. This variable needs to be created in conjunction with the imp_wat_rec dummy to identify where the improved water source is available within 30 minutes. Categories after harmonization:

1=collection time of imp_wat_rec less than or equal to 30 mins;
0=collection time of imp_wat_rec more than 30 mins

w_avail
w_avail is a categorical variable that specifies whether improved water is available when needed. This variable needs to be created in conjunction with the imp_wat_rec dummy to identify where the improved water source is available reliably 24/7. Categories after harmonization:

1 = improved water is available continuously, reliable source
0 = improved water source is unreliable

sanitation_source
sanitation_source is a categorical variable that specifies the source of sanitation facilities. The best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect. Harmonizers should refer to the survey questionnaire to assess the best matches.
Main sanitation source, fourteen categories after harmonization:

1 = A flush toilet
2 = A piped sewer system
3 = A septic tank
4 = Pit latrine
5 = Ventilated improved pit latrine (VIP)
6 = Pit latrine with slab
7 = Composting toilet
8 = Special case

9 = A flush/pour flush to elsewhere
10 = A pit latrine without slab
11 = Bucket
12 = Hanging toilet or hanging latrine
13 = No facilities or bush or field
14 = Other
Category 8 applies to improved sanitation facilities for which the respondent does not know whether the facility is connected to a sewer or septic tank.

sanitation_original
sanitation_original is a string variable that specifies the original survey response for the sanitation_source variable. It is a country-specific variable. It must follow the naming convention: "1 – Flush toilet " (as string).

toilet_acc
toilet_acc is a categorical variable that indicates type of access to a flush toilet.
Access to flush toilet, four categories after harmonization:
0 = No;
1 = Yes, in premise;
2 = Yes, but not in premise including public toilet,
3 = Yes, unstated whether in or outside the premise.

sewer
sewer is a categorical variable that specifies whether a household has access to a toilet connected to a piped sewer system.
Access to sewer, two categories after harmonization
0=No
1=flush/poor flush to piped sewer system

open_def
open_def is a categorical variable that specifies whether a household has access to any sanitation facility.
Access to any sanitation facility, two categories after harmonization:
0=availability of any facility (from list of categories in sanitation_source including unimproved options)
1=no facility, or bush, or field (13)

imp_san_rec
imp_san_rec is a categorical variable that estimates the categorization for access to improved sanitation facilities in each country, or how evidence suggests that the expected error might be minimized. If the relevant survey was on file in the SDG calculations, this would be considered 1 if the majority of the problematic category was estimated therein to be of an improved type at the rural level, and otherwise considered 0. If the survey was not already in the SDG calculations, recommendations are based on the standard international classifications plus any relevant insights from other surveys on file for the specific country. In the few instances where there was no evidence, 0 is used.

This variable is created with the help of the WASH team (Luis Alberto Andres landres@worldbank.org), the poverty GP is in charge of proposing the classification of the categories, the WASH team will be in charge of reviewing the proposal, then approved it or recommend another classification following the next work flow.

1. The Regional TSD will fill out the excel template including all categories for access to sanitation and the actual question in the questionnaire as it is shown below, each country year must be submitted in a separate worksheet.

Improved sanitation

Harmonization

Category	Share of population (%)			Improved	
	National	Rural	Urban	Recommended	WaterGP review
Flush Toilet	2.79	0.53	7.49	Y	
VIP Latrine	6.85	3.50	13.81	Y	
Covered Pit Latrine with a slab	27.73	20.01	43.79	Y	
Covered Pit Latrine without a slab	39.29	47.15	22.95	N	
Uncovered Pit Latrine with a slab	3.02	3.01	3.03	Y	
Uncovered Pit Latrine without a slab	11.22	13.97	5.52	N	
Ecosan (compost toilet)	0.27	0.26	0.30	Y	
No facility/bush/ polythene bags/ bucket/ etc	8.49	11.21	2.82	N	
Other (specify)	0.34	0.36	0.30	N	
Missing	0.00	0.00	0.00	Missing	

Share improved	Share of population (%)			JMP National	JMP Rural	JMP Urban
	National	Rural	Urban			
Recommended estimate	40.7	27.3	68.4			
- unshared	18.9	16.3	24.4			

Snippet from questionnaire

What type of toilet facility does this household mainly use?

01= Flush Toilet
02= VIP Latrine
03= Covered Pit Latrine with a slab
04= Covered Pit Latrine without a slab
05= Uncovered Pit Latrine with a slab
06= Uncovered Pit Latrine without a slab
07= Ecosan (compost toilet)
08= No facility/bush/ polythene bags/ bucket/ etc.
96= Other (specify)

Does the household share this toilet facility with other households?

1= Yes
2= No **(>> HC17)**
97= N/A (For "No facility/bush/ polythene bags/bucket/ etc). **(>> HC18)**

2. The WASH team reviews and proposes a classification in the section "Water GP review"

Improved sanitation

Harmonization

Category	Share of population (%)			Improved	
	National	Rural	Urban	Recommended	WaterGP review
Flush Toilet	2.79	0.53	7.49	Y	Y
VIP Latrine	6.85	3.50	13.81	Y	Y
Covered Pit Latrine with a slab	27.73	20.01	43.79	Y	Y
Covered Pit Latrine without a slab	39.29	47.15	22.95	N	N
Uncovered Pit Latrine with a slab	3.02	3.01	3.03	Y	Y
Uncovered Pit Latrine without a slab	11.22	13.97	5.52	N	N
Ecosan (compost toilet)	0.27	0.26	0.30	Y	Y
No facility/bush/ polythene bags/ bucket/ etc	8.49	11.21	2.82	N	N
Other (specify)	0.34	0.36	0.30	N	N
Missing	0.00	0.00	0.00	Missing	Missing

Share improved	Share of population (%)			JMP National	JMP Rural	JMP Urban
	National	Rural	Urban			
Recommended estimate	40.7	27.3	68.4			
- unshared	18.9	16.3	24.4	19%	16%	26%

Snippet from questionnaire

What type of toilet facility does this household mainly use?

01= Flush Toilet
07= Ecosan (compost toilet)
08= No facility/bush/ polythene bags/ bucket/ etc.
96= Other (specify)

Does the household share this toilet facility with other households?

1= Yes
2= No **(>> HC17)**
97= N/A (For "No facility/bush/ polythene bags/bucket/ etc). **(>> HC18)**

3. Regional TSD will create the harmonized variables based on the recommendation of WASH team (based on "Water GP review").

The recommended access, two categories after harmonization:
0 = No
1 = Yes

Note: Any ***sanitation facility that is shared among households is not considered as recommended*** access to improved sanitation.

waste
waste is a categorical variable that indicates the type of solid waste disposal. This variable contains information on the usual manner of collection and disposal of solid waste or garbage generated by occupants of the housing unit. Type of solid waste disposal is categorized by the manner of disposal, such as collection, disposal, burial or compost and by the administrator of the waste disposal, such as authorized collectors, self-appointed collectors, and dump supervised by authorities.
Main types of sewage disposal system, ten categories after harmonization:

1 = Solid waste collected on a regular basis by authorized collectors;
2 = Solid waste collected on an irregular basis by authorized collectors;
3 = Solid waste collected by self-appointed collectors;
4 = Occupants dispose of solid waste in a local dump supervised by authorities;
5 = Occupants dispose of solid waste in a local dump not supervised by authorities;
6 = Occupants burn solid waste;
7 = Occupants bury solid waste;
8 = Occupants dispose solid waste into river, sea, creek, pond;
9 = Occupants compost solid waste;
10 = Other arrangement.

6.2.1.2 Access to Energy

central_acc
central_acc is a dummy variable that indicates the access to central heating in the dwelling. Categories after harmonization:
0=No
1-Yes

heatsource
heatsource is a categorical variable that indicates the main source of heating. Main source of heating refers to the type of system used to provide heating for most of the space. It may be central heating covering all or parts of living quarters, or it may not be central, in which case the heating will be provided separately within the living quarters by a stove, fireplace or some other heating body.

As for the energy used for heating purposes, it is closely related to the type of heating and refers to the predominant source of energy, such as solid fuels (coal, lignite, and products of coal and lignite, wood), oils, gaseous fuels (natural or liquefied gas), or electricity.

Main sources of heating, seven categories after harmonization:

1 = Firewood;
2 = Kerosene;
3 = Charcoal;
4 = Electricity;
5 = Gas;
6 = Central;
9 = Other;

10= No heating

gas
gas is a categorical variable that identifies type of gas usage. The categories after harmonization are:
0 = No
1 = Yes, piped gas (LNG)
2 = Yes, bottled gas (LPG)
3 = Yes, but don't know

cooksource
cooksource is a categorical variable that identifies the source of cooking. The categories after harmonization are:

1 = Firewood
2 = Kerosene
3 = Charcoal
4 = Electricity
5 = Gas
9 = Other
10=No cook source

lightsource
lightsource is a categorical variable that identifies the source of light. The categories after harmonization are:
1 = Electricity
2 = Kerosene
3 = Candles
4 = Gas
9 = Other
10=No light source

elec_acc
elec_acc is a categorical variable that identifies type of connection to electricity. For instance, access to electricity ('Yes') may be public/quasi-public referring to mains electricity (i.e. the term used to refer to the electricity supply from power stations to households) or private referring to electricity from generator or solar or private company. The quality of electricity is assessed by other Tier 3 variables, such as number of electricity hours per day (elechr_acc).

