POLICY NOTE ISSN 2792-1468

THE GAP BETWEEN POLICY AND PRACTICE

A REVIEW OF THE DATA DISSEMINATION PRACTICES OF SRI LANKA'S DEPARTMENT OF CENSUS AND STATISTICS

SUMMARY

Government data collection and dissemination, although not the most glamorous of tasks, plays a crucial role in policymaking. In the absence of accurate and timely data, policymakers effectively make decisions with incomplete information. This policy note evaluates the data dissemination policy of Sri Lanka's Department of Census and Statistics (DCS), the primary government agency tasked with the collection, compilation, and dissemination of official statistics.

This review sets out two key findings. First, there is a significant mismatch between the DCS's data dissemination policy and its application in practice. Specifically, the entity is deficient in terms of its commitments regarding access to Public User Files (PUFs), adhering to timelines, and the high cost of Licensed Files (LFs). This mismatch undermines the DCS's ability to achieve its stated objectives.[1] It also precludes the public from readily accessing data, to make informed and evidence-based decisions.

Second, Sri Lanka falls behind South Asian countries such as India, a country where data collection and dissemination is far more complicated and costly. In contrast to Sri Lanka, India has implemented micro-data dissemination policies that are more progressive, inclusive, and successfully operationalized, making it the regional leader in this area that Sri Lanka aspires to be.

1. INTRODUCTION

Reliable data is a fundamental prerequisite for a wide range of developmental interventions. For example, accurate and timely data is important to critically evaluate project alternatives, design evidence-based policy, and target policy interventions. Sound statistics are also crucial for: businesses to develop new products and markets; the public to hold their representatives accountable; civil society and non-governmental organizations to better advocate and promote societal well-being; and, researchers and academics to develop high quality research outputs.

1.1. The role of the DCS

National Statistical Offices (NSOs) routinely collect and maintain a vast amount of data to produce statistics on the state of a country's economy or society. The rich data collected by NSOs has the potential to assist in building a robust evidence base to improve the quality of decisions taken by legislators, policy makers, businesses, and the public.

In Sri Lanka, the DCS serves as the NSO responsible for the collection and preparation of statistics pertaining to the country's economic conditions.[2]

The DCS's vision statement emphasizes its aspiration to be the regional leader in terms of generating timely statistical information. Further, its mission statement sets out its objective of providing accurate data to facilitate socio-economic development and enhance Sri Lanka's prosperity in the context of globalization.[3]

The DCS collects data through the administration of periodic surveys on nationally representative samples of the population. This entity is also responsible for

1 Website of the Department of Census and Statistics, Sri Lanka, 'About Us', at http://www.statistics.gov.lk/about_us [last accessed 19th November 2020].

2 The Statistics Ordinance provides for the establishment of a Bureau of Statistics in Sri Lanka, for the collection and preparation of statistics pertaining to agriculture, emigration, immigration, factories, manufacture, industries, labour, and cost of living. Website of the Department of Census and Statistics, Sri Lanka, 'Statistics Ordinance' at: http://www.statistics.gov.lk/PopHouSat/StatOrd_1956_Revision.pdf [last accessed 19th November 2020].

3 Website of the Department of Census and Statistics, Sri Lanka, 'About Us', at http://www.statistics.gov.lk/about_us [last accessed 19th November 2020].

administering the decennial census of households and the population. As such, the DCS is a repository of rich micro and macro socioeconomic data.

1.2. Microdata vs. macrodata

Microdata refer to single objects or observations of data,[4] defined as the "atomic data element describing the individual object being studied".[5] An individual object may comprise a person, a household, a firm, an agricultural holding, a school, a health facility, etc.[6] By contrast, macrodata are aggregate, or summary, statistics[7] that capture the highlights of survey findings.[8] Macrodata are formed by aggregating micro level data observations.

The value of microdata lies in its ability to identify distinct relationships and interactions between variables by allowing the user to compute statistics at different levels of disaggregation. Such analysis is only possible with this type of data, underscoring the enormous research potential that can be unlocked by making this data available for in-depth analyses.

