Guidelines for the harmonising of census and survey microdata

Guidelines for the harmonising of census and survey microdata

Abimbola Sylvester Young, Michael Sharp, Jolly May Catalan, Scott Pontifex, and Olivier Menaouer

Pacific Community
Communauté du Pacifique

Noumea, New Caledonia, 2020

CONTENTS

TABLES

ACRONYMS AND ABBREVIATIONS

CAPI	computer-assisted personal interview
COICOP	Classification of Individual Consumption According to Purpose
CSPro	Census and Survey Processing System
EAP	East Asia and Pacific
HH	household
HIES	household income and expenditure survey
HM	household member
ICSE	International Classification of Status in Employment
ILO	International Labour Organization
IPUMS	Integrated Public Use Microdata Series
ISCED	International Standard Classification of Education
ISCED-A	ISCED Attainment
ISCED-F	ISCED Fields of Education and Training
ISCO	International Standard Classification of Occupations
ISIC	International Standard Industrial Classification of All Economic Activities
LFS	labour force surveys
MICS	Multiple Indicator Cluster Survey
NPISH	non-profit institutions serving households
PACCOI	Pacific Classification of Income
PACSCO	Pacific Standard Classification of Occupations
PACSIC	Pacific Standard Industrial Classification of All Economic Activities
PHC	population and housing census
PICTs	Pacific Island countries and territories
RIGA	Rural Income Generating Activities project
RIGA-L	RIGA individual wage employment dataset
SDD	Statistics for Development Division
SDG	Sustainable Development Goals
SNA	System of National Accounts
SPC	Pacific Community
UIS	UNESCO Institute for Statistics
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations International Children's Fund

ACKNOWLEDGEMENTS

FOREWORD

I am delighted to present the guidelines for the harmonising of census and survey microdata.

Measurement of progress against development indicators, and evaluation of policy, requires the production of data that allow the computation of statistics and indicators that are comparable over time and space. SPC, its development partners and the statistical agencies of the Pacific region have made significant progress in statistical collection standardisation initiatives. This includes standardisation of methods and instruments for specific statistical collections, but also standardisation across statistical collections.

These guidelines take standardisation initiatives a step further and include a series of instruments designed to help facilitate harmonisation of census and survey microdata. They practically describe how the structuring of existing and future microdata sets can be made using data harmonisation to more rapidly and consistently produce comparable statistical outputs to guide and monitor policy.

It is envisaged that data harmonisation initiatives will help reduce time and costs in developing data collections, reduce respondent burden and unnecessary duplication across sources, while maximising the opportunities for the cross-analysis, exchange and reuse of data.

We hope that these guidelines will be both a useful resource for the users and producers of Pacific Island statistics and help to improve data dissemination and use in the Pacific region.

Mr Epeli Waqavonovono

Director
Statistics for Development Division (SDD)
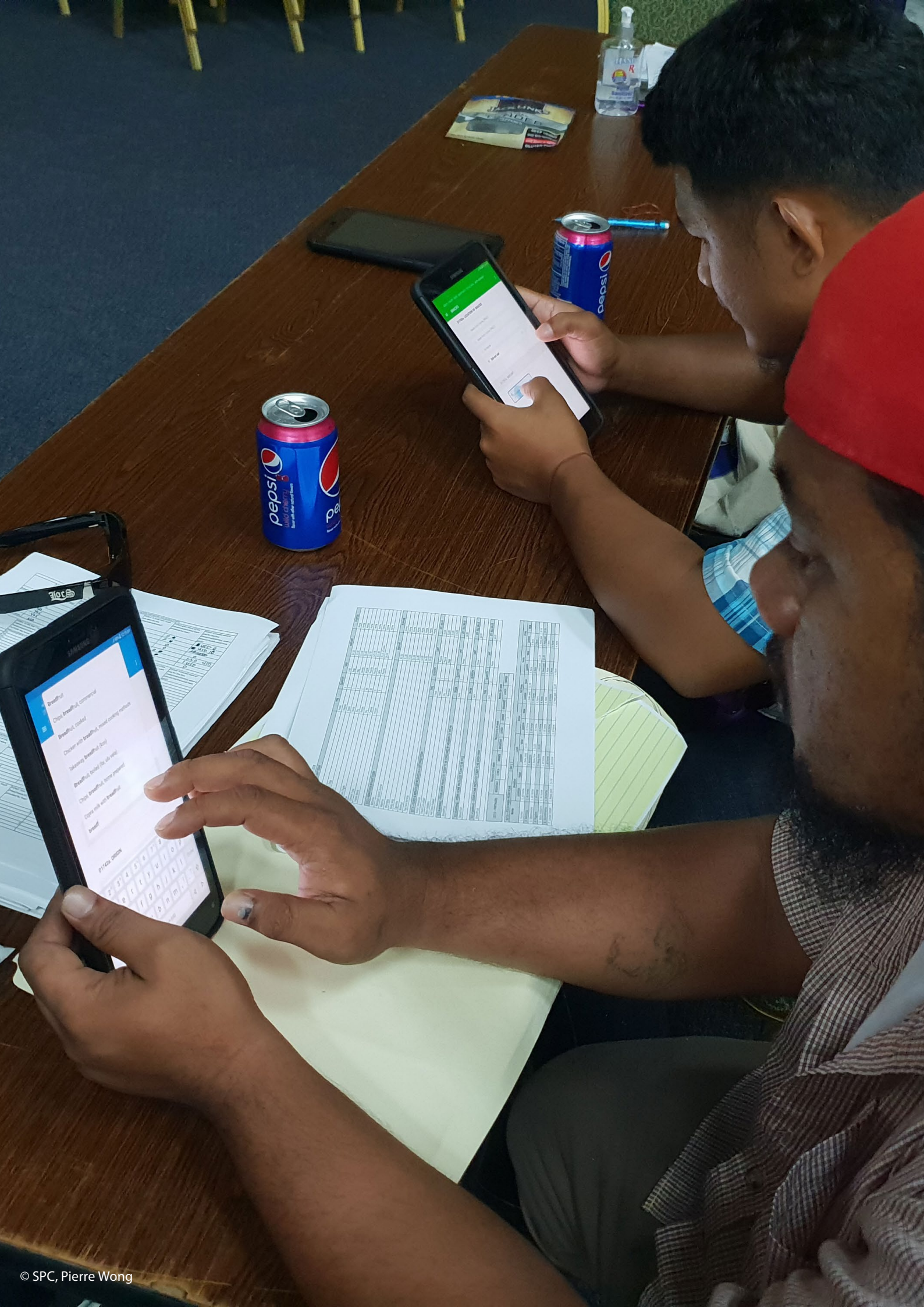
CHAPTER 1: INTRODUCTION

Background

In 2017, the World Bank's Trust Fund for Statistical Capacity Building approved a grant to the Pacific Community (SPC) to improve data dissemination and use in the Pacific region. The goal of the project is to strengthen the capacity of national statistical systems and social and economic planning agencies to provide policy-makers and analysts with important demographic, economic and social indicators for planning and decision-making. It is envisaged that one way this can be achieved is by producing and disseminating a collection of harmonised datasets obtained by standardising data from existing surveys and censuses that are available in the Pacific.

As a first step towards achieving this ambition, these guidelines have been developed by SPC to facilitate the harmonisation of Pacific Island census and survey microdata. The guidelines provide guidance and metadata for the standardisation of 28 census and household income and expenditure survey (HIES) datasets from 14 Pacific Island countries and cover multiple thematic areas, such as demographic, household expenditure, labour, education and health.

