

# Guidelines for the Harmonising Census and Survey Microdata

## Introduction

Harmonization is about making statistics and data more comparable, consistent and coherent. Data harmonization refers to the processes that enable data from different sources and studies to be combined, queried, and analysed. Harmonization also seeks to bring together various types, levels and sources of data in such a way that they can be made compatible and comparable, and thus useful for decision making.

A key consideration is to find pragmatic ways of making compatible and integrable datasets which have been collected for similar or slightly different purposes under different collection approaches and using different standards and methodologies. This means avoiding the need to convert all the data to a single standard, but rather finding ways to make it usable at some higher level of aggregation or generalisation.

## 1. Background

### 1.1. Defining data harmonisation

An act of reconciling the definition and representation formats (syntax) of data elements in a domain of interest. It entails a set of activities that improves the consistency in the use of data elements in terms of their meaning and representation format.

Harmonization differs from standardisation in that it does not impose a single methodology or norm, but rather seeks to find ways of integrating or making "an agreeable effect" from information gathered through disparate methodologies.

Data harmonization is a process used to standardise the data elements that are used frequently shared across multiple applications or are selected for inclusion in a master or reference data. Data harmonization will identify a set of core data elements (data elements expressed using different descriptions but with identical meaning).

### 1.2. Objective and Goal of Data Harmonization

The Pacific Community, Statistics for Development Division (SDD) along with Pacific Island National Statistics Offices produce a wide number of statistics on a range of topics. Harmonization of Pacific Island statistics involves bringing together of statistical methodologies and tools to produce comparable, accurate and up-to-date statistics to inform both national development processes and the process of regional integration. Alignment with international standards will help reduce duplication and enable data interoperability among stakeholders in the region, leading to improved data quality over time.

Currently, there is a need to optimise, and further develop the data collection and data dissemination methods in the Pacific Islands region. This is why harmonization is such an important aspect of statistical quality and building trust. Harmonising data elements reduces ambiguity and provides consistency. It enables us to improve the coherence of statistical outputs and the consistency of our statistical inputs. And through this, increase their value and efficiency. More comparable and coherent statistics leads to better decisions.

This guide provides information about harmonization of data produced via the Pacific Community, Statistics for Development Divisions (SDD)'s priority statistical collections, namely Population and Housing Census (PHC) and Household Income and Expenditure Survey (HIES). This guide also covers the harmonization of

variables and variable labels (data dictionary: common topics and variables), input datasets and translation tables used

The overall aim of these technical guidelines is to provide a credible and practical harmonization methodology on how to improve data collection methods for data providers, governments, policy makers and other interested stakeholders in the Pacific Islands region. It outlines the systematic approach, standards and tools to facilitate a data harmonization exercise, practical recommendations for data harmonization, cases and application of the results of data harmonization exercises and details on the organizational and procedural processes necessary to achieve data harmonization.

### 1.3. Interest of SPC

The development of guidelines for harmonisation of census and survey development microdata, and ultimately to harmonise existing datasets, is a result of the significant progress in *ex ante* statistical collection harmonization (or standardisation) of census and survey methodologies in the Pacific region. That is, regionally standardized instruments and methodologies have been designed for the 2020 round of population and housing censuses (PHC), household income and expenditure surveys (HIES) and labour force surveys (LFS).

These harmonization activities were implemented to ensure that Pacific statistical collections meet international standards, to achieve economies of scale in SDD's support to its members, and to meet data user needs for regionally comparable statistics.

In consideration of the significant progress made in harmonizing *ex ante* statistical collection, to achieve an objective of improving accessibility and dissemination of data, these guidelines aim for *ex post* statistical collection harmonization of census and survey datasets. This involves harmonising relevant data sets to facilitate, among other things, regional reporting against SDGs indicators and comparable statistical analysis over time and across the Pacific region.

The *ex post* statistical collection harmonization initiative aim to meet a wide array of data user needs and they will provide an efficient platform to disseminate data for rapid longitudinal and spatial analysis, and to report against various indicators, including the SDGs.

### 1.4. Overview of Data Harmonization

Standardisation of census and survey methodologies and instruments

The Pacific region has made significant advancements in the standardisation of census and survey methodologies and instruments – *ex ante* statistical collection standardisation. The standardisation has largely been orchestrated by the Statistics for Development Division (SDD) of the Pacific Community (SPC), who has acted as a mediator between international standards and classifications and the social, economic and cultural context of the Pacific Island countries and territories (PICTs). Standardisation of methods and instruments enables SDD to be a more efficient statistical resource to the Pacific region as economies of scale can be achieved through the implementation of established methods and instruments. Furthermore, the production and dissemination of data produced under Pacific standardised methodologies ensures the data are:

- i. of high quality as they are produced using tried and tested methodologies;
- ii. consistent with international recommendation as SDD closely collaborates with development partners and subject matter specialists to ensure the data meet user requirements and standards;
- iii. quickly processed as established data processing systems are exploited;
- iv. quickly released as established tabulation and reporting templates are used;
- v. comparable over time, between surveys and across the Pacific region; and
- vi. efficient to analyse as data structures and variable names are standardised.

The regionally standardised census instrument was developed during a [2020 World Round of Population and Housing Census](#) meeting, which included statisticians and planners of PICTs, development partners and SPC. The meeting covered areas including: field management, core and supplementary census questionnaire, data processing, computer assisted personal interview (CAPI), cartography and uses of census data. One of the main outputs of the meeting was an agreed census instrument (core and supplementary questionnaire) that has been used in the Pacific region since 2015.

The regionally standardised HIES instrument was developed in 2011 under the guidance of a Technical Working Group (TWG), which included Pacific statisticians and development partners. The TWG held three meetings to develop various aspects of the HIES methodology. The first meeting, held in 2011, reviewed and endorsed the HIES instruments and field implementation protocols. The second meeting, held in 2013, guided the method for construction of the household consumption expenditure and income aggregates. The third, held in 2015 and inclusive of various data users, guided the HIES report structure. In 2018/19, a HIES experiment was conducted to test:

- i. whether the use of CAPI is appropriate given the complex nature of the HIES instrument and the low-electrification and low-connectivity status of many Pacific Islands;
- ii. the inclusion of complementary modules in the HIES questionnaire, such as food away from home, meal partakers, food insecurity experiences, labour market and disability; and
- iii. the optimal methodology for the collection of consumption data.

The results of the HIES experiment have guided the development of a new regional HIES methodology.

There are numerous other examples of standardised methodologies that are being implemented in the Pacific region, including:

- i. the [Multiple Indicator Cluster Survey](#) programme that is implemented by PICTs with technical contribution from UNICEF, UNFPA and SDD;
- ii. the [Demographic and Health Survey](#) programme that is implemented by PICTs with technical contribution from UNFPA and SDD;
- iii. labour market data where standardised modules have been developed for their inclusion in PHC and HIES, and Labour Force Surveys that are implemented by PICTs with technical contribution from ILO and SDD;
- iv. Disability surveys that are implemented by PICTs with technical contribution from UNICEF and SDD; and
- v. Agricultural censuses and surveys that are implemented by PICTs with technical contribution from FAO and SDD.

Standardised data capture systems have been developed for both PHC and HIES in Census and Survey Processing System ([CSPPro](#)) for paper-based surveys and [Survey Solutions](#) for CAPI-based surveys. It is noted that MICS uses a CSPPro CAPI system, however this was not developed by SDD. SDD processes PHC, HIES, LFS, disability and agriculture data using Stata while UNICEF uses SPSS for MICS.

Pacific-relevant classifications have been developed to ensure consistent classification across collections and over time. These include: i) the Pacific Standard Classification of Occupations (PACSCO) 2016; ii) the Pacific Standard Industrial Classification (PACSIC) 2014; and iii) the Pacific Classification of Individual Consumption According to Purpose (PACCOICOP) 2012. An unpublished classification of income (the Pacific Classification of Income, PACCOI) is also used in the classification of income data collected through HIES.

Generally speaking, standardisation of census and survey methodologies make for a more efficient statistical collection, from planning through to data use, and ensures the data that are produced meet user requirements for policy and planning purposes. Furthermore, ex ante statistical collection standardisation greatly facilitates, and improved the efficiency of, ex post statistical collection microdata harmonisation initiatives, such as those for which these guidelines are being written.

## 2. Designing a harmonized dataset

### 2.1. Introduction

Many organizations are involved in some form of output harmonization exercise of microdata sets with the aim of having a standardized set of data files, from existing household survey and PHC datasets, that are consistent and preferably comparable across both space (countries) and time. Amongst these organizations are:

- the World Bank (Global Consumption Database) to derive comparable consumption expenditure weights for the International Comparison Programme (ICP);
- UNICEF (MICS Tabulator) for easy analysis of MICS surveys across countries and time;
- IPUMS to have a set of data from PHCs in which the same code has the same meaning in all times and places;
- ILO (ILOSTAT) to derive comparable microdata set on labour market statistics from labour force and other related household surveys and PHCs;
- World Bank's EAP Team for Statistical Development (EAPTS) to have a consistent database for the analysis of poverty, shared prosperity, and other socio-demographic developments across countries in the East Asia and Pacific region;
- RIGA to create household-level income aggregates using a consistent methodology and surveys from more than 15 countries. In particular, RIGA-L produces comparable labour market data on wages and working time for rural wage employment across several countries.

There are two approaches used by these organizations for their harmonization processes:

- (a) One approach is based on using the original national microdata set as captured from questionnaires (raw survey data). The variables, codes and metadata are then prepared from scratch directly by the organization. So harmonization is automatic and there is no need for translation tables. (ILO, RIGA and World Bank).
  - World Bank imports household characteristics, with mapping when necessary for some characteristics, e.g. educational attainment, but directly computes consumption data values from original questionnaire values. So it is necessary to deal with imputation of missing values and treating outliers for the consumption data.
  - ILO re-processes original data (entered from the questionnaire), in consultation with national authorities, using definitions and classifications consistent with international standards. Similarly missing values are imputed by the ILO and outliers detected and treated.
  - For both World Bank and ILO values computed and disseminated could be different from those produced by national authorities. They are however closer to being comparable across space and time.
  - RIGA-L also directly computes variables relating to labour time characteristics and employment wages from raw survey data of specified countries. Again missing values and outliers are handled directly by the organization. RIGA-L imports unique household and individual identification variables which are already available in the raw survey data. These are renamed for consistency.
- (b) Another approach uses pre-processed national microdata set with their own original variables, codes and metadata. Common variables with codes and metadata are determined and then translation tables used to map each set to the integrated, common dataset. (IPUMS, MICS).
  - IPUMS receives anonymized microdata files from National Statistical Offices with data coded into a wide variety of classification schemes. Often the data are only samples from the full population data

set to preserve anonymity. Firstly the dataset is standardized through the production of a data dictionary specifying the variable names, variable codes and value labels with links to their metadata and even the census questions. Then the datasets are searched to identify variables that are 'common' across them. The harmonization is done by use of a translation table for each variable which aligns the codes from each dataset for this variable to a set of unique output codes and labels. The original codes pertaining to the variable in each dataset are identified (electronically) and then these are used manually to assign a set of unique output codes and labels for the variable. Usually a single- or double-digit code is sufficient to cover information available across all the datasets. However, sometimes this has to extend to one or two more digits to represent information available in only a subset of datasets.

- With this method, it is easy to expand the columns of the table for a new dataset or to expand the rows to create a new output code and label for a situation that did not exist previously. Changes in grouping of a continuous variable are however more challenging to handle.
  - The process adopted by UNICEF is being developed in conjunction with IPUMS using the MICS datasets. The above IPUMS approach is thus the one used.
- (c) World Bank's EAP Team for Statistical Development (EAPSD) has developed a harmonized database of socio-economic statistics constructed from microdata of household surveys across 19 countries in the East Asia and the Pacific (EAP) region. The method used for its construction is not immediately available but it is likely to be the same as that used for the World Bank Global Consumption Database. It has however produced 4 files of harmonized variables: (a) Basic information on survey methods used in surveys; (b) Standardized data on basic demographic and other socio-economic variables; (c) Standardized information on households; and (v) Poverty aggregates.

The choice for designing the harmonization datasets by SPC is second approach (b): that is pre-processed input datasets in SPC's custody will be used to generate the output harmonized datasets.

In this chapter, we will:

- i. Describe the input datasets used to form these microdata harmonisation guidelines;
- ii. Discuss how common topics among PHC and HIES were identified and how the data dictionary was produced, including identification of common variables amongst PHC and HIES datasets – these are the variables that are to be harmonised; and
- iii. Explain the construct and use of the translation tables, which provide the metadata for pre-processing input datasets to generate output harmonised datasets, as described above.

## 2.2. Input datasets

The Statistics for Development Division (SDD) of the Pacific Community (SPC) obtains datasets from countries and territories of the SPC collected through surveys and population and housing censuses. These datasets, hosted in the Pacific Microdata Library, are to be used as inputs to produce harmonized output datasets using the present guidelines. The harmonization process translates each input dataset into a harmonized output dataset which has the same structure, variables, variable names, variable labels, codes and values across all input datasets.

There are currently 85 datasets, 50 population and housing censuses and 35 household income and expenditure surveys earmarked for harmonization. The process however will be implemented in stages. In this first stage, only a core of 29 datasets are used as input datasets (see Annex 3 for details). These are the most recent datasets for each data source (HIES, PHC) and each country in scope. When it was not possible to use the most recent dataset, the next most recent useable dataset is selected. Other datasets will be added in subsequent stages.



Each dataset is fully described through a codebook which describes each variable in terms of its variable name, variable label, answer labels and codes, variable type and sometimes even location. An extract from a codebook for a PHC dataset is in Table 1. These codebooks are essential documents for the design and implementation of the harmonization process.

**Table 1: An extract from a dataset codebook**

Variable Name	Variable Label	Answer Label	Answer Code	Variable Type
ld		Open ended		Numeric
r0_hh_members		Open ended		String
r2_sex	Sex			long
		Male	1	
		Female	2	
r3_day_dob	Day of birth	Open ended		long
r3_month_dob	Month of birth	Open ended		long
r3_year_dob	Year of birth	Open ended		long
r3_age	Age	Open ended		integer
r4_rel	Relationship			long
		Head	1	
		Spouse	2	
		Biological son/daughter	3	
		Adopted son/daughter	4	
		Son in law/daughter in law	5	
		Brother/Sister	6	
		Grandchild	7	
		Parent of head	8	
		Parent of spouse	9	
		Child of spouse	10	
		Other relatives	11	
		No relation	12	

There are however instances in which codebooks do not exist or are not available. In this case, the relevant questionnaires, if available, are mined for this information. Sometimes, even this is not possible and so the input dataset cannot feature in the harmonization process. Annex 3 indicates the situation for each of the core datasets with respect to the existence, or otherwise, of codebooks.

It should be noted that each input dataset is in fact presented as sub-datasets (3 for PHC and 5 for HIES): cover (PHC and HIES), household (PHC and HIES), person (PHC and HIES), income (HIES only) and expenditure (HIES only). However, the corresponding harmonized dataset will have only 4 sub-datasets:

- 1 household sub-dataset (PHC and HIES) incorporating relevant variables from the 'cover' sub-dataset;
- 1 person sub-dataset (PHC and HIES) incorporating some of the relevant variables from the cover sub-dataset;
- 1 income sub-dataset (HIES) and 1 expenditure sub-dataset (HIES), both also incorporating some relevant variables from the cover dataset.

### 2.3. Common topics

In the first instance, a set of topics that are common across the different types of input datasets is identified for use in the harmonized output dataset. This does not require that the same topic is covered in every dataset but that it is in a sufficient number according to the type of the dataset. For example, the topic 'livelihood choices' appears in only a few person datasets and so was not selected for the harmonized person

dataset. In contrast, income is a topic in every HIES person dataset and so was identified even though it hardly appears in the PHC person dataset. The topics identified for the harmonization dataset are listed in Annex 4, along with some different types of datasets used in their identification.

## 2.4. Data dictionary

Variables are then identified relating to each of the common topics for use in the harmonized dataset. For each variable, a variable name, variable label, variable values (codes or numeric) are specified. Also relevant notes, more detailed variable description, peculiar circumstances, risks to comparability between datasets, and so on, are noted for each variable. All of these are put in the form of a 'Data dictionary' (Annex 5).

There are in fact four data dictionaries in all. The first relates to variables relevant to the household including basic characteristics, income and expenditure deciles, facilities and utilities, assets, economic activities and mortality (156 variables in total). Similarly, the person data dictionary deals with variables relating to individuals in households such as basic demographic characteristics, migration, education, health, disability, employment, economic activities and fertility (97 variables in total). The income data dictionary contains details of variables relating to individual and/or household income items (13 variables in total) whilst the expenditure data dictionary deals with expenditure items again of both individuals and households (24 variables in total). Each dictionary has at the beginning the same basic household characteristics including household identification, location, household composition and sampling details (where applicable). As much as possible consideration is given to maintaining variable names and labels that are consistent with those used in other regional or international harmonization exercises. The data dictionaries are the full and complete description of the harmonized dataset in much the same way as the codebooks describe the input datasets.