Access to electricity, categories after harmonization

1 = Yes, public/quasi-public
2 = Yes, private
3 = Yes, source unstated
4 = No

electricity
electricity is a dummy variable that specifies the access to electricity in the household. Categories after harmonization:
0 = No
1 = Yes

elechr_acc

elechr_acc is a numeric continuous variable that specifies the access to electricity in hours per day.

electyp
electyp is a categorical variable that ***specifies the source of energy when cooksource and lightsource variables are not available*** and there is only one question about the type of energy source in the household; when *cooksource* and *lightsource* are available this variable has to be created prioritizing electricity, then Gas, then Lamp. Categories after harmonization:

1 = Electricity
2 = Gas
3 = Lamp
4 = Others
10=No cook and light source

When cooksource and lightsource are available, electyp can be created using the following code:

```
gen electyp=.
replace electyp=1 if  cooksource==4 | lightsource==1
replace electyp=2 if (cooksource==5 | lightsource==4) & mi(electyp)
replace electyp=3 if (cooksource==2 | inlist(lightsource,2,3)) & mi(electyp)
replace electyp=4 if inlist(cooksource,1,3,9) | lightsource==9) & mi(electyp)
replace electyp=10 if  cooksource==10 & lightsource==10
```

6.2.2 Affordability

6.2.2.1 Essential variables – Water, Sanitation and Hygiene (WASH)

GMD utilities affordability variables are monetary variables expressed at current prices in the local currency unit (LCU) and non-deflated either temporal nor spatial

pwater_exp
pwater_exp is a continuous variable that refers to total annual household expenditures on water supply/piped water. It includes associated expenditure such as hire of meters, reading of meters, standing charges, etc. GMD water consumption variables include an aggregate water variable comprising water supply (pwater_exp) and hot water (hwater_exp) and defined as water_exp. As in the case of the COICOP classification, the variable excludes household expenditures on hot water. Drinking water sold in bottles or containers is also excluded from water supply.

hwater_exp
hwater_exp is a continuous variable that refers to total annual household expenditure on hot water supply.

water_exp

water_exp is a continuous variable that refers to total annual household expenditure on water supply and hot water supply. This variable specifies the sum of expenditure of water supply (pwater_exp) and hot water supply (hwater_exp).

garbage_exp
garbage_exp is a continuous variable that refers to total annual household expenditures on collection and disposal of garbage or refuse.

sewage_exp
sewage_exp is a continuous variable that refers to total annual household expenditures on collection and disposal of wastewater.

waste_exp
waste_exp is a continuous variable that refers to the total annual household expenditure on garbage (garbage_exp) and sewage (sewage_exp) collection.

dwelothsvc_exp
dwelothsvc_exp is a continuous variable that refers to total annual household expenditures on other services relating to the dwelling. These expenditures typically include co-proprietor charges in multi-occupied buildings, security services, and other miscellaneous services. Co-proprietor charges include charges for caretaking, gardening, stairwell cleaning, heating and lighting, maintenance of lifts and refuse disposal chutes, etc. This variable does not include household services such as window cleaning, disinfecting, fumigation and pest extermination[10]; bodyguards[11]. Maintenance and repair of the dwelling[12] is also excluded from other services relating to the dwelling (dwelothsvc_exp) but included as additional variables defined as dwelmat_exp and dwelsvc_exp.

6.2.2.2 Essential variables - Energy[13]

elec_exp
elec_exp is a continuous variable that refers to total annual household expenditures on electricity and other associated expenditures such as hire of meters, reading of meters and standing charges.

ngas_exp
ngas_exp is a continuous variable that refers to total annual household expenditure on town gas and natural gas.

LPG_exp
LPG_exp is a continuous variable that refers to total annual household expenditure on LPG that includes butane, propane, “bottled gas” etc.

gas_exp

[10] Also known as COICOP 5.6.2.
[11] Also known as COICOP 12.7.0.
[12] Also known as COICOP 4.3
[13] It also serves as indices for 5. Affordability for “Access to Cooking Solutions” and “Access to Space Heating” in MTF.

gas_exp is a continuous aggregate variable comprised of total annual household expenditures on network/natural gas and liquefied gas (LPG). Due to differences in characteristics and price patterns, two types of gas are recorded as separate variables under gas: 1) Town gas and natural gas (ngas_exp); and 2) LPG (liquefied petroleum gas (LPG_exp): includes butane, propane, “bottled gas”, etc.). Associated expenditure such as hire of meters, reading of meters, storage containers, standing charges, etc. are included in the construction of the variable.

diesel_exp
diesel_exp is a continuous variable that refers to total household expenditure on diesel or gasoil. Mostly use on electricity generators, SUV, Trucks, buses, very few sedan cars use this type of fuel.

kerosene_exp
kerosene_exp is a continuous variable that refers to total annual household expenditure on kerosene.

gasoline_exp
gasoline_exp is a continuous variable that refers to total annual household expenditure on gasolines. Use mostly in sedan cars and motorcycles.

othliq_exp
othliq_exp is a continuous variable that refers to total annual household expenditure on other liquid fuels such as heating oil, black oil and lighting oil.

liquid_exp
liquid_exp is a continuous aggregate variable comprised of total annual household expenditures on all liquid fuels. Liquid fuels are subcategorized into: gasoline/petrol (gasoline_exp), diesel (diesel_exp), kerosene (kerosene_exp), gasoline (gasoline_exp), and other liquid fuels (othliq_exp). Other liquid fuels category includes all other liquid fuels other than diesel and kerosene. Examples include “heating oil”, “black oil” and “lighting oil”.

wood_exp
wood_exp is a continuous variable that refers to total annual household expenditure on firewood.

coal_exp
coal_exp is a continuous variable that refers to total annual household expenditure on coal.

peat_exp
peat_exp is a continuous variable that refers to total annual household expenditure on peat.

othsol_exp
othsol_exp is a continuous variable that refers to total annual household expenditure on other solid fuels such as agricultural residue and charcoal.

solid_exp
solid_exp is a continuous aggregate variable comprised of total annual household expenditures on all solid fuels. Solid energy is subcategorized into expenditures on coal (coal_exp), firewood (wood_exp) and peat (peat_exp), and other solid fuels (othsol_exp). Other solid fuels category includes all other solid fuels not included in the above three categories. Examples include “pressed dung, corn brans, brushwood”, and “other solid”.

othfuel_exp
othfuel_exp is a continuous variable that refers to total annual household expenditure on other fuels that are not captured under othliq_exp and othsol_exp.

central_exp
central_exp is a continuous variable that refers to total annual household expenditure on central heating.

heating_exp
heating_exp is a continuous aggregate variable comprised of total annual household expenditures on heating. These expenditures can be subcategorized into expenditures on central heating (central_exp) and hot water (hwater_exp). It is worth to note that COICOP narrowly defines heat energy to purchase from district heating plant only, but GMD includes heat energy from building or other sources. Note that expenditure for central heating is frequently combined either with expenditures pm hot water or rent. Hot water is also often combined with cold water. Also note that COICOP categorizes hot water under 4.5.5 Heat energy, while cold water is reflected under 4.4.1 Water supply.

utl_exp
utl_exp is a continuous aggregate variable comprised of total annual household expenditure on all utilities excluding telecom and other housing expenses. Utilities expenditure in this case is sum of the following variables: electricity (elec_exp), gas (gas_exp), liquid fuels (liquid_exp), solid fuels (solid_exp), central heating (central_exp), water (water_exp), waste (waste_exp) and other fuels (othfuel_exp). Excludes expenditures for other housing (othhousing_exp), fuel for transportation (transfuel_exp), telecommunication services (comm_exp) and tv services (tv_exp).

