

Dissemination Policy on Microdata
Department of Census and Statistics (DCS)
Sri Lanka

01. Rationale for dissemination

In compliance with the Statistics Law of Sri Lanka, the Department of Census and Statistics, as well as statistical units in other Ministries and other institutions of the Government, publish and disseminate the statistical data they produce to all users, in order to help in policy formulation and decision making.

Appendix A presents a list of all statistics for dissemination, their periodicity as well as the type of micro data available for these statistics.

02. Statistical Information for Public Dissemination

All designated official statistics which include basic and sectoral statistics, either solely produced by DCS or in cooperation with other statistics units of the Government, shall be disseminated to the public on a regular basis.

In addition to the summary or aggregate statistics derived from the surveys, and Censuses microdata will also be disseminated. Metadata relating to these microdata shall also be provided in order to help researches understand what the data are measuring and how they have been created. This will allow users to assess quality of the data and guide them to its correct use.

03. The objective of this policy

This policy on microdata dissemination has been formulated with the objective of defining the nature of the anonymised microdata files that will be released, the intended use of these files and the conditions under which these files will be released.

04. Technical Terms

For the purposes of this policy, the following terms are defined:

4.1 Anonymization. "Anonymization" refers to the process of removing direct and indirect identifiers from the survey file to conceal the characteristics and the identity of individual respondents.

4.2 Anonymised Microdata files. These are microdata files that have been anonymised for dissemination purposes.

4.3 Direct identifiers. These include such information as names, addresses or other direct personal identifiers which must be removed from all files made available to users.

4.4 Indirect identifiers. These refer to characteristics which are shared with several other respondents and when combined with other information, can lead to compromising the identity of the respondent.

4.5 Census. A census is a survey conducted on the full set of observation objects belonging to a given population or universe. A census is the complete enumeration of a population or groups at a point in time with respect to well defined characteristics: for example, population, production, traffic on particular roads. In some connection the term is associated with the data collected rather than the extent of the collection so that the term sample census has a distinct meaning. A Census, therefore, involves the listing and recording of the characteristics of each individual person or household as of a specified time. It is the source of information on the size and distribution of the population, farming and establishment as well as its demographic, social, economic, and cultural characteristics. These information are vital for making rational plans and programs for national and local development. The statistical data that will be collected by the Census will be for use of the government, business and industry, social scientists and other researchers, and the public in general.

4.6 Survey. Survey is a scientific statistical collection of data on individuals, households, establishments or other organizational units where only a sample of units in the population is enumerated.

4.7 Dissemination. "Dissemination" refers to the act of making microdata files, with supporting metadata, available for access and use.

4.8 Microdata. Data collected on an individual object - statistical unit. Units of observations can be households, individuals, agricultural holding, etc.

4.9 Metadata. A description of the data for the users to understand the data in-depth. This include, among others, description of the source, compilation methodology, time of dissemination, institution and persons responsible for the compilation.

4.10 Domain. In terms of geographic coverage, domain refers to the lowest level of geographic disaggregation where the data is representative.

4.11 Sectoral statistics. These are statistics collected by Ministries or Government institutions for their internal needs and reporting purposes.

4.12 Basic statistics. These are economic, social, environment and demographic statistics that are cross sectoral in nature, national and sub-national, that are required by the Government for policy and program formulation and evaluation, as well as for use by the wider national and international communities.

05. Degree of anonymisation

There are three main types of anonymised files that may be produced under the terms of this policy. The major differences among these files are the levels of geographic and characteristic detail.

- Files that have less geographic details and less characteristic details (i.e., geographic details and data dictionary are hidden geography has been collapsed and variable detail has been reduced)
- Files that have only higher levels of geographic details but the data dictionary is provided less geographic details and more variable details and
- Files that have full geographic details and the data dictionary more or full geographic detail and more variable detail. (Such files are generally only available through data enclaves).

To the extent possible, geographic detail included in the anonymised microdata files will be at the domain level. Moreover, suppression of variable details will be kept at a minimum. It is however, within discretion of DCS to determine which variables to suppress/collapse in view of the risk of identity disclosure of its respondents.

