Data Acquisition

Data Acquisition – describes ways to collect and collate microdata and its associated metadata. Microdata and metadata are generated from various data collection activities such as household surveys, population censuses, and administrative recording systems. Many organisations in the Pacific as part of their work not only capture their own microdata, but also acquire microdata from other sources as well. They can be generated by many official and non-official producers for example Pacific Island National Statistics Offices (NSO), line ministries, researchers, and the private sector.

Proper and secure microdata require National Statistical Offices (NSO) to establish policies and procedures that formally define the conditions for acquisition of their microdata in the form of data sharing or license agreements and terms of use. Transparent licensing of datasets is also necessary for SPC to protect both the data owner and potential users and acquiring data with clearly stated licensing and terms of use reduces uncertainty and thus increases accessibility for new users.

Microdata and metadata are also acquired from other sources including existing data repositories such as an existing inventory of data at SPC and external data collection deposits, that include Microdata collected and compiled by different international, regional and national agencies.

Data archives need to define the scope of their collections. The broader the scope, the greater the usefulness of the collection will be. When established by an NSO, the priority will be to archive the data collected by the agency itself. However, those agencies can play an important role by expanding the scope of their data archive to other official sources. For example, the amount of data the *Pacific Data Hub-Microdata Library* microdata.pacificdata.org houses depends greatly on data producers' such as Pacific Island NSO willingness to share their data.

Initially, preparing datasets for inclusion in the SPC *Pacific Data Hub-Microdata Library* has focused on datasets managed or held by Pacific Island National Statistics Offices (NSO) where the preparation, documentation and cataloguing have been carried out in accordance with the Data Documentation Initiative (DDI) standard. These datasets generally consist of census and household survey data. Much of this work was undertaken since 2012 with the assistance of the Paris 21 and Accelerated Data Program (ADP) funded by the World Bank.

The approach in developing the Pacific Data Hub – Microdata Library has been to only include those data sets where complete metadata are available and where access to

data sets can also be provided. Procedures for getting access to the data are set out in Data License Agreement and Terms of Use. For those data sets held and managed by the SPC, where possible open access will be provided as far as possible in line with the open data initiative. However, for more other datasets and more sensitive data, where SPC does not have ownership of the data, the procedures for obtaining access and the conditions of use are determined by the owner or licensor.

The PDH will continue to acquire new data sets for inclusion as they become available and as other agencies are prepared to publish metadata in the catalogue. Future acquisitions will focus on the backlog of datasets held and managed by the SPC and discussions will be held with other dataset owners to identify material that will be suitable for inclusion. It is expected that new acquisitions will be added to the Library on a regular basis.

Steps involved in Dataset Acquisition for Pacific Island microdata

SPC shall undertake the following activities:

1) The process of identifying suitable and dependable data can be complicated. SPC locates suitable datasets and ensures whether the dataset exists and is available for use, and therefore may be acquired;

2) SPC staff establish a Data License Agreement and Terms of Use for acquiring the Dataset that is agreed to by the Data Provider or owner and SPC. SPC and Data Provider must evaluate the License/Terms-of-Use to ensure it meets their requirements, and that the both parties can comply with them; or

3) SPC Staff acquire the Dataset pursuant to either, (i) an MoU (ii) a Legal Agreement, or (iii) an informal document (email) from an authorised representative of the Data Provider to SPC.

4) The Data License Agreement and Terms of Use include information about how data will be shared, including when the data will be accessible, how long the data will be available, how access can be gained, and any rights that the data provider reserves for using data. They also describe any obligations that exist for sharing data collected and address any ethical or privacy issues and legal requirements with data sharing.

5) Data ownership: Data providers should ensure they are the data owners with rights to deposit data to be shared with SPC Pacific Data Hub. People submitting datasets must have the legal authority to do so.

6) Metadata documentation (such as questionnaires, data descriptions, classifications and definitions) are important pieces of information and must be acquired along with microdata.