6.2.2.3 Additional variables

dwelmat_exp
dwelmat_exp is a continuous variable that refers to total annual household expenditures on product and materials for maintenance and repair of the dwelling. Products and materials for minor maintenance and repair typically include expenditures on paints and varnishes, renderings, wallpapers, fabric wall coverings, window panes, plaster, cement, putty, wallpaper pastes. Fitted carpets and linoleum (5.1.2); hand tools, door fittings, power sockets, wiring flex and lamp bulbs (5.5.2); brooms, scrubbing brushes, dusting brushes and cleaning products (5.6.1); products, materials and fixtures used for major maintenance and repair (intermediate consumption) or for extension and conversion of the dwelling (capital formation) are excluded.

dwelsvc_exp
dwelsvc_exp is a continuous variable that refers to total annual household expenditures on services for minor maintenance and repair of the dwelling. This variable generally includes expenditures on services of plumbers, electricians, carpenters, glaziers, painters, decorators, floor polishers, etc as well as total value of the service (that is, both the cost of labor and the cost of materials are covered). It excludes separate purchases of materials made by the household with the intention of undertaking the maintenance or repair by themselves (4.3.1); services engaged for major maintenance and repair (intermediate consumption) or for the extension and conversion of the dwelling (capital formation).

othhousing_exp

othhousing_exp is a continuous variable that refers to total annual household expenditures on other materials and services for minor maintenance and repair of the dwelling.

transfuel_exp
transfuel_exp is a continuous variable that refers to total annual household expenditures on fuels for personal transportation. According to COICOP, fuels use for transportation purposes are classified under Fuels and lubricants for personal transport equipment (COICOP 7.2.2). COICOP 7.2.2 also includes lubricants, which are excluded from this GMD indicator.

landphone_exp
landphone_exp is a continuous variable that refers to total annual household expenditures on landphone. This includes installation, subscription and service usage fees. Expenditure on equipment are not included.

cellphone_exp
cellphone_exp is a continuous variable that refers to total annual household expenditures on cellphone. This includes installation, subscription and service usage fees. Expenditure on equipment are not included.

tel_exp
tel_exp is a continuous aggregate variable comprised of total annual household expenditures on landline phone (landphone_exp) and cell phone (cellphone_exp) which may include (i) Installation and subscription costs of personal telephone equipment, (ii) telephone calls from a private line or from a public line (public telephone box, post office cabin, etc.); telephone calls from hotels, cafés, restaurants and the like, (iii) hire of telephones, telefax machines, telephone-answering machines and telephone loudspeakers. Expenditures on relevant equipment are not included.
Telephone and telefax services (COICOP 8.3.0) are subcategorized into 4 categories: landline phone, cell phone, internet and telefax services.

internet_exp
internet_exp is a continuous variable that refers to total annual household expenditures on information transmission and Internet connection services. This variable also includes installation, subscription, and service usage fees and costs, but excludes consumption for equipment. Telefax services (telefax_exp) includes telegraphy, telex and telefax services, as well as radio-telephony, radio-telegraphy and radiotelex services. Expenditures on relevant equipment are not included.

telefax_exp
telefax_exp is a continuous variable that refers to total annual household expenditures on telegraphy, telex and telefax services. This includes: radio-telephony, radio-telegraphy and radiotelex services.

comm_exp
comm_exp is a continuous variable comprised of total annual household expenditures on all telephone and telefax services, including expenditures on landline phone (landphone_exp), cell phone (cellphone_exp), internet (internet_exp) and telefax services (telefax_exp).

tv_exp
tv_exp is a continuous variable that refers to total annual household expenditures on television broadcasting services, license fees for television equipment and subscriptions to television networks. This variable is

compatible with COICOP 9.4.2 Cultural services but does not include spending on such services as theatres, museums and historic monuments.

tvintph_exp

tvintph_exp is a continuous aggregate variable comprised of total annual household expenditures on internet (internet_exp), telephone (tel_exp) and television broadcasting services (tv_exp).

Table 5 below provides the summary of all the utilities expense variables. The variables highlighted in yellow are secondary variables that are aggregated using primary variables. However, there might be surveys that report expenditures on secondary level only. For example: waste expenditure (waste_exp) is sum of garbage expenditure (garbage_exp) and sewage expenditure (sewage_exp). In surveys where expenditures are reported on disaggregated level will include values for garbage expenditure and sewage expenditure and then waste_exp is created by adding garbage and sewage expenditures. However, some surveys will report expenditure only for total waste i.e. waste_exp, leading to missing values for garbage_exp and sewage_exp.

Table 5 Utilities Expenditure Variables (rows in yellow are aggregated variables)

pwater_exp	Total annual consumption of water supply
hwater_exp	Total annual household consumption of hot water supply
water_exp	Total annual consumption of water supply and hot water
garbage_exp	Total annual consumption of garbage collection
sewage_exp	Total annual consumption of sewage collection
waste_exp	Total annual consumption of garbage and sewage collection
dwelothsvc_exp	Total annual consumption of other services relating to the dwelling
elec_exp	Total annual consumption of electricity
ngas_exp	Total annual consumption of network/natural gas
LPG_exp	Total annual consumption of liquefied gas
gas_exp	Total annual consumption of network/natural and liquefied gas
gasoline_exp	Total annual consumption of gasoline
diesel_exp	Total annual consumption of diesel
kerosene_exp	Total annual consumption of kerosene
othliq_exp	Total annual consumption of other liquid fuels
liquid_exp	Total annual consumption of all liquid fuels
wood_exp	Total annual consumption of firewood
coal_exp	Total annual consumption of coal
peat_exp	Total annual consumption of peat
othsol_exp	Total annual consumption of other solid fuels
solid_exp	Total annual consumption of all solid fuels
othfuel_exp	Total annual consumption of all other fuels
central_exp	Total annual consumption of central heating
hwater_exp	Total annual consumption of hot water
heating_exp	Total annual consumption of hot and cold water
garbage_exp	Total annual consumption of garbage collection
sewage_exp	Total annual consumption of sewage collection
waste_exp	Total annual consumption of garbage and sewage collection
other_exp	Total annual consumption of all other utilities

utl_exp	Total annual consumption of all utilities excluding telecom and other housing, current year prices
dwelmat_exp	Total annual consumption of materials for the maintenance and repair of the dwelling
dwelsvc_exp	Total annual consumption of services for the maintenance and repair of the dwelling
othhousing_exp	Total annual consumption of maintenance and repair of the dwelling
transfuel_exp	Total annual consumption of fuels for personal transportation
landphone_exp	Total annual consumption of landline phones
cellphone_exp	Total annual consumption of cell phones
tel_exp	Total annual consumption of all phones
internet_exp	Total annual consumption of internet
telefax_exp	Total annual consumption of other telefax services
comm_exp	Total annual consumption of telecommunication services
tv_exp	Total annual consumption of tv broadcasting services
tvintph_exp	Total annual consumption of tv, internet and telephone services

6.3 Missing Value Codes

Harmonizers need to clearly differentiate missing values of variables from variables that were present in the survey but could not be harmonized due to reasons such as time unavailability. This will help the future harmonizers to focus on the unharmonized variables. The missing value code for these two scenarios are:

- For variables unavailable in survey = .
- For variables available in the survey but not harmonized = .a

6.4 Main challenges/ common mistakes (Module specific)

6.4.1 Data harmonization

The main challenge in this module is inconsistency between what is available in the survey and variable and allowed codes of the harmonization. This needs to be considered carefully in the harmonization process and will differ between the affordability and access submodules.

Distinguishing missing, value 0 and skip patterns also needs careful attention. Some variables or questionnaires may contain 0 in the raw data when the value should be missing. y. This could lead to underestimation of the variable for the survey, distorting the snapshot of energy-related information. Skip patterns should also be distinguished from both missing and value 0.

6.4.2 Affordability

The harmonization process must carefully consider Inconsistencies between what is available in the survey and variable and the codes specified in the harmonization. To solve this problem, the affordability submodule includes several variables that are defined as a sum of other variables. For instance, some surveys provide separate expenditures for solid fuels such as wood or coal, while others provide total expenditure on solid fuels only. The GMD provides both levels of expenditures variables such as solid fuel, wood and coal, so that surveys with more detailed information retains their information in the harmonization and also can be easily compared with other surveys without the information.

How to annualize monthly or quarterly expenditures while taking seasonality into account is also a point of discussion. GMD annualizes variables by annualizing expenditures. For example, if the value is from 3 months recall, it will be annualized by multiplying 4. Though this is unlikely to cause when the surveyed month is evenly distributed across the year. But it could cause biases when considering households interviewed at a particular time. When analyzing these variables, it is typically useful to interpret them in the context of the interview month.

Valuing items that were not purchased – such as collected, received as gift or in-kind – is also challenging. GMD adds the value of purchased items and non-purchased items for the expenditure for the item when the value of non-purchased item is given. However, when only other measures, such as quantity, weight, or volume is given for non-purchased items, it is not calculated as a part of the expenditure.

6.4.2.1 Access to Services

Recommendations or framework for utilities access to variables distinguishes access to services within and outside the premise, though many existing surveys do not provide the item. GMD allowed codes include a third category that "unstated" the raw data do not provide the location of the access point for the utility.

Box 2 Countries with extreme seasonality in utilities consumption

Notice that the consumption aggregate should reflect the *normal* consumption of a household over the period considered (in our case, one year). In some countries, and particularly in the ECA region, there is an extreme fluctuation of energy consumption between winter and summer. Beware that households interviewed over the summer months might have a very different consumption than households interviewed during winter months (for example, in Tajikistan). Nevertheless, under the assumption that households are interviewed throughout the year and assuming that there is no sample bias, this should not be an issue.