Table 2 is an extract from the data dictionary for the household basic file. The actual location details (column number, width) and type of variable (numeric/interval, string, categorical) are omitted at this stage. This extract is in fact common to all 4 data dictionaries, the full details of which are in Annex 5. The data dictionaries are living documents which may expand to take on new variables or contract to eliminate some variables, as all datasets are examined. There may also be changes in codes as the process develops.

**Table 2: An extract from a data dictionary**

Topic	Variable name	Variable label and Codes	Description & Notes
<b>HH Identification &amp; location of HH</b>	country	Country ID <i>ISO 3166 3-digit code</i>	Name of country.
	datasourc	Name of survey or PHC	Name of survey or PHC.
	year	Year of survey or PHC	Year survey or PHC conducted.
	hid	Household identifier	Unique identification of household. Use as in source dataset or construct as a concatenation of variables in source file.
	geolev1	Sub-national code level 1	Highest sub-national administrative level at which sample is representative such as island.
	geolev2	Sub-sub-national code level 2	Second highest sub-national administrative level for which sample is representative such as district.
	rururb	Area of residence <i>1 = Urban 2 = Rural</i>	Urban-rural differentiation as determined at national level.
<b>Sampling details</b>	ea	Enumeration Area	Identification of area of sampling to which household belongs.
	stratum	Stratum	Code of stratum from sample design.

<b>HH Composition</b>	psu	Primary sampling unit	An identification of the PSU to which the household belongs. Important for computation of sampling errors.
	hhwt	Household weight	Weight assigned to each household for use with household level data.
	indwt	Individual weight	Weight assigned to each unit for use in expanding to population estimates. This is the same value for all units in the same household.
	hhsiz	Household size	Number of regular members of household, excluding domestic help, paying boarders and visitors.
	hhsiz_m	Number of male HM	Total number of male household members.
	hhsiz_f	Number of female HM	Total number of female household members.

## 2.5. Translation tables

The next step in the harmonization process is to determine how variables and their values (codes and numeric) from the input dataset are best ‘transported’ to equivalent variables and values in the harmonized dataset.

For a variable with numeric values, the transportation is direct from the input dataset to the output dataset, unless the variable is in fact a grouped numeric variable. This means that for monetary numeric variables, such as income, the values in the harmonized dataset will be in the same currency as in the input dataset.

For a categorical variable, it is necessary to determine how the codes in the input dataset are to be recorded against a unique code of the harmonized dataset. This is done through what is referred to as a ‘translation table’.

There is one translation table for each categorical variable in each data dictionary. The codes for the variable are taken from the relevant data dictionary. Each code of the equivalent variable in the input dataset is taken from the codebook and a decision is made as to which code in the harmonized dataset is ‘closest’ to it. This is a judgemental call which is not always straightforward. It is made on the basis of the information about the input code from its description, the corresponding question in the questionnaire, the relevant metadata and knowledge about the variable from other sources including other input datasets. It is important always that the decision made in each case is consistent across all datasets. The relationship is many (input codes) to one (output code). Each input code should be translated to one, and only one output code but several input codes could be translated to the same output code. In some instances, even when the relationship is unique, the translation may not be exact. Any differences should be noted for the benefit of users of the harmonized database. Also when more than one input code is assigned to the same output code, the specification for each original input code should also be noted. Any other information relating to the variable, the code and the source of the information which is different from what is expected should also be noted: for example if the codes for the input dataset are taken from a questionnaire and not a codebook. All these should be included as part of the metadata for the harmonized dataset.

An example of a translation table is given in Annex 6 and the full set of translation tables are accessible upon request. The statements in the note boxes are to be used as metadata for the source, variable or code to which they are attached. For example in cell (18, Y) there are two input codes ‘5’ and ‘6’ from the input dataset (FJI, HIES, 2017) coded into the harmonized code ‘54’ in the harmonized 2017 HIES dataset for Fiji. When the input data code is ‘5’, the note explains that the output code ‘54’ in the harmonized dataset should be understood as referring to hostels. The researcher or user of the harmonized dataset can then decide whether or not to use this data point in their analysis.

For some variables, e.g. sex, it is relatively straightforward to construct the translation table. Other variables, such as materials for walls, roofs and floors, are more challenging given the limited knowledge of the local materials and the language used to describe them. A single digit is used for the harmonized codes when there



are no sub-categories. In some instances however an additional 1 or 2 digits have been added to allow for the various input sub-categories of some datasets.

Translation tables are living documents that may change on the basis of new experience. For example, the translation table for the variable 'urbrur' so far has details only for very few of the 29 core datasets. It could therefore be a candidate for dropping unless there is reason to believe that the remaining datasets or new ones are likely to have this information. Similarly, there are 3 translation tables for the fishing location reef (submerged reef, outer reef and coastal reef). However the information available from the codebooks rarely contains this detailed information. These 3 tables could in fact be collapsed into 1 for reef location with the current 3 variables treated as sub-categories. This method of constructing the translation tables makes it easy to add and delete codes or change their structure and /or to expand coverage to other datasets.

There are 98 translation tables for the household dataset, 66 for the individual dataset, 4 for the income and 8 for the expenditure datasets.

### 3. Creating a harmonized dataset

In this chapter we discuss the different stages required to create a harmonised dataset, including preparation of input datasets and use of required metadata. We also provide example of linking charts to demonstrate the interaction between the translation tables, the variables in the input dataset and the harmonised output variables.

#### 3.1. Preparatory stages

The input datasets have been used by the respective countries for analysis and preparation of their own survey/census reports. They have therefore been processed and cleaned but to varying degrees. So, prior to the start of creating harmonized datasets from them, each input dataset has to be: (i) examined to determine to what extent they have been processed; and then (ii) further processed to ensure that all the datasets are reasonably at the same level of readiness for harmonization. This examination and further processing should be for each variable in the dataset. The elements include:

1. Ascertaining that information is available for the harmonization software application to efficiently access each relevant data element in the dataset. This includes but is not limited to:
  - a. appropriate naming of the variable and specification of its values and their location in the dataset;
  - b. availability of metadata relevant to the variable itself and/or its values;
  - c. source documents relevant to the survey/census such as questionnaires, coding books, data collection and processing manuals;
  - d. information on the survey/census methods used including sample design, where relevant – in most cases, survey/census metadata is available via the [Pacific Microdata Library](#).
2. The datasets come in various application formats, e.g. Stata or SPSS, and should be converted to a common format. The pdf and other static information should be converted to XML format so that they can be easily accessed and aligned using software.
3. The variable values should as much as possible be clean and complete. This requires:
  - a. checking for and treating outliers and suspect values using an appropriate methodology (Annex 1);
  - b. separating out missing values from genuine 0 values, where appropriate, and using imputation methods for missing values, wherever necessary (Annex 1);
  - c. using a suitable process to convert interval grouped variables into individual values (Annex 2).

The aim should be to have at the end of these steps input datasets that are clean and ready for the harmonization process. It is expected that these steps would not be extensive and complicated since the datasets have already been processed at national level.

4. These data preparation activities would in the first instance be limited to the core set of datasets. However, they should be well established and documented so that they can be easily applied to the secondary datasets as well as to new ones, as these become available. Some older datasets may however fail to meet these standards and so have to be excluded from the harmonization process.

### 3.2. Metadata

As mentioned earlier, some metadata for the harmonized dataset is generated through the process itself. Some others come through the metadata of the input datasets. Key elements of the harmonization metadata are:

- information about the codes (from the translation tables),
- issues of comparability, detailed description of the variable and its codes, specification of the universe (e.g. all persons, persons above 15 years, etc.) issues that may affect comparability across space (i.e. countries) and time (from the data dictionaries),
- details from the original data generation process for the input datasets including description of source variable(s), question text in questionnaire, survey/census design, data collection, editing and data processing manuals.

These should be in a form in which they are electronically processable through use of XML-tagging and XML mark-up. This way the metadata can be easily linked to the variables and their values (including codes) and accessible to users to make intelligent choices of the data used for analysis.

### 3.3. Linking charts

Each input dataset generates a harmonized output dataset in which the codes of input categorical variables are translated into the harmonized codes through the use of the translation tables. The values of numeric and string variables of input datasets are reproduced as is in the output dataset. The structure and codes of the harmonized output dataset is the same for all input datasets. Given the share size of the input datasets, especially the PHCs, this process of going from input to output has to be done using an appropriate application software.

To aid in the development of such an application software, it would be useful to tabulate for each variable in the harmonized dataset to identify how it links up to the corresponding variable(s) in the input dataset. This can be done using linking charts. Table 6 is an extract from the table for the household sub-dataset Cook Islands 2017 HIES. It indicates the variable name and data file of the input variable(s) that should be used to populate the output variable. For a numeric or string variable, the output variable is populated directly using the values of the input variable. For example the output variable 'ea, Enumeration Area' is populated using values of the input variable 'anon\_ea\_code' in the Cover data file. For a categorical variable, values of the output variable are determined from the relevant translation table. For example, for the output variable 'dw\_type, Type of dwelling' its values are obtained using the translation table TT1 relating to values of the input variable hq20111.

An extended version of the linkage chart is given in Annex 13 for the harmonized household sub-dataset of 3 input datasets: COK, 2017 HIES; FSM, 2013 HIES, and PLW, 2015 PHC.

Table 3: An extract of a linking chart

Harmonized household dataset (COK, HIES, 2015)			Input household dataset (COK, HIES, 2015)	
Variable number	Variable name	Variable label	Data file	Variable name
H1	Country	Country ID		COK
H2	datasourc	Name of survey or PHC		HIES
H3	Year	Year of survey or PHC		2015
H4	Hid	Household identifier	HHD	id07
H5	geolev1	Sub-national code level 1	Cover	anon_island_code
H6	geolev2	Sub-sub-national code level 2	Cover	anon_district_code
H7	Rururb	Area of residence		*
H8	Ea	Enumeration Area	Cover	anon_ea_code
H9	Stratum	Stratum	Cover	strata
H10	Psu	Primary sampling unit		*
H11	Hhwt	Household weight	HHD	weight
H12	Indwt	Individual weight		*
H13	Hhsize	Household size	Cover	total
H14	hhsizem	Number of male HM	Cover	males
H15	hhsizef	Number of female HM	Cover	females
H16	pcinc_cl	Per capita annual income quintile	Cover	quint_pctot_inc
H17	inch	Total annual income	Cover	pc_tot_inc
H18	pcexp_cl	Per capita annual expenditure quintile	Cover	quint_pctot_exp
H19	Exphh	Total annual household expenditure	Cover	pc_tot_exp
H20	dw_type	Type of dwelling	HHD	TT1 - hq20111
H21	Tenure	Type of tenure	HHD	TT3
H22	Builtyr	Year of construction	HHD	hq20116
H23	rooms	Number of habitable rooms	HHD	hq20115
H24	cookarea	Cooking area	HHD	TT6 - hq20117
H25	roof	Main material used for roof	HHD	TT7 - hq20112

## 4. Using the harmonized datasets

In this chapter, we discuss ongoing demand for statistics produced from microdata, which can be met through the use of the harmonised output datasets. We also discuss international classifications that the harmonised output datasets will correspond with and some limitations in terms of the construct and use of the harmonised datasets.

### 4.1. Major regional and international data requests

The harmonized datasets will be a useful tool to respond to data requests from regional and international agencies. These requests cover data needed for SDG indicators but also for other direct needs of these

agencies. Two agencies in particular will benefit from use of these harmonized datasets: the International Labour Organization (ILO) and the United Nations Educational Scientific and Cultural Organization (UNESCO).

#### 4.1.1. Labour market data required by the ILO

The ILO through its Statistics Department requires labour market and other data and metadata of member States. The data is obtained:

- a. using an annual electronic Excel questionnaire covering 17 labour-related topics (Annex 7) including the working age population, labour force, employed population and its characteristics, labour underutilization including unemployment, industrial relations, occupational safety, labour administration and the working poor;
- b. by directly harvesting data from websites of member States and other repositories;
- c. by compiling and re-processing microdata from labour force surveys of member States to populate a harmonized database of labour market indicators.

As can be seen from Annex 7, over 70% of the required indicators in a) are obtainable from labour force surveys, PHCs and HIES. Therefore, these could be accessed from the SPC Harmonized Datasets through ILO's mining activities, or directly from SPC on request. The ILO is also the focal point (custodian) for labour-related SDG indicators (Annex 8). Of the 15 SDG indicators, 10 will be available in the SPC Harmonized Datasets.

The data obtained through a) and b) are disseminated by the ILO directly as collected, with minor editing, as national statistics. Given the above harmonization process, the data and indicators in the SPC Harmonized Datasets are also national statistics. The statistics and indicators generated from c) are disseminated as ILO statistics.

#### 4.1.2. Regular data required by UNESCO

The UIS, the Statistics Department of UNESCO, collects data from countries on literacy and on educational attainment using questionnaires. The data are disaggregated by sex, age group and urban/rural location. Data on literacy and educational attainment are available from PHC whilst most household surveys, especially HIES and LFS, have educational attainment data. The SPC harmonized database should therefore serve as a source for these statistics.

The UIS also has a set of 44 indicators on which it expects to have country data. Annex 9 lists these indicators by their likely data sources. About 40% of these indicators can be computed from PHC or household surveys. The others are sourced mainly from administrative sources of the educational system of the country including school censuses and surveys. UNESCO is also custodian for the 11 SDG global indicators (Annex 10) 4 of which could be obtainable from a PHC or household survey. Thus, to some extent, the SPC Harmonized Datasets could be a source to compute these indicators.

## 4.2. Relevant Classifications

Statistical classifications are an essential tool in the description and analysis of variables. It is important therefore that in a harmonized system, the classifications used should be standard across all the PICTs. Hence the proposal is to use regional classifications for the harmonization exercise at a level in which they are consistent with the international classifications.

International classifications that are useful for the description and analysis of statistics, including labour market, education statistics, income statistics and expenditure statistics are:

- The International Standard Industrial Classification of All Economic Activities (ISIC)
- The International Standard Classification of Occupations (ISCO)
- The International Classification of Status in Employment (ICSE)

- The International Standard Classification of Education (ISCED)
- The Classification of Individual Consumption According to Purpose (COICOP)

#### 4.2.1. *The International Standard Industrial Classification of All Economic Activities (ISIC)*

ISIC classifies the production units in an economy according to the economic activities carried out in the unit. The main aim is to produce categories of economic activities for use in the collection and reporting of statistics. The classification is based on the principal activity of the production unit, usually an establishment, determined using the main goods and services produced in the unit. Thus the classification groups establishments into detailed industries. Individuals are categorized into industries on the basis of the classification of the unit in which they work. In multi-job situations, the same individual can then have several classifications. In this case, depending on the planned use, one job has to be selected (main job) to have a unique classification for each person.

The structure of the most recent version of this classification, ISIC Rev 4<sup>1</sup>, uses a hierarchical 4-level structure consisting of:

- 21 **sections** at the highest level, labelled alphabetically (see Annex 11), then organized into successively more detailed categories, which are numerically coded:
  - 88 **divisions** (2-digit level),
    - 238 **groups** (3-digit level), and
      - 419 **classes** (4-digit level).

This structure exhaustively covers all economic activities in the economy from the detailed levels of industries (classes) and industry groups (groups) to the more aggregated levels of divisions and sections. It thus facilitates presentation and analysis at detailed levels of the economy in an internationally standardized way. The 21 sections of ISIC are sometimes further aggregated for description, analysis and presentation. They are re-grouped into 10 categories for use in connection with System of National Accounts (SNA) and into three broad areas for easier reporting and/or to avoid many cells that are empty or have only a few values (Annex 12).

Some countries still use older versions of ISIC such as Rev 3.1 but the UN Statistical Division (UNSD), the custodian of the classification, has correspondence tables to go from the older to the newest version<sup>2</sup>.

The regional equivalent developed for the Pacific region is the Pacific Standard Industrial Classification of All Economic Activities (PACSIC) 2014. It is consistent with ISIC, Rev 4, and has the same structure particularly at the section and division levels. These are usually the levels of particular interest in the harmonization process. It is thus convenient that there are no differences. The main differences between the two classifications are at the class level where:

- a. some classes in ISIC have been split into sub-classes to reflect regional priorities, and
- b. some others have been dropped as such activities do not take place in the PICTs.

Also PACSIC uses the new International Standard Classification of Education (ISCED) 2011 for the section on education instead of the old ISCED 1997 currently used in ISIC Rev 4.

#### 4.2.2. *The International Standard Classification of Occupations (ISCO)*

An occupational classification is important for the statistical description and analysis of the labour market and the social and economic structure of society.