06. Its coverage

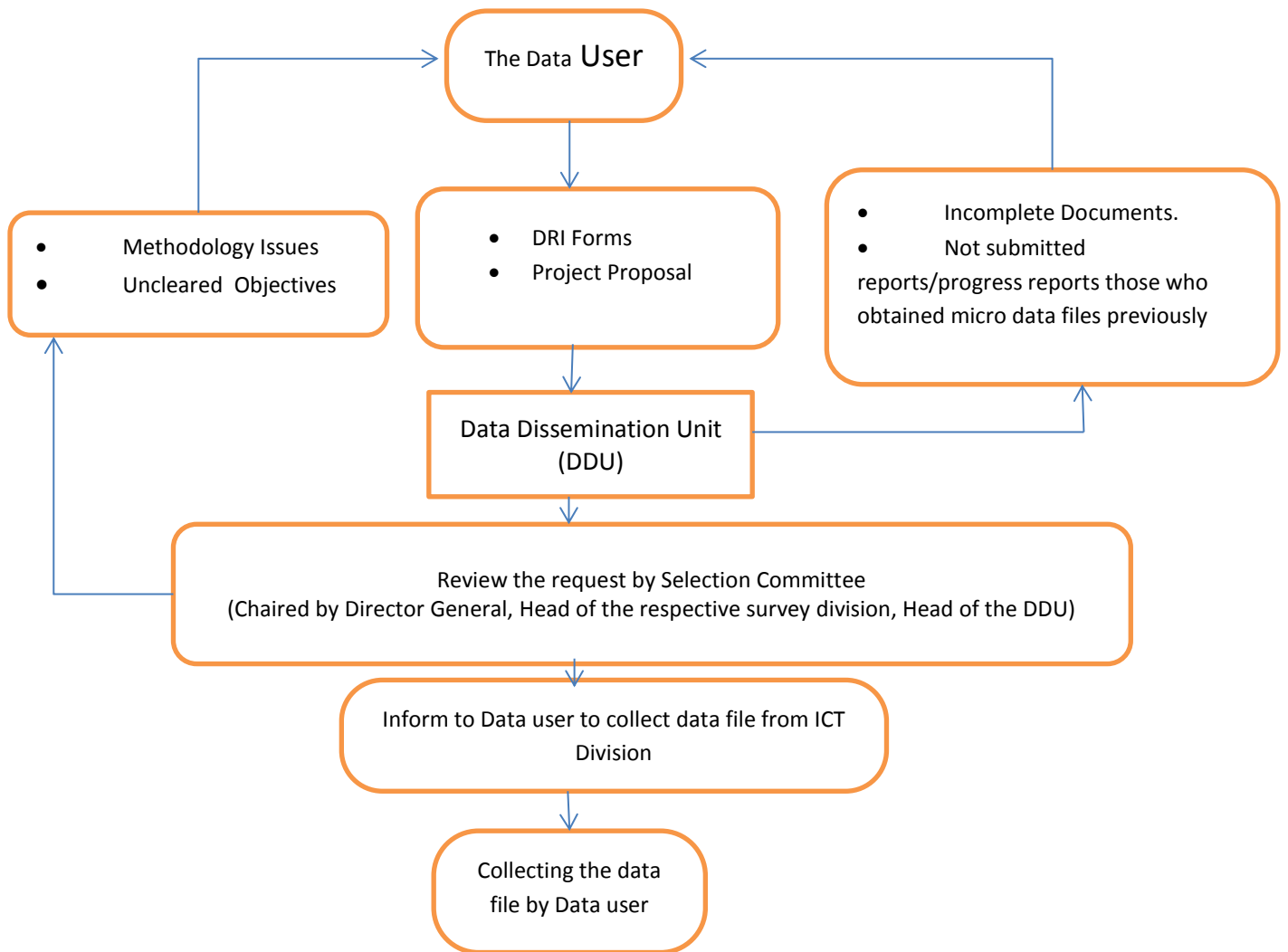
The type of microdata file that will be produced will depend on the users' needs and their research objectives. This policy covers two types:

- **Public Use Files (PUFs):** Microdata files that are disseminated for general public use outside DCS. They have been highly anonymised by removing the names and addresses and by suppressing/ collapsing geographic and respondent characteristic details to ensure that identification of individuals is highly unlikely. These files are made available for downloading from the DCS site to individuals who identify themselves by name, provide their email addresses and agree to abide by the terms and conditions appropriate for a PUF. See Appendix B.
- **Licensed files:** These files require that there be a signed agreement between the DCS and major users, to permit them to access data files that are more sensitive than PUFs e.g. datasets which make available geographic variables beyond the domain level or full details of some indirect identifier variables. For these files, all direct identifiers have been removed and some characteristic detail may be collapsed or removed. Licensing agreements are only entered into with bona fide users working for registered organizations. The primary and secondary researchers must be identified by name and a responsible officer of the organization must endorse co-sign the license agreement. See Appendix C.

07. Confidentiality of Information

Information collected under the Statistical Law is to be used only for statistical purposes. The Law further requires that all levels of DCS staff, as well as designated officers of other statistical units in Ministries and State institutions, shall ensure confidentiality of all individual information obtained from respondents.

08. Procedure for applying and obtaining licensed microdata file



09. Authorized users

The DCS recognizes that microdata files are intended for specialised users with advanced quantitative skills, such as :

- Policymakers and researchers employed by line-ministries and planning departments who are engaged in the development of regional and national strategies and programs, including the monitoring and evaluation of these programs.
- International agencies involved in the conduct of special studies aimed at identifying development and support opportunities and the development programs and infrastructures within DCS
- Research and academic institutes involved in social and economic research
- Students and professors mainly engaged in educational activities, and
- Other users who are involved in conducting scientific research (to be approved on a case by case basis).