Box 3 Inclusion of self-collected or received in-kind items as a part of consumption

Energy expenditures include both purchased items and also self-collected or received in-kind items when they are available and are reported in monetary terms. Non-purchased items that are only reported in quantity, volume or any other non-monetary terms are not valued in monetary terms to be added as a part of energy consumption. Estimating value per unit for monetary conversion can be biased, even if value per unit purchased for the same item for the same household is given, due to differences in quality, inaccuracy in unit measurement, and other factors not listed here.

Box 4 Note on access to services inferred from expenditures

When there is no direct question on access to services, expenditure on the item if often used as a proxy for the access. However, it should be noted that access inferred from expenditure generally underestimates actual access, because it leaves out cases such as when the household do not pay for the item due to the subsidy, or the expense is charged along with other expenditure such as rent.

Not all surveys have detailed information that would allow for better understanding of access to water and sanitation. For example, round trip collection time and location for water and whether facilities are shared and if there is safe disposal of excreta for sanitation faculties, would allow for better understanding of the service levels.

Table 6 GMD Utilities Module – key variables and descriptions

Module Code	Variable label	Variable name	Allowed codes after standardization
ID	Country code	countrycode	String
ID	Year	year	Numeric discrete
ID	Household identifier	hhid	String
Survey Characteristics & Welfare	Weight	weight	Numeric
Utilities	Sources of drinking water	water_source	Numeric categorical 1 = Piped water into dwelling 2 = Piped water to yard/plot 3 = Public tap or standpipe 4 = Tubewell or borehole 5 = Protected dug well 6 = Protected spring 7 = Bottled water 8 = Rainwater 9 = Unprotected spring 10 = Unprotected dug well 11 = Cart with small tank/drum 12 = Tanker-truck 13 = Surface water 14 = Other
Utilities	Improved water recommended estimate	imp_wat_rec	Numeric categorical 0 = No; 1 = Yes
Utilities	Original survey response in string for water_source variable	water_original	String
Utilities	Type of water questions used in the survey	watertype_quest	Numeric categorical 1=drinking water; 2=general water; 3=both; 4=others
Utilities	Access to piped water	piped	Numeric categorical 0 = No; 1 = water_source 1,2 or 3
Utilities	Access to piped water on premises	piped_to_prem	Numeric categorical 0 = No; 1 = water_source 1 or 2
Utilities	Collection time of imp_wat_rec within 30 minutes	w_30m	Numeric categorical 1=collection time of imp_wat_rec less than or equal to 30 mins; 0=collection time of imp_wat_rec more than 30 mins
Utilities	Availability of water when needed	w_avail	Numeric categorical 1= water is available continuously, reliable source; 0=water source is unreliable

Utilities	Main sanitation source	sanitation_source	Numeric categorical 1 = A flush toilet 2 = A piped sewer system 3 = A septic tank 4 = pit latrine 5 = ventilated improved pit latrine (VIP) 6 = pit latrine with slab 7 = composting toilet 8 = Special case 9 = A flush/pour flush to elsewhere 10 = A pit latrine without slab 11 = Bucket 12 = hanging toilet or hanging latrine 13 = No facilities or bush or field 14 = Other
Utilities	Original survey response in string for sanitation_source variable	sanitation_original	String
Utilities	Access to flushed toilet	toilet_acc	Numeric categorical 0 = No; 1 = Yes, in premise 2 = Yes, but not in premise including public toilet 3 = Yes, unstated whether in or outside premise
Utilities	Access to sewer	sewer	Numeric categorical 0=No; 1=flush/pour flush to piped sewer system
Utilities	Access to any sanitation facility	open_def	Numeric categorical 0 = availability of any facility (from responses in sanitation_source, including unimproved); 1 = no facility, or bush, or field (13)
Utilities	Improved sanitation facility recommended estimate (not considering sharing)	imp_san_rec	Numeric categorical 0 = No; 1 = Yes
Utilities	Main types of sewage disposal system	waste	Numeric categorical 1 = Solid waste collected on a regular basis by authorized collectors 2 = Solid waste collected on an irregular basis by authorized collectors 3 = Solid waste collected by self-appointed collectors 4 = Occupants dispose of solid waste in a local dump supervised by authorities 5 = Occupants dispose of solid waste in a local dump not supervised by authorities 6 = Occupants burn solid waste 7 = Occupants bury solid waste 8 = Occupants dispose solid waste into river, sea, creek, pond 9 = Occupants compost solid waste 10 = Other arrangement
Utilities	Access to central heating	central_acc	Numeric categorical 0 = No; 1 = Yes

Utilities	Main source of heating	heatsource	Numeric categorical 1 = Firewood; 2 = Kerosene; 3 = Charcoal; 4 = Electricity; 5 = Gas; 6 = Central; 9 = Other; 10 = No heating
Utilities	Connection to gas/Usage of gas	gas	Numeric categorical 0 = No 1 = Yes, piped gas (LNG) 2 = Yes, bottled gas (LPG) 3 = Yes, but don't know
Utilities	Main source of cooking fuel	cooksource	Numeric categorical 1 = Firewood; 2 = Kerosene; 3 = Charcoal; 4 = Electricity; 5 = Gas; 9 = Other; 10 = No cook source
Utilities	Main source of lighting	lightsource	Numeric categorical 1 = Electricity; 2 = Kerosene; 3 = Candles; 4 = Gas; 9 = Other; 10= No light source
Utilities	Access to electricity	elec_acc	Numeric categorical 1 = Yes, public/quasi-public; 2 = Yes, private 3 = Yes, source unstated; 4 = No
Utilities	Access to electricity in dwelling	electricity	Numeric categorical 0 = No; 1 = Yes
Utilities	Electricity availability (hr/day)	elechr_acc	Numeric continuous
Utilities	Lighting and/or electricity – type of	electyp	Numeric categorical 1 = Electricity; 2 = Gas; 3 = Lamp; 4 = Others; 10 = No cook and light source
Utilities	Total annual household expenditures on water supply/piped water	pwater_exp	Numeric continuous
Utilities	Total annual household expenditure on hot water supply	hwater_exp	Numeric continuous
Utilities	Total annual household expenditure on water supply and hot water supply	water_exp	Numeric continuous
Utilities	Total annual household expenditures on collection and disposal of garbage or refuse	garbage_exp	Numeric continuous
Utilities	Total annual household expenditures on collection and disposal of wastewater	sewage_exp	Numeric continuous
Utilities	The total annual household expenditure on garbage and sewage collection	waste_exp	Numeric continuous
Utilities	Total annual household expenditures on other services relating to the dwelling	dwelothsvc_exp	Numeric continuous
Utilities	Total annual household expenditures on electricity and other associated expenditures	elec_exp	Numeric continuous
Utilities	Total annual household expenditure on town gas and natural gas	ngas_exp	Numeric continuous
Utilities	Total annual household expenditure on LPG	LPG_exp	Numeric continuous
Utilities	Total annual household expenditures on network/natural gas and liquefied gas (LPG)	gas_exp	Numeric continuous
Utilities	Total household expenditure on diesel	diesel_exp	Numeric continuous

Utilities	Total annual household expenditure on kerosene	kerosene_exp	Numeric continuous
Utilities	Total annual household expenditure on gasolines	gasoline_exp	Numeric continuous
Utilities	Total annual household expenditure on other liquid fuels	othliq_exp	Numeric continuous
Utilities	Total annual household expenditures on all liquid fuels	liquid_exp	Numeric continuous
Utilities	Total annual household expenditure on firewood	wood_exp	Numeric continuous
Utilities	Total annual household expenditure on coal	coal_exp	Numeric continuous
Utilities	Total annual household expenditure on peat	peat_exp	Numeric continuous
Utilities	Total annual household expenditure on other solid fuels	othsol_exp	Numeric continuous
Utilities	Total annual household expenditures on all solid fuels	solid_exp	Numeric continuous
Utilities	Total annual household expenditure on other fuels that are not captured under othliq_exp and othsol_exp	othfuel_exp	Numeric continuous
Utilities	Total annual household expenditure on central heating	central_exp	Numeric continuous
Utilities	Total annual household expenditures on heating	heating_exp	Numeric continuous
Utilities	Total annual household expenditure on all utilities excluding telecom and other housing expenses	utl_exp	Numeric continuous
Utilities	Total annual household expenditures on product and materials for maintenance and repair of the dwelling	dwelmat_exp	Numeric continuous
Utilities	Total annual household expenditures on services for minor maintenance and repair of the dwelling	dwelsvc_exp	Numeric continuous
Utilities	Total annual household expenditures on other materials and services for minor maintenance and repair of the dwelling	othhousing_exp	Numeric continuous
Utilities	Total annual household expenditures on fuels for personal transportation	transfuel_exp	Numeric continuous
Utilities	Total annual household expenditures on landphone	landphone_exp	Numeric continuous
Utilities	Total annual household expenditures on cellphone	cellphone_exp	Numeric continuous
Utilities	Total annual household expenditures on landline phone and cell phone	tel_exp	Numeric continuous
Utilities	Total annual household expenditures on information transmission and Internet connection services	internet_exp	Numeric continuous
Utilities	Total annual household expenditures on telegraphy, telex and telefax services	telefax_exp	Numeric continuous
Utilities	Total annual household expenditures on all telephone and telefax services	comm_exp	Numeric continuous

Utilities	Total annual household expenditures on television broadcasting services, license fees for television equipment and subscriptions to television networks	tv_exp	Numeric continuous
Utilities	Total annual household expenditures on internet, telephone and television broadcasting services	tvintph_exp	Numeric continuous

7 Assets and Dwellings (DWL)

7.1 Framework for Harmonization

The GMD Assets and Dwellings module contains information on housing conditions and asset ownership. The chapter introduces the concepts, definitions and data requirements for examining asset/dwelling ownership and condition and provide guidance for producing statistics on housing indicators that are commonly derived from household survey data sets or that can be constructed using existing variables. The overall objective of this indicator harmonization is to generate a unified data source of globally comparable indicators for assets and dwellings to support routine analytical and corporate business functions of the World Bank.