ISCO organizes jobs into groups based on the tasks and duties usually undertaken in the job, taking into consideration the similarity of skills required for the job. The latest version (ISCO 08) was established in 2008

<sup>1</sup> Approved by the UN Statistical Commission in March 2006

<sup>2</sup> Department of Economic and Social Affairs Statistics Division, *International Standard Industrial Classification of All Economic Activities Revision 4*, United Nations, New York, 2008



by the Governing Body of the ILO and then endorsed by the International Conference of Labour Statisticians and the UN Statistical Commission also in 2008<sup>3</sup>.

In this classification,

- An occupation is defined as a “set of jobs whose main tasks and duties are characterized by a high degree of similarity”.
- Jobs are classified into an occupation group on the basis of the similarity of skills required to perform them.

The classification has a 4-level hierarchical structure as follows:

- At its most disaggregated level (4-digit level), jobs are classified into 463 **unit groups** which are then aggregated upwards into:
  - 130 **minor groups** (3-digit level). These in turn are further aggregated into
    - 43 **sub-major groups** (2-digit level) and finally into
      - 10 **major groups** (1-digit level).

The major groups are listed in Table 4.

**Table 4: International Standard Classification of Occupations (2008)**

<b>ISCO 08 (Major Groups)</b>	
<b>Code</b>	<b>Title</b>
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

Except for the last, armed forces occupations, the major groups are in descending order of skill level.

Along with the structure, ISCO-08 also includes a Dictionary of occupations and an Index of occupational titles. The latter is particularly useful to cross-walk from one classification system to another. The usual presentation of occupational statistics is at the 1-digit major group level. Statistics at a more disaggregated level such as the 2-digit level, are used in more in-depth analysis of the working force such as the analysis of occupational disaggregation.

Individuals are classified on the basis of the occupational classification of their job. In multi-job situations, the same individual can then have multiple occupational classifications. Depending on the context of the analysis, one job has to be selected (main job) to have a unique classification for each person.

The regional SPC equivalent is the Pacific Standard Classification of Occupations (PACSCO) 2016 which was developed on the basis of ISCO-08. PACSCO has an additional hierarchical level called Occupation below the ISCO-08 unit group to allow for classification at the regional level. It is at this level when, in some instances, ISCO-08 unit group categories are either expanded or contracted depending on the economic situation in PICTs. An additional difference is that PACSCO uses the new International Standard Classification of Education (ISCED) 2011 to determine skill levels instead of the old ISCED 1997 currently used in ISCO-08. These apart, at the four-digit level (unit group), PACSCO 2016 is comparable to ISCO-08 and can be easily converted to ISCO-

<sup>3</sup> ILO, ISCO-08, <https://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf>

08 by aggregating the categories at the most detailed level of PACSCO 2016 (i.e. six-digit level known as Occupation).

#### 4.2.3. *The International Classification of Status in Employment (ICSE)*

An employed person's status in employment is the nature of that person's work relationship with the economic units in which or for which the work is performed. The International Classification of Status in Employment (ICSE) assesses this in terms of (a) the authority relationships between the person and the economic unit; and (b) the economic risks faced by the person from the contractual or other conditions of the work. The most recent version (ICSE-18) was adopted by the International Conference of Labour Statisticians in November 2018<sup>4</sup>.

ICSE-18 classifies jobs in employment for pay or profit into ten detailed categories based on the concepts of type of authority and type of economic risk. They are as follows:

##### **INDEPENDENT WORKERS**

##### **A Employers**

11 – Employers in corporations

12 – Employers in household market enterprises

##### **B Independent workers without employees**

21 – Owner-operators of corporations without employees

22 – Own-account workers in household market enterprises without employees

##### **DEPENDENT WORKERS**

##### **C Dependent contractors**

30 – Dependent contractors

##### **D Employees**

41 – Permanent employees

42 – Fixed-term employees

43 – Short-term and casual employees

44 – Paid apprentices, trainees and interns

##### **E Contributing family workers**

51 – Contributing family workers

Given its adoption is recent, many countries still use the older version, ICSE-93, which classifies individuals as paid employees, employers, own-account workers, contributing family workers and members of producers cooperatives. Thus this is likely to be the one that features in the SPC harmonization system and it will be a limitation for harmonisation of future surveys/censuses as they will adopt ICSE-18.

As individuals are classified on the basis of their work relationship with the economic units in which they work, in multi-job situations, the same individual can then have multiple classifications of status in employment. Depending on the context of the analysis, one job has to be selected (main job) to have a unique classification for each person.

ICSE is used for statistical analysis of the labour market, classification of socio-economic status as well as for the analysis of statistics on wages, earnings and labour costs and on wage employment. It is also used in national accounts and economic analysis.

There is as yet no regional equivalent for ICSE, however it is noted that labour market modules in most Pacific surveys/censuses will adopt ICSE-18 from 2019 onwards.

<sup>4</sup> ILO, ICSE-18, <https://www.ilo.org/public/english/bureau/stat/icse/docs/resol.pdf>

#### 4.2.4. The International Standard Classification of Education (ISCED)

The International Standard Classification of Education (ISCED) categorizes education programmes<sup>5</sup> as well as classifies a population separately in terms of their level and field of education<sup>6</sup>. It is thus a family of three classifications. The classification according to level of education, ISCED-A, and the one according to programme, ISCED-P, were adopted formally by the General Conference of UNESCO Member States in 2011. That according to field of education, ISCED-F, was adopted in 2013. ISCED was also presented to the UN Statistical Commission as it is one of the United Nations International Family of Economic and Social Classifications.

The classification by level of education is the one that is mostly applied in household surveys and PHC. The level of education (referred to as educational attainment) is usually assessed as the highest level successfully completed by the person, as evidenced either by an appropriate qualification or being able to move to a higher level. It classifies the educational attainment of a population into 9 separate levels as follows:

- |   |                                       |
|---|---------------------------------------|
| 0 | Less than primary education           |
| 1 | Primary education                     |
| 2 | Lower secondary education             |
| 3 | Upper secondary education             |
| 4 | Post-secondary non-tertiary education |
| 5 | Short-cycle tertiary education        |
| 6 | Bachelor's or equivalent level        |
| 7 | Master's or equivalent level          |
| 8 | Doctoral or equivalent level          |
| 9 | Not elsewhere classified.             |

These levels are in ascending order from 0 to 8 based on a combination of cumulative duration and level of qualification.

Each of these levels is further sub-divided into categories and sub-categories. For categories, a distinction is made with respect to the first level (less than primary) in terms of no schooling and partial pre-primary schooling of some form. The other levels are differentiated in terms of orientation – general/academic versus vocational/professional. Sub-categories are determined in terms of the degree of completion of the level. A full list is available in the ISCED 2011 publication.

ISCED-F uses ten categories to classify fields of education at the first level, as follows:

- |    |   |
|----|---|
|    | <b>Broad field</b>                                |
| 00 | Generic programmes and qualifications             |
| 01 | Education   |
| 02 | Arts and humanities                               |
| 03 | Social sciences, journalism and information       |
| 04 | Business, administration and law                  |
| 05 | Natural sciences, mathematics and statistics      |
| 06 | Information and Communication Technologies (ICTs) |
| 07 | Engineering, manufacturing and construction       |
| 08 | Agriculture, forestry, fisheries and veterinary   |
| 09 | Health and welfare                                |
| 10 | Services  |
| 99 | Field unknown                                     |

<sup>5</sup> UNESCO Institute of Statistics, 2012, *International Standard Classification of Education ISCED 2011*, <http://www.uis.unesco.org>

<sup>6</sup> UNESCO Institute of Statistics, 2015, *INTERNATIONAL STANDARD CLASSIFICATION OF EDUCATION Fields of education and training 2013 (ISCED-F 2013)*, <http://www.uis.unesco.org>

The categories are not in any hierarchical order. They are then further sub-divided into about 56 narrow areas and a large number of detailed fields<sup>7</sup>. This is a relatively newer classification than ISCED-A and so not yet widely applied. It should however be possible to collect data for it through household surveys.

Both ISCED-A and ISCED-F are of particular relevance in the description and analysis of the labour market. Educational attainment is an important component of the skill composition of the labour force, although not the only one.

While there is as yet no regional equivalent of ISCED, a mapping of national educational programmes in Pacific Island countries to the ISCED2011 classification has been undertaken.

#### 4.2.5. *The Classification of Individual Consumption According to Purpose (COICOP)*

COICOP groups individual household consumption expenditures according to the purpose of their corresponding goods and services. Household consumption expenditure are those expenditures made for the benefit of an individual or a household. These could be incurred by three institutional sectors: households themselves, Non-profit Institutions Serving Households (NPISHs) and government. All household consumption expenditures incurred by households are individual and are made by households to satisfy their needs and wants for goods and services. All consumption expenditures of NPISHs are also individual and made to satisfy the needs and wants of individuals and households. They are referred to as transfers to households from NPISHs. However, only some of the consumption expenditures of government are considered individual, e.g. housing. They are also transfers to households from government.

COICOP classifies expenditures, which are the basic unit of classification, into a purpose classification by putting together all expenditures on goods and services that are serving a common purpose such as nourishing the body. The latest version, COICOP 2018, was considered and endorsed by the 49th Session of the United Nations Statistical Commission in March 2018. It is a hierarchical structure of

- 15 divisions (2-digit), at the highest level,
  - 63 groups (3-digit), at the next level;
    - 186 classes (4-digit), at the following level; and
    - 338 subclasses (5-digit), at the lowest level.

**Table 5: The Divisions of COICOP**

Code	Description
01	Food and non-alcoholic beverages
02	Alcoholic beverages, tobacco and narcotics
03	Clothing and footwear
04	Housing, water, electricity, gas and other fuels
05	Furnishings, household equipment and routine household maintenance
06	Health
07	Transport
08	Information and communication
09	Recreation, sport and culture
10	Education services
11	Restaurants and accommodation services
12	Insurance and financial services
13	Personal care, social protection and miscellaneous goods and services
14	Individual consumption expenditure of non-profit institutions serving households (NPISHS)
15	Individual consumption expenditure of general government

<sup>7</sup> UNESCO Institute of Statistics, 2015, *INTERNATIONAL STANDARD CLASSIFICATION OF EDUCATION Fields of education and training 2013 (ISCED-F 2013)*, Appendix 1: List of possible codes <http://www.uis.unesco.org>

These divisions cover the entire spectrum of consumption expenditures and so facilitate economic analysis.

COICOP is an essential classification for the description and analysis of:

- expenditure statistics from household income and expenditure surveys;
- living standards, etc.

It is also used in the system of national accounts, in the computation of weights for CPI and in the International Comparison Programme.

The regional SPC equivalent is the Pacific Classification of Individual Consumption According to Purpose 2012 (PACCOICOP). It was derived from COICOP and maintains the basic framework of the classification up to the subclass level. Some modifications were made to reflect regional specificity by expanding some classes and deleting those that were not relevant for the region. The classifications are consistent.

#### 4.3. Important Limitations

- Census v survey and variance
- Limited frequency of collections hinders longitudinal analysis
- Only recent ex ante standardisation limits ex post harmonisation
- Access to microdata – SPC does not have all datasets
- Scope creep where it may have been wiser to harmonise a few variables over more datasets (“in demand variables” such as those needed for SDG reporting), rather than many variables over few datasets with cases of few datasets containing the harmonised variables
- Limited ability to compare income/expenditure and requirement to convert into a PPP equivalent currency to compare income/expenditure indicators (e.g., average per capita expenditure)
- Survey methodology and implementation protocols can limit comparability
- Limited capacity of PICTs to analyse
- Reluctance of PICTs to be compared with each other (naming and shaming) and regional approaches not necessarily representing national development challenges
- International/regional statistics demand does not necessarily represent national statistics demand.

### 5. Expanding the harmonized datasets

Please include future work:

- Programming required to operationalise/automate the translation tables
- Ongoing development – adding new variables/datasets
- Start with a sub-set of themes to test the use ability of the harmonised dataset.

Explore adding other datasets (MICS, DHS, LFS).



## Annex

Annex 1\_Outliers & Missing Values

Annex 2\_Grouped Variables

Annex 3\_Input datasets

Annex 4\_Common Topics

Annex 5\_Data Dictionary

Annex 6\_Translation Table Example

Annex 7\_ILO Data Requests

Annex 8\_ILO SDGs & source

Annex 9\_UIS Data Requests

Annex 10\_UIS SDGs & Source

Annex 11\_ISIC divisions

Annex 12\_ISIC aggregated

Annex 13\_Linking table

Annex 14\_ISCED progression and mappings

## Annex 1: Identifying and Treating Outliers, Other suspect values and Missing values

### Suspect Values

These are usually inconsistent with the expected range or type of values, for example a reported age of 2 years for labour force participation or a value of 10 for the number of days in a week. They are logically incorrect. One choice is to treat the value as a missing value (see below). This is what would most likely be done in the former case. Another, used by some organizations, is to replace the value with one that is logical such as the largest/smallest possible value: in the case of the latter example, to replace the value with 7, the largest possible number of days in a week (approach taken by RIGA for a time variable). This latter approach could be premised on a wrong assumption that the correct value is the largest/smallest value. It is possible that the value is suspect due to other reasons for example having an incorrect trailing zero added.

In the absence of knowledge from other sources, the safer approach is to treat suspect values as missing (approach taken by IPUMS). This is the recommendation for the current exercise by SPC. It can be easily done

- for each numeric variable which has logical limits; and
- for each categorical variable (value outside the scope of the codes) by running a frequency check.

### Outliers - Identification

These are possibly legitimate values which however look unexpectedly too large or too small. For example, an annual income of \$100 for a professional in finance aged 43 years looks too small. Identifying outliers require the setting of an upper and/or lower limit for the value beyond which values are considered for treatment as outliers. As implied in the example given, such limits would have to take into consideration factors that influence the value of the variable such as industry, occupation, age, sex, etc. in the case of income.

Two examples that have been used elsewhere as limits are:

- 3 standard deviations either side of the mean or median value of the variable or a transformed value of the variable such as its log (used by RIGA with log income);
- The sum of the 75<sup>th</sup> percentile and 3 times the interquartile range as upper limit and the 25<sup>th</sup> percentile less 3 times the interquartile range for the lower limit (the World Bank Global Database uses the former to identify consumption expenditure that is too large at both the item and household levels).

Both these suggestions are distributional and could be used with any distribution conditioned on any factor of interest. For example the former could be determined separately for individual industry sectors whilst the latter separately for the urban and rural areas. Zeros and missing values are excluded before applying any of these methods.

The approaches are mechanistic and serve just to identify possible outliers. Indeed, since the limit is a cut-off of a distribution, for example values below 3 standard deviations in the first example, there must be values outside of it, some of which could be legitimate. Further investigation, usually manual, has to be done to confirm which ones are in fact outliers.

Either approach can be easily applied to the datasets in focus for harmonization. The choice, per variable, should depend on the nature of the variable and the shape of its distribution.

## Missing values – Identification

They are values that are supposed to exist but do not. This could be due to the fact that the value was not collected during data collection, misplaced during data processing or struck out during editing. In some instances, particularly for categorical variables, a code is assigned to missing values. The challenge is with numeric variables when zeros could be legitimate and the data capture programme could have treated blanks also as zeros. A manual examination is then necessary to separate out these two instances.

## Treating Outliers and Missing values

For numeric variables, a simple method is to replace the outlier with the median or mean value of non-outliers in the group used to identify the outlier (RIGA and World Bank Global Database). Such non-outliers are referred to as donors. For example, for an income outlier identified within the agriculture sector, its value is replaced by the median income of donors in that industry. In computing the median or mean, zeros and missing values are excluded as donors. Weights may be used.

A similar method can be used to replace a missing value for a numeric variable. It is advisable to restrict the computation to values falling within a group defined by some characteristic relevant to the variable of interest. For example use only rural consumption expenditure values to estimate a missing consumption expenditure value in the rural areas.

A more sophisticated method is to use regression of the variable of interest on key characteristics that are influential in determining the variable values (ILO). For example, a regression model of log income on industry, occupation, age and sex could be used to produce estimates of missing income values.

An important consideration when using any of these methods is the number of donor values that are available to derive the estimates. This is particularly relevant if using the mean of donor values or the regression method. When this occurs, an alternative grouping with a sufficiently large set of observations should be used, e.g. rural/urban instead of industry (RIGA).

For categorical variables, it is sometimes possible to use other information within the dataset to estimate a missing value. For example, educational attainment could be used to determine a replacement for a missing value in literacy.

In the current exercise, it may be preferable to leave it to the users of the harmonized dataset to choose their own imputation method. The information available to use any of the above methods is also available to them.

## Annex 2: Dealing with grouped numeric variables

In some data sets, a numeric variable may be available only as a grouped variable. Examples of this are age groups instead of actual ages, working time bands instead of actual hours worked and grouped income instead of actual income. These could be handled within a harmonization process if the same standard grouping is used in all datasets. For example since age groups tend to be standard 5-year or 10-year age groups, individual ages in some datasets could be grouped according to the standardized values at the preparatory stage. Also datasets use standard classifications e.g. ISCO or convertible to ISCO. There are however other grouped variables that are based on various grouping intervals or methods across datasets, e.g. income. They cannot therefore be easily standardized for input into a harmonized dataset. Some method is thus required to replace the intervals in these datasets by individual values that can then be used along with those from other datasets in a harmonized way.