The guidelines include the following elements:

1. a database of topics that are common across censuses and HIES;

2. a review of other microdata harmonisation initiatives and relevant international standards/ classifications;

3. a data dictionary, which includes a list of harmonised variables, their labels and codes;

4. a translation for every harmonised variable for all 14 Pacific Island countries and all 28 datasets (a translation table example is included in Annex 6); and

5. useful information on how to handle data cleaning issues such as dealing with missing values/outliers, database structure and dataset naming conventions in the construction of harmonised datasets.

Measurement of progress against development indicators, and evaluation of policy in the Pacific region, requires the production of data that allow for statistics and indicators to be consistent and comparable over time and space. It also requires the production of these over the same space and time but by different populations of interest, such as gender, and disability or poverty status. To do this, microdata are needed.

While microdata are enormously flexible and powerful, in the Pacific region they are often not readily comparable across other datasets (i.e. from different times and places). Some examples of these differences include different collection approaches and use of different standards and methodologies, such as how the questionnaire was designed, or the way the data are stored and structured.

To overcome some of these interoperability challenges, SPC, its development partners and the statistical agencies of the Pacific region have invested in standardisation of methods and instruments for specific statistical collections, but also standardisation across statistical collections. Regionally standardised instruments and methodologies have been designed for the 2020 round of population and housing censuses (PHC), HIES and labour force surveys (LFS). These activities were implemented to ensure that Pacific statistical collections meet international standards and data user needs for regionally comparable statistics and data analysis. Regional standardisation has also provided a solid platform for microdata harmonisation to be exploited.

Harmonised datasets – datasets that are structurally the same, stored in the same format and utilise the same variable name, coding and classification conventions – present an opportunity for rapid analysis of Pacific development microdata to produce comparable indicators over time and across different statistical collections and countries. Commonly structured and coded microdata facilitates efficient analysis, as analysis processes undertaken for one dataset can be applied across all harmonised datasets.

Data harmonisation refers to all efforts to combine data from different sources and provide users with a comparable view of data from different studies.[1] Harmonisation is about making statistics and data more comparable, consistent and coherent.[2] A key consideration in data harmonisation is to find pragmatic ways of making compatible and integrable datasets.

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.[3] Harmonisation also seeks to bring together various types, levels and different sources of data in such a way that they can be made compatible and comparable, so that they can be queried and analysed and thus useful for decision-making. In addition, harmonisation may also facilitate more efficient analysis as the results of analysis can be generated across multiple harmonised datasets rather than individually.

Harmonised microdata will provide an efficient and cost-effective solution to meet data needs regularly expressed by various stakeholders. This approach will allow SPC to be more efficient in response to data user requests for regionally comparable statistics and in providing data to various data dissemination platforms. This approach will also be highly complementary to other ongoing data initiatives of SPC, including the Pacific Data Hub, PopGIS[4] and .Stat Suite,[5] and will foster research work and reporting against Sustainable Development Goal (SDG) indicators.

Defining data harmonisation

Data harmonisation is defined as 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.

Harmonisation 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 harmonisation 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 harmonisation will identify a set of core data elements (data elements expressed using different descriptions but with identical meaning).

Objective and goal of data harmonisation

SPC's Statistics for Development Division (SDD) along with Pacific Island national statistics offices produce a large number of statistics on a range of topics. Harmonisation of Pacific Island statistics involves bringing together 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

[1] https://www.icpsr.umich.edu/web/pages/DSDR/harmonization.html
[2] https://gss.civilservice.gov.uk/guidance/harmonisation/
[3] http://www.fao.org/3/X0587E/x0587e09.htm
[4] https://sdd.spc.int/mapping-popgis
[5] SPC has set up a standards-based, user-friendly regional database (https://stats.pacificdata.org/) to facilitate the storage and dissemination

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 harmonisation 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 the harmonisation of data produced via SDD's priority statistical collections, namely PHC and HIES. This guide also covers the harmonisation 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 harmonisation 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 harmonisation exercise, practical recommendations for data harmonisation, cases and application of the results of data harmonisation exercises and details on the organisational and procedural processes necessary to achieve data harmonisation.

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 harmonisation (or standardisation) of census and survey methodologies in the Pacific

of key development indicators.

region. That is, regionally standardised instruments and methodologies have been designed for the 2020 round of PHC, HIES and LFS.

These harmonisation 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 *ex ante* statistical collection harmonisation, to achieve an objective of improving accessibility and dissemination of data, these guidelines aim for *ex post* statistical collection harmonisation of census and survey datasets. This involves harmonising relevant datasets to facilitate, among other things, regional reporting against SDG indicators and comparable statistical analysis over time and across the Pacific region.

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

Overview of data harmonisation

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 SDD, which 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:

1. of high quality, as they are produced using tried and tested methodologies;

2. consistent with international recommendations, as SDD closely collaborates with development partners and subject matter specialists to ensure the data meet user requirements and standards;

3. quickly processed, as established data processing systems are exploited;

4. quickly released, as established tabulation and reporting templates are used;

5. comparable over time, between surveys and across the Pacific region; and

6. 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 Censuses](#) 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, which included Pacific statisticians and development partners. The Technical Working Group 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:

1. 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;

2. 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

3. 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:

1. the [Multiple Indicator Cluster Surveys](#) (MICS) programme that is implemented by PICTs with technical contribution from UNICEF, the United Nations Population Fund and SDD;

2. the [Demographic and Health Surveys](#) programme that is implemented by PICTs with technical contribution from the United Nations Population Fund and SDD;

3. labour market data where standardised modules have been developed for their inclusion in PHC and HIES, and LFSs that are implemented by PICTs with technical contribution from the International Labour Organization (ILO) and SDD;

4. disability surveys that are implemented by PICTs with technical contribution from UNICEF and SDD; and

5. 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 the Census and Survey Processing System (CSPro) for paper-based surveys and Survey Solutions for CAPI-based surveys. It is noted that MICS uses a CSPro 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 (a) the Pacific Standard Classification of Occupations (PACSCO) 2016, (b) the Pacific Standard Industrial Classification of All Economic Activities (PACSIC) 2014, and (c) 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 makes 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 improves the efficiency of, *ex post* statistical collection microdata harmonisation initiatives, such as those for which these guidelines are being written.

© SPC_Mecki Kronen

CHAPTER 2: DESIGNING A HARMONISED DATASET

Introduction

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

- the World Bank (Global Consumption Database) – to derive comparable consumption expenditure weights for the International Comparison Programme;

- UNICEF (MICS Tabulator) – for easy analysis of MICS surveys across countries and time;

- IPUMS (Integrated Public Use Microdata Series) – to have a set of data from PHCs in which the same code has the same meaning in all times and places;

- ILO (ILOSTAT – ILO's online database on labour statistics) – to derive comparable microdata sets on labour market statistics from labour force and other related household surveys and PHCs;

- the World Bank's East Asia and Pacific (EAP) Team for Statistical Development – to have a consistent database for the analysis of poverty, shared prosperity, and other socio-demographic developments across countries in the EAP region; and

- RIGA (Rural Income Generating Activities project, a collaboration between the Food and Agriculture Organization, the World Bank, and American University [Washington, DC]) – to create household-level income aggregates using a consistent methodology and surveys from more than 15 countries. In particular, the RIGA individual wage employment dataset (RIGA-L) produces comparable

labour market data on wages and working time for rural wage employment across several countries.

There are three approaches used by these organisations for their harmonisation processes:

1. 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 organisation. So harmonisation is automatic and there is no need for translation tables. (World Bank, ILO and RIGA).

   - The 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 treat outliers for the consumption data.

   - ILO reprocesses 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 the 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 organisation. RIGA-L imports unique house-

hold and individual identification variables that are already available in the raw survey data. These are renamed for consistency.