In this module, the primary unit of analysis is the level of household.

Since there is no commonly agreed-upon framework for assets and dwellings survey indicators, these guidelines draw on various major projects and internationally accepted principles for the collection and production of assets and dwellings statistical indicators based on household survey data. The mentioned harmonization projects include United Nations Principles and Recommendations for Population and Housing Censuses Revision 3 (2015), World Bank Living Standards Measurement Study (LSMS), International Income Distribution Database (I2D2) and other relevant regional and global harmonization efforts of the World Bank are referenced as appropriate.

Note that the GMD Dwellings and Assets module adopts narrow definition of Housing/Dwelling, and only include dwelling description and ownership variables in the dwellings module among the following five categories commonly used by UN, WB and Eurostat: 1) Description of Dwellings 2) Dwelling services 3) Dwelling expenditures 4) Assets in their Housing/Dwellings module and 5) Others.

This module consists of Dwellings and Assets subsections. The Dwellings section covers variables that describe dwelling condition and ownership, which can be categorized into three groups covering (i) materials, (ii) facilities and characteristics and (iii) ownership.[14] The Assets section is sub-divided into three parts: (i) household appliances, (ii) means of transportation and (iii) household animals.

Assets serve multiple functions. In their productive capacity, they generate income and facilitate access to capital and credit. They also strengthen a household's capacity to cope with and respond to shocks by enhancing its ability to diversify income and ease liquidity constraints. Moreover, assets comprise a store of wealth that can be sold to generate income or passed on to future generations. Finally, assets may provide status and security to individuals or households. Assets are therefore an important indicator of economic welfare that is complementary to consumption or income.

Understanding assets and dwellings is also critical to forming indicators under the SDGs. SDG 1 and 5, poverty and gender respectively, both require data on land access and ownership. For example, SDG 1.4 is requires knowing the proportion of total adult population with secure tenure rights to land, with legally

[14] Note that access to and expenditure on utilities services, such as energy, water and urban services are included in GDM Utilities module. Rent and/or imputed rent is discussed in GDM Welfare.

recognized documentation and who perceive their rights to land as secure, by sex and by type of tenure. Indicator 5.a.1 requires further information, including the proportion of total agricultural population with ownership or secure rights over agricultural land, disaggregated by sex. Proxies for these indicators can be created using harmonized data in this module.

7.2 Mapping and Description of Variables

GMD Assets and Dwellings module contains a large amount of metadata that provides a wealth of information about variables, including their types, descriptions, and sources. In order to improve readability, only the most significant information has been included in this section. For a complete list of all variables captured in the module, including Tier 3 and Tier 4 variables, please consult Table at the end of the chapter.

7.3 Assets

Assets consist of household appliances and means of transportation assessed at the household level. Household animals are tier 3 indicators and are not included in this version.

7.4 Appliances

All appliance variables are yes/no dummies and do not take the number of items into account.

landphone
landphone is a dummy variable indicating whether the household owns a landline phone. It is generally defined as landline phone, home telephone, or fixed phone. Landline phone ownership does not depend on who owns the phone within the household, nor on its condition.
Two categories after harmonization:

0 = No
1 = Yes

cellphone
cellphone is a dummy variable indicating whether anyone in the household owns a cell phone. Cell phone ownership does not depend on who owns the cellphone is within the household, nor on its condition. Two categories after harmonization:

0 = No
1 = Yes

phone
phone is a dummy variable indicating whether the household owns *either a land phone or a cell phone*. It should only be coded in cases where the survey does not distinguish between ownership of landline and cell phones. In other cases, it may be coded as missing. Phone ownership does not depend on who owns the phone within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

computer
computer is a dummy variable indicating whether the household owns a computer, including desktop and laptop computer. Computer ownership does not depend on who owns the computer within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

etablet
etablet is a dummy variable indicating the ownership of an electronic tablet. Two categories after harmonization:
0 = No
1 = Yes

internet
internet is a categorical variable indicating whether anyone in the household can use a device that is connected to the internet within the home or have access to internet outside the house. Connection to the Internet can be both wired and wireless and does not depend on who manages it within the household. Four categories after harmonization:

1 = Subscribed in the house
2 = Accessible outside the house (includes internet cafes and smartphones with internet access)
3 = Either (Use this category when the questionnaire does not specify whether the access is in the house or outside the house)
4 = No internet

radio
radio is a dummy variable indicating whether the household owns a radio (i.e. radio, radio cassette, and 3-in-1 radio cassette player (radio). Radio ownership does not depend on who owns the radio within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

tv
tv is a dummy variable indicating whether the household owns a TV set. This includes both color and black and white TVs. TV set ownership does not depend on who owns the TV set within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

tv_cable
tv_cable is a dummy variable indicating whether the household owns a cable or dish antenna services. Only for households that reported having a TV (**tv=1**).
Two categories after harmonization:

0 = No
1 = Yes

video
video is a dummy variable indicating whether the household owns a videocassette player and/or video cassette recorder. Video cassette player ownership does not depend on who owns the player within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

fridge
fridge is a dummy variable indicating whether the household owns a refrigerator (i.e. refrigerator or freezer). It does not include cooler, icebox or ice chest. Refrigerator ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

sewmach
sewmach is a dummy variable indicating whether the household owns a sewing machine. Sewing machine ownership does not depend on who owns the sewing machine within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

washmach
washmach is a dummy variable indicating whether the household owns a machine for washing clothes and household linen; but does not include non-electric washing machine. Washing machine ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

stove
stove is a dummy variable indicating whether the household owns a stove. Stove generally refers to a portable or fixed apparatus that burns fuel or uses electricity to provide heat for cooking or heating purposes and includes a cooker (stove). Stove ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

ricecook
ricecook is a dummy variable indicating whether the household owns a rice cooker. Rice cooker ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:

0 = No
1 = Yes

fan
fan is a dummy variable indicating whether the household owns a fan operated by electricity. Fan ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

ac
ac is a dummy variable indicating whether the household owns a central or wall air conditioner. Air conditioner ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

ewpump
ewpump is a dummy variable indicating the ownership of an electric water pump. Two categories after harmonization:
0 = No
1 = Yes

7.4.1 Means of transportation

bcycle
bcycle is a dummy variable indicating whether the household owns a bicycle. Note that motored bicycles are classified as motorcycle regardless of motor type. Bicycle ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

mcycle
mcycle is a dummy variable indicating whether the household owns a motorcycle. Motorcycle refers to an automotive vehicle with two in-line wheels, including motorbike or moped. Motorcycle ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

oxcart
oxcart is a dummy variable indicating whether the household owns an animal cart, which is used as a means of transport or a farm tool. Animal cart ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:

0 = No
1 = Yes

boat
boat is a dummy variable indicating whether the household owns a boat. Boat ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

car
car is a dummy variable indicating whether the household owns a car or truck for household use, excluding commercial vehicle. Car ownership does not depend on who owns the asset within the household, nor on its condition. Two categories after harmonization:
0 = No
1 = Yes

canoe
canoe is a dummy variable indicating the ownership of a canoe. Two categories after harmonization:
0 = No
1 = Yes

7.4.2 Dwellings

Dwellings variables are mapped into three sections: Materials; Facilities and Characteristics; and Ownership. Materials, characteristics, ownership variables are categorical variables, and most facility variables are yes/no dummies. All variables are assessed at household level.

7.4.3 Materials

roof[15]
roof is a categorical variable that indicates ***type of material used for roof***, such as adobe, thatch, iron, and tiles. The roof material is categorized into 3 broad categories namely: Natural, rudimentary and finished. For cases that cannot be covered in the above three categories, please use code 15 = Other – "Specific".