One simple method of doing this is to use the mid-point of the interval as the individual values for all records in that interval (IPUMS). Another method is to assign a value that takes account of the frequency of data values in the neighbouring intervals. Thus if the length of the interval to be used to determine the assigned value is  $x$  and  $y$  is the difference between the interval and  $x$  (i.e.  $x + y = \text{interval}$ ) then the ratio of  $x$  to  $y$  is given by

$$x(|f_{i+1} - f_i|) = y(|f_i - f_{i-1}|),$$

where  $f_i$  is the frequency in the interval,  $f_{i-1}$  is the frequency in the preceding interval,  $f_{i+1}$  is the frequency of the subsequent interval, at least one of these frequencies is different from the others and  $||$  is the absolute value sign. Note that the first method (IPUMS) is a special case of this second method.

The disadvantage in both these methods is that all observations in the same interval are assigned the same value. A way around this is to randomly assign points in the interval to the observations in that interval. For example, divide the interval by the number of observations in it and then assign these points randomly as the values of the observations. This is however a more complex and complicated approach to use.

## Annex 3: Input datasets

No.	Country	Country Code	Data source	Year	Available codebook	Selected as core dataset
1	Cook Islands	COK	HIES	2015	with codebook	Substitute
2	Cook Islands	COK	PHC	2011	with codebook	Most recent
3	Federated States of Micronesia	FSM	HIES	2013	with codebook	Most recent
4	Federated States of Micronesia	FSM	PHC	2010	with codebook	Most recent
5	Fiji (Republic of)	FJI	HIES	2008	with codebook	Substitute
6	Fiji (Republic of)	FJI	PHC	2017	with codebook	Most recent
7	Kiribati	KIR	HIES	2019	with codebook	Substitute
8	Kiribati	KIR	PHC	2015	with codebook-from CAPI questionnaire	Most recent
9	Marshall Islands (Republic of)	MHL	HIES	2019	with codebook-from CAPI questionnaire	Most recent
10	Marshall Islands (Republic of)	MHL	PHC	2011	with codebook	Most recent
11	Nauri	NRU	HIES	2012	with codebook	Most recent
12	Nauri	NRU	PHC	2011	with codebook	Most recent
13	Niue	NIU	HIES	2015	with codebook	Substitute
14	Niue	NIU	PHC	2011	with codebook	Most recent
15	Palau	PLW	HIES	2014	with codebook	Most recent
16	Palau	PLW	PHC	2015	with codebook	Most recent
17	Papau New Guinea	PNG	PHC	2000	with codebook	Substitute
18	Samoa	WSM	HIES	2013	with codebook	Substitute
19	Samoa	WSM	PHC	2016	with codebook	Most recent
20	Solomon Islands	SLB	HIES	2012	with codebook	Substitute
21	Solomon Islands	SLB	PHC	2009	with codebook	Most recent
22	Tokelau	TKL	HIES	2015	with codebook	Substitute
23	Tokelau	TKL	PHC	2016	with codebook	Most recent
24	Tonga	TON	HIES	2015	with codebook	Most recent
25	Tonga	TON	PHC	2016	with codebook	Most recent
26	Tuvalu	TUV	HIES	2015	with codebook	Most recent
27	Tuvalu	TUV	PHC	2017	with codebook- from questionnaire	Most recent
28	Vanuatu (Republic of)	VUT	HIES	2019	with codebook-from CAPI questionnaire	Most recent
29	Vanuatu (Republic of)	VUT	PHC	2016	with codebook- incomplete	Most recent
Secondary datasets						
30	Cook Islands	COK	HIES	2005	no codebook	
31	Cook Islands	COK	PHC	2001	no codebook	
32	Cook Islands	COK	PHC	2006	with codebook	
33	Cook Islands	COK	PHC	2016	with codebook	
34	Federated States of Micronesia	FSM	HIES	2005	with codebook	
35	Federated States of Micronesia	FSM	PHC	2000	with codebook	
36	Fiji (Republic of)	FJI	HIES	2002	no file	
37	Fiji (Republic of)	FJI	HIES	2013	with codebook	
38	Fiji (Republic of)	FJI	HIES	2019	no codebook	
39	Fiji (Republic of)	FJI	PHC	2007	with codebook	
40	Kiribati	KIR	HIES	2006	no file	
41	Kiribati	KIR	PHC	2000		
42	Kiribati	KIR	PHC	2005		
43	Kiribati	KIR	PHC	2010		



44 Kiribati	KIR	PHC	2018
45 Marshall Islands (Republic of)	MHL	HIES	2002
46 Marshall Islands (Republic of)	MHL	PHC	1999
47 Nauri	NRU	HIES	2006
48 Nauri	NRU	PHC	2002
49 Nauri	NRU	PHC	2006
50 Niue	NIU	HIES	2002 no codebook
51 Niue	NIU	PHC	2001
52 Niue	NIU	PHC	2006
53 Niue	NIU	PHC	2017
54 Palau	PLW	HIES	2006
55 Palau	PLW	PHC	2000
56 Palau	PLW	PHC	2005
57 Palau	PLW	PHC	2012
58 Papua New Guinea	PNG	HIES	2009 no codebook
59 Papua New Guinea	PNG	PHC	2011 no codebook
60 Samoa	WSM	HIES	2002 no file
61 Samoa	WSM	HIES	2008
62 Samoa	WSM	HIES	2018
63 Samoa	WSM	PHC	2001
64 Samoa	WSM	PHC	2006
65 Samoa	WSM	PHC	2011
66 Solomon Islands	SLB	HIES	2005 no file
67 Solomon Islands	SLB	PHC	1999
68 Solomon Islands	SLB	PHC	2019
69 Tokelau	TKL	PHC	2001 no file
70 Tokelau	TKL	PHC	2006 with codebook
71 Tokelau	TKL	PHC	2011 with codebook
72 Tokelau	TKL	PHC	2013 with codebook
73 Tokelau	TKL	PHC	2019 no file
74 Tonga	TON	HIES	2000
75 Tonga	TON	HIES	2009
76 Tonga	TON	PHC	2006
77 Tonga	TON	PHC	2011
78 Tuvalu	TUV	HIES	2004
79 Tuvalu	TUV	HIES	2010
80 Tuvalu	TUV	PHC	2002
81 Tuvalu	TUV	PHC	2012
82 Vanuatu (Republic of)	VUT	HIES	2006
83 Vanuatu (Republic of)	VUT	HIES	2010
84 Vanuatu (Republic of)	VUT	PHC	1999
85 Vanuatu (Republic of)	VUT	PHC	2009

## Annex 4: Common Topics

### Identification of common topics for HIES and PHC (selected PICTs)

Included:



Excluded:



Unit	Topic	Category	HIES				PHC		
			COK-2015	FSM-2013	Tuvalu-2015	Fiji-2009	FSM-2010	Tuvalu-2017	Kirbati-2018
Household	HH Identification & location of HH	Other							
	Sampling details	Other							
	HH Composition	Demographic							
	Income & Expenditure classes	Economic							
	Dwelling characteristics - type, year	Other							
	Construction materials	Other							
	Electricity and utilities	Other							
	Sewerage and sanitation	Other							
	Housing characteristics - number of rooms, etc.	Other							
	Tenure and Rent	Other							
	Assets	Economic							
	Communication - Access to Phone & IT	Social							
	Agriculture, Fisheries, Livestock activities	Economic							
	HH Livelihood standards re choices of G & S	Economic							
	Fertility	Demographic							
	Mortality	Demographic							
	Family labour force status	Economic							
	Location of family members and remittances	Economic							
	Natural disasters	Social							
	Household Income	Economic							
Person	Person Id - HH Member id, EA and municipality	Other							
	Sampling details	Other							
	Basic demographics - age, sex, marital status, relationship, residence etc.	Demographic							
	Family structure	Demographic							
	Fertility and mortality	Demographic							
	Ethnicity, citizenship and nationality	Social							
	Migration	Demographic							
	Employment	Economic							
	Other work	Social							
	Education	Social							
	Literacy	Social							
	Health inc. anthropometrics	Social							
	Disability	Social							
	Communication status - IT and phone use	Social							

[illegible]

Household Basic file	Topic	Variable name	Variable label and Codes	Description & Notes
	<b>HH Identification &amp; location of HH</b>	country	Country ID <i>ISO 3166 3-digit code</i>	Name of country.
		datasourc	Name of survey or PHC	Name of survey or PHC.
		year	Year of survey or PHC	Year survey or PHC conducted.
		hid	Household identifier	Unique identification of household. Use as in source dataset or construct as a concatenation of variables in source file.
		geolev1	Sub-national code level 1	Highest sub-national administrative level at which sample is representative such as island.
		geolev2	Sub-sub-national code level 2	Second highest sub-national administrative level for which sample is representative such as district.
	<b>Sampling details</b>	rururb	Area of residence <i>1 = Urban                      2 = Rural</i>	Urban-rural differentiation as determined at national level.
		ea	Enumeration Area	Identification of area of sampling to which household belongs.
		stratum	Stratum	Code of stratum from sample design.
		psu	Primary sampling unit	An identification of the PSU to which the household belongs. Important for computation of sampling errors.
		hhwt	Household weight	Weight assigned to each household for use with household level data.
		indwt	Individual weight	Weight assigned to each unit for use in expanding to population estimates. This is the same value for all units in the same household.
	<b>HH Composition</b>	hhsiz	Household size	Number of regular members of household, excluding domestic help, paying boarders and visitors.
		hhsiz_m	Number of male HM	Total number of male household members.
		hhsiz_f	Number of female HM	Total number of female household members.
	<b>Income classes</b>	pcinc_cl	Per capita annual income quintile <i>1 = 1st quintile                      2 = 2nd quintile</i> <i>3 = 3rd quintile                      4 = 4th quintile</i> <i>5 = 5th quintile</i>	Per capita annual household income quintile (local currency).
	<b>Income amount</b>	inchh	Total annual income	Total annual household income from all sources and all HMs (local currency).
	<b>Expenditure classes</b>	pcexp_cl	Per capita annual expenditure quintile <i>1 = 1st quintile                      2 = 2nd quintile</i> <i>3 = 3rd quintile                      4 = 4th quintile</i> <i>5 = 5th quintile</i>	Per capita annual household expenditure quintile (local currency).
	<b>Expenditure amount</b>	exphh	Total annual household expenditure	Total annual household expenditure (local currency) including purchased, home produced and received items.

Standardized Household File	Topic	Variable name	Variable label and Codes	Description & Notes
	<b><i>Dwelling characteristics</i></b>	dw_type	Type of dwelling <i>10 = One family house detached from any other house</i> <i>20 = One family house attached to one or more houses</i> <i>30 = Building with 2 or more apartments</i> <i>31 = Sharing kitchen/toilet</i> <i>40 = Building attached to a business / other non-resident building</i> <i>50 = Institutional building</i> <i>52 = Hospital</i> <i>90 = Other</i> <i>99 = Missing</i>	Type of household dwelling unit or living quarters.
		tenure	Type of tenure <i>1 = Renting the dwelling</i> <i>2 = Living rent free by employer</i> <i>3 = Living rent free by others</i> <i>4 = Own the dwelling without mortgage</i> <i>5 = Own the dwelling with a mortgage</i>	
		builtyr	Year of construction	Year dwelling was constructed. Some datasets may present this only as a grouped variable. For example: <i>1 = Older than 1970, 2 = Between 1970 and 1989, 3 = Recent: 1990-2014; and 4 = Not stated</i> . The issue is complicated by the fact that countries may use different groupings when doing this. It may then be necessary to use some imputation method to convert these intervals into values for use in the harmonized dataset. The simplest method is to use the mid-points of the intervals.
		rooms	Number of habitable rooms	Number of rooms in the whole household dwelling unit which may consist of one or more structures, excluding bathrooms, stores and kitchens.
		cookarea	Cooking area <i>1 = Yes, inside &amp; outside the house</i> <i>2 = Yes, inside only</i> <i>3 = Kitchen outside the house</i> <i>4 = No designated cooking area</i>	This is the place where the household do their regular cooking.



<b>Construction materials</b>	roof	<p>Main material used for roof</p> <p>10 = Earth, Clay, Mud, Wattle  20 = Thatch, Bamboo  40 = Tin, Metal sheets  60 = Cement, Concrete, Stone  70 = Leaves, Grass, Plastic sheet, salvaged materials  90 = Other</p>	<p>30 = Wood  50 = Tiles, Bricks, Shingles</p>	The second digit can be used to further differentiate between the 1st digit classes or bring in additional categories according to national circumstances.
	wall	<p>Main material used for external walls</p> <p>10 = Clay, Mud, Wattle  30 = Wood  50 = Cement, Concrete, Stone  60 = Brick  90 = Other</p>	<p>20 = Bamboo  40 = Tin, Metal sheets</p>	The second digit can be used to further differentiate between the 1st digit classes or bring in additional categories according to national circumstances.
	floor	<p>Main material used for floor</p> <p>10 = Clay, Mud, Earth  30 = Wood  50 = Tiles  70 = Vinyl</p>	<p>20 = Bamboo  40 = Cement  60 = Brick  90 = Other</p>	The second digit can be used to further differentiate between the 1st digit classes or bring in additional categories according to national circumstances.
<b>Electricity and utilities</b>	watersrc	<p>Main source of water</p> <p>1 = Piped (own tap)  3 = Protected well  5 = Surface water  7 = Vendor, Truck  9 = Other  11 = Missing</p>	<p>2 = Public standpipe  4 = Unprotected well  6 = Rain water  8 = Bottled water  10 = Don't know</p>	<p>Drinking water, if specified. 'Piped': to tap in house or compound. 'Public' to tap outside compound. 'Wells': include boreholes, springs. 'Surface water': rivers, lakes, ponds. We can distinguish between improved and not-improved water sources as follows: Codes 1, 2, 3 and 8 correspond to 'improved'. Codes 4, 5, 6, 7 &amp; 9 are 'not-improved'.</p>
	lighting	<p>Main source of lighting in dwelling</p> <p>10 = Electricity  12 = private  14 = private generator  16 = other electric  20 = No, no electricity  22 = Battery lamp  98 = Don't know</p>	<p>11 public/quasi public  13 = private solar/wind  15 = private company  Text = Other note  21 = Kerosine lamp  23 = Other type  99 = Missing</p>	Both electric and non-electric. Does not reflect actual electrical service received by household. Only that there is a connection to the dwelling.
	cookenerg	<p>Main cooking energy</p> <p>1 = Firewood  3 = Charcoal  5 = Gas</p>	<p>2 = Kerosine  4 = Electricity  9 = Other</p>	'Electricity': mains, generator, solar; 'Other': fuel from coffee waste, saw dust, crop residue, cow dung, etc.