2. Another approach uses preprocessed national microdata sets 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 an integrated, common dataset. (IPUMS and MICS).

- IPUMS receives anonymised microdata files from national statistics offices with data coded into a wide variety of classification schemes. Often, the data are only samples from the full population dataset to preserve anonymity. First, the dataset is standardised 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 using a software application to identify variables that are "common" across them. The harmonisation is done by use of a translation table for each variable that 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.

3. World Bank's EAP Team for Statistical Development has developed a harmonised database of socio-economic statistics constructed from microdata of household surveys across 19 countries in the 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's Global Consumption Database. It has, however, produced four files of harmonised variables: (a) basic information on survey methods used in surveys, (b) standardised data on basic demographic and other socio-economic variables, (c) standardised information on households, and (d) poverty aggregates.

The choice for designing the harmonisation datasets by SPC is the approach described in 2. above: that is, preprocessed input datasets in SPC's custody will be used to generate output harmonised datasets.

In this chapter, we will:

1. describe the input datasets used for the harmonisation process outlined in these guidelines;

2. discuss how common topics among PHC and HIES were identified and how the data dictionary was produced, including identification of common variables among PHC and HIES datasets – these are the variables that are to be harmonised; and

3. explain the construct and use of the translation tables, which provide the metadata for preprocessing input datasets to generate output harmonised datasets, as previously described.

Input datasets

The SDD obtains datasets from Pacific Island countries and territories (PICTs) collected through surveys and PHCs. These datasets, hosted in the Pacific Data Hub – Microdata Library, are to be used as inputs to produce harmonised output datasets using the present guidelines. The harmonisation process translates each input dataset into a harmonised output dataset that has the same structure, variables, variable names, variable labels, codes and values across all input datasets.

There are currently 85 datasets, 50 PHC and 35 HIES, earmarked for harmonisation. The process, however, will be implemented in stages. In this first stage, only 29 core 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 usable dataset was selected. Other datasets will be added in subsequent stages.

A full description of a dataset is done through a codebook that 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 given in Table 1. The name of the variable "relationship to head of household" in the input dataset is "r4_relat" and its answer labels include "1" for the head, "4" for an adopted son/daughter and so on. Such codebooks are essential documents for the design and implementation of the harmonisation process.

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 harmonisation 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 (three for PHC and five for HIES): cover (PHC and HIES), household (PHC and HIES), person (PHC and HIES), income (HIES only) and expenditure (HIES only). However, the corresponding harmonised dataset will have only four sub-datasets:

- one household sub-dataset (PHC and HIES) covering data from the input household dataset but also incorporating relevant variables from the "cover" sub-dataset;

Table 1: An extract from a dataset codebook

Variable name	Variable label	Answer label	Answer code	Variable type
Id		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_relat	Relationship			long
		Head	1	
		Spouse	2	
		Biological son/daughter	3	
		Adopted son/daughter	4	
		Son in law/daughter in law	5	
		Brother/Sister	6	

© SPC_Gaelle Le Gall-Quegueineur

- one person sub-dataset (PHC and HIES) covering data from the input person dataset but also incorporating some of the relevant variables from the cover sub-dataset;

- one income sub-dataset (HIES) and one expenditure sub-dataset (HIES) covering, respectively, data from the input income and expenditure datasets, both also incorporating some relevant variables from the household dataset.

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 harmonised 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 dataset. For example, the topic "livelihood choices" appears in only a few person datasets and so was not selected for the harmonised person dataset. In contrast, income is a topic in every HIES

person dataset and so was identified even though it rarely appears in the PHC person dataset. The topics identified for the harmonisation dataset are listed in Annex 4, along with some different types of datasets used in their identification.

Data dictionary

Variables are then identified relating to each of the common topics for use in the harmonised dataset. For each variable, a variable name, variable label, and variable values (codes or numeric) are specified. Also, relevant notes, more detailed variable descriptions, 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), while the expenditure data dictionary deals with expenditure items 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 harmonisation exercises. The data dictionaries are the full and complete description of the harmonised 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. The variable "area of residence" has "rururb" as its name and the codes are "1" for urban and "2" for rural. This extract is in fact common to all four data dictionaries, the full details of which are in Annex 5. The data dictionaries are living documents that 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 and notes
HH identification and location of HH	country	Country ID *ISO 3166 3-digit code*	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	HH identifier	Unique identification of HH. 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 for 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 *1 = Urban* *2 = Rural*	Urban–rural differentiation as determined at national level
Sampling details	ea	Enumeration area	Identification of area of sampling to which HH belongs
	stratum	Stratum	Code of stratum from sample design
	psu	Primary sampling unit	An identification of the PSU to which the HH belongs. Important for computation of sampling errors
	hhwt	Household weight	Weight assigned to each HH for use with HH-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 HH
HH composition	hhsize	Household size	Number of regular members of HH, excluding domestic help, paying boarders and visitors
	hhsize_m	Number of male HM	Total number of male HH members
	hhsize_f	Number of female HM	Total number of female HH members

Translation tables (categorical variables)

The next step in the harmonisation process is to determine how variables and their values (codes and numeric values) from the input dataset are best "transported" to equivalent variables and values in the harmonised dataset.

For a categorical variable (one whose values are categories represented by codes), it is necessary to determine how the codes for that variable in the input dataset are to be recorded against a unique code for the equivalent variable in the harmonised 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 in the harmonised dataset 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 of the corresponding variable in the harmonised dataset is "closest" to it. This is a judgemental call that 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 harmonised 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 as metadata. Any other information relating to the variable, the code

and the source of the information if it 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 harmonised dataset. Knowledge of these metadata is important to users of the harmonised dataset to enable their decision-making on whether or not to use the data element in their analysis.

An example of a translation table is given in Annex 6. 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, PHC, 2017) coded into the harmonised code "54" in the harmonised 2017 HIES dataset for Fiji. When the input data code is "5", the note explains that the output code "54" in the harmonised dataset should be understood as referring to lodging houses or hostels. The researcher or user of the harmonised 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 variation of the local materials, the language used to describe them and the limited local language skills of the person preparing the table. A single digit is used for the harmonised codes when there are no sub-categories. In some instances, however, an additional one or two 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 for only 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 were initially three transla-

tion tables for the fishing location reef (submerged reef, outer reef and coastal reef). However, the information available from the codebooks rarely contains this level of detailed information. Thus the three tables were subsequently collapsed into one for reef location. 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 96 translation tables for the household dataset, 66 for the individual dataset, four for the income and eight for the expenditure datasets. These tables were prepared going from the harmonised output dataset to each of the corresponding input datasets. They could be prepared the other way round, going from each of the input datasets into the harmonised output dataset. This would result in one table per variable for each input dataset. The numbers of tables would then be much larger than at present but there would be less need for metadata notes in the tables.

Numeric-valued variables

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 harmonised dataset will be in the same currency as in the input dataset. Grouped numeric variables that have standard groupings across all input datasets can be directly imported as they are into the harmonised dataset, just like classifications. Others have to be "individualised" and then transported as individual data points.

Special cases

Standard classifications

Special treatment is reserved for standard regional/international classifications (e.g. PACSCO, PACSIC and the International Standard Classification of Education or ISCED). In all cases, at the level of analytic interest in these classifications, the values from the input datasets are the same across all datasets. There is thus no need to translate them, and they are used as they are in the harmonised dataset.