Main source of material used for roof:

1 = Natural – Thatch/palm leaf
2 = Natural – Sod
3 = Natural – Other
4 = Rudimentary – Rustic mat
5 = Rudimentary – Palm/bamboo
6 = Rudimentary – Wood planks
7 = Rudimentary – Other

[15] Hereafter, the best possible match is sought, but in many cases the correspondence between country-specific values and these standardized codes is imperfect. If a perfect match is not found, users could classify the type of roofs as Natural-Other, Rudimentary-Other, Finished-Other or ultimately as Other – "Specific".

8 = Finished – Wood
9 = Finished – Asbestos
10 = Finished – Tile
11 = Finished – Concrete
12 = Finished – Metal
13 = Finished – Roofing shingles
14 = Finished – Other
15 = Other

wall

wall is a categorical variable that indicates ***type of material used for walls***. The wall material is categorized into 3 broad categories namely: Natural, rudimentary and finished. For cases that cannot be covered in the above three categories, please use code 19 = Other – "Specific". Main source of material used for walls, 19 categories after harmonization:

1 = Natural – Cane/palm/trunks
2 = Natural – Dirt
3 = Natural – Other
4 = Rudimentary – Bamboo with mud
5 = Rudimentary – Stone with mud
6 = Rudimentary – Uncovered adobe
7 = Rudimentary – Plywood
8 = Rudimentary – Cardboard
9 = Rudimentary – Reused wood
10 = Rudimentary – Other
11 = Finished – Woven Bamboo
12 = Finished – Stone with lime/cement
13 = Finished – Cement blocks
14 = Finished – Covered adobe
15 = Finished – Wood planks/shingles
16 = Finished – Plaster wire
17 = Finished – GRC/Gypsum/Asbestos
18 = Finished – Other
19 = Other

floor

floor is a categorical variable that indicates ***type of material used for floors***. The floor material is categorized into 3 broad categories namely: Natural, rudimentary and finished. For cases that cannot be covered in the above three categories, please use code 14 = Other – "Specific".

Main source of material used for floors, 14 categories after harmonization as shown below.

1 = Natural – Earth/sand;
2 = Natural – Dung;
3 = Natural – Other
4 = Rudimentary – Wood planks

5 = Rudimentary – Palm/bamboo
6 = Rudimentary – Other
7 = Finished – Parquet or polished wood
8 = Finished – Vinyl or asphalt strips
9 = Finished – Ceramic/marble/granite
10 = Finished – Floor tiles/terrazzo
11 = Finished – Cement/red bricks
12 = Finished – Carpet
13 = Finished – Other
14 = Other

7.4.4 Facilities and characteristics

kitchen
kitchen is a dummy variable indicating whether the household has a separate kitchen in the dwelling*, implying an independent space is set aside for cooking inside the dwelling* (kitchen). Any other space reserved for cooking, such as kitchenette or an outer space for kitchen, is not considered as a kitchen. The unit of enumeration for this topic is the housing unit. However, some countries may find it useful to collect information on the availability of kitchen facilities for the use of occupants in collective living quarters, such as hotels, lodging houses, institutions camps and workers' quarters, though people living in these places are generally not captured in a household survey. Two categories after harmonization:
0 = No
1 = Yes

bath
bath is a dummy variable indicating whether the household has a separate bathing facility such a shower or bathroom in the dwelling. Fixed bath or shower outside housing unit is not considered. Two categories after harmonization:
0 = No
1 = Yes

rooms
rooms is a integer variable that refers to the number of habitable rooms in the whole household dwelling unit. It may consist of one or more structure(s) (rooms), including all rooms used for living, sleeping and eating. It excludes storerooms, bathrooms, kitchens and rooms used for business or professional purposes. In the case of a one-room dwelling this variable will have the value of one.

areaspace
areaspace is a continues variable that refers to the total floor area (in square meters) of all rooms and auxiliary premises (kitchen, vestibule, cloakroom, hallway, toilet room, sauna that is within the dwelling, pantry, interstice, bathroom, storeroom, porch, integrated wall closets) in the whole household dwelling unit.

The area of the dwelling does not include cellars, garages (incl. in private houses), boiler rooms, attics (if they are not suitable for permanent habitation) and common rooms (such as stairways, corridors, saunas,

etc.) in buildings with multiple dwellings. Open areas (loggias, balconies and terraces) are not included in the area of the dwelling. However, if such areas have been closed in and insulated, they should be added to the total area of the dwelling. If a household lives in an uncompleted residential building, enter the area of the finished part of the house.

ybuilt
This is an integer variable that indicates the year when the dwelling was built. This information should be gathered, when available, for all HHs, irrespective of ownership status.

7.4.5 Ownership

ownhouse
ownhouse is a categorical variable that specifies whether a household owns, rents, is provided for free, or squats in their house. Ownership (1) includes ownership or other equivalent of secure tenure, whether or not full-payment has been made yet. Rental (2) denotes that regular payment is made to the owner (which could be private, corporate, or government) with or without formal agreement.

This variable has four categories after harmonization:

1 = ownership/ secure rights
2 = renting
3 = provided for free
4= without permission

acqui_house
acqui_house is a categorical variable that specifies the mode of acquisition for their dwellings. Only for household owners (Category 1 in ***ownhouse*** variable)

Acquisition of house, three categories after harmonization:

1 = Purchased
2 = Inherited
3 = Other

Category 3 would apply to cases if the members built their own homes or obtained it from other means specific to countries.

dwelownlti
dwelownlti is a dummy variable specifying whether a household has legal evidence for ownership (yes/no). See Lessons learned/Challenges in the next section for more information on what can be considered legal evidence. Two categories after harmonization:
0 = No
1 = Yes

fem_dwelownlti

fem_dwelownlti is a dummy variable that specifies whether the names of female household members are listed on the legal document specifying ownership of the dwelling (yes/no). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for the dwelling. Two categories after harmonization:
0 = No
1 = Yes

dwelownti
dwelownti is a categorical variable that specifies the type of legal document the household has as evidence for ownership of their dwelling.

Type of legal document, six categories after harmonization:

1= Title, deed, freehold
2= Government issued leasehold
3= Occupancy certificate – govt issued
4= legal document in the name of group (community; cooperative)
5= condominium (apartment)
6= Other

selldwel
selldwel is a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their dwelling (yes/no). Two categories after harmonization:
0 = No
1 = Yes

transdwel
transdwel is a dummy variable that specifies whether the respondent has the right to bequeath the dwelling to the next generation of their family (yes/no). Two categories after harmonization:
0 = No
1 = Yes

ownland
ownland is a dummy variable that specifies whether a household owns residential land (yes/no). Ownership for property versus residential land on which property is constructed can be different in certain jurisdictions (land vested in a state or municipality). Two categories after harmonization:
0 = No
1 = Yes

acqui_land
acqui_land is a categorical variable that specifies the mode of acquisition for any residential land that the household uses. Only for the main residence. Only for land owners (category 1 in ***ownland*** variable)

Acquisition of residential land, categories after harmonization:

1 = Purchased
2 = Inherited
3 = Other

doculand
doculand is the dummy variable specifying whether the household has a legal document for their residential land (yes/no). See the main challenges/lessons learned in the next section for more information on what can be considered legal evidence. Only for land owners (category 1 in ***ownland*** variable). Two categories after harmonization:
0 = No
1 = Yes

fem_doculand
fem_doculand is the dummy variable specifying whether the household has the name of female household members listed on a legal document for their residential land (yes/no). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for residential land. Only for land owners (category 1 in **ownland** variable). Two categories after harmonization:
0 = No
1 = Yes

landownti
landownti is a categorical variable that specifies the type of document that a household has to prove land ownership. The two customary rights categories (3 and 4) differentiate whether issued by plot or as a joined group title. Customary groups and cooperatives are differentiated, as well. Customary groups not required to have formal membership declared, while cooperative members have formalized status.Land ownership type of document. Only for land owners (category 1 in **ownland** variable). If household have pieces of land under several type of ownership, the harmonizer should collapse the plots area by title type and then pick the type of ownership for the largest area. Six categories after harmonization:
1 = Title; deed
2 = leasehold (govt issued)
3 = Customary land certificate/plot level
4 = Customary based / group right
5 = Cooperative group right
6 = Other

sellland
sellland is a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their residential land (yes/no). Only for land owners (category 1 in **ownland** variable). Two categories after harmonization:

0 = No

1 = Yes

transland
transland is a dummy variable that specifies whether the respondent has the right to bequeath residential land to the next generation of their family (yes/no). Only for land owners (category 1 in **ownland** variable). Two categories after harmonization:

0 = No

1 = Yes

agriland
agriland is a dummy variable that specifies whether a household is using agricultural land according to the classification of the World Census of Agriculture 2020 (yes/no)[16]. Two categories after harmonization:

0 = No

1 = Yes

area_agriland
area_agriland is a numeric, continuous variable that specifies the ***total area of agricultural land used*** by household members in hectares. This could be land that is owned, rented, or sharecropped, or some combination.