<b>Sewerage and sanitation</b>	toilet	<p>Main toilet facility</p> <p>10 = Flush toilet                      11 = public sewer</p> <p>12 = septic tank                      13 = pour-flush</p> <p>20 = Improved pit latrine            21 = ventilated</p> <p>30 = Pit latrine                      31 = pit</p> <p>32 = bucket                          33 = public</p> <p>34 = open                              40 = Ocean, etc.</p> <p>50 = No facility                      90 = Other</p> <p>text = other (note)                  98 = Don't know</p> <p>99 = Missing</p>	<p>Flush toilet': flush to main sewer or septic tank;'Improved pit latrine': includes ventilated pits, suilabh or double vault compost latrine, slab toilet; 'No facility': open fields, bush, seaside, streams, mangrove; 'Other': bucket, pan etc.</p> <p>We can distinguish between improved and not-improved toilet as follows: Improved toilet = 10 - 21; Not-improved toilet = 30 -90.</p>
<b>Phone &amp; IT Connectivity</b>	phone	<p>Access to a landline phone</p> <p>10 = Yes                                  11 = Yes, functioning</p> <p>20 = No                                  98 = Don't know</p> <p>99 = Missing</p>	<p>Presence of a fixed landline telephone in house. Some countries impose the requirement of 'functionning', which may thus affect comparability.</p>
	cell	<p>Access to a cell phone</p> <p>10 = Yes, Own                      11 = Yes, signal exists</p> <p>20 = No                                  98 = Don't know</p> <p>99 = Missing</p>	<p>Presence of a cell or mobile phone or signal in house.</p>
	internet	<p>Access to internet</p> <p>10 = Yes, access                      11 = physical line</p> <p>12 = satellite                      13 = public WIFI</p> <p>14 = mobile phone                  15 = other access</p> <p>20 = No access                      98 = Don't know</p> <p>99 = Missing</p>	<p>Access within or outside the house.</p>
	cabletv	<p>Access to cable TV service</p> <p>10 = Yes                                  11 = Yes, satellite</p> <p>12 = Yes, cable                      20 = No</p> <p>98 = Don't know                      99 = Missing</p>	<p>Access within or outside the house.</p>
<b>Mortality</b>	anymort	<p>Any HM died last 3 years</p> <p>1 = Yes                                  2 = No</p> <p>8 = Don't know                      9 = Missing</p>	
	mortnum	<p>Number HM died last 3 years</p>	<p>Only if death = 1. Provision made for up to 3.</p>
	sexmort1	<p>Sex of 1st dead person</p> <p>1 = Male                                  2 = Female</p> <p>8 = Don't know                      9 = Missing</p>	<p>Sex of first dead person.</p>
	agemort1	<p>Age of 1st dead person</p>	<p>Age in completed years at death.</p>
	cod_1	<p>Cause of death - 1st person</p> <p>10 = Illness                              20 = Accident</p> <p>30 = Other                              98 = Don't know</p> <p>99 = Missing</p>	<p>Main reason person died, if known. The second digit of the code may be used for additional categories according to national standards.</p>

reprmort1	Reproductive status of 1st person at time of death <i>1 = Within 6 weeks of pregnancy</i> <i>2 = Later in pregnancy</i> <i>3 = Within 6 weeks of giving birth</i> <i>4 = Giving birth</i> <i>5 = Other</i> <i>8 = Don't know</i> <i>9 = Missing</i>	Only if female (sexmort1 = 2) between ages of 15 and 50 (agemort1 = 15 or more but less than 50).
sexmort2	Sex of 2nd dead person <i>1 = Male</i> <i>2 = Female</i> <i>8 = Don't know</i> <i>9 = Missing</i>	Sex of second dead person.
agemort2	Age of 2nd dead person	Age in completed years at death.
cod_2	Cause of death - 2nd person <i>10 = Illness</i> <i>20 = Accident</i> <i>30 = Other</i> <i>98 = Don't know</i> <i>99 = Missing</i>	Main reason person died, if known. The second digit of the code may be used for additional categories according to national standards.
reprmort2	Reproductive status of 2nd person at time of death <i>1 = Within 6 weeks of pregnancy</i> <i>2 = Later in pregnancy</i> <i>3 = Within 6 weeks of giving birth</i> <i>4 = Giving birth</i> <i>5 = Other</i> <i>8 = Don't know</i> <i>9 = Missing</i>	Only if female (sexmort2 = 2) between ages of 15 and 50 (agemort2 = 15 or more but less than 50).
sexmort3	Sex of 3rd dead person <i>1 = Male</i> <i>2 = Female</i> <i>8 = Don't know</i> <i>9 = Missing</i>	Sex of third dead person.
agemort3	Age of 3rd dead person	Age in completed years at death.
cod_3	Cause of death - 3rd person <i>10 = Illness</i> <i>20 = Accident</i> <i>30 = Other</i> <i>98 = Don't know</i> <i>99 = Missing</i>	Main reason person died, if known. The second digit of the code may be used for additional categories according to national standards.
reprmort3	Reproductive status of 3rd person at time of death <i>1 = Within 6 weeks of pregnancy</i> <i>2 = Later in pregnancy</i> <i>3 = Within 6 weeks of giving birth</i> <i>4 = Giving birth</i> <i>5 = Other</i> <i>8 = Don't know</i> <i>9 = Missing</i>	Only if female (sexmort3 = 2) between ages of 15 and 50 (agemort3 = 15 or more but less than 50).

Household assets Topic  
file

Variable name Variable label and Codes

Description & Notes

<b>Assets (Furniture)</b>	beds	Ownership of beds 1 = Yes 2 = No	Including mattresses and bedroom sets (headboards, frames, etc.).
	chairs	Ownership of chairs 1 = Yes 2 = No	Including sofas, lounge chairs, sofa sets, couches.
	tables	Ownership of tables 1 = Yes 2 = No	Including dining & coffee tables, desks, etc.
	lamps	Ownership of lamps 1 = Yes 2 = No	Including floor and wall lamps, fittings and hanging lights.
	oth_ind	Ownership of other indoor furniture 1 = Yes 2 = No	Significant other indoor furniture including book shelves and stands.
	outfurn	Ownership of outdoor furniture 1 = Yes 2 = No	Including deck furniture, folding chairs/tables, etc.
	flcover	Ownership of floor coverings 1 = Yes 2 = No	Including carpets, rugs, mats, local mats, etc.
	othfurn	Ownership of any other furniture 1 = Yes 2 = No	Any other indoor, outdoor, furnishings, floor coverings.
<b>Assets (Textiles)</b>	sheets	Ownership of sheets 1 = Yes 8 = Don't know 2 = No 9 = Missing	Including blankets, pillow cases, pillows, etc.
	curtain	Ownership of curtains 1 = Yes 8 = Don't know 2 = No 9 = Missing	Including drapes, mini blinds, etc.
	towels	Ownership of towels 1 = Yes 8 = Don't know 2 = No 9 = Missing	Including bath, hand, tea towels, etc.
	othertext	Ownership of other textiles 1 = Yes 8 = Don't know 2 = No 9 = Missing	Including table cloth, table covers, tarps, etc.
<b>Assets (Household appliances)</b>	watertanks	Ownership of water tanks 1 = Yes 8 = Don't know 2 = No 9 = Missing	Presence of a water tank. Some countries impose the requirement of functioning, which may thus affect comparability.
	refrig	Ownership of refrigerator 1 = Yes 8 = Don't know 2 = No 9 = Missing	Presence of a refrigerator or freezer in house. Some countries impose the requirement of functioning, which may thus affect comparability.
	stove	Ownership of stove 11 = Yes, electric 13 = Yes, kerosine 12 = Yes, gas 20 = No	Presence of a stove in the house. Some countries impose the requirement of functioning, which may thus affect comparability.

burner	Ownership of gas burner 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of a gas burner or cylinder in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
microwave	Ownership of microwave oven 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of a microwave oven in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
washer	Ownership of washing machine 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of a washing machine in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
dryer	Ownership of clothes dryer 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of a clothes dryer in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
aircon	Ownership of air conditioner 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of a air conditioner in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
generator	Ownership of generator 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of a generator in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
solar	Ownership of solar power unit 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of a solar power unit in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
hotwater	Ownership of water heater 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of a water heater in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
othappl	Ownership of other major appliances 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of other major appliances in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
smallelec	Ownership of small electrical appliances 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of small appliances such as rice cooker, vacuum, sewing machine, blender, etc. in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
tv	Ownership of television 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of television in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
radio	Ownership of radio 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of radio, stereos, radio cassette, etc. in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
video	Ownership of video 1 = Yes 8 = Don't know	2 = No 9 = Missing	Presence of video & DVD player in the house. Some countries impose the requirement of functioning, which may thus affect comparability.

	oth_aud	Ownership of other audio devices 1 = Yes 2 = No 8 = Don't know 9 = Missing	Presence of iPod, MP3 player, etc in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
	consoles	Ownership of game consoles 1 = Yes 2 = No 8 = Don't know 9 = Missing	Presence of PlayStation, Nintendo, Xbox, PSP, DS, etc. in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
	photo	Ownership of photo equipment 1 = Yes 2 = No 8 = Don't know 9 = Missing	Presence of still or video cameras. Some countries impose the requirement of functioning, which may thus affect comparability.
	sports	Ownership of sports & camping equipment 1 = Yes 2 = No 8 = Don't know 9 = Missing	Ownership of sports and camping equipment such as tents, basketball, etc.
	o Equip	Ownership of other recreational equip 1 = Yes 2 = No 8 = Don't know 9 = Missing	Ownership of other recreational equipment but excluding boats.
	computer	Ownership of computers 1 = Yes 2 = No 8 = Don't know 9 = Missing	Presence of a computer (desktop, laptop, tablet, etc.) in the house which is not for commercial use. Some countries impose the requirement of functioning, which may thus affect comparability.
	printer	Ownership of printer 1 = Yes 2 = No 8 = Don't know 9 = Missing	Presence of printer and printer supplies in the house. Some countries impose the requirement of functioning, which may thus affect comparability.
	softpkg	Ownership of software packages 1 = Yes 2 = No 8 = Don't know 9 = Missing	Excluding games for consoles.
	port_dr	Ownership of portable media drives 1 = Yes 2 = No 8 = Don't know 9 = Missing	Such as flash drive, mini hard drive, etc.
	otherpceq	Ownership of other computer equipment 1 = Yes 2 = No 8 = Don't know 9 = Missing	Such as scanner, mouse, speaker.
<b>Assets (Transport)</b>	car	Number of cars	Number of cars or station wagons owned by one or more household members. Value is 0 if none.
	usv	Number of utility vehicles	Number of utility vehicles or pick-up trucks owned by one or more household members. Value is 0 if none.
	truk_bus	Number of trucks or buses	Number of trucks, buses or vans owned by one or more household members. Value is 0 if none.
	mcycle	Number of motorcycles	Number of motorcycles, quad-cycle, scooters, etc. owned by one or more household members. Value is 0 if none.
	bcycle	Number of bicycles	Number of bicycles or any other pedal powered vehicles owned by one or more household members. Value is 0 if none.



**Household  
Agricultural file**

**Topic**

mboat	Number of motor boats
othboat	Number of other boats
othveh	Number of other vehicles
out_brd	Number of outboard motors
trailer	Number of trailers

Number of boats with a motor owned by one or more household members. Value is 0 if none.

Number of other boats without a motor owned by one or more household members. Value is 0 if none.

Number of other vehicles not mentioned above owned by one or more household members. Value is 0 if none.

Number of outboard motors owned by one or more household members. Value is 0 if none.

Number of trailers owned by one or more household members. Value is

**Description & Notes**

***Agriculture***

hmag	HH members involved in agriculture <i>List of HH member ID</i>
nohmag	Number of hh members involved in agriculture
paidhpag	Number of paid help in agriculture
landtyp	Land type <i>1 = Custom ownership      2 = Own land</i> <i>3 = Rent or lease land      4 = Other</i> <i>8 = Don't know      9 = Missing</i>
parcel own	Number of parcels own
parcel size	Size of parcels
extvisit	Number of extension visits
coopmemb	Participation as cooperative member <i>1 = Yes      2 = No</i> <i>8 = Don't know      9 = Missing</i>
credit	Microcredit use <i>1 = Yes      2 = No</i> <i>8 = Don't know      9 = Missing</i>

HM aged 13 and above involved in agricultural activities in the household (start with owner first).

Number of HM aged 13 and above involved in agricultural activities in the household.

Number of paid helpers involved in agricultural activities.

Ownership of agricultural land as defined in country. May include arable land (under temporary crops, temporary meadows, market or kitchen gardens, temporarily fallow), land under permanent crops and land under permanent pastures.

The number of parcels owned by HH. If none, put 0.

Area of parcels (in acres)

The number of visits received in last 3 months.

Any HM participate in producers' cooperative.

Use of microcredit to develop agricultural activities.

***Livestock***

hmlv	HH members involved in livestock <i>List of HH member ID</i>
nohmlv	Number of hh members involved in livestock
paidhplv	Number of paid help in livestock

HM aged 13 and above involved in livestock activities in the household (start with owner first).

Number of HM aged 13 and above involved in livestock activities in the household.

Number of paid helpers involved in livestock activities.

pigs	Pig number	Number of pigs owned by household. If none, value is 0. Some datasets may present this only as a grouped variable. For example: 1 = 1 or 2 pigs; 2 = 3 to 7 pigs; 3 = 8 to 16 pigs; 4 = 17 or more pigs; 0 = None. The issue is complicated by the fact that countries may use different groupings when doing this. It may then be necessary to use some imputation method to convert these intervals into values for use in the harmonized dataset. The simplest method is to use the mid-points of the intervals.
chicken	Chicken number	Number of chickens owned by household. If none, value is 0. Some datasets may present this only as a grouped variable. For example: 1 = 1 or 2 chickens; 2 = 3 to 9 chickens; 3 = 10 to 24 chickens; 4 = 25 or more chickens; 0 = None. The issue is complicated by the fact that countries may use different groupings when doing this. It may then be necessary to use some imputation method to convert these intervals into values for use in the harmonized dataset. The simplest method is to use the mid-points of the intervals.
ducks	Ducks number	Number of ducks owned by household. If none, value is 0. Some datasets may present this only as a grouped variable. For example: 1 = 1 or 2 ducks; 2 = 3 to 9 ducks; 3 = 10 ducks or more; 0 = None. The issue is complicated by the fact that countries may use different groupings when doing this. It may then be necessary to use some imputation method to convert these intervals into values for use in the harmonized dataset. The simplest method is to use the mid-points of the intervals.
cows	Cow number	Number of cows owned by household. If none, value is 0. Some datasets may present this only as a grouped variable. For example: 1 = 1 or 2 cows; 2 = 3 to 9 cows; 3 = 10 cows or more; 0 = None. The issue is complicated by the fact that countries may use different groupings when doing this. It may then be necessary to use some imputation method to convert these intervals into values for use in the harmonized dataset. The simplest method is to use the mid-points of the intervals.
horses	Horse number	Number of horses owned by household. If none, value is 0. Some datasets may present this only as a grouped variable. For example: 1 = 1 horse; 2 = 2 to 4 horses; 3 = 5 or more horses; 0 = None. The issue is complicated by the fact that countries may use different groupings when doing this. It may then be necessary to use some imputation method to convert these intervals into values for use in the harmonized dataset. The simplest method is to use the mid-points of the intervals.

# **Aquaculture**

dogs	Dog number	Number of dogs owned by household. If none, value is 0. Some datasets may present this only as a grouped variable. For example:1 = 1 dog; 2 = 2 to 4 dogs; 3 = 5 or more dogs; 0 = None. The issue is complicated by the fact that countries may use different groupings when doing this. It may then be necessary to use some imputation method to convert these intervals into values for use in the harmonized dataset. The simplest method is to use the mid-points of the intervals.
goats	Goat number	Number of goats owned by household. If none, value is 0. Some datasets may present this only as a grouped variable. For example:1 = 1 goat; 2 = 2 to 4 goats; 3 = 5 or more goats; 0 = None. The issue is complicated by the fact that countries may use different groupings when doing this. It may then be necessary to use some imputation method to convert these intervals into values for use in the harmonized dataset. The simplest method is to use the mid-points of the intervals.
othlv	Number of other livestock	Number of other livestock owned by household. If none, value is 0.
hmaq	HH members involved in aquaculture <i>List of HH member ID</i>	HM aged 13 and above involved in aquaculture activities in the household (start with owner first).
nohmaq	Number of hh members involved in aquaculture	Number of HM aged 13 and above involved in aquaculture activities in the household.
paidhpaq	Number of paid help in aquaculture	Number of paid helpers involved in aquaculture activities.
giantclam	Ownership & number of of giant clam <i>1 = Yes                      Actual = Number</i> <i>2 = No                        8 = Don't know</i> <i>9 = Missing</i>	
crab	Ownership & number of of crab <i>1 = Yes                      Actual = Number</i> <i>2 = No                        8 = Don't know</i> <i>9 = Missing</i>	
milkfish	Ownership & number of of milkfish <i>1 = Yes                      Actual = Number</i> <i>2 = No                        8 = Don't know</i> <i>9 = Missing</i>	
eel	Ownership & number of of eel <i>1 = Yes                      Actual = Number</i> <i>2 = No                        8 = Don't know</i> <i>9 = Missing</i>	
seacucum	Ownership & number of sea cucumber <i>1 = Yes                      Actual = Number</i> <i>2 = No                        8 = Don't know</i> <i>9 = Missing</i>	

	lobster	Ownership & number of lobster 1 = Yes 2 = No 9 = Missing	Actual = Number 8 = Don't know	
	trochus	Ownership & number of trochus 1 = Yes 2 = No 9 = Missing	Actual = Number 8 = Don't know	
	reef fish	Ownership & number of reef fish 1 = Yes 2 = No 9 = Missing	Actual = Number 8 = Don't know	
	tuna	Ownership & number of tuna 1 = Yes 2 = No 9 = Missing	Actual = Number 8 = Don't know	
	shellfish	Ownership & number of shellfish 1 = Yes 2 = No 9 = Missing	Actual = Number 8 = Don't know	
	freshwater fish	Ownership & number of freshwater fish 1 = Yes 2 = No 9 = Missing	Actual = Number 8 = Don't know	
	othaq	Ownership & number of other aqua 1 = Yes 2 = No 9 = Missing	Actual = Number 8 = Don't know	
	hmfsh	HH members involved in fisheries List of HH member ID		HM aged 13 and above involved in fisheries activities in the household (start with owner first).
	nohmfsh	Number of hh members involved in fisheries		Number of HM aged 13 and above involved in fisheries activities in the household.
	paidhpfsh	Number of paid help in fisheries		Number of paid helpers involved in fisheries activities.
<b>Fisheries</b>	trolling	Trolling 1 = Yes 8 = Don't know	2 = No 9 = Missing	Fishing method used by household members.
	bottom	Bottom fishing 1 = Yes 8 = Don't know	2 = No 9 = Missing	Fishing method used by household members.

casting	Casting <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
spearfsh	Spear fishing <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
netfsh	Net fishing <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
gather	Gathering <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
trapping	Trapping <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
hook&line	Hook & line <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Including handline, vertical long line and drop line.
othfsh	Other fishing method <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
oceanloc	Open ocean <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
FADloc	FAD location <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
subreef	Submerged reef <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
outreef	Outer reef <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
costreef	Coastal reef <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.
lagoon	Lagoon <i>1 = Yes</i> <i>8 = Don't know</i>	<i>2 = No</i> <i>9 = Missing</i>	Fishing method used by household members.

mangrove	Mangrove 1 = Yes 8 = Don't know 2 = No 9 = Missing	Fishing method used by household members.
river	River 1 = Yes 8 = Don't know 2 = No 9 = Missing	Fishing method used by household members.
othloc	Other fishing locations 1 = Yes 8 = Don't know 2 = No 9 = Missing	Such as Fishing: outside or over reef; Gathering: outside or over reef; Fishing: freshwater; Gathering: freshwater; Fishing: combination of locations; Gathering: combination of locations; fishing other; gathering other.
notrips	Number of fishing trips	Average number of times members of this household fish in a typical month.
nohours	Number of hours per trip	Average number of hours members of the household fish per trip.
travmode	Main mode of travel 1 = Motorised boat 3 = Walk 5 = Other 9 = Missing 2 = Non-Motorised boat 4 = Swim 8 = Don't know	Main mode of transport to fishing site.