Employment

The original set of employment variables, though relevant and mostly available, has been downscaled to a more manageable level at this initial stage. An important factor in this is the use of different frameworks for employment in the input datasets. The majority of them presently use the main activity framework for employment data. In this, activity is categorised based on a combination of activity (what was done or not done), status in employment (employee, employer, self-employed) and institutional sector (private, public). The other input datasets, particularly the more recent ones, use the labour force framework. In this, employment data are collected based on the activity principle of what the person was "currently" doing during the reference period – work or not at work. This difference affects the employment data structure in the input datasets. For example, in the labour force framework, data are separately collected on status in employment and on institutional sector for each job. The main activity framework mixes these up for each job in a way that is not easily separable. So the variables employment status and institutional sector (public or private) cannot be used for the majority of input datasets that use the main activity framework. They have therefore been dropped. One consequence of this is that the classification by employment status is no longer needed. Variables relating to secondary activity/second job have also been dropped to keep the process manageable. These can be added in later stages of the harmonisation process along with others such as time-related underemployment and working time.

CHAPTER 3: CREATING A HARMONISED 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 an example of a linking chart to demonstrate the interaction between the translation tables, the variables in the input dataset and the harmonised output variables.

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 harmonised datasets from them, each input dataset has to be (a) examined to determine to what extent they have been processed, and then (b) further processed to ensure that all the datasets are reasonably at the same level of readiness for harmonisation.

This examination and further processing should be for each variable in the dataset. The elements include:

1. ascertaining that information is available for the harmonisation software application to efficiently access each relevant data element in the dataset. This includes but is not limited to:

   a. appropriate naming of the input 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; and

   d. information on the survey/census methods used including sample design, where relevant – in most cases, survey/census metadata are available via the Pacific Data Hub – Microdata Library.

2. converting datasets to a common format, as datasets come in various application formats (e.g. Stata or SPSS). The choice in SDD is Stata. The pdf and other static information should be converted to XML format so that they can be easily accessed and aligned using software.

3. cleaning and completing the variable values as much as possible. 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); and

   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 harmonisation process. It is expected that these steps would not be extensive and complicated since the datasets have already been processed at national level.

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 will have to be excluded from the harmonisation process.

Metadata

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

harmonisation metadata are:

- information about the codes (from the translation tables);

- issues of comparability, detailed description of the variable and its codes, and 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); and

- 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 the 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 so that they can make intelligent choices of the data used for analysis.

Linking charts

Each input dataset generates a harmonised output dataset in which the codes of input categorical variables are translated into the harmonised codes through the use of the translation tables. The values of numeric and string variables of input datasets are reproduced as they are in the output dataset. The structure and codes of the harmonised output dataset are 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 identify for each variable in the harmonised dataset how it links up to the corresponding variable(s) in the input dataset. This can be done using linking charts. Table 3 is an extract from the table for the household sub-data-

set Cook Islands HIES 2015. 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 harmonised household sub-dataset of three input datasets: COK, HIES, 2015; FSM, HIES, 2013; and PLW, PHC, 2015.

© SPC_Gaelle Le Gall-Queguineur

Table 3: An extract of a linking chart

Harmonised 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	hhsize_m	Number of male HM	Cover	males
H15	hhsize_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 HH 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

CHAPTER 4: USING THE HARMONISED 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 are used in the harmonised output datasets and some limitations in terms of the construct and use of the harmonised datasets.

Major regional and international data requests

The harmonised 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 the other direct needs of these agencies. Two agencies in particular will benefit from use of these harmonised datasets: the ILO and the United Nations Educational, Scientific and Cultural Organization (UNESCO).

Labour market data required by the ILO

The ILO, through its Department of Statistics, requires labour market and other data and metadata of member states. The data are obtained:

1. 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 underutilisation including unemployment, industrial relations, occupational safety, labour administration and the working poor;

2. by directly harvesting data from websites of member states and other repositories;

3. by compiling and reprocessing microdata from LFS of member states to populate a harmonised database of labour market indicators.

As can be seen from Annex 7, over 70% of the required indicators in 1. are obtainable from LFS,

PHCs and HIES. Therefore, these could be accessed from the SPC harmonised 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 from the SPC harmonised datasets.

The data obtained through 1. and 2. are disseminated by the ILO directly as collected, with minor editing, as national statistics. Given the above harmonisation process, the data and indicators in the SPC harmonised datasets are also national statistics. The statistics and indicators generated from 3. are disseminated as ILO statistics.

Regular data required by UNESCO

The statistics department of UNESCO, the UNESCO Institute for Statistics (UIS), 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, while most household surveys, especially HIES and LFS, have educational attainment data. The SPC harmonised 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 either as preferred or secondary sources. 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), four of which could be obtainable from a PHC or household survey. Thus, to some extent, the SPC harmonised datasets could be a source to compute these indicators.

Relevant classifications

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

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)

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 categorised 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 most recent version of this classification, ISIC Rev. 4,[6] uses a 4-level hierarchical structure consisting of:

- 21 **sections** at the highest level, labelled alphabetically (see Annex 11), then organised into successively more detailed categories, which are numerically coded into:
    - 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 standardised 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 the 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 United Nations Statistics Division, the custodian of the classification, has correspondence tables to go from the older to the newest version.[7]

The regional equivalent developed for the Pacific region is 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 harmonisation process. It is thus convenient that there are no differences. The main differences between the two classifications are at the class level where:

---

[6]Approved by the United Nations Statistical Commission in March 2006.
[7]Department of Economic and Social Affairs, Statistics Division, International Standard Industrial Classification of All Economic Activities Revision 4, United Nations, New York, 2008.

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

Also, PACSIC uses the new ISCED 2011 for the section on education instead of the old ISCED 1997, currently used in ISIC Rev. 4.

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 organises 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 by the Governing Body of the ILO and then endorsed by the International Conference of Labour Statisticians and the United Nations Statistical Commission, also in 2008.[8]

In this classification:

- an occupation is defined as a "set of jobs whose main tasks and duties are characterised by a high degree of similarity"; and
- 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).

[8]ILO, ISCO-08, https://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf

The major groups are listed below in Table 4.

Table 4: International Standard Classification of Occupations (2008)

ISCO-08 (Major groups)	
Code	Title
1	Managers
2	Professionals
3	Technicians and associate professionals
4	Clerical support workers
5	Service and sales workers
6	Skilled agricultural, forestry and fishery workers
7	Craft and related trades workers
8	Plant and machine operators, and assemblers
9	Elementary occupations
0	Armed forces occupations

Except for the last, armed forces occupations, the major groups are in descending order of skill level; hence, professionals are higher skilled than clerical support staff.

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 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 classifi-

cation 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 ISCED 2011 to determine skill levels instead of the old ISCED 1997 currently used in ISCO-08. These apart, at the 4-digit level (unit group), PACSCO 2016 is comparable to ISCO-08 and can be easily converted to ISCO-08 by aggregating the categories at the most detailed level of PACSCO 2016 (i.e. 6-digit level known as Occupation).

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 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.[9]

ICSE-18 classifies jobs in employment for pay or profit into 10 detailed categories based on the concepts of type of authority and type of economic risk. They are as follows:

INDEPENDENT WORKERS
A   Employers
11 – Em ployers in corporations
12 – Employers in household market enterprises

Independent workers without
B   employees
21 – Owner-operators of corporations without employees
22 – Own-account workers in household market enterprises without employees

---

[9]https://ilostat.ilo.org/resources/concepts-and-definitions/classification-status-at-work/

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.

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. At this initial stage of the harmonisation process, this classification is in fact not used, as the variable status in employment has been dropped. However, there are indications that from 2019 onwards labour market modules in most Pacific surveys/censuses are not only moving to the labour force framework but also adopting ISCE-18. So at the next stage of harmonisation, this variable and its classification will surely be included.