A hectare is equal to 10,000 square meters or equivalent to 2.471 acres

ownagriland
ownagriland is a dummy variable that specifies whether a household owns agricultural land (yes/no). Owned land can be by freehold, deed, customary, or government leasehold. Only those households that declared using agricultural land (category 1 in **agriland** variable). Two categories after harmonization:

[16] FAO (2015). "WORLD PROGRAMME FOR THE CENSUS OF AGRICULTURE 2020". Paragraph (8.2.35) FAO's recommended land use classification in the Figure 1 includes the following aggregate classes:

- **Arable land** is land that is used in most years for growing temporary crops. It includes land used for growing temporary crops during a twelve-month reference period, as well as land that would normally be so used but is lying fallow or has not been sown due to unforeseen circumstances. Arable land does not include land under permanent crops or land that is potentially cultivable but is not normally cultivated. Such land should be classified as "permanent meadows and pastures" if used for grazing or haying, "forest and other wooded land" if overgrown with trees and not used for grazing or haying, or "other area not elsewhere classified" if it becomes wasteland.
- **Cropland** is the total of arable land and land under permanent crops.
- **Agricultural land** is the total of cropland and permanent meadows and pastures.
- **Land used for agriculture** is the total of "agricultural land" and "land under farm buildings and farmyards".

0203 Area of holding according to land tenure types

- Legal ownership or legal owner-like possession
- Non-legal ownership or non-legal owner-like possession
- Rented from someone else
- Other types of land tenure

0 = No

1 = Yes

area_ownagriland
area_ownagriland is a numeric, continuous variable that specifies the ***total area of agricultural land owned*** in hectares. Only for agriculture land owners (category 1 in **ownagriland** variable).

A hectare is equal to 10,000 square meters or equivalent to 2.471 acres

purch_agriland
purch_agriland is a dummy variable specifying whether a household has purchased the agricultural land they own (yes/no). Only for agriculture land owners (category 1 in **ownagriland** variable). Two categories after harmonization:

0 = No

1 = Yes

areapurch_agriland
areapurch_agriland is a numeric, continuous variable that specifies the total area of agricultural land purchased in hectares. Only for category 1 in **purch_agriland** variable.

A hectare is equal to 10,000 square meters or equivalent to 2.471 acres

inher_agriland
inher_agriland is a dummy variable specifying whether a household has inherited the agricultural land they own (yes/no). Only for agriculture land owners (category 1 in **ownagriland** variable). Two categories after harmonization:

0 = No

1 = Yes

areainher_agriland
areainher_agriland is a numeric, continuous variable that specifies the total area of agricultural land inherited in hectares. Only for category 1 in **inher_agriland** variable.

A hectare is equal to 10,000 square meters or equivalent to 2.471 acres

rentout_agriland
rentout_agriland is a dummy variable that specifies whether any of the agricultural land a household uses is rented–out land or sharecropped (yes/no). Only for agriculture land owners (category 1 in **ownagriland** variable). This refers to land (or use rights) owned by the household but cultivated or utilized by someone else irrespective of the type of the tenant (individual, household, legal entity, etc.) and contractual arrangements (fixed rental, sharecropping, etc.). Two categories after harmonization:

0 = No

1 = Yes

arearentout_agriland
arearentout_agriland is a numeric, continuous variable that specifies the total area of agricultural land rented out or share cropped in hectares. Only for category 1 in **rentout_agriland** variable.

A hectare is equal to 10,000 square meters or equivalent to 2.471 acres

rentin_agriland
rentin_agriland is a dummy variable that specifies whether any of the agricultural land a household uses is rented–in land or sharecropped (yes/no). This refers land owned by others (not members of the household) but cultivated or used by the household under fixed rental, sharecropped or similar arrangements. We agree that this question should apply to all households using agricultural land (agriland==1). Two categories after harmonization:

0 = No

1 = Yes

arearentin_agriland
arearentin_agriland is a numeric, continuous variable that specifies the total area of agricultural land rented in or share cropped in hectares. Only for category 1 in **rentin_agriland** variable.

A hectare is equal to 10,000 square meters or equivalent to 2.471 acres

docuagriland
docuagriland is the dummy variable specifying whether the household has a legal document for their agricultural land (yes/no). See main challenges/lessons learned in the next section for more information on what can be considered legal evidence. Only for agriculture land owners (category 1 in **ownagriland** variable). Two categories after harmonization:

0 = No
1 = Yes

area_docuagriland
area_docuagriland is a numeric, continuous variable that specifies the total area of agricultural land owned with legal documentation in hectares. Only for category 1 in **docuagriland** variable.

A hectare is equal to 10,000 square meters or equivalent to 2.471 acres

fem_agrilandownti
fem_agrilandownti is the dummy variable specifying whether the household has the name of female household members listed on a legal document for their agricultural land (yes/no). This will be derived from questions asking about the roster ID of the household member(s) whose name(s) are on the legal document for agricultural land. Only for category 1 in **docuagriland** variable. Two categories after harmonization:

0 = No
1 = Yes

agrilandownti

agrilandownti is a categorical variable that specifies the type of document that a household has to prove agricultural land ownership. The two customary rights categories (3 and 4) differentiate whether issued by plot or as a joined group title. Customary groups and cooperatives are differentiated, as well. Customary groups not required to have formal membership declared, while cooperative members have formalized status. Agricultural land ownership type of document. Only for category 1 in **docuagriland** variable. If household have several plots under several type of ownership, the harmonizer should collapse the plots by area and then pick the type of ownership for the largest area. Categories after harmonization:

1 = Title; deed
2 = leasehold (govt issued)
3 = Customary land certificate/plot level
4 = Customary based / group right
5 = Cooperative
6 = Other

sellagriland

sellagrilandis a dummy variable that specifies whether the respondent has alienation rights (i.e. the right to sell) for their agricultural land (yes/no). Only for agricultural land owners, category 1 in **ownagriland** variable**.** Two categories after harmonization:

0 = No
1 = Yes

transagriland

transagriland is a dummy variable that specifies whether the respondent has the right to bequeath agricultural land to the next generation of their family (yes/no). Only for agricultural land owners, category 1 in **ownagriland** variable**.** Two categories after harmonization:

0 = No
1 = Yes

dweltyp

dweltyp is a categorical variable that specifies the type of dwelling. Categories after harmonization are:

1 = Detached house
2 = Multi-family house
3 = Separate apartment
4 = Communal apartment
5 = Room in a larger dwelling
6 = Several buildings connected
7 = Several separate buildings
8 = Improvised housing unit
9 = Other

typlivqrt

typlivqrt is a categorical variable that specifies the type of living quarters. Categories after harmonization are:

1 = Housing units, conventional dwelling with basic facilities
2 = Housing units, conventional dwelling without basic facilities
3 = Other housing units

7.5 Missing Value Codes

Harmonizers need to clearly differentiate missing values of variables from variables that were present in the survey but could not be harmonized due to reasons such as time unavailability. This will help the future harmonizers to focus on the unharmonized variables. The missing value code for these two scenarios are:

- For variables unavailable in survey = .
- For variables available in the survey but not harmonized = .a

7.6 Main Challenges/ Lessons learned (Module specific)

The dwelling and land variables will differ in different country surveys. In particular, documentation of legal ownership and types of dwelling/land ownership will be different depending on the country context and will often be missing in household budget surveys. To help harmonizers to understand what can be considered documentation across countries, see Table 2.

Table 2. Legal documentation of property and land ownership and use

Examples of title or deed – full ownership	Title deed
	Registered title
	Landhold title
	Real property title based on court decision
	Donation deed
	Registered conveyance of property/land in country with deeds system
Long term leasehold with government	Registered leasehold (Rwanda, Vietnam, Ethiopia etc.)
	Right of use and benefit of land (DUAT in Mozambique)
Occupancy agreement/semi-formal title for dwelling	Certificate of localization
	Occupancy permit
	Certificate of occupancy
	Grant of admission
	Bails (French system)
	Rural/ urban concession
Occupancy agreement/semi-formal title for land (including registration of customary land)	Land use certificate
	Certificate of customary ownership
	Certificate of hereditary acquisition listed in registry
	Provisional concession
	Official “petits papier”
Collective agreement of occupancy / semi-formal	Collective land certificate
	Community DUAT (Right of use and benefit of land)
	Tribal/Land certificate
	Council tax letter

Examples of **NOT** possessing a legal document as evidence of ownership for dwelling or land	Minutes of palaver
	Mortgage agreement
	Mortgage bond
	Private land sale contract
	Promise of purchase document
	Property tax receipt
	Simple petits papier
	Survey plan
	Utility bills
	Private Rental contract (with property owner)

Furthermore, harmonizers should checks to ensure data has been harmonized correctly. Dummy variables should have a value of 0 or 1 (or be missing). They should also have variation as not all households will have certain items. Using landphone as an example, the Stata code to check that the value is 0 or 1:

```
landphone!=. & landphone!= 1 & landphone!= 0
```

code to check variation:

```
su landphone if sdlandphone==0, d
```

Table 7 GMD Assets and Dwellings Module – key variables and descriptions.