## Annex 6: Translation Table Example

[illegible]





WSM PHC 2011	WSM PHC 2016	SLB HIES 2012	TKL HIES 2015	TON HIES 2015	TON PHC 2016	TUV HIES 2015	VUT HIES 2019	VUT PHC 2016	NIU PHC 2017
	*	h1421	hc_q20111	hq20111	g1a_building	hq20111	h801	*	h9_building
		1	1	1	1	1	1		1
		2	2	2	2	2	2		2,3
		3	3	3	3	3	3		4
			4	4		4	4		
		4	5	5	4-5	5	6		5
									6
		5	6	6	8	6	7		8

## Annex 7: Topics and sources – ILO data requests

No	Topics	Preferred source	Other sources
1	Working-age population	PHC	LFS, HIES
2	Labour force	LFS	PHC, HIES
3	Employment	LFS	PHC, HIES
4	Informal employment	LFS	HIES
5	Time-related underemployment	LFS	HIES
6	Employees	LFS	PHC, HIES
7	Public employment	LFS	HIES, ES, Admin
8	Unemployment	LFS	PHC, HIES
9	Other labour underutilization	LFS	HIES
10	Hours of work	LFS	HIES
11	Employment income	LFS	HIES, Admin
12	Labour costs	ES	National accounts
13	Industrial relations	Admin	LFS
14	Occupational injuries	Admin	LFS
15	Strikes	Admin	
16	Labour administration	Admin	
17	Working poor	HIES	

## Annex 8: ILO SDG Indicators and sources

Indicator ID	Indicator description	Preferred source	Secondary source
1.1.1	Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)	Poverty Survey (PS)	HIES
1.2.1	Proportion of population living below the national poverty line, by sex and age	PS	HIES
1.2.2	Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	PS	HIES
1.3.1	Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable	Admin	
5.4.1	Proportion of time spent on unpaid domestic and care work, by sex, age and location	TUS	LFS, HIES
5.5.2	Proportion of women in managerial positions	LFS	PHC, HIES
8.2.1	Annual growth rate of real GDP per employed person	Admin	National accounts
8.3.1	Proportion of informal employment in non-agriculture employment, by sex	LFS	HIES
8.5.1	Average hourly earnings of female and male employees, by occupation, age and persons with disabilities	LFS	HIES
8.5.2	Unemployment rate, by sex, age and persons with disabilities	LFS	PHC, HIES
8.6.1	Proportion of youth (aged 15–24 years) not in education, employment or training	LFS	HIES
8.7.1	Proportion and number of children aged 5–17 years engaged in child labour, by sex and age	CLS	LFS
8.8.1	Frequency rates of fatal and non-fatal occupational injuries, by sex and migrant status	Admin	LFS
8.8.2	Level of national compliance with labour rights (freedom of association and collective bargaining) based on International Labour Organization (ILO) textual sources and national legislation, by sex and migrant status	Admin	
8.b.1	Existence of a developed and operationalized national strategy for youth employment, as a distinct strategy or as part of a national employment strategy	Admin	

## Annex 9: List of UIS Indicators and sources

Indicator ID	Education indicators	Preferred Source	Alternative Sources
1	ADULT LITERACY OR ILLITERACY RATE	PHC	LFS
2	GROSS INTAKE RATIO (GIR) IN THE FIRST GRADE OF PRIMARY	Admin	PHC
3	NET INTAKE RATE (NIR) IN THE FIRST GRADE OF PRIMARY	Admin	PHC
4	SCHOOL-LIFE EXPECTANCY (SLE)	Admin	PHC
5	AGE SPECIFIC ENROLMENT RATE (ASER)	Admin	PHC
6	EDUCATIONAL ATTAINMENT OF THE POPULATION AGED 25 YEARS AND ABOVE	PHC	HIES & LFS
7	YOUTH LITERACY RATE	PHC	HIES & LFS
8	NUMBER OF ADULT ILLITERATES		
9	TRANSITION RATE (TR)	Admin	
10	GROSS ENROLMENT RATIO (GER)	Admin	
11	NET ENROLMENT RATE (NER)	Admin	
12	REPETITION RATE BY GRADE (RR)	Admin	
13	SURVIVAL RATE BY GRADE (SR)	Admin	
14	COEFFICIENT OF EFFICIENCY	Admin	
15	YEARS-INPUT PER GRADUATE	Admin	
16	PERCENTAGE OF REPEATERS	Admin	
17	PUBLIC EXPENDITURE ON EDUCATION AS % OF GROSS NATIONAL INCOME	Admin	
18	PUBLIC EXPENDITURE ON EDUCATION AS % OF TOTAL GOVERNMENT EXPENDITURE	Admin	
19	PERCENTAGE DISTRIBUTION OF PUBLIC CURRENT EXPENDITURE ON EDUCATION BY LEVEL	Admin	
20	PUBLIC CURRENT EXPENDITURE PER PUPIL (STUDENT) AS % OF GROSS NATIONAL INCOME (GNI) PER CAPITA	Admin	
21	PUPIL-TEACHER RATIO (PTR)	Admin	
22	PERCENTAGE OF FEMALE TEACHERS	Admin	
23	PERCENTAGE DISTRIBUTION OF STUDENTS IN TERTIARY EDUCATION BY ISCED LEVEL	Admin	
24	PERCENTAGE OF FEMALE STUDENTS IN EACH ISCED LEVEL OF TERTIARY EDUCATION	Admin	
25	PERCENTAGE DISTRIBUTION OF STUDENTS IN TERTIARY EDUCATION BY ISCED FIELDS OF EDUCATION.	Admin	
26	PERCENTAGE DISTRIBUTION OF GRADUATES BY ISCED FIELDS OF EDUCATION AT THE TERTIARY LEVEL	Admin	
27	PERCENTAGE OF PRIVATE ENROLMENT	Admin	
28	PERCENTAGE OF TEACHING STAFF IN PRIVATE EDUCATIONAL INSTITUTION	Admin	
29	NUMBER OF STUDENTS IN TERTIARY EDUCATION PER 100,000 INHABITANTS	Admin	

30	PERCENTAGE DISTRIBUTION OF ENROLMENT IN SECONDARY EDUCATION BY ORIENTATION OF EDUCATION PROGRAMME	Admin	
31	PUBLIC CURRENT EXPENDITURE ON EDUCATION AS % OF TOTAL PUBLIC EXPENDITURE ON EDUCATION	Admin	
32	PERSONNEL COMPENSATION AS % OF PUBLIC CURRENT EXPENDITURE ON EDUCATION	Admin	
33	GROSS INTAKE RATIO IN THE LAST GRADE OF PRIMARY (GIRLG)	Admin	
34	EXPECTED GROSS INTAKE RATIO IN THE LAST GRADE OF PRIMARY (EGIRLG)	Admin	
35	GROSS PRIMARY GRADUATION RATIO (GPGR)	Admin	
36	EXPECTED GROSS PRIMARY GRADUATION RATIO (EGPGR)	Admin	
37	OUT-OF-SCHOOL CHILDREN-PRIMARY (OOS)	Admin	
38	PROMOTION RATE BY GRADE (PR)	Admin	
39	DROPOUT RATE BY GRADE (DR)	Admin	
40	GROSS ENROLMENT RATIO IN EARLY CHILDHOOD CARE AND EDUCATION (GER/ECCE)	Admin	
41	PERCENTAGE OF NEW ENTRANTS TO PRIMARY EDUCATION WITH ECCE EXPERIENCE.	Admin	
42	PERCENTAGE OF TRAINED TEACHERS	Admin	
43	PUBLIC EXPENDITURE ON A SPECIFIC ISCED LEVEL AS % OF TOTAL PUBLIC EXPENDITURE ON EDUCATION	Admin	
44	GENDER PARITY INDEX (GPI)	Each indicator	

## Annex 10: List of UIS SDG Indicators and sources

Indicator ID	Indicator description	Preferred source	Secondary source
4.1.1	Proportion of children and young people (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex	Admin	
4.2.1	Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex	Admin	
4.2.2	Participation rate in organized learning (one year before the official primary entry age), by sex	Admin	PHC, HIES, LFS
4.3.1	Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex	Admin	PHC, HIES, LFS
4.4.1	Proportion of youth and adults with information and communications technology (ICT) skills, by type of skill	Admin	
4.5.1	Parity indices (female/male, rural/urban, bottom/top wealth quintile and others such as disability status, indigenous peoples and conflict-affected, as data become available) for all education indicators on this list that can be disaggregated	Same as for indicator	
4.6.1	Proportion of population in a given age group achieving at least a fixed level of proficiency in functional (a) literacy and (b) numeracy skills, by sex	PHC	HIES, LFS
4.7.1	Extent to which (i) global citizenship education and (ii) education for sustainable development, including gender equality and human rights, are mainstreamed at all levels in: (a) national education policies, (b) curricula, (c) teacher education and (d) student assessment	Admin	
4.a.1	Proportion of schools with access to: (a) electricity; (b) Internet for pedagogical purposes; and (c) computers for pedagogical purposes	Admin	
	Proportion of schools with access to: (d) adapted infrastructure and materials for students with disabilities	Admin	
	Proportion of schools with access to: (e) basic drinking water; (f) single-sex basic sanitation facilities; and (g) basic handwashing facilities (as per the WASH indicator definitions)	Admin	
4.b.1	Volume of official development assistance flows for scholarships by sector and type of study	Admin	
4.c.1	Proportion of teachers in: (a) pre-primary education; (b) primary education; (c) lower secondary education; and (d) upper secondary education who have received at least the minimum organized teacher training (e.g., pedagogical training) pre-service or in-service required for teaching at the relevant level in a given country, by sex	Admin	



9.5.1	Research and development expenditure as a proportion of GDP	Admin	
9.5.2	Researchers (in full-time equivalent) per million inhabitants	Admin	
11.4.1	Total expenditure (public and private) per capita spent on the preservation, protection and conservation of all cultural and natural heritage, by type of heritage (cultural, natural, mixed and World Heritage Centre designation), level of government (national, regional and local/municipal), type of expenditure (operating expenditure/investment) and type of private funding (donations in kind, private non-profit sector and sponsorship)	Admin	
12.8.1	Extent to which (i) global citizenship education and (ii) education for sustainable development (including climate change education) are mainstreamed in (a) national education policies; (b) curricula; (c) teacher education; and (d) student assessment	Admin	

Annex 11: Sections, divisions and descriptions (ISIC Rev 4)

<b>International Standard Industrial Classification of all economic activities (2008)</b>		
<b>ISIC Rev 4 (Sections, Divisions and Descriptions)</b>		
Section	Divisions	Description
A	01–03	Agriculture, forestry and fishing
B	05–09	Mining and quarrying
C	10–33	Manufacturing
D	35	Electricity, gas, steam and air conditioning supply
E	36–39	Water supply; sewerage, waste management and remediation
F	41–43	Construction
G	45–47	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	49–53	Transportation and storage
I	55–56	Accommodation and food service activities
J	58–63	Information and communication
K	64–66	Financial and insurance activities
L	68	Real estate activities
M	69–75	Professional, scientific and technical activities
N	77–82	Administrative and support service activities
O	84	Public administration and defence; compulsory social security
P	85	Education
Q	86–88	Human health and social work activities
R	90–93	Arts, entertainment and recreation
S	94–96	Other service activities
T	97–98	Activities of households as employers; undifferentiated goods-and services-producing activities of households for own use
U	99	Activities of extraterritorial organizations and bodies

## Annex 12: Higher level aggregations of ISIC, rev. 4 (for Labour statistics and in SNA)

Labour statistics	SNA	Sections	Description
Agriculture	1	A	Agriculture, forestry and fishing
Industry	2	B, C, D, E	Manufacturing, mining and quarrying and other industrial activities
	3	F	Construction
Services	4	G, H, I	Wholesale and retail trade, transportation and storage, accommodation and food service activities
	5	J	Information and communication
	6	K	Financial and insurance activities
	7	L	Real estate activities a
	8	M, N	Professional, scientific, technical, administrative and support service activities
	9	O, P, Q	Public administration and defence, education, human health and social work activities
	10	R, S, T, U	Other service activities

Harmonized household dataset (COK, HIES, 2015)			Input household dataset (COK, HIES, 2015)	
Variable number	Variable name	Variable label	Data file	Variable name
H1	country	Country ID		COK
H2	datasourc	Name of survey or PHC		HIES
H3	year	Year of survey or PHC		2015
H4	hid	Household identifier	HHD	id07
H5	geolev1	Sub-national code level 1	Cover	anon_island_code
H6	geolev2	Sub-sub-national code level 2	Cover	anon_district_code
H7	rururb	Area of residence		*
H8	ea	Enumeration Area	Cover	anon_ea_code
H9	stratum	Stratum	Cover	strata
H10	psu	Primary sampling unit		*
H11	hhwt	Household weight	HHD	weight
H12	indwt	Individual weight		*
H13	hhsiz	Household size	Cover	total
H14	hhsiz_m	Number of male HM	Cover	males
H15	hhsiz_f	Number of female HM	Cover	females
H16	pcinc_cl	Per capita annual income quintile	Cover	quint_pctot_inc
H17	inchh	Total annual income	Cover	pc_tot_inc
H18	pcexp_cl	Per capita annual expenditure quintile	Cover	quint_pctot_exp
H19	exphh	Total annual household expenditure	Cover	pc_tot_exp
H20	dw_type	Type of dwelling	HHD	TT1 - hq20111
H21	tenure	Type of tenure	HHD	TT3
H22	builtyr	Year of construction	HHD	hq20116
H23	rooms	Number of habitable rooms	HHD	hq20115
H24	cookarea	Cooking area	HHD	TT6 - hq20117
H25	roof	Main material used for roof	HHD	TT7 - hq20112
H26	wall	Main material used for external walls	HHD	TT8 - hq20113
H27	floor	Main material used for floor	HHD	TT9 - hq20114
H28	watersrc	Main source of water	HHD	TT10 - Various Variables
H29	lighting	Main source of lighting in dwelling	HHD	TT11 - Various variables
H30	cookenrg	Main cooking energy	HHD	TT12 - *
H31	toilet	Main toilet facility	HHD	TT14 - *
H32	phone	Access to a landline phone	HHD	TT15 - hq20151
H33	cell	Access to a cell phone	HHD	TT16 - hq20157
H34	internet	Access to internet	HHD	TT17 - Various variables

