

Date: 15 October 2019

Ref: RFQ19-084

Request for Quotation

Terms of reference: Preparation of guidelines for the harmonization of census and survey datasets (ex-post statistical collection standardization)

1. INTRODUCTION

The Pacific Community

The Pacific Community (SPC) (www.spc.int) is the principal scientific and technical organization supporting development in the Pacific region. It is an international organization established by treaty (the Canberra Agreement) in 1947 and is owned and governed by its 26 members, including all 22 Pacific Island countries and territories (PICTs). SPC's headquarters are in Noumea, New Caledonia, and it is one of nine member agencies of the Council of Regional Organizations of the Pacific (CROP). In 2018, SPC had six divisions and four cross-cutting programs.

The Statistics for Development Division

The Statistics for Development Division (SDD) (<http://sdd.spc.int/en/>) of SPC is the main stakeholder in the implementation of the [Ten-Year Pacific Statistics Strategy](#) (TYPSS), which was adopted by PICT leaders in 2009 in recognition of the need for a comprehensive plan to drive improvement and development of statistics in the Pacific region. TYPSS proposed a regional approach to provide regional strategic leadership to improve the scope and quality of national statistics. The 10 year strategy (2010 - 2020) outlined a series of short and long-term programs to contribute to national statistics capacity development and regional coordination and capability. SDD's strategic framework is guided by TYPSS and its current work program by the Business Plan 2018 - 2020.

SDD's goal 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.

SDD plays a significant role for TYPSS as a large and comprehensive technical provider to national statistics offices (NSOs) in the Pacific region. SDD's technical assistance role has been strongly endorsed by NSOs, in particular the Small Islands Developing States (SIDS) where resources are limited and technical assistance and supplementation is essential. As part of a regional organization with a broad technical and social mandate operating across 22 PICTs,

SDD has been able to translate and mediate between international standards and frameworks and Pacific contexts and focus.

SPC Headquarters: Noumea, New Caledonia. Regional offices: Suva, Fiji, and Pohnpei, Federated States of Micronesia.
Country office: Honiara, Solomon Islands.

For contact details – Website: www.spc.int Email: spc@spc.int

Siège de la CPS : Nouméa (Nouvelle-Calédonie), Antennes régionales : Suva (Fidji) et Pohnpei (États fédérés de Micronésie)
Bureau de pays : Honiara, (Iles Salomon).
Site Web : www.spc.int Courriel : spc@spc.int

2. BACKGROUND & PURPOSE

Improving Data Dissemination and Use in Pacific Island Countries

In 2017, the World Bank's Trust Fund for Statistical Capacity Building (TFSCB) approved a grant to improve data dissemination and use in the Pacific region. The TFSCB grant aims to support the achievement of the TYPSS objectives and the implementation of various recommendations formulated in recent reviews of statistical programmes in the Pacific region. The goal 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. This will be achieved by pursuing two specific objectives:

- i. The first objective is to promote and increase the access to, and use of, existing data for research and policy making. This will be achieved by:
 - a. Establishing the [Pacific Microdata Library](#) (a regional data hub in charge of the preservation, cataloguing and dissemination of Pacific development microdata); and
 - b. Producing and disseminating a collection of harmonized datasets obtained by standardizing data from existing surveys and censuses.
- ii. The second objective is to improve the quality and relevance of future data collection by developing a sampling strategy for socio-economic surveys conducted in the region. The emerging and changing data needs (including the demand for new and more disaggregated indicators to monitor the Sustainable Development Goals / SDGs), and opportunities offered by new technologies (such as computer-assisted personal interviewing / CAPI), justify the development of a new approach to sampling adapted to the specificity of the PICTs.

The expected key results of the project are:

- i. The establishment and sustained operation of the Pacific Microdata Library. This central data repository, hosted at SDD, will provide a detailed and searchable inventory of microdata available in the region, and facilitate their access and use. By establishing this data archive, a sustainable centre of excellence in microdata curation and dissemination will be established in the region.
- ii. The production of protocols and technical guidelines for the production of harmonized datasets (ex-post statistical collection data assessment and standardization of survey and census datasets covering multiple fields, such as: demographic, household expenditure, labour, education, health, and others). These data will be particularly useful for comparable spatial (across the Pacific region) and longitudinal analysis and indicator production. They will provide an efficient and cost-effective solution to meet data needs regularly expressed by various stakeholders, and will foster research work and reporting against SDG indicators.
- iii. An optimized sampling strategy for Pacific Island countries and territories, providing a cost-effective approach to meet emerging and changing data needs.

The Contractor will be tasked with the production of protocols and technical guidelines for the production of harmonized datasets (project objective i.b and key result ii).