Module Code	Variable label	Variable name	Allowed codes after standardization
ID	country code	countrycode	string
ID	Year	year	numeric
ID	Household identifier	hhid	string or numeric
Survey Characteristics & Welfare	Weight	weight	numeric
Assets & Dwellings	Ownership of a land phone	landphone	0 = No 1 = Yes
Assets & Dwellings	Ownership of a cell phone	cellphone	0 = No 1 = Yes
Assets & Dwellings	Ownership of a telephone	phone	0 = No 1 = Yes
Assets & Dwellings	Ownership of a computer	computer	0 = No 1 = Yes
Assets & Dwellings	internet connexion	internet	1 = Subscribed in the house 2 = Accessible outside the house 3 = Either 4 = No internet
Assets & Dwellings	Ownership of a radio	radio	0 = No 1 = Yes
Assets & Dwellings	Ownership of a tv	tv	0 = No 1 = Yes

Assets & Dwellings	Ownership of a cable tv	tv_cable	0 = No 1 = Yes
Assets & Dwellings	Ownership of a video	video	0 = No 1 = Yes
Assets & Dwellings	Ownership of a refrigerator	fridge	0 = No 1 = Yes
Assets & Dwellings	Ownership of a sewing machine	sewmach	0 = No 1 = Yes
Assets & Dwellings	Ownership of a washing machine	washmach	0 = No 1 = Yes
Assets & Dwellings	Ownership of a stove	stove	0 = No 1 = Yes
Assets & Dwellings	Ownership of a rice cooker	ricecook	0 = No 1 = Yes
Assets & Dwellings	Ownership of an electric fan	fan	0 = No 1 = Yes
Assets & Dwellings	Ownership of an air conditioner	ac	0 = No 1 = Yes
Assets & Dwellings	Ownership of a bicycle	bcycle	0 = No 1 = Yes
Assets & Dwellings	Ownership of a motorcycle	mcycle	0 = No 1 = Yes
Assets & Dwellings	Ownership of an oxcart	oxcart	0 = No 1 = Yes
Assets & Dwellings	Ownership of a boat	boat	0 = No 1 = Yes
Assets & Dwellings	Ownership of a Car	car	0 = No 1 = Yes
Assets & Dwellings	Ownership of a electronic tablet	etablet	0 = No 1 = Yes
Assets & Dwellings	Ownership of a electric water pump	ewpump	0 = No 1 = Yes
Assets & Dwellings	Ownership of a canoes	canoe	0 = No 1 = Yes
Assets & Dwellings	Main material used for roof	roof	1 = Natural – Thatch/palm leaf 2 = Natural – Sod 3 = Natural – Other 4 = Rudimentary – Rustic mat 5 = Rudimentary – Palm/bamboo 6 = Rudimentary – Wood planks 7 = Rudimentary – Other 8 = Finished – Roofing 9 = Finished – Asbestos 10 = Finished – Tile 11 = Finished – Concrete 12 = Finished – Metal tile 13 = Finished – Roofing shingles 14 = Finished – Other 15 = Other – “Specific”

Assets & Dwellings	Main material used for external walls	wall	1 = Natural – Cane/palm/trunks 2 = Natural – Dirt 3 = Natural – Other 4 = Rudimentary – Bamboo with mud 5 = Rudimentary – Stone with mud 6 = Rudimentary – Uncovered adobe 7 = Rudimentary – Plywood 8 = Rudimentary – Cardboard 9 = Rudimentary – Reused wood 10 = Rudimentary – Other 11 = Finished – Woven Bamboo 12 = Finished – Stone with lime/cement 13 = Finished – Cement blocks 14 = Finished – Covered adobe 15 = Finished – Wood planks/shingles 16 = Finished – Plaster wire 17 = Finished – GRC/Gypsum/Asbestos 18 = Finished – Other 19 = Other – “Specific”
Assets & Dwellings	Main material used for floor	floor	1 = Natural – Earth/sand; 2 = Natural – Dung; 3 = Natural – Other 4 = Rudimentary – Wood planks 5 = Rudimentary – Palm/bamboo 6 = Rudimentary – Other 7 = Finished – Parquet or polished wood 8 = Finished – Vinyl or asphalt strips 9 = Finished – Ceramic/marble/granite 10 = Finished – Floor tiles/terazzo 11 = Finished – Cement/red bricks 12 = Finished – Carpet 13 = Finished – Other 14 = Other – “Specific”
Assets & Dwellings	Separate kitchen in the dwelling	kitchen	0 = No 1 = Yes
Assets & Dwellings	Bathing facility in the dwelling	bath	0 = No 1 = Yes
Assets & Dwellings	Number of habitable rooms	rooms	numeric, continuous
Assets & Dwellings	Area	areaspace	numeric, continuous
Assets & Dwellings	Year the dwelling built	ybuilt	numeric, continuous
Assets & Dwellings	Ownership of house	ownhouse	1 = ownership/secure rights 2 = renting 3 = provided for free 4 = without permission
Assets & Dwellings	Acquisition of house	acqui_house	1 = Purchased 2 = Inherited 3 = Other

Assets & Dwellings	Acquisition of land	acqui_land	1 = Purchased 2 = Inherited 4 = Other
Assets & Dwellings	Legal title for Ownership	dwelownlti	0 = No 1 = Yes
Assets & Dwellings	Legal title for Ownership - Female	fem_dwelownlti	0 = No 1 = Yes
Assets & Dwellings	Type of Legal document	dwelownti	1 = Title, deed, freehold 2 = Government issued leasehold 3 = Occupancy certificate – govt issued 4 = legal document in the name of group (community; cooperative) 5 = condominium (apartment) 6 = Other
Assets & Dwellings	Right to sell dwelling	selldwel	0 = No 1 = Yes
Assets & Dwellings	Right to transfer dwelling	transdwel	0 = No 1 = Yes
Assets & Dwellings	Ownership of land	ownland	0 = No 1 = Yes
Assets & Dwellings	Legal document for residential land	doculand	0 = No 1 = Yes
Assets & Dwellings	Legal document for residential land - female	fem_doculand	0 = No 1 = Yes
Assets & Dwellings	Land Ownership	landownti	1 = Title; deed 2 = leasehold (govt issued) 3 = Customary land certificate/plot level 4 = Customary based / group right 5 = Cooperative group right 6 = Other
Assets & Dwellings	Types of living quarters	typlivqrt	1 = Housing units, conventional dwelling with basic facilities 2 = Housing units, conventional dwelling without basic facilities 3 = Other housing units
Assets & Dwellings	Types of Dwelling	dweltyp	1 = Detached house 2 = Multi-family house 3 = Separate apartment 4 = Communal apartment 5 = Room in a larger dwelling 6 = Several buildings connected 7 = Several separate buildings 8 = Improvised housing unit 9 = Other
Assets & Dwellings	Agriculture Land	agriland	0=No 1= Yes
Assets & Dwellings	Type Agri Land ownership doc	agrilandownti	0 =No 1= Yes

Assets & Dwellings	Area of Agriculture land	area_agriland	numeric, continuous
Assets & Dwellings	Area of documented agri land	area_docuagriland	numeric, continuous
Assets & Dwellings	Area of agriculture land owned	area_ownagriland	numeric, continuous
Assets & Dwellings	Area of inherited agriculture land	areainher_agriland	numeric, continuous
Assets & Dwellings	Area of purchased agriculture land	areapurch_agriland	numeric, continuous
Assets & Dwellings	Area of rent in agri land	arearentin_agriland	numeric, continuous
Assets & Dwellings	Area of rent out agri land	arearentout_agriland	numeric, continuous
Assets & Dwellings	Documented Agri Land	docuagriland	0 =No 1= Yes
Assets & Dwellings	Ownership Agri Land - Female	fem_agrilandownti	0 =No 1= Yes
Assets & Dwellings	Inherit agriculture land	inher_agriland	0 =No 1= Yes
Assets & Dwellings	Ownership of agriculture land	ownagriland	0 =No 1= Yes
Assets & Dwellings	Purchased agri land	purch_agriland	0 =No 1= Yes
Assets & Dwellings	Rent in Land	rentin_agriland	0 =No 1= Yes
Assets & Dwellings	Rent Out Land	rentout_agriland	0 =No 1= Yes
Assets & Dwellings	Right to sell agri land	sellagriland	0 =No 1= Yes
Assets & Dwellings	Right to sell land	sellland	0 =No 1= Yes
Assets & Dwellings	Right to transfer agri land	transagriland	0 =No 1= Yes
Assets & Dwellings	Right to transfer land	transland	0 =No 1= Yes