The International Standard Classification of Education (ISCED)

The ISCED categorises education programmes[10] as well as classifies a population separately in terms of their level and field of education.[11] 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. The classification according to fields of education and training, 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. At the 1-digit level, it classifies the educational attainment of a population into nine 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.

---

[10]UNESCO Institute for Statistics, 2012, International Standard Classification of Education, ISCED 2011, http://www.uis.unesco.org
[11]UNESCO Institute for Statistics, 2015, International Standard Classification of Education, ISCED Fields of Education and Training 2013 (ISCED-F 2013), http://www.uis.unesco.org

Each of these levels is further subdivided 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.[12]

ISCED-F uses 10 categories to classify fields of education at the first level, as follows:

	Broad field
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

The categories are not in any hierarchical order. They are then further subdivided into about 56 narrow areas and a large number of detailed fields.[13] 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.

There is as yet no regional equivalent for ISCED.

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 expenditures 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 has a 4-level hierarchical structure of:

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

Table 5 describes the different divisions of COICOP. These divisions cover the entire spectrum of consumption expenditures and so facilitate economic analysis.

---

[12]http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf
[13]UNESCO Institute for Statistics, 2015, International Standard Classification of Education, ISCED Fields of Education and Training 2013 (ISCED-F 2013), Appendix I: ISCED fields of education and training, http://www.uis.unesco.org

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 (NPISH)
15	Individual consumption expenditure of general government

COICOP is an essential classification for the description and analysis of expenditure statistics from HIES and living standards, etc. It is also used in the SNA, in the computation of weights for consumer price indices and in the International Comparison Programme.

The regional SPC equivalent is the PACCOICOP 2012. It was derived from COICOP and maintains the basic framework of the classification up to the sub-class 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.

The Pacific Classification of Income (PACCOI)

There is at present no international classification of income. As mentioned earlier, there is an unpublished regional classification of income (PACCOI) that is used for income data in the region. This is the reference classification in the harmonisation exercise.

Important limitations and risks

Although the ideal for harmonisation exercises such as this one is comparability of data across countries and across time, this may not be achievable as many pre-output and other factors play a role. There are limitations due to differences in the sources generating the input datasets. Even when the surveys have the same label (e.g. HIES), differences in concepts, definitions, survey design and methodology, field work, editing and data processing methods, other implementation protocols and so on can impact negatively on comparability. The differences between different sources such as HIES and PHC are even more glaring. The recent drive by SDD towards *ex ante* standardisation would be helpful in addressing some of these issues and limiting the need for *ex post* harmonisation.

The quality of the data in the harmonised datasets depends completely on that of the input datasets. Consideration must therefore be given to the identification and treatment of statistical errors, both sampling and non-sampling, missing values, outliers, etc., on the statistical estimates from these different input sources and the impact of their importation into the harmonised dataset on quality. Increased ability of PICTs to analyse their own survey and census data would improve the quality of the input datasets, and thus that of the harmonised dataset.

The analysis of the income and expenditure data across space will require some standardisation of currencies, such as the use of purchasing power parity, and across time will require some adjustments for inflation. This is an added complication to the well-known difficulty of and differences in estimating income and expenditure aggregates and indicators (e.g. average per capita expenditure) even within a given input dataset. It is, however, anticipated that whatever methods are used within the input dataset will also have transferred to the output harmonised dataset. The challenge is the comparability across different harmonised datasets

that may be based on different methods for these aggregations and indicators in the input datasets.

The input datasets are the driving force for the harmonisation exercise. The anticipated analysis of data from the harmonised datasets would require a reasonably high frequency in their production for longitudinal analysis, reasonable access to the microdata and their documentation by SPC, especially their codebooks and metadata, and an openness and support from PICTs for the construction of harmonised datasets. The lack of openness from PICTs is evidenced by the lack of codebooks and the data itself for many input datasets in the current exercise. It is understandable that differences between national needs and development challenges as well as the requirements from regional and international organisations may contribute to these issues. However, an appreciation by PICTs of the importance of the work being done at these levels by SPC and others would contribute to a better understanding between them.

Thus the end product of harmonisation that can be expected is a standard set of variables, codes, classifications and metadata that can be consistently used with different sources across space and time. In other words, it would produce a common language for describing and displaying data as a first step towards full comparability.

The use of the harmonised datasets should be determined and clearly spelt out. Will records from them be combined for analysis or will only statistics and indicators produced from each be used? In the case of the former, the risk should be assessed of possible duplication of records from, say, a PHC in a HIES that is conducted quickly afterwards.
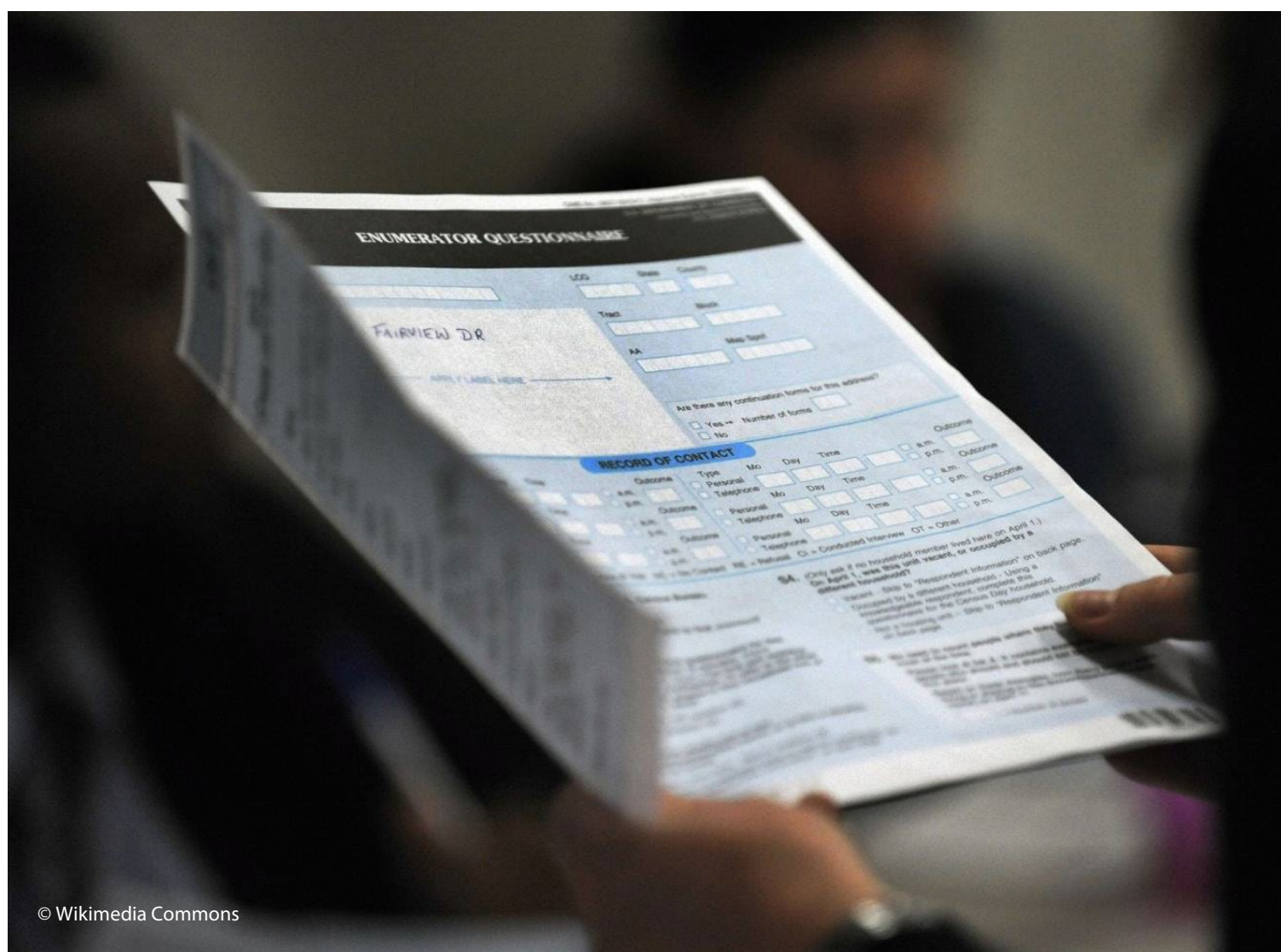
CHAPTER 5: FUTURE WORK

In this final chapter, we identify in steps the future work to be carried out to complete this stage of the harmonisation exercise, including finalising the translation tables, preparing the input datasets, developing software for automating the use of the translation tables and preparing the harmonised dataset, using the dataset and extending the harmonisation process to other/new datasets and new variables.