7) Identify the data format of the dataset being acquired to ensure safe transfer and integrity of the data. For example, the Pacific Data Hub Microdata Library itself, data are provided in Stata, SPSS and SAS formats. In many cases ASCII versions are also
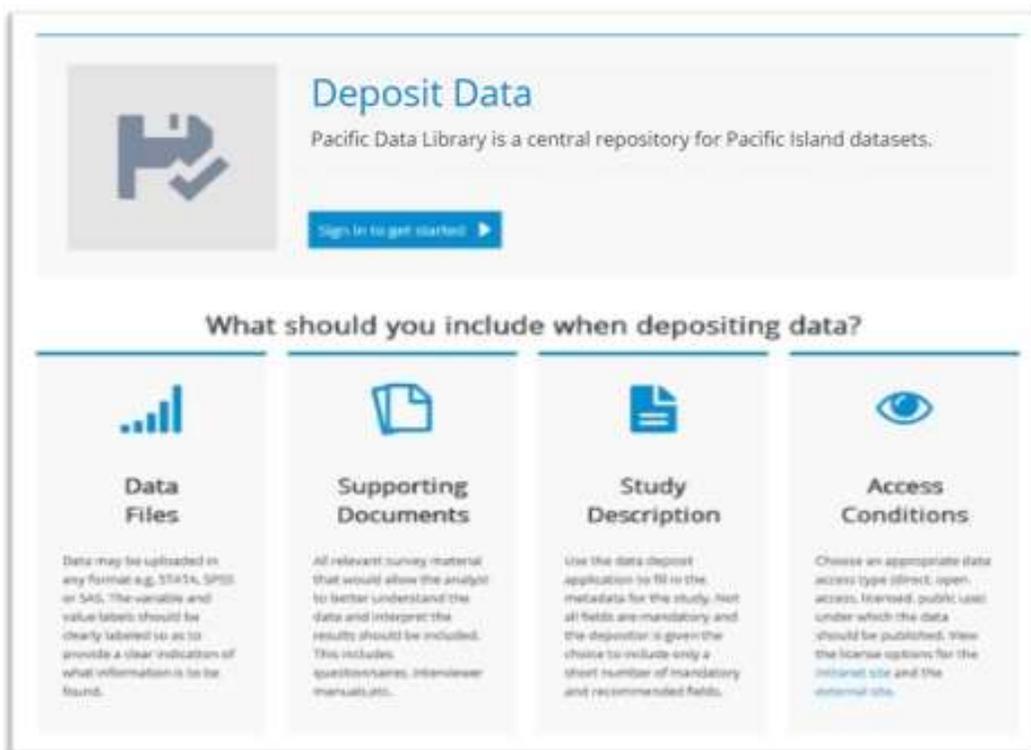
provided with syntax files included for reading the data into SPSS and SAS. If the demand justifies it, we may consider adding other formats. Data provided by external catalogs are under their control. SPC does not offer a service for data conversion however Software like StatTransfer, or the Nesstar Publisher (which is freeware) can be used to convert datasets into other formats.

8) Indicate how the data should be cited by others and address intellectual property and copyright issues.

9) Assure data (conduct data quality checks, undertake disclosure control). Double-check the microdata for completeness and quality (tabulation of aggregates and checking for integrity of dataset). Ensuring the quality of the data is a high priority, for we know that good research is only possible with reliable data.

10) Maintain preservation copies of the data in the long term.

11) Create and publish metadata to assist researchers to discover and use the data.

12) If necessary, modifying data to reduce disclosure risk including r emoval of any direct identifiers such as names, addresses, telephone numbers or any other linkable variables that point explicitly to particular individuals or units and removal of any indirect identifiers. These are variables that can be problematic as they may be used together or in conjunction with other information to identify individual respondents.

13) Limiting access to datasets for which modifying the data would substantially limit their utility or the risk of disclosure remains high

NOTE: SPC reviews all datasets to assess potential disclosure risk. In addition, SPC trains Pacific Island counterparts to apply specified procedures to protect respondent confidentiality in all of the data curates, archives, and distributes. For example, the Pacific Data Hub Microdata team checks each study for identifiers present in the data.

Depositing data:

Transfer from data producer to any type of data archive like the *Pacific Data Hub-Microdata Library* can be a tedious process and therefore any tools which can streamline and support the smooth transfer of digital resources from suppliers into their collections. The use of such tools and processes are likely to benefit both sides. With this in mind, SPC has developed a Data Deposit tool which allows data providers to share their data. http://pdl.spc.int/data-deposit The Data Deposit is DDI-compliant, open-source software and simple, user-friendly, yet powerful and secure on-line system which can be used to securely deposit data with SPC.

Deposit Data

Pacific Data Library is a central repository for Pacific Island datasets.

Sign in to get started ▶

What should you include when depositing data?