The broader picture of this initiative is that SPC will standardise census and survey microdata – following the guidelines resulting from this contract – so it can more rapidly produce comparable longitudinal and spatial statistics. For example, if a data user requests statistics on the mean years of schooling across all PICTs, rather than interrogating 22 datasets (which are often of different structure and format), harmonising them and producing the statistics one-by-one, a short programme can be written to extract the relevant data from each harmonised dataset to rapidly produce these statistics. Harmonised microdata will allow SPC to be more responsive to ad hoc requests for regional statistics and will facilitate the development of an interactive user interface that allows for online dynamic data manipulation and tabulation. This will be highly complementary to other ongoing data initiatives of SPC, including the [Pacific Data Hub](#) and [Stat Suite](#).

Harmonization of survey datasets

This component of the project was developed in recognition of significant progress in *ex ante* harmonization of statistical collections in the Pacific region. Regionally standardized instruments and methodologies have been designed for the 2020 round of population and housing censuses (PHC), household income and expenditure surveys (HIES) and labour force surveys (LFS).

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

In consideration of the significant progress made in harmonizing *ex ante* statistical collections, to achieve the objective of improved accessibility and dissemination of data, it is proposed that *ex post* statistical collection harmonization of census and survey data sets is achieved. This will harmonize relevant data sets to facilitate, among other things, regional reporting against SDGs indicators and comparable regional data analysis.

The *ex post* harmonization initiative will develop guidelines for harmonization of the numerous census and survey data sets that are available in the Pacific. The harmonization initiative will aim to meet a wide array of data user needs and will provide an efficient platform to disseminate data and to report against various indicators, including the SDGs.

The objective of the consultancy is to produce a set of guidelines for harmonisation of census and survey microdata, including: data cleaning; dealing with missings/outliers; and construction of harmonized data sets.

3. TERMS OF REFERENCE

Scope of work

The scope of work covers the writing of guidelines for harmonisation of microdata produced by SPC's [22 Pacific Island country and territory](#) (PICTs) member states. The guidelines will focus on harmonisation of data produced via SDD's priority statistical collections, [Population and Housing Census](#) (PHC) and [Household Income and Expenditure Survey](#) (HIES). We would, however, welcome the inclusion of other collections, including: Multiple Indicator Cluster Survey (MICS); Demographic and Health Surveys (DHS); Labour Force Surveys (LFS); Disability Surveys (Disab.); and Agricultural Census/Survey (Agric.).

The scope of work is broken down into 5 tasks, including:

- i. Identify topics that are common across PHC, HIES and other datasets (i.e., topics/subjects that are common in the cover, person, household, income and expenditure records) to include in the harmonisation guidelines;
- ii. Learn from other microdata harmonisation initiatives, demand for harmonised microdata data, and international standards/classifications;
- iii. Identify variables that are conceptually the same in PHC, HIES and other collections to be harmonised and included under each topic and establish a common variable naming convention (ideally concurring with other harmonised datasets or guidelines and/or with user naming conventions), label, code, description and metadata;
- iv. For every variable, prepare a translation table (i.e., harmonised output values) for all relevant collections (PHC, HIES, others) for every PICT – the idea is to retain the original variable while preparing a truly integrated database;
- v. Preparation of guidelines for harmonisation of census and survey microdata, including data cleaning, dealing with missings/outliers, database structure and dataset naming convention, and construction of harmonised datasets.

More detail on each task is provided below.

Task 1: Identify topics that are common across census and surveys (approximately 5 working days)

Task 1 will require the review of census and survey questionnaires to gain an understanding of the topics that are common across datasets resulting from PHC and HIES. The topics are to be listed in a database (in Excel, or other) and may be generally be categorised as: demographic, migration, social, economic, and other. The data source (PHC, HIES, other) for each topic should be identified for each PICT (e.g., if the topic is education, the database should identify for each PICT the specific PHC and/or HIES that collected education-related data). The list of topics within each category should be comprehensive, which will allow for prioritisation based on data demand and availability.

In order to confine the scope of this task, the PHC and HIES questionnaires from the following PICTs are to be reviewed: Cook Islands, Federated States of Micronesia, Fiji, Kiribati, Marshall Islands, Nauru, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu and Vanuatu; and the period should be from 1999 to present. Based on the [Census and Survey Calendar](#), there are 50 PHC and 35 HIES instruments to review (refer Annex) and identify common topics. Questionnaires can be access via the [Pacific Microdata Library](#), or provided directly by SPC.

The output for Task 1 will be a comprehensive database of topics with data source by PICT.