10. Dissemination Channel

Metadata for DCS studies are provided to public through the DCS NADA <http://statistics.sltidc.lk/index.php/home>. Access to PUFs is granted to users provided they adhere to the conditions prescribed in Appendix B. For licensed files, Data Requester Application form (**D.R.A.**) can be downloaded and send it by postal mail, email(signed, scanned copy) to Director General of DCS. Requests will be evaluated over a period of two (2) weeks. NADA provides a facility where users can check the status of their requests.

11. Timing and Release

Recognizing the importance of user needs and timely data, the DCS will strive to release microdata after 45 days of the first release of final results. Survey data may need to be released in phases to ensure that the survey data have been reviewed and analysed by the DCS staff prior to their being anonymised for use outside the DCS.

12. Cost Recovery

It is the policy of the DCS to encourage broad use of its products by making them affordable for users.

The following cost-recovery principles will apply to the dissemination of microdata files.

- Public use files posted in the DCS data server for broad download will be available without charge.
- There is a nominal cost associated to each micro dataset (Licensed use). This is just to cover the costs of supply of microdata and is not intended at all to cover the all activities including cost of collection of data. The charges are depended on capacity of the data file/s and the user type as shown in the following table.

| User type | Charges per 50KB |
|---------------|------------------|
| Local Users | Rs. 100.00 |
| Foreign Users | 2 US\$ |

- If the requester is a Sri Lanka government institutes (State and Provincial public Sector), a student engaged in higher education, and selected international agency or sponsoring agency, the requester will not be charged for the supply of microdata.

13. Dispute Resolution

In case of any dispute between the Requester and DCS, or between any two parties related to the microdata, the Director General of DCS will give opportunity to all the aggrieved party(ies) to present their case(s). The decision of the Director General will be final.

14. Policy version

This is version 1.1 of the policy, which was adopted on 16 October 2014 and became effective on 16 October 2014.

15. Contact Person

For further information about access to microdata, please contact:

The Head,

Data Dissemination Unit,

Department of Census and Statistics.

Tel: +94112147445

email: data.requests@statistics.gov.lk

Appendix A

Statistics for dissemination

Appendix B

Terms and Conditions on the use of public use files:

1. The data and other materials provided by the Department of Census and Statistics (DCS) will not be redistributed or sold to other individuals, institutions, or organizations without the written agreement of the DCS.
2. The data will be used for statistical and scientific research purposes only. They will be used solely for reporting of aggregated information or the development of statistical models.
3. The user must not attempt to link the micro data with any personal data or seeks to identify observation unit (Individuals or establishment or other) in the microdata set.
4. An electronic copy of all reports and publications based on the requested data will be sent to the DCS.
5. Users alone will be responsible for all their analyses and interpretations derived from microdata. DCS by no means can be held responsible for any interpretation or analysis by the user or any consequences thereof.

Terms and Conditions on the use of licensed data files:

1. All the data requests should be made to Director General (DG) of the DCS as the sole authority of releasing data is vested with the DG, DCS. DCS of Sri Lanka reserves sole right to approve or reject any data request made depending on the confidential nature of the data set and intended purpose of the study or analysis.
2. Requests for micro data should be made through the Data Requester Application form (**D.R.A.**) with triplicate, designed by DCS along with study/projects proposal. If requests are made for the micro data of more than one survey, a separate agreement should be signed.
 - 2.1 When Officers in an organization or institute, request micro data files, those requests should be made through head of the institute/country representatives.
 - 2.2 If Data requester is a student their request should be made through respective heads or supervisors.
3. If the request is approved, full data file is released only for the surveys while **5%** sample data files is released for the census based on the researcher purpose of the study.
4. The released data file should be used only for the specific study/analysis mentioned in the agreement form and shall not be used for any other purpose without prior approval of the DCS.
5. Before releasing the Microdata files to the users, those who have already obtained any data files, it is needed to submit final report or should be reported the progress with time frame of the previous study(ies) to consider the request.
6. The data and other materials provided by the Department of Census and Statistics (DCS) will not be redistributed or sold to other individuals, institutions, or organizations without the written agreement of the DCS.
7. The data will be used for statistical and scientific research purposes only. They will be used solely for reporting of aggregated information or the development of statistical models.
8. The user must not attempt to link the micro data with any personal data or seeks to identify observation unit (Individuals or establishment or other) in the microdata set.
9. Users alone will be responsible for all their analyses and interpretations derived from microdata. DCS by no means can be held responsible for any interpretation or analysis by the user or any consequences thereof.
10. An electronic copy of all reports and publications based on the requested data should be sent to the DCS within one (1) week after publication as DCS publishes summary of projects and provides links to papers produced by users on its website.

11. The primary and other researchers who will be involved in using the data must be identified.
12. The researchers' organization must be identified as must a suitable representative of the organization who must be a signatory to the license.
13. Datasets must be kept fully secure, so that no unauthorized use takes place.