H35	cabletv	Access to cable TV service	HHD	TT18 - hq20159
H36	anymort	Any HM died last 3 years	HHD	TT19 - *
H37	mortnum	Number HM died last 3 years	HHD	*
H38	sexmort1	Sex of 1st dead person	HHD	TD42:D48T20 - *
H39	agemort1	Age of 1st dead person	HHD	*
H40	cod_1	Cause of death	HHD	TT21 - *
H41	reprmort1	Reproductive status of 1st person at time of death	HHD	TT22 - *
H42	sexmort2	Sex of 2nd dead person	HHD	TT23 - *
H43	agemort2	Age of 2nd dead person	HHD	*
H44	cod_2	Cause of death - 2nd person	HHD	TT24 - *
H45	reprmort2	Reproductive status of 2nd person at time of death	HHD	TT25 - *
H46	sexmort3	Sex of 3rd dead person	HHD	TT26 - *
H47	agemort3	Age of 3rd dead person	HHD	*
H48	cod_3	Cause of death - 3rd person	HHD	TT27 - *
H49	reprmort3	Reproductive status of 3rd person at time of death	HHD	TT28 - *
H50	beds	Ownership of beds	HHD	TT29 - hq20502_1
H51	chairs	Ownership of chairs	HHD	TT30 - hq20502_2
H52	tables	Ownership of tables	HHD	TT31 - hq20502_3
H53	lamps	Ownership of lamps	HHD	TT32 - hq20502_4
H54	oth_ind	Ownership of other indoor furniture	HHD	TT33 - hq20502_5
H55	outfurn	Ownership of outdoor furniture	HHD	TT35 - hq20502_6
H56	flcover	Ownership of floor coverings	HHD	TT34 - hq20502_7
H57	othfurn	Ownership of any other furniture	HHD	TT36 - hq20502_8
H58	sheets	Ownership of sheets	HHD	TT38 - hq20502_10
H59	curtain	Ownership of curtains	HHD	TT39 - hq20502_11
H60	towels	Ownership of towels	HHD	TT40 - hq20502_12
H61	othertext	Ownership of other textiles	HHD	TT41 - hq20502_13
H62	watertanks	Ownership of water tanks	HHD	TT42 - hq20502_14
H63	refrig	Ownership of refrigerator	HHD	TT43 - hq20502_15
H64	stove	Ownership of stove	HHD	TT44 - Various variables
H65	burner	Ownership of gas burner	HHD	TT45 - hq20502_19
H66	microwave	Ownership of microwave oven	HHD	TT46 - hq20502_20
H67	washer	Ownership of washing machine	HHD	TT47 - hq20502_21
H68	dryer	Ownership of clothes dryer	HHD	TT48 - hq20502_22
H69	aircon	Ownership of air conditioner	HHD	TT49 - hq20502_23
H70	generator	Ownership of generator	HHD	TT50 - hq20502_24

H71	solar	Ownership of solar power unit	HHD	TT51 - hq20502_25
H72	hotwater	Ownership of water heater	HHD	TT52 - hq20502_26
H73	othappl	Ownership of other major appliances	HHD	TT53 - hq20502_27
H74	smallelec	Ownership of small electrical appliances	HHD	TT54 - hq20502_28
H75	tv	Ownership of television	HHD	TT55 - hq20502_29
H76	radio	Ownership of radio	HHD	TT56 - hq20502_30
H77	video	Ownership of video	HHD	TT57 - hq20502_31
H78	oth_aud	Ownership of other audio devies	HHD	TT58 - hq20502_32
H79	consoles	Ownership of game consoles	HHD	TT59 - hq20502_33
H80	photo	Ownership of photo equipment	HHD	TT60 - hq20502_34
H81	sports	Ownership of sports & camping equipment	HHD	TT61 - hq20502_35
H82	o_equip	Ownership of other recreational equip	HHD	TT62 - hq20502_36
H83	computer	Ownership of computers	HHD	TT63 - hq20502_37
H84	printer	Ownership of printer	HHD	TT64 - hq20502_38
H85	softpkg	Ownership of software packages	HHD	TT65 - hq20502_39
H86	port_dr	Ownership of portable media drives	HHD	TT66 - hq20502_40
H87	otherpceq	Ownership of other computer equipment	HHD	TT67 - hq20502_41
H88	car	Number of cars	HHD	hq20602_1
H89	usv	Number of utility vehicles	HHD	hq20602_2
H90	truk_bus	Number of trucks or buses	HHD	hq20602_3
H91	mcycle	Number of motorcycles	HHD	hq20602_4
H92	bcycle	Number of bicycles	HHD	hq20602_5
H93	mboat	Number of motor boats	HHD	hq20602_6
H94	othboat	Number of other boats	HHD	hq20602_7
H95	othveh	Number of other vehicles	HHD	hq20602_8
H96	out_brd	Number of outboard motors	HHD	hq20602_9
H97	trailer	Number of trailers	HHD	*
H98	hmag	HH members involved in agriculture	HHD	*
		<i>List of HH member ID</i>		
H99	nohmag	Number of hh members involved in agriculture	HHD	*
H100	paidhpag	Paid help in agriculture 1 = Yes 2 = No 8 = Don't know 9 = Missing	HHD	TT71 - 40202
H101	landtyp	Land type	HHD	TT68 - hq40203
H102		Number of parcels own	HHD	*
H103		Size of parcels	HHD	*

H104	extvisit	Number of extension visits	HHD	*
H105	coopmemb	Participation as cooperative member	HHD	TT69 - *
H106	credit	Microcredit use	HHD	TT70 - *
H107	hmlv	HH members involved in livestock <i>List of HH member ID</i>	HHD	*
H108	nohmlv	Number of hh members involved in livestock	HHD	*
H109	paidhplv	Existence/Number of paid help in livestock	HHD	hq40302
H110	pigs	Pig number	HHD	hq40404_3b
H111	chicken	Chicken number	HHD	hq40404_4b
H112	ducks	Ducks number	HHD	hq40404_5b
H113	cows	Cow number	HHD	hq40404_1b
H114	horses	Horse number	HHD	hq40404_2b
H115	dogs	Dog number	HHD	*
H116	goats	Goat number	HHD	hq40404_6b
H117	othlv	Number of other livestock	HHD	hq40404_7b
H118	hmaq	HH members involved in aquaculture <i>List of HH member ID</i>	HHD	*
H119	nohmaq	Number of hh members involved in aquaculture	HHD	*
H120	paidhpaq	Existence/Number of paid help in aquaculture	HHD	hq40502
H121	giantclam	Ownership & number of of giant clam	HHD	TT72 - Various variables
H122	crab	Ownership & number of of crab	HHD	TT73 - Various variables
H123	milkfish	Ownership & number of of milkfish	HHD	TT74 - Various variables
H124	eel	Ownership & number of of eel	HHD	TT75 - Various variables
H125	seacucum	Ownership & number of sea cucumber	HHD	TT76 - *
H126	lobster	Ownership & number of lobster	HHD	TT77 - *
H127	trochus	Ownership & number of trochus	HHD	TT78 - *
H128	reef fish	Ownership & number of reef fish	HHD	TT79 - *
H129	tuna	Ownership & number of tuna	HHD	TT80 - *
H130	shellfish	Ownership & number of shellfish	HHD	TT81 - *
H131	freshwater fish	Ownership & number of freshwater fish	HHD	TT82 - *
H132	othaq	Ownership & number of other aqua	HHD	TT83 - Various variables

H133	hmfsh	HH members involved in fisheries <i>List of HH member ID</i>	HHD	*
H134	nohmfsh	Number of hh members involved in fisheries	HHD	*
H135	paidhpfsh	Existence/Number of paid help in fisheries	HHD	hq40302
H136	trolling	Trolling	HHD	TT84 - hq40303e
H137	bottom	Bottom fishing	HHD	TT85 - hq40303f
H138	casting	Casting	HHD	TT86 - *
H139	spearfsh	Spear fishing	HHD	TT87 - hq40303g
H140	netfsh	Net fishing	HHD	TT88 - 40303a
H141	gather	Gathering	HHD	TT89 - Various variables
H142	trapping	Trapping	HHD	TT90 - *
H143	hook&line	Hook & line	HHD	TT91 - Various variables
H144	othfsh	Other fishing method	HHD	TT92 - Various variables
H145	oceanloc	Open ocean	HHD	TT93 - *
H146	FADloc	FAD location	HHD	TT94 - *
H147	subreef	Submerged reef	HHD	TT95 - ?Various variables
H148	outreef	Outer reef	HHD	TT96 - ?
H149	costreef	Coastal reef	HHD	TT97 - ?
H150	lagoon	Lagoon	HHD	TT98 - Various variables
H151	mangrove	Mangrove	HHD	TT99 - *
H152	river	River	HHD	TT4a - *
H153	othloc	Other fishing locations	HHD	TT5a - Various variables
H154	notrips	Number of fishing trips	HHD	hq40305
H155	nohours	Number of hours per trip	HHD	hq40306
H156	travmode	Main mode of travel	HHD	TT5b - hq40307



## Harmonized household dataset (FSM, HIES, 2013)

## Input household dataset (FSM, HIES, 2013)

Variable number	Variable name	Variable label	Data file	Variable name
H1	country	Country ID		FSM
H2	datasourc	Name of survey or PHC		HIES
H3	year	Year of survey or PHC		2013
H4	hid	Household identifier	HHD	id07
H5	geolev1	Sub-national code level 1	Cover	state
H6	geolev2	Sub-sub-national code level 2	Cover	anon_municipality_code *
H7	rururb	Area of residence		
H8	ea	Enumeration Area	Cover	anon_ea_code
H9	stratum	Stratum	Cover	strata
H10	psu	Primary sampling unit		*
H11	hhwt	Household weight	HHD	weight
H12	indwt	Individual weight		*
H13	hhsiz	Household size	Cover	tot_p
H14	hhsiz_m	Number of male HM	Cover	tot_m
H15	hhsiz_f	Number of female HM	Cover	tot_f
H16	pcinc_cl	Per capita annual income quintile	Cover	*
H17	inchh	Total annual income	Cover	t_inc_quintile
H18	pcexp_cl	Per capita annual expenditure quintile	Cover	*
H19	exphh	Total annual household expenditure	Cover	t_exp_quintile
H20	dw_type	Type of dwelling	HHD	TT1 - q20111
H21	tenure	Type of tenure	HHD	TT3 - *
H22	builtyr	Year of construction	HHD	q20116
H23	rooms	Number of habitable rooms	HHD	q20115
H24	cookarea	Cooking area	HHD	TT6 - q20117
H25	roof	Main material used for roof	HHD	TT7 - q20112
H26	wall	Main material used for external walls	HHD	TT8 - q20113
H27	floor	Main material used for floor	HHD	TT9 - H4_VS1
H28	watersrc	Main source of water	HHD	TT10 - q20134_1
H29	lighting	Main source of lighting in dwelling	HHD	TT11 - Various variables
H30	cookenerg	Main cooking energy	HHD	TT12 - q20126
H31	toilet	Main toilet facility	HHD	TT14 - q20144_1
H32	phone	Access to a landline phone	HHD	TT15 - q20151 & q20152
H33	cell	Access to a cell phone	HHD	TT16 - q20153 & q20154
H34	internet	Access to internet	HHD	T17 - Various variables
H35	cabletv	Access to cable TV service	HHD	TT18 - *

H36	anymort	Any HM died last 3 years	HHD	TT19 - *
H37	mortnum	Number HM died last 3 years	HHD	*
H38	sexmort1	Sex of 1st dead person	HHD	TD42:D48T20 - *
H39	agemort1	Age of 1st dead person	HHD	*
H40	cod_1	Cause of death	HHD	TT21 - *
H41	reprmort1	Reproductive status of 1st person at time of death	HHD	TT22 - *
H42	sexmort2	Sex of 2nd dead person	HHD	TT23 - *
H43	agemort2	Age of 2nd dead person	HHD	*
H44	cod_2	Cause of death - 2nd person	HHD	TT24 - *
H45	reprmort2	Reproductive status of 2nd person at time of death	HHD	TT25 - *
H46	sexmort3	Sex of 3rd dead person	HHD	TT26 - *
H47	agemort3	Age of 3rd dead person	HHD	*
H48	cod_3	Cause of death - 3rd person	HHD	TT27 - *
H49	reprmort3	Reproductive status of 3rd person at time of death	HHD	TT28 - *
H50	beds	Ownership of beds	HHD	TT29 - *
H51	chairs	Ownership of chairs	HHD	TT30 - *
H52	tables	Ownership of tables	HHD	TT31 - *
H53	lamps	Ownership of lamps	HHD	TT32 - *
H54	oth_ind	Ownership of other indoor furniture	HHD	TT33 - *
H55	outfurn	Ownership of outdoor furniture	HHD	TT34 - *
H56	flcover	Ownership of floor coverings	HHD	TT35 - *
H57	othfurn	Ownership of any other furniture	HHD	TT36 - *
H58	sheets	Ownership of sheets	HHD	TT38 - *
H59	curtain	Ownership of curtains	HHD	TT39 - *
H60	towels	Ownership of towels	HHD	TT40 - *
H61	othertext	Ownership of other textiles	HHD	TT41 - *
H62	watertanks	Ownership of water tanks	HHD	TT42 - Various variables
H63	refrig	Ownership of refrigerator	HHD	TT43 - *
H64	stove	Ownership of stove	HHD	TT44 - *
H65	burner	Ownership of gas burner	HHD	TT45 - *
H66	microwave	Ownership of microwave oven	HHD	TT46 - *
H67	washer	Ownership of washing machine	HHD	TT47 - *
H68	dryer	Ownership of clothes dryer	HHD	TT48 - *
H69	aircon	Ownership of air conditioner	HHD	TT49 - *
H70	generator	Ownership of generator	HHD	TT50 - Various variables

H71	solar	Ownership of solar power unit	HHD	TT51 - Various variables
H72	hotwater	Ownership of water heater	HHD	TT52 - *
H73	othappl	Ownership of other major appliances	HHD	TT53 - *
H74	smallelec	Ownership of small electrical appliances	HHD	TT54 - *
H75	tv	Ownership of television	HHD	TT55 - *
H76	radio	Ownership of radio	HHD	TT56 - *
H77	video	Ownership of video	HHD	TT57 - *
H78	oth_aud	Ownership of other audio devies	HHD	TT58 - *
H79	consoles	Ownership of game consoles	HHD	TT59 - *
H80	photo	Ownership of photo equipment	HHD	TT60 - *
H81	sports	Ownership of sports & camping equipment	HHD	TT61 - *
H82	o_equip	Ownership of other recreational equip	HHD	TT62 - *
H83	computer	Ownership of computers	HHD	TT63 - *
H84	printer	Ownership of printer	HHD	TT64 - *
H85	softpkg	Ownership of software packages	HHD	TT65 - *
H86	port_dr	Ownership of portable media drives	HHD	TT66 - *
H87	otherpceq	Ownership of other computer equipment	HHD	TT67 - *
H88	car	Number of cars	HHD	*
H89	usv	Number of utility vehicles	HHD	*
H90	truk_bus	Number of trucks or buses	HHD	*
H91	mcycle	Number of motorcycles	HHD	*
H92	bcycle	Number of bicycles	HHD	*
H93	mboat	Number of motor boats	HHD	*
H94	othboat	Number of other boats	HHD	*
H95	othveh	Number of other vehicles	HHD	*
H96	out_brd	Number of outboard motors	HHD	*
H97	trailer	Number of trailers	HHD	*
H98	hmag	HH members involved in agriculture	HHD	*
		<i>List of HH member ID</i>		
H99	nohmag	Number of hh members involved in agriculture	HHD	*
H100	paidhpag	Paid help in agriculture 1 = Yes 2 = No 8 = Don't know 9 = Missing	HHD	TT71 - *
H101	landtyp	Land type	HHD	TT68 - *
H102		Number of parcels own	HHD	*
H103		Size of parcels	HHD	*

H104	extvisit	Number of extension visits	HHD	*
H105	coopmemb	Participation as cooperative member	HHD	TT69 - *
H106	credit	Microcredit use	HHD	TT70 - *
H107	hmlv	HH members involved in livestock <i>List of HH member ID</i>	HHD	*
H108	nohmlv	Number of hh members involved in livestock	HHD	*
H109	paidhplv	Existence/Number of paid help in livestock	HHD	*
H110	pigs	Pig number	HHD	*
H111	chicken	Chicken number	HHD	*
H112	ducks	Ducks number	HHD	*
H113	cows	Cow number	HHD	*
H114	horses	Horse number	HHD	*
H115	dogs	Dog number	HHD	*
H116	goats	Goat number	HHD	*
H117	othlv	Number of other livestock	HHD	*
H118	hmaq	HH members involved in aquaculture <i>List of HH member ID</i>	HHD	*
H119	nohmaq	Number of hh members involved in aquaculture	HHD	*
H120	paidhpaq	Existence/Number of paid help in aquaculture	HHD	*
H121	giantclam	Ownership & number of of giant clam	HHD	TT72 - *
H122	crab	Ownership & number of of crab	HHD	TT73 - *
H123	milkfish	Ownership & number of of milkfish	HHD	TT74 - *
H124	eel	Ownership & number of of eel	HHD	TT75 - *
H125	seacucum	Ownership & number of sea cucumber	HHD	TT76 - *
H126	lobster	Ownership & number of lobster	HHD	TT77 - *
H127	trochus	Ownership & number of trochus	HHD	TT78 - *
H128	reef fish	Ownership & number of reef fish	HHD	TT79 - *
H129	tuna	Ownership & number of tuna	HHD	TT80 - *
H130	shellfish	Ownership & number of shellfish	HHD	TT81 - *
H131	freshwater fish	Ownership & number of freshwater fish	HHD	TT82 - *
H132	othaq	Ownership & number of other aqua	HHD	TT83 - *