Task 2: Learning from others (approximately 5 working days)

The objective of this task is to learn from: i) other harmonisation initiatives; ii) data users to identify regular demand for statistics sourced from microdata; and iii) international classifications.

The first objective of this task will be to review approaches to other harmonisation initiatives. We envisage the review would, among others, include: [IPUMS International](#) (particularly the [technical note](#)); East Asia and the Pacific Team for Statistical Development Harmonization Reference Manual (2016; unpublished); the [Global Consumption Database](#) (particularly the technical note: Dupriez, O. (2007) Building a household consumption database for the calculation of poverty PPPs, World Bank); the [Rural Income Generating Activities](#) (particularly the [methodological note](#)); the [ILOSTAT Microdata Processing Quick Guide](#) (incl. Labour Force Survey microdataset pre-processing: a standard ILO methodology); and the [MICS tabulator](#). The review should include: methodological aspects (e.g., IPUMS construct of translation tables); variable name and label conventions; data dictionaries; metadata; and data cleaning and dealing with missings/outliers.

The second objective under this task is to review regular statistics demands that are sourced from census and survey microdata. In order to maintain a reasonable scope for this task (and because it aligns with other activities under the project), the data demand review will focus on: i) labour market data required by [ILO](#); and ii) education data required by [UIS](#). The review should include regular demand for statistics sourced from microdata and data to populate SDGs which ILO and UIS are custodian of.

The third objective of this task is to identify relevant international classifications (e.g., [ISIC](#), [ISCO](#), [ISCED](#), [COICOP](#), PACCOI, and their [Pacific derivatives](#)) to ensure that harmonised variables that are to be developed under Task 3 accord to international classification and standard.

Task 2 will help to guide Tasks 3, 4 and 5. As an output from this task, it is expected that the guidelines (Task 5) will include a chapter or annex documenting the findings of the review of: i) other harmonisation initiatives, including approaches to data cleaning, and dealing with missings/outliers; ii) data users to identify regular demands for statistics sourced from microdata; and iii) international classifications.

Task 3: Identify variables to harmonise (approximately 10 working days)

Task 3 involves identifying common variables, which are conceptually the same, across PHC and HIES. These are the variables for which the harmonisation guidelines will be prepared. By theme, the output of this task will be the preparation of a database (in Excel, or other) including the harmonised variable name, label, description, code, metadata and any other necessary information (e.g., data cleaning, treatment of missings/outliers).

Harmonised coding structures may require differing levels according to the level of disaggregation available across PICT datasets and user needs.

To facilitate this task, SPC will, where available, furnish the Contractor with the codebook for each PHC and HIES that is within the scope of this assignment.

The output from Task 3 will be a comprehensive database of variables, by theme, and corresponding metadata to be included in the harmonisation guidelines.

Task 4: Preparation of translation table (approximately 15 working days)

For each variable identified in Task 3, translation tables (incl. code map from the original variable to the harmonised variable) will be prepared. The translation table will require the use of the codebooks provided under Task 3 and will “map” variables and corresponding codes to the harmonised variable and code structure that is defined in Task 3. The translation tables will form the metadata for the creation of harmonised variables (and eventually, a harmonised database) and are necessary to allow future users of the harmonised database to understand the harmonisation decisions and the potential loss of data granularity.

The output of this task will be a translation table for every variable identified in Task 3, which maps all PHC and HIES variables for all in-scope PICTS to the harmonised variable.

Task 5: Prepare guidelines (approximately 10 working days)

The harmonisation guidelines will direct users to the database that is prepared under Tasks 1, 3 and 4 and it will include the information produced under Task 2. The harmonisation guidelines should highlight the benefit of harmonising microdata and serve as a guide for someone who is physically preparing a harmonised dataset. The guidelines will include data cleaning, dealing with missings/outliers, database structure and dataset naming convention, and construction of harmonised datasets.

Institutional arrangements

The Contractor will work under the direct supervision of Mr. Michael Sharp, Economic Statistics & Microdata Specialist, in consultation with Mr. Epeli Waqavonovono, Director, and Mr. Scott Pontifex, Microdata Curator and Dissemination Adviser.

Place of Assignment

Majority of the work will be home based. The consultancy may require travel to Noumea, New Caledonia, however this is to be mutually agreed between the Contractor and SPC.

Duration of assignment

This nine week (45 working days) short-term consultancy contract will be completed over the period of 1 November 2019 through to 16 February 2020.

Qualifications of the successful Contractor

The Successful Contractor will have:

- i. A relevant degree in statistics, demography, economics, data science, or equivalent, preferably at the post-graduate level.
- ii. 5-years of experience working in implementation of census and surveys, preferably in an international (ideally, Pacific) context.