Finalising the translation tables

Step 1

The current set of translation tables has quite a few variables that exist for only a limited set of input datasets. In fact for one, there is no input dataset with data for it. This is the result of having to choose common variables manually using only a limited set of types of input datasets. The challenge is that there is no easy way to manually get details about the availability of information for all the input datasets in scope in advance. IPUMS in fact identifies the variables for inclusion using a software application. The consequence is having a wide range of variables but with limited range of input datasets rather than a smaller set of variables with a wide range of input datasets. The latter is clearly preferable. To achieve this, an examination of all the translation tables should be carried out to assess those that should be retained in terms of the availability of data for them in the core datasets. A cut-off of 60% could be used. So any table with data for less than 60% of the core input dataset should be put aside unless they are "demand variables".

Step 2

The current set of tables was chosen using a broad sweep of all variables in the input datasets. There is, however, a need at this early stage of the harmonisation exercise for parsimony in the number of tables.

Those tables that can be deferred to the next stage of harmonisation should be excluded in a process similar to what was done for the employment tables.

Step 3

A thorough in-depth review of the remaining translation tables should be implemented to ascertain the correctness of the details of the input datasets used for them. In particular, the possible use of a more appropriate input variable than the one currently used should be assessed.

Preparation of the input datasets

Step 4

Each core dataset should undergo the process of preparation, as outlined earlier, for each of the variables in scope. The interval or continuous variables as well as the string variables in the data dictionaries should be reassessed in terms of whether they should be included in the harmonisation process à la the above step for the categorical variables. Are they "in demand" variables? Should they be included at this initial stage?

Automate the use of the translation tables and prepare the harmonised output datasets

Step 5

Software development for:

- identifying the variables selected in steps 1, 2 and 4 in the input datasets;
- translating the codes for the categorical variables according to the translation tables;
- transporting the selected numeric and string variables into the output harmonised dataset; and
- associating and incorporating all relevant metadata for the input dataset, the variables

and the codes into the harmonised dataset using XML.

Testing the new harmonised dataset

Step 6

After development of the harmonised databases, some analysis should be carried out to assess the usefulness of these datasets to produce (a) the desired SDG indicators, (b) some of the indicators required by regional and international agencies, and (c) statistics and indicators needed for SPC policy-making.

As part of the overall process, documentation of the work should be prepared.

Extending the work

Step 7

Based on the above analysis, (a) add the variables removed in steps 1, 2 and 4 for the core input datasets; (b) extend the exercise to the rest of the 85 input datasets based on the same variables; (c) bring in new input datasets from other exercises, especially LFS and then MICS and Demographic and Health Surveys; and (d) add new variables for the core, extended and new datasets.

© TDoS

ANNEXES

Annex 1: Identifying and treating outliers, other suspect values and missing values

Suspect values

Suspect values are usually inconsistent with the expected range or type of values; for example, a reported age of 2 years for a value of 120% for labour force participation rate 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 organisations, 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 (the 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, such as having an incorrect trailing zero added.

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

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

Outliers: Identification

Outliers are possibly legitimate values that, 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 requires 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); and
- the sum of the 75th percentile and 3 times the interquartile range as the upper limit and the 25th percentile less 3 times the interquartile range for the lower limit (the World Bank's Global Consumption 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, while the latter could be determined 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 (e.g. values below 3 standard deviations in the first example), there must be values outside of it, some of which could be legitimate. Indeed, they are such that values outside the range will always exist. 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 harmonisation. The choice, per variable,

should depend on the nature of the variable and the shape of its distribution.

Missing values: Identification

Missing values 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's Global Consumption 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. 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 harmonised 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 datasets, 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 harmonisation 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 standardised 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 standardised for input into a harmonised 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 harmonised way.

One simple method of doing this is to use the midpoint 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 = 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 midpoint) 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	Nauru	NRU	HIES	2012	with codebook	Most recent
12	Nauru	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	Papua 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

Annex 4: Identification of common topics for HIES and PHC (selected PICTs)

Included: (yellow) Excluded: (black)

Unit	Topic	Category	HIES COK 2015	HIES FSM 2013	HIES TUV 2015	HIES FJI 2009	PHC FSM 2010	PHC TUV 2017	PHC KIR 2018
Household	HH identification and location of HH	Other	Included	Included	Included	Included	Included	Included	Included
	Sampling details	Other	Included	Included	Included	Included	Excluded	Excluded	Excluded
	HH composition	Demographic	Included	Included	Included	Excluded	Included	Included	Excluded
	Income and expenditure classes	Economic	Included	Included	Included	Included	Excluded	Excluded	Included
	Dwelling characteristics – type, year	Other	Included	Included	Included	Included	Included	Included	Excluded
	Construction materials	Other	Included	Included	Included	Included	Included	Excluded	Excluded
	Electricity and utilities	Other	Included	Included	Included	Included	Included	Included	Included
	Sewerage and sanitation	Other	Excluded	Included	Included	Included	Included	Included	Included
	Housing characteristics – number of rooms, etc.	Other	Included	Included	Included	Included	Included	Excluded	Excluded
	Tenure and rent	Other	Included	Excluded	Included	Included	Included	Included	Included
	Assets	Economic	Included	Included	Included	Included	Included	Included	Included
	Communication – access to phone and IT	Social	Included	Included	Included	Included	Included	Included	Excluded
	Agriculture, fisheries, livestock activities	Economic	Included	Included	Included	Excluded	Included	Included	Included
	HH livelihood standards re choices of goods and services	Economic	Excluded	Excluded	Included	Excluded	Excluded	Included	Included
	Fertility	Demographic	Excluded	Excluded	Excluded	Excluded	Excluded	Included	Included
	Mortality	Demographic	Excluded	Excluded	Excluded	Excluded	Included	Included	Included
	Family labour force status	Economic	Excluded	Excluded	Excluded	Excluded	Included	Excluded	Included
	Location of family members and remittances	Economic	Excluded	Excluded	Excluded	Excluded	Included	Included	Included
	Natural disasters	Social	Excluded	Excluded	Excluded	Excluded	Excluded	Included	Included
	HH income	Economic	Excluded	Excluded	Excluded	Excluded	Excluded	Included	Included
Person	Person ID – HH member ID, EA and municipality	Other	Included	Included	Included	Included	Included	Included	Included
	Sampling details	Other	Included	Included	Included	Included	Excluded	Excluded	Excluded
	Basic demographics – age, sex, marital status, relationship, residence, etc.	Demographic	Included	Included	Included	Included	Included	Included	Included
	Family structure	Demographic	Excluded	Excluded	Excluded	Excluded	Included	Included	Excluded
	Fertility and mortality	Demographic	Excluded	Excluded	Excluded	Excluded	Included	Included	Included
	Ethnicity, citizenship and nationality	Social	Included	Included	Included	Included	Included	Included	Included
	Migration	Demographic	Excluded	Excluded	Excluded	Excluded	Included	Included	Included
	Employment	Economic	Included	Included	Included	Included	Included	Included	Included
	Other work	Social	Included	Included	Included	Excluded	Included	Excluded	Included
	Education	Social	Included	Included	Included	Included	Included	Included	Included