H133	hmfsh	HH members involved in fisheries <i>List of HH member ID</i>	HHD	*
H134	nohmfsh	Number of hh members involved in fisheries	HHD	*
H135	paidhpfsh	Existence/Number of paid help in fisheries	HHD	*
H136	trolling	Trolling	HHD	TT84 - *
H137	bottom	Bottom fishing	HHD	TT85 - *
H138	casting	Casting	HHD	TT86 - *
H139	spearfsh	Spear fishing	HHD	TT87 - *
H140	netfsh	Net fishing	HHD	TT88 - *
H141	gather	Gathering	HHD	TT89 - *
H142	trapping	Trapping	HHD	TT90 - *
H143	hook&line	Hook & line	HHD	TT91 - *
H144	othfsh	Other fishing method	HHD	TT92 - *
H145	oceanloc	Open ocean	HHD	TT93 - *
H146	FADloc	FAD location	HHD	TT94 - *
H147	subreef	Submerged reef	HHD	TT95 - *
H148	outreef	Outer reef	HHD	TT96 - *
H149	costreef	Coastal reef	HHD	TT97 - *
H150	lagoon	Lagoon	HHD	TT98 - *
H151	mangrove	Mangrove	HHD	TT99 - *
H152	river	River	HHD	TT4a - *
H153	othloc	Other fishing locations	HHD	TT5a - *
H154	notrips	Number of fishing trips	HHD	*
H155	nohours	Number of hours per trip	HHD	*
H156	travmode	Main mode of travel	HHD	TT5b - *

## Harmonized household dataset (PLW, PHC, 2015)

Variable number	Variable name	Variable label
H1	country	Country ID
H2	datasourc	Name of survey or PHC
H3	year	Year of survey or PHC
H4	hid	Household identifier
H5	geolev1	Sub-national code level 1
H6	geolev2	Sub-sub-national code level 2
H7	rururb	Area of residence
H8	ea	Enumeration Area
H9	stratum	Stratum
H10	psu	Primary sampling unit
H11	hhwt	Household weight
H12	indwt	Individual weight
H13	hhsz	Household size
H14	hhsz_m	Number of male HM
H15	hhsz_f	Number of female HM
H16	pcinc_cl	Per capita annual income quintile
H17	inchh	Total annual income
H18	pcexp_cl	Per capita annual expenditure quintile
H19	exphh	Total annual household expenditure
H20	dw_type	Type of dwelling
H21	tenure	Type of tenure
H22	builtyr	Year of construction
H23	rooms	Number of habitable rooms
H24	cookarea	Cooking area
H25	roof	Main material used for roof
H26	wall	Main material used for external walls
H27	floor	Main material used for floor
H28	watersrc	Main source of water
H29	lighting	Main source of lighting in dwelling
H30	cookenerg	Main cooking energy
H31	toilet	Main toilet facility
H32	phone	Access to a landline phone
H33	cell	Access to a cell phone
H34	internet	Access to internet
H35	cabletv	Access to cable TV service
H36	anymort	Any HM died last 3 years

## Input household dataset (PLW, PHC, 2015)

Data file	Variable name
	PLW
	PHC
	2015
HHD	HLDNUM_NUM_VS1
Cover	STATE_VS1
Cover	HAMLET_VS1
	*
Cover	EA_VS1
Cover	*
	*
HHD	*
	*
Cover	CS8_3_VS1
Cover	CS8_1_VS1
Cover	CS8_2_VS1
Cover	*
Cover	*
Cover	*
Cover	*
HHD	TT1 - A2_VS1 & A6_VS1
HHD	TT3 - A3_VS1
HHD	A8_VS1
HHD	A12_VS1
HHD	TT6 - A26_VS1
HHD	TT7 - A10_VS1
HHD	TT8 - A9_VS1
HHD	TT9 - *
HHD	TT10 - Various variables
HHD	TT11 - A18_VS1
HHD	TT12 - Various variables
HHD	TT14 - A17_VS1
HHD	TT15 - A19_1_VS1
HHD	TT16 - A19_2_VS1 & A19_2_VS2
HHD	TT17 - *
HHD	TT18 - *
HHD	TT19 - *

H37	mortnum	Number HM died last 3 years	HHD	*
H38	sexmort1	Sex of 1st dead person	HHD	TD42:D48T20 - *
H39	agemort1	Age of 1st dead person	HHD	*
H40	cod_1	Cause of death	HHD	TT21 - *
H41	reprmort1	Reproductive status of 1st person at time of death	HHD	TT22 - *
H42	sexmort2	Sex of 2nd dead person	HHD	TT23 - *
H43	agemort2	Age of 2nd dead person	HHD	*
H44	cod_2	Cause of death - 2nd person	HHD	TT24 - *
H45	reprmort2	Reproductive status of 2nd person at time of death	HHD	TT25 - *
H46	sexmort3	Sex of 3rd dead person	HHD	TT26 - *
H47	agemort3	Age of 3rd dead person	HHD	*
H48	cod_3	Cause of death - 3rd person	HHD	TT27 - *
H49	reprmort3	Reproductive status of 3rd person at time of death	HHD	TT28 - *
H50	beds	Ownership of beds	HHD	TT29 - *
H51	chairs	Ownership of chairs	HHD	TT30 - *
H52	tables	Ownership of tables	HHD	TT31 - *
H53	lamps	Ownership of lamps	HHD	TT32 - *
H54	oth_ind	Ownership of other indoor furniture	HHD	TT33 - *
H55	outfurn	Ownership of outdoor furniture	HHD	TT34 - *
H56	flcover	Ownership of floor coverings	HHD	TT35 - *
H57	othfurn	Ownership of any other furniture	HHD	TT36 - *
H58	sheets	Ownership of sheets	HHD	TT38 - *
H59	curtain	Ownership of curtains	HHD	TT39 - *
H60	towels	Ownership of towels	HHD	TT40 - *
H61	othertext	Ownership of other textiles	HHD	TT41 - *
H62	watertanks	Ownership of water tanks	HHD	TT42 - A22_VS1
H63	refrig	Ownership of refrigerator	HHD	TT43 - A19_5_VS1
H64	stove	Ownership of stove	HHD	TT44 - *
H65	burner	Ownership of gas burner	HHD	TT45 - *
H66	microwave	Ownership of microwave oven	HHD	TT46 - A19_4_VS1
H67	washer	Ownership of washing machine	HHD	TT47 - *
H68	dryer	Ownership of clothes dryer	HHD	TT48 - *
H69	aircon	Ownership of air conditioner	HHD	TT49 - A21_VS1
H70	generator	Ownership of generator	HHD	TT50 - *
H71	solar	Ownership of solar power unit	HHD	TT51 - *
H72	hotwater	Ownership of water heater	HHD	TT52 - *

H73	othappl	Ownership of other major appliances	HHD	TT53 - *
H74	smallelec	Ownership of small electrical appliances	HHD	TT54 - *
H75	tv	Ownership of television	HHD	TT55 - A19_6_VS1
H76	radio	Ownership of radio	HHD	TT56 - A20_VS1
H77	video	Ownership of video	HHD	TT57 - *
H78	oth_aud	Ownership of other audio devies	HHD	TT58 - *
H79	consoles	Ownership of game consoles	HHD	TT59 - *
H80	photo	Ownership of photo equipment	HHD	TT60 - *
H81	sports	Ownership of sports & camping equipment	HHD	TT61 - *
H82	o_equip	Ownership of other recreational equip	HHD	TT62 - *
H83	computer	Ownership of computers	HHD	TT63 - A19_3_VS1
H84	printer	Ownership of printer	HHD	TT64 - *
H85	softpkg	Ownership of software packages	HHD	TT65 - *
H86	port_dr	Ownership of portable media drives	HHD	TT66 - *
H87	otherpceq	Ownership of other computer equipment	HHD	TT67 - *
H88	car	Number of cars	HHD	*
H89	usv	Number of utility vehicles	HHD	*
H90	truk_bus	Number of trucks or buses	HHD	*
H91	mcycle	Number of motorcycles	HHD	*
H92	bcycle	Number of bicycles	HHD	*
H93	mboat	Number of motor boats	HHD	*
H94	othboat	Number of other boats	HHD	*
H95	othveh	Number of other vehicles	HHD	*
H96	out_brd	Number of outboard motors	HHD	*
H97	trailer	Number of trailers	HHD	*
H98	hmag	HH members involved in agriculture <i>List of HH member ID</i>	HHD	*
H99	nohmag	Number of hh members involved in agriculture	HHD	*
H100	paidhpag	Paid help in agriculture    1 = <i>Yes</i> 2 = <i>No</i> <i>8 = Don't know</i> 9 = <i>Missing</i>	HHD	TT71 - *
H101	landtyp	Land type	HHD	TT68 - BS20_VS1 & BS1_4_VS2
H102		Number of parcels own	HHD	B3
H103		Size of parcels	HHD	*
H104	extvisit	Number of extension visits	HHD	*



H105	coopmemb	Participation as cooperative member	HHD	TT69 - *
H106	credit	Microcredit use	HHD	TT70 - *
H107	hmlv	HH members involved in livestock <i>List of HH member ID</i>	HHD	*
H108	nohmlv	Number of hh members involved in livestock	HHD	*
H109	paidhplv	Existence/Number of paid help in livestock	HHD	*
H110	pigs	Pig number	HHD	BS10_VS1
H111	chicken	Chicken number	HHD	BS12_VS1
H112	ducks	Ducks number	HHD	BS13_VS1
H113	cows	Cow number	HHD	BS9_VS1
H114	horses	Horse number	HHD	*
H115	dogs	Dog number	HHD	*
H116	goats	Goat number	HHD	BS11_VS1
H117	othlv	Number of other livestock	HHD	BS14_VS1
H118	hmaq	HH members involved in aquaculture <i>List of HH member ID</i>	HHD	*
H119	nohmaq	Number of hh members involved in aquaculture	HHD	*
H120	paidhpaq	Existence/Number of paid help in aquaculture	HHD	*
H121	giantclam	Ownership & number of of giant clam	HHD	TT72 - BS15_VS1
H122	crab	Ownership & number of of crab	HHD	TT73 - *
H123	milkfish	Ownership & number of of milkfish	HHD	BS16_VS1
H124	eel	Ownership & number of of eel	HHD	TT75 - *
H125	seacucum	Ownership & number of sea cucumber	HHD	TT76 - *
H126	lobster	Ownership & number of lobster	HHD	TT77 - *
H127	trochus	Ownership & number of trochus	HHD	TT78 - *
H128	reef fish	Ownership & number of reef fish	HHD	TT79 - *
H129	tuna	Ownership & number of tuna	HHD	TT80 - *
H130	shellfish	Ownership & number of shellfish	HHD	TT81 - *
H131	freshwater fish	Ownership & number of freshwater fish	HHD	TT82 - *
H132	othaq	Ownership & number of other aqua	HHD	TT83 - BS17_VS1

H133	hmfish	HH members involved in fisheries <i>List of HH member ID</i>	HHD	BS21
H134	nohmfish	Number of hh members involved in fisheries	HHD	*
H135	paidhpfsh	Existence/Number of paid help in fisheries	HHD	*
H136	trolling	Trolling	HHD	TT84 - BS23_2_VS1
H137	bottom	Bottom fishing	HHD	*
H138	casting	Casting	HHD	*
H139	spearfish	Spear fishing	HHD	TT87 - BS23_3_VS1
H140	netfish	Net fishing	HHD	TT88 - BS23_1_VS1
H141	gather	Gathering	HHD	TT89 - BS23_5_VS1
H142	trapping	Trapping	HHD	TT90 - *
H143	hook&line	Hook & line	HHD	TT91 - BS23_4_VS1
H144	othfish	Other fishing method	HHD	TT92 - BS23_6_VS1
H145	oceanloc	Open ocean	HHD	TT93 - BS24_4_VS1
H146	FADloc	FAD location	HHD	TT94 - *
H147	subreef	Submerged reef	HHD	TT95 - BS24_3_VS1
H148	outreef	Outer reef	HHD	TT96 - ?
H149	costreef	Coastal reef	HHD	TT97 - ?
H150	lagoon	Lagoon	HHD	TT98 - BS24_2_VS1
H151	mangrove	Mangrove	HHD	TT99 - BS24_1_VS1
H152	river	River	HHD	TT4a - *
H153	othloc	Other fishing locations	HHD	TT5a - BS24_5_VS1
H154	notrips	Number of fishing trips	HHD	*
H155	nohours	Number of hours per trip	HHD	*
H156	travmode	Main mode of travel	HHD	TT5b - *

MAPPING OF PACIFIC ISLAND NATIONAL EDUCATION SYSTEMS TO ISCED CLASSIFICATION

Produced by Statistics for Development Division in association with Education Quality and Assessment Programme, 2019

ISCED Level	0	1	2	3	4	5	6	7	8
-------------	---	---	---	---	---	---	---	---	---

Age / Country	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27							
Cook Islands	6-4-3			Pre-School		Primary Ed. (Years 1-6)						Secondary Ed. (Forms 1-4)			Secondary Ed. (Forms 5-7)		Degree or Bachelor			Master Degree															
																Vocational																			
Fiji	6-4-3			Pre-School			Primary Ed. (Class 1-6)						Junior Secondary Ed. (Forms 1-4)		Senior Secondary Ed. (Forms 5-7)		Diploma or Certificate		Degree or Bachelor		Master Degree		Doctorate												
Kiribati	6-3-4			Pre-school		Primary Ed. (Class 1-6)						Junior Secondary Ed. (Years 7-9)					TVET Certificate																		
																	Senior Secondary (Years 10-13)													Bachelor					
Marshall Islands	6-4-4			Pre-Kindi Headstart Program		Elementary Ed. (Grades 1-6)						Elementary (Years 7-8) Secondary School (Grades 9-10)			Secondary School (Years 11-12)		Diploma		Bachelor			Master Degree													
																	Associate Degree																		
FSM	6-2-4			Pre-school		ECE/ Kindergarten		Elementary Ed. (Grades 1-6)						Elementary Ed. (Grades 7-8)		High School (Grades 9-12)			Certificate																
																			Associate Degree																
																			Bachelor Degree												Master Degree		Doctorate		
Nauru	6-3-3				Play center	Pre-school	Prepara-tory class	Primary Ed. (Class 1-6)					Junior Secondary Ed. (Form 1-3)			Senior Secondary Ed. (Form 4-6)		USP Foundation																	
Niue	6-4-3	Playschool/ECE			Primary Ed. (Years 1-6)						Lower Secondary Ed. (Years 7-10)			Upper Secondary (Years 10-12)		OP Certificate		Diploma										Master Degree							
																Bachelor																			
Palau	6-2-4			Headstart program			Elementary Ed. (Grades 1-6)						Elementary Ed. (Grades 7-8)		High School (Grades 9-12)			Certificate																	
																		Associate Degree																	
Papua New Guinea	6-3-3				Kindergarten (Years 1-2)		Primary Ed. (Grades 1-6)						Lower Secondary Ed. (Form 1-3)		Upper Secondary Ed. (Form 4-6)																				
Samoa	6-2-5			Pre-school		Primary Ed. (Years 1-6)						Primary Ed. (Years 7-8)		Secondary Ed. (Years 9-13)															Certificate		Master Degree		Doctorate		
																													Diploma						
																		Bachelor																	
Solomon Islands	6-3-4			Pre-primary Ed.			Standard Ed. (Std 1-6)						Junior Secondary Ed. (Form 1-3)			Upper Secondary Ed. (Form 4-7)			SINU Certificate		Diploma														
																			University Diploma																
																			Bachelor			Master Degree													
Tokelau	6-5-2			Early Childhood Ed.		Primary Ed. (Years 1-6)						Secondary Ed. (Years 7-11)				Senior Secondary (Years 12-13)		Certificate or Diploma		Bachelor and Posgraduate															
Tonga	6-5-2				Pre-primary		Primary Ed. (Years 1-6)						Lower Secondary Ed. (Years 7-11)				Upper Secondary Ed.		USP Certificate		C or D or AD							(Certificate or Diploma or Associate Degree)							
																					Diploma														
																	Vocational Certificate		Bachelor of Theology or Divinity		Bachelor of Divinity with honor or of Arts		Master Degree		Doctorate										
Tuvalu	6-4-3			Pre-school			Primary Ed. (Years 1-6)						Primary Ed. (Years 7-8)		Secondary Ed. (Years 9-10)		Senior Ed. (Years 11-13)		Certificate																
Vanuatu	6-4-3				ECCE/Kindergarten		Basic Ed. (Years 1-6)						Basic Ed. (Years 7-8)		Lower Secondary Ed. (Years 9-10)		Upper Secondary Ed. (Year 11-13)			Licence or Bachelor											Master Degree				

# AGE PROGRESSION BY EDUCATION LEVEL

Produced by Statistics for Development Division in association with Education Quality and Assessment Programme, 2019

ISCED Level	0	1	2	3	5-8
-------------	---	---	---	---	-----