- iii. 5-years of experience analysing microdata generated through census and survey and an understanding of how harmonised datasets are constructed (preferably in Stata or similar software).
- iv. Experience in the preparation of harmonised datasets, mapping microdata, or work-related to microdata harmonisation is desirable.
- v. A demonstrated understanding of the TORs and the required outputs.
- vi. Good time management and the ability to meet deadlines.
- vii. Meticulous attention to detail and good understanding of code structures and mapping

4. EXPECTED OUTPUTS

The expected outputs and timeline are provided in the below table.

Task	Output	Deadline	Estimated number of days payment
Signing of contract	Upon signing of contract (20% advance payment)		9 working days
Identify common topics in PHC and HIES	Database of topics with data source by PICT	20 November 2019	5 working days
Learning from others	Documentation of: <ul style="list-style-type: none"> i. other harmonisation initiatives, including approaches to data cleaning, and dealing with missings/outliers; ii. data users to identify regular demands for statistics sourced from microdata; and iii. international classifications. 	30 November 2019	10 working days
Identify variables to harmonise	Database of variables, by theme, and corresponding metadata to be included in the harmonisation guidelines	15 December 2019	6 working days
Draft translation tables	Provide draft translation tables for a sub-set of variables for methodological review by SPC	15 January 2020	
Final translation tables	Translation table for every variable identified in Task 3, which maps all PHC and HIES variables for all in-scope PICTS to the harmonised variable	31 January 2020	10 working days
Draft harmonisation guidelines	Submission of Draft harmonisation guidelines to SDD for review	7 February 2020	
Final harmonisation guidelines	Final harmonisation guidelines	16 February 2020	5 working days

Submission of quotations, remuneration and travel

If you are interested in this assignment, please send in a proposal to Mr Michael Sharp, Economic Statistics and Microdata Specialist (michaels@spc.int) with copy to Mrs Sandra Gianini, Finance and Administration Officer (sandrag@spc.int), **before 4.00 pm Noumea time on the 31 October 2019**. In your proposal, please:

- i. Outline your understanding of the TOR and provide a brief work plan;
- ii. Report against the qualification requirements; and
- iii. Provide a daily rate for consultancy services.

The rate must be commensurate with the qualifications and experience required for the role and represent value for money.

Please treat this document as *Business-in-confidence*; it is to assist you in deciding if you wish to register your interest in taking up this offer.

NOTE: Where required economy class travel via the most economic route and a daily per-diem will be covered under SPC travel rules, and will be arranged on your behalf.

Annex 1

Table 1: Census and surveys conducted by PICT and year

PICT / Year	HIES	PHC	TOTAL	PICT / Year	HIES	PHC	TOTAL
Cook Islands	2	4	6	Papua New Guinea	1	2	3
2001		1	1	2000		1	1
2005	1		1	2009	1		1
2006		1	1	2011		1	1
2011		1	1	Samoa	4	4	8
2015	1		1	2001		1	1
2016		1	1	2002	1		1
Federated States of Micronesia	2	2	4	2006		1	1
2000		1	1	2008	1		1
2005	1		1	2011		1	1
2010		1	1	2013	1		1
2013	1		1	2016		1	1
Fiji (Republic of)	4	2	6	2018	1		1
2002	1		1	Solomon Islands	2	3	5
2007		1	1	1999		1	1
2008	1		1	2005	1		1
2013	1		1	2009		1	1
2017		1	1	2012	1		1
2019	1		1	2019		1	1
Kiribati	2	5	7	Tokelau	1	6	7
2000		1	1	2001		1	1
2005		1	1	2006		1	1
2006	1		1	2011		1	1
2010		1	1	2013		1	1
2015		1	1	2015	1		1
2018		1	1	2016		1	1
2019	1		1	2019		1	1
Marshall Islands (Republic of)	2	2	4	Tonga	3	3	6
1999		1	1	2000	1		1
2002	1		1	2006		1	1
2011		1	1	2009	1		1
2019	1		1	2011		1	1
Nauru	2	3	5	2015	1		1
2002		1	1	2016		1	1
2006	1	1	2	Tuvalu	3	3	6
2011		1	1	2002		1	1
2012	1		1	2004	1		1
Niue	2	4	6	2010	1		1
2001		1	1	2012		1	1
2002	1		1	2015	1		1
2006		1	1	2017		1	1
2011		1	1	Vanuatu (Republic of)	3	3	6
2015	1		1	1999		1	1
2017		1	1	2006	1		1
Palau	2	4	6	2009		1	1
2000		1	1	2010	1		1
2005		1	1	2016		1	1
2006	1		1	2019	1		1
2012		1	1	TOTAL	35	50	85
2014	1		1				
2015		1	1				