Literacy	Social	
Health including anthropometrics	Social	
Disability	Social	
Communication status – IT and phone use	Social	
Personal habits – alcohol, tobacco, etc.	Social	
Livelihood choices – basic goods, services and other financial resources	Social	
Involvement in activities – agriculture, fishing, livestock, HH business	Economic	
Details relating to previous HH members	Economic	
Income and expenditure classes	Economic	
Income and remittances	Economic	

Expenditure

HH identification	Other	
Sampling details	Other	
Item identification – line number, HIES code, description, etc.	Other	
Classification – COICOP, expenditure category	Other	
Value – raw, amended, annual, original, etc.	Economic	
Type of expenditure and product	Economic	
Location, durability and beneficiary	Economic	
Transaction date	Economic	
Survey details – date of interview, diary schedule	Other	
HH composition	Demographic	
Income and expenditure classes	Economic	

Income

HH identification	Other	
Sampling details	Other	
Item identification – line number, HIES code, description, etc.	Other	
Classification – PACCOI, income category, COICOP	Other	
Value	Economic	
Type of income	Economic	
Location of provider and beneficiary	Economic	
Transaction date	Economic	
Survey details – date of interview, diary schedule	Other	
HH composition	Demographic	
Income and expenditure classes	Economic	

Annex 5: Identification of common topics for HIES and PHC (selected PICTs)

Expenditure item file	Topic	Variable name	Variable label and codes	Description and notes
	HH identification and location of HH	country	Country ID *ISO 3166 3-digit code*	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 for 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 *1 = Urban* *2 = Rural*	Urban–rural differentiation as determined at national level
	Sampling details	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 unit 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
	HH composition	hhsize	Household size	Number of regular members of household, excluding domestic help, paying boarders and visitors
		hhsize_m	Number of male HM	Total number of male household members
		hhsize_f	Number of female HM	Total number of female household members
	Item identification	line_num	Line number	
		exp_id	Expenditure item identification	Concatenation of hid and line_num
		item_descr	Description of item	Description of the expenditure item
		exp_cat	Category of expenditure *100 Consumption expenditure of households* *101 Food and non-alcoholic beverages* *102 Alcoholic beverages, tobacco and narcotics* *103 Clothing and footwear* *104 Housing, water, electricity, gas and other fuels* *105 Furnishings, household equipment and routine household maintenance* *106 Health* *107 Transport* *108 Information and communication* *109 Recreation, sport and culture* *110 Education services* *111 Restaurants and accommodation services* *112 Personal care, social protection and miscellaneous goods and services* *113 Income in kind/Own consumption* *114 Imputed rent*	

				115 Other consumption expenditure 200 Non-consumption expenditure of households 201 Donation to other households 202 Donation to church and other religious organisations 203 Other donation 204 Mortgage payment 206 Taxes and fines 207 Purchase of land or house 208 House construction and major improvements to house 300 Other expenditure 301 Plant or equipment 302 Primary production operation costs 303 Non-subsistence business operating costs 998 Don't know 999 Missing	
		exp_class	Expenditure classification *COICOP or PACCOICOP (4-digit code)*	Classification of Individual Consumption According to Purpose (COICOP), 2018. Has 15 divisions, 63 groups, 186 classes and 338 sub-classes. At this level, same as PACCOICOP	
		eType	Expenditure type *10 = Cash purchases* *11 = Own use home production* *12 = Cash – gifts* *13 = Home produced – gifts* *14 = Imputed rents* *15 = In-kind receipts (employer)* *16 = Gift received* *20 = Intermediate expenditures* *21 = Int exp – home production* *22 = Int exp – primary activities* *23 = Int exp – business* *98 = Don't know* *99 = Missing*		
		code	HIES code		
		coicop	Commodity code		
		provider	Location of provider *1 = In-country* *2 = Overseas*		
		rawamt	Amount as collected	This relates to the period covered as indicated in questionnaire for non-diary items	
		multiply	Annual multiplier	The required value to expand to a year	
		annualamt	Annual amount	The original value expanded to cover a year	
		flagimp	Imputation occurred *1 = Yes* *2 = No*		
		impute	Note on imputation	As in original dataset	
		transdate	Transaction date *DD/MM/YYYY*	Date of the expenditure	
		qntyfood	Raw quantity from diary	For food items from diary	
		unitfood	Original unit from diary	For food items from diary	
		convertfood	Quantity converted to grams	For food items from diary	
		estvaluefood	Estimated value of expenditure	For food items from diary	
		diaryday	Weekday of transaction *1 = Monday* *2 = Tuesday* *3 = Wednesday* *4 = Thursday* *5 = Friday* *6 = Saturday* *7 = Sunday*		

		diarymonth	Month of transaction *1 = January* *2 = February* *3 = March* *4 = April* *5 = May* *6 = June* *7 = July* *8 = August* *9 = September* *10 = October* *11 = November* *12 = December*	
		diaryyear	Year of transaction	
		diaryweek	Week of transaction *1 = Week 1* *2 = Week 2*	
		benefit	Beneficiaries *00 = None* *51 = All* *List of HH member ID* *52 = Other HH* *53 = Church* *57 = Previous HH member 1* *58 = Intermediate expenditure* *61 = Previous HH member 2* *62 = Previous HH member 3*	If no HH member benefits from expenditure, benefit = 00. If all benefit, benefit = 51. Otherwise, indicate HMid for all HM benefiting from income (start with owner first).

Annex 6: Translation table

Harmonised dataset		Input datasets																							
Country		COK	FSM	NIU	PLW	FSM	SLB	FSM	FSM	FJI	FJI	KIR	MHL	MHL	NRU	NRU	PLW	SLB	TKL	TON	TON	TUV	VUT		NIU
Name of survey or PHC		HIES	HIES	HIES	PHC	PHC	PHC	HIES	PHC	HIES	PHC	HIES	HIES	PHC	HIES	PHC	HIES	HIES	HIES	HIES	PHC	HIES	HIES		PHC
Year of survey or PHC		2015	2013	2015	2015	2010	2009	2005	2000	2002	2017	2019	2019	2011	2012	2011	2014	2012	2015	2015	2016	2015	2019		2017
Variable label		*Type of dwelling (HHD)*																							
Variable name	dw_type	hq20111	q20111	hq20111	A2_VS1	A6_VS1	H1_VS1	H1_VS1	TYPE_VS1	units_structure	tlqtrs	g1_building	h1101	h11_resp	type_of_dwelling	H1414_VS2	building	recd_q20111	h1421	hc_q20111	hq20111	g1a_building	hq20111	h801	h9_building
Answer labels		Answer codes																							
One family house detached from any other house	10	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
One family house attached to one or more houses	20	2	2	2	2	2	2	2	2	2	2	2			2	2	2	2	2	2	2	2	2	2	2, 3
Building with 2 or more apartments	30	3	3	3	3–7	2	3,4	3	3–7	3–7		3	2	2		3	3	3	3	3	3	3	3	3	4
sharing kitchen/toilet	31	4	4	4														4		4	4		4	4	
Building attached to a business/other non-resident building	40	5	5	5	11	1	5	5		8	4	4	3	3	3	4	4	5	4	5	5	4–5	5	6	5
Institutional building	50				9																				
School dorm	51																								
Hospital	52																								
Prison	53																								
Hotel	54						6				5, 6						5	5							6
Other	90	6	6	6	8, 10, 11	2	6	7	8–9	9	3	7	4	4	4	6–7, 98	6–8	6	5	6	6	8	6	7	8
Don't know	98															97									
Missing	99															99									

