



## What is Microdata?

### Context

Microdata are the units of data that aggregate statistics are compiled from. Microdata is data about individual people, households, or organisations and consist of sets of records containing information on individual respondents or other entities. Microdata is the raw data about individual objects such as people, households, events, transactions or organisations as opposed to the aggregated statistics appearing in a published report. Objects have properties which are often expressed as values of variables of the objects. For example, a 'person' object may have values of variables such as 'name', 'address', 'age', 'income'. Microdata represent observed or derived values of certain variables for certain objects. Microdata may also be data about other characteristics of the Pacific Islands such as geographical data.

National microdata is usually available from censuses, surveys and administrative and register data. These data are most commonly collected by the national government or Pacific Island National Statistics Offices (NSO) and access provided by the NSO or the national archive. The data are collected at an individual, household, or institution level as appropriate (Desai and Cowell, 2006).

SPC provides safe access to microdata via its *Pacific Data Hub-Microdata Library* <https://microdata.pacificdata.org/> to enable research and analysis that benefits Pacific Island people. Microdata is a level of data that creates a risk of recognition/identification of individual people, households or organisations and as such must be managed carefully to protect against this risk.

Micro-datasets are tremendously rich sources of information and the scope for analysing them to answer different research questions is considerable. Dissemination of microdata facilitates more sophisticated analysis than is possible based on tables of aggregated data. Allowing researchers from outside the statistical office to access this data dramatically increases the number of analysts who can study the data thus making much greater use of this valuable statistical resource.

Where an NSO goes to the huge expense of running a census or survey, publishing the results in a traditional analytical report, but without subsequently allowing analysts from other organizations access the microdata, a valuable resource is being wasted. Pacific Island NSOs can only ever hope to carry out a small fraction of the analysis that could potentially be of great value in social research and policymaking. The task of the official statistician is to design procedures and select methods which strike a balance between these conflicting imperatives: granting some form of access for researchers to

census microdata while still protecting the confidentiality of the information provided by individual census respondents.

The demand for microdata is growing and becoming more diverse. This diversity is visible in a broadening of the audience but also in the way in which the microdata is being used. Increasingly, as more Open Data information becomes available, users are seeking to combine data from diverse sources for better measurement and impact. Limited quantifiable measures such as web usage statistics obtained from microdata repositories provide anecdotal evidence that this demand is significant. While still predominantly originating from rich countries and international organizations, it appears to be increasing in developing countries. A more systematic and comprehensive assessment of the demand (i.e., by category of users, country of origin, datasets of interest, purpose of use, level of satisfaction, e.t.c) would allow data producers to improve their data dissemination services.

Over the last twenty-five years, statistical offices worldwide have increasingly sought to meet the demand from researchers for greater access to statistical microdata. Pacific Island census and survey microdata however are not as easily accessible to researchers. Although there are some well-established and commonly used procedures technical, administrative and legal which are used to disseminate anonymized census and survey microdata to researchers, they have not been widely used in the Pacific. The small size of many Pacific Island countries such as Small Island States makes anonymization relatively more difficult and standard methods are not always directly applicable.

Microdata is a level of data that creates a risk of recognition/identification of individual people, households or organisations and as such must be managed carefully to protect against this risk. These guidelines have been developed to support producers and users of microdata. Statistical disclosure control methods make it possible for statistical offices to anonymize microdata so that there is a low risk of individuals and households being identified within the data. Such methods make it possible to disseminate microdata to researchers in universities or in government thus more fully exploiting its potential value for social research and policy analysis.

## What is the Pacific Data Hub Microdata Library

The Pacific Data Hub - Microdata Library is a central repository for Pacific Island statistical microdata, reports and documents. It is an online cataloguing and dissemination system of survey and census metadata and microdata. It is a service established to facilitate access to microdata that provide information about people living in Pacific Island developing countries, their institutions, their environment, their communities and the operation of their economies

### *The Pacific Data Hub - Microdata Library*



Discovery  
and re-use



Governance  
and access



Raise  
awareness



Identify  
documentation

**The components of the *Pacific Data Hub - Microdata Library (PML)* include:**

- Acquisition, documentation and dissemination of data (survey and census microdata, geospatial data, and selected administrative datasets).
- Compliance with international standards and best practice (Data Documentation Initiative).
- Standardization of survey and census datasets.
- Access to Pacific Island Microdata for enhanced analysis and research.

### What it does

PML allows researchers to browse, search, compare, apply for access, and download relevant survey and census information from the Pacific Islands region.

PML allows data producers to disseminate survey information in a secured environment, in compliance with their policies and regulations.

The Pacific Data Hub - Microdata Library facilitates access to data collected through sample surveys of households, business establishments or other facilities, population, housing or agricultural censuses or through an administrative data collection processes.

The Library contains supporting documentation from censuses and surveys conducted or supported by SPC, as well as by other international organizations, statistical agencies and other agencies in low and middle-income countries. Metadata are prepared using the Data Documentation Initiative (DDI) standard and the catalog is built using open source software developed by the World Bank.