Data Files

Data may be uploaded in any format e.g. STATA, SPSS or SAS. The variable and value labels should be clearly labeled so as to provide a clear indication of what information is to be found.

Supporting Documents

All relevant survey material that would allow the analyst to better understand the data and interpret the results should be included. This includes questionnaires, interviewer manuals etc.

Study Description

Use the data deposit application to fill in the metadata for the study. Not all fields are mandatory and the depositor is given the choice to include only a short number of mandatory and recommended fields.

Access Conditions

Choose an appropriate data access type (direct, open access, licensed, public use) under which the data should be published. View the license options for the intranet site and the external site.

Sharing your data with the Pacific Data Hub ensures that your data will be professionally curated and easily accessible to users now and in the future and increases the impact and visibility of your own data collection or research. Sharing your data will ensure your research is cited more widely. We will manage and publish your research data according to global best practice.

The Data Deposit application allows data producers or "depositors" to describe a study and upload the data and supporting documents to a secure file server. While the project is in draft status, the depositor can enter information and edit the project. Once submitted, the deposit joins the curation workflow, gets assigned a reference number, making it trackable. The Depositor can also monitor its progress until it is processed. The application facilitates email communication between the administrator, the depositor and any collaborators while the project is being processed. Finally, the online deposit form has fields that follow the DDI metadata standard and match the fields in the metadata editor. This makes it convenient to import information provided by the Depositor into the metadata editor if available.

Formal acquisition:

The microdata curation process begins when microdata and other information such as relevant supporting documents are acquired from agencies and organisations such as SPC and Pacific Island NSO's that manage or collect data. For example, microdata is often acquired as part of ongoing SPC statistical work in the region with data providers

or sourced from external data collections e.g. national data archives, international organizations that collect survey data e.g. WHO, UNICEF-MICS, DHS

There is a formal acquisition process that needs to take place so that microdata and metadata can be deposited. It is important to organize these data deposits to:

- Ensure that all necessary data and metadata are provided
- Maintain information on what is deposited, by whom, and when
- Provide the person or agency who deposits the data with the opportunity to inform the data archive about access restrictions
- Provide the depositor with a "receipt" indicating what has been deposited
- Acquire datasets from data providers.
- Maintain and build relationships with key data providers

Why deposit microdata?

Sharing your data with the Pacific Data Hub – Microdata ensures that your data will be professionally curated and easily accessible to users now and in the future and increases the impact and visibility of your own data collection or research.

Research data are a valuable resource, usually requiring much time and money to be produced. Many data have a significant value beyond the original research or policy objective. Sharing data:

- encourages scientific enquiry and debate
- enables scrutiny of research outcomes
- facilitates research beyond the scope of the original research
- leads to new collaborations between data users and data creators
- increases the impact and visibility of research
- promotes and credits the research that created the data and its outcomes
- reduces the cost of duplicating data collection
- provides important resources for education and training
- encourages the improvement and validation of research methods

The data curation process is enhanced by managing and sharing microdata. Sharing data permits new and innovative research built on existing information but also allows reliable verification of results. Managing data so that it can be shared should not be onerous, but form part of good practice. Depositing and sharing data with a specialist data-catalog such as the Pacific Data Hub Microdata Library at SPC ensures:

- safe keeping of research data in a secure environment
- long-term preservation and back-up of data
- resource discovery of data through inclusion in online catalogues
- rights management of data (licensing issues)
- administration for data access
- enhancing the visibility of your data and enabling more use and citation

- management and monitoring of usage of data
- promotion of data to user groups

How do I deposit my data with Pacific Data Hub Microdata Library?

Contact us at pdl@spc.int and we will assist you to deposit your data with us for publishing on our Pacific Data Hub Microdata Library site. Read our depositor MOU, which indicates our responsibilities.

What type of data can I deposit with Pacific Data Hub Microdata Library?
Pacific Data Hub Microdata Library accepts administrative records, survey research data, and any other data that could be useful for research

What formats should the data be in?
Pacific Data Hub Microdata Library accepts data files in all formats, but preferred formats for data files are Stata and SPSS. Data deposited in other formats will be converted to our standard formats.

What should be included with the data files?
Background documentation helps other researchers to understand how to re-use your data. Any documents pertaining to the research should be deposited with the data files, such as administrative forms, questionnaires, code lists, manuals, enumerator manuals, and reports.

How to use the data deposit application