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, Employment and Unemployment and Employment Survey, Admin
8	Unemployment	LFS	PHC, HIES
9	Other labour underutilisation	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)	HIES	
1.2.1	Proportion of population living below the national poverty line, by sex and age	HIES	
1.2.2	Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	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	Time use survey	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 nonagriculture 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	Child labour survey	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 operationalised 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	PHC, HHS
11	Net enrolment rate (NER)	Admin	PHC, HHS
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	PHC, HHS
24	Percentage of female students in each ISCED level of tertiary education	Admin	PHC, HHS
25	Percentage distribution of students in tertiary education by ISCED fields of education	Admin	PHC, HHS
26	Percentage distribution of graduates by ISCED fields of education at the tertiary level	Admin	PHC, HHS
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	PHC, HHS
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	PHC, HHS
36	Expected gross primary graduation ratio (EGPGR)	Admin	PHC, HHS
37	Out-of-school children-primary (OOS)	Admin	PHC, HHS
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	PHC, HHS
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 sources
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	HHS
4.2.2	Participation rate in organised 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	HHS	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) the 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 organised 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)

International Standard Industrial Classification of All Economic Activities (ISIC), Rev. 4 (2008)		
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 activities
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 HHs as employers; undifferentiated goods- and services-producing activities of HHs for own use
U	99	Activities of extraterritorial organisations 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

[a] *of which imputed rental services of owner-occupied dwellings.*

Annex 13: Linking chart for household dataset

Harmonised 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	hhsize_m	Number of male HM	Cover	males
H15	hhsize_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 HH 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	cookenerg	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 – 1st person	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	TT34 - hq20502_6
H56	flcover	Ownership of floor coverings	HHD	TT35 - 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 devices	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 equipment	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 *List of HH member ID*	HHD	*

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 owned	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 *List of HH member ID*	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 *List of HH member ID*	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 giant clam	HHD	TT72 - Various variables
H122	crab	Ownership & number of crab	HHD	TT73 - Various variables
H123	milkfish	Ownership & number of milkfish	HHD	TT74 - Various variables
H124	eel	Ownership & number 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 aquatic animals	HHD	TT83 - Various variables
H133	hmfsh	HH members involved in fisheries *List of HH member ID*	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

Harmonised 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	HH 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	hhsize	Household size	Cover	tot_p
H14	hhsize_m	Number of male HM	Cover	tot_m
H15	hhsize_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 HH 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	TT17 - 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 – 1st person	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 devices	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 equipment	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 *List of HH member ID*	HHD	*
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 owned	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 *List of HH member ID*	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 *List of HH member ID*	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 giant clam	HHD	TT72 - *
H122	crab	Ownership & number of crab	HHD	TT73 -*
H123	milkfish	Ownership & number of milkfish	HHD	TT74 - *
H124	eel	Ownership & number 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 aquatic animals	HHD	TT83 - *
H133	hmfsh	HH members involved in fisheries *List of HH member ID*	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 - *

Harmonised household dataset (PLW, HIES, 2015)			Input household dataset (PLW, HIES, 2015)	
Variable number	Variable name	Variable label	Data file	Variable name
H1	country	Country ID		PLW
H2	datasourc	Name of survey or PHC		PHC
H3	year	Year of survey or PHC		2015
H4	hid	HH identifier	HHD	HLDNUM_NUM_VS1
H5	geolev1	Sub-national code level 1	Cover	STATE_VS1
H6	geolev2	Sub-sub-national code level 2	Cover	HAMLET_VS1
H7	rururb	Area of residence		*
H8	ea	Enumeration area	Cover	EA_VS1
H9	stratum	Stratum	Cover	*
H10	psu	Primary sampling unit		*
H11	hhwt	Household weight	HHD	*
H12	indwt	Individual weight		*
H13	hhsize	Household size	Cover	CS8_3_VS1
H14	hhsize_m	Number of male HM	Cover	CS8_1_VS1
H15	hhsize_f	Number of female HM	Cover	CS8_2_VS1
H16	pcinc_cl	Per capita annual income quintile	Cover	*
H17	inchh	Total annual income	Cover	*
H18	pcexp_cl	Per capita annual expenditure quintile	Cover	*
H19	exphh	Total annual HH expenditure	Cover	*
H20	dw_type	Type of dwelling	HHD	TT1 - A2_VS1 & A6_VS1
H21	tenure	Type of tenure	HHD	TT3 - A3_VS1
H22	builtyr	Year of construction	HHD	A8_VS1
H23	rooms	Number of habitable rooms	HHD	A12_VS1
H24	cookarea	Cooking area	HHD	TT6 - A26_VS1

H25	roof	Main material used for roof	HHD	TT7 - A10_VS1
H26	wall	Main material used for external walls	HHD	TT8 - A9_VS1
H27	floor	Main material used for floor	HHD	TT9 - *
H28	watersrc	Main source of water	HHD	TT10 - Various variables
H29	lighting	Main source of lighting in dwelling	HHD	TT11 - A18_VS1
H30	cookenerg	Main cooking energy	HHD	TT12 - Various variables
H31	toilet	Main toilet facility	HHD	TT14 - A17_VS1
H32	phone	Access to a landline phone	HHD	TT15 - A19_1_VS1
H33	cell	Access to a cell phone	HHD	TT16 - A19_2_VS1 & A19_2_VS2
H34	internet	Access to internet	HHD	TT17 - *
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 – 1st person	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 devices	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 equipment	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 *List of HH member ID*	HHD	*
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 - BS20_VS1 & BS1_4_VS2
H102		Number of parcels owned	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 *List of HH member ID*	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 *List of HH member ID*	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 giant clam	HHD	TT72 - BS15_VS1
H122	crab	Ownership & number of crab	HHD	TT73 -*
H123	milkfish	Ownership & number of milkfish	HHD	BS16_VS1

H124	eel	Ownership & number 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 aquatic animals	HHD	TT83 - BS17_VS1
H133	hmfsh	HH members involved in fisheries *List of HH member ID*	HHD	BS21
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 - BS23_2_VS1
H137	bottom	Bottom fishing	HHD	*
H138	casting	Casting	HHD	*
H139	spearfsh	Spear fishing	HHD	TT87 - BS23_3_VS1
H140	netfsh	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	othfsh	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) / Vocational			Degree or Bachelor				Master Degree					
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)			Senior Secondary (Years 10-13) / TVET Certificate			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 / Associate Degree	Bachelor				Master Degree					
FSM	6-2-4			Pre-school	ECE/ Kinder-garten		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 / Bachelor			Master Degree									
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 / Diploma / Bachelor			Master Degree		Doctorate					
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. / Vocational Certificate	USP Certificate	C or D or AD *(Certificate or Diploma or Associate Degree)* / Diploma / 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						

Pacific Community
Communauté du Pacifique

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 |

ISCED 5-8
Tertiary Ed.
(3 years Degree)

ISCED 3
Senior Secondary School
(Years 11-13)

ISCED 2
Junior Secondary School
(Years 7-10)

ISCED 1
Primary School
(Years 1-6)

ISCED 02
Kindergarten

ISCED 01
Early Childhood Care

Age: 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

Duration in years: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21