Microdata is pivotal to economic research and development. It helps to answer complex questions when aggregate statistics are insufficient, and to calculate marginal rather than average effects of key variables.[9] Governments around the world, looking towards evidence-based policies, are increasingly advocating for open data sharing in both public and private spheres.

1.3. Sri Lanka's microdata dissemination policy

In recognition of the value of microdata, the DCS introduced a Microdata Dissemination Policy that came into effect on 16 October 2014. This policy was instituted to "publish and disseminate the statistical data they [DCS and government] produce to all users, in order to help in policy formulation and decision making".[10]

The policy defines the nature of anonymized microdata files available for dissemination, their intended use, and the conditions under which they can be released.[11]

The policy also differentiates between two types of microdata files: Public User Files (PUFs) and Licensed Files (LFs). The main difference between the two types of files is the level of anonymization. According to the DCS's microdata dissemination policy, PUFs can be freely downloaded by the general public from the DCS website, while LFs require a signed agreement between the DCS and major users and are provided at a cost (see Exhibit 1).

4 Hand D.J., 'Microdata, Macrodata and Metadata', (1992) In: Dodge Y., Whittaker J. (eds) Computational Statistics. Physica-Verlag HD, at: https://doi.org/10.1007/978-3-642-48678-4_41 [last accessed 19th November 2020].

5 Ibid.

6 OECD, 'Expert Group for international collaboration on microdata access Final Report', (2014), at: https://www.oecd.org/sdd/microdata.htm [last accessed 19th November 2020].

7 Hand D.J. (1992) Microdata, Macrodata and Metadata. In: Dodge Y., Whittaker J. (eds) Computational Statistics. Physica-Verlag HD, at https://doi.org/10.1007/978-3-642-48678-4_41 [last accessed 19th November 2020].

8 OECD, 'Expert Group for international collaboration on microdata access Final Report', (2014), at: https://www.oecd.org/sdd/microdata.htm [last accessed 19th November 2020].

9 Julia Lane, "The Uses of Micro-data", The Urban Institute, 12th June 2003, at http://webarchive.urban.org/Uploaded-PDF/1000564_microdata.pdf [last accessed 10th September 2020].

10 Website of The Department of Census and Statistics, 'Dissemination Policy on Microdata', at: http://www.statistics.gov.lk/Datadessimination/DataDissaPolicy_2007Oct26 [last accessed 19th November 2020].

11 Ibid.

Exhibit 1: Types of microdata files in the DCS's microdata dissemination policy

Types of Microdata Files	
1. Public User Files	**2. Licensed Files**
• Disseminated for the use of the general public. • Highly anonymized by removing names and addresses and by suppressing/collapsing geographic and respondent characteristic details to ensure that identification of individuals is highly unlikely. • 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.	• Signed agreement between the DCS and major users required to permit access to data files that are more sensitive than PUF - e.g. datasets that make available geographic variables beyond the domain level or full details of some indirect identifier variables. • All direct identifiers are removed, and some characteristic details may be collapsed or removed. • Licensing agreements are only entered into with bona fide users working for registered organizations. • Primary and secondary researchers must be identified by name and a responsible officer of the organization must endorse and co-sign the license agreement.

Source: Dissemination Policy on Microdata 2007[12]

12 Website of the Department of Census and Statistics, 'Dissemination Policy on Microdata 2007', at: http://www.statistics.gov.lk/Datadessimination/DataDissaPolicy_2007Oct26 [last accessed 3rd February 2021]

1.4. Increasing access to data will increase returns to public funds invested

The data collected by the DCS is essentially a public good; they are collected using public funds (see Exhibit 2), and are "non-rival" in nature (i.e. the data collected can be used by multiple users at the same time, for different purposes, without diminishing its value). Increasing access to data is similar to public investment in tangible goods such as a road, where greater usage determines the return to investment. In addition, the cost of implementing and administering surveys is fixed. As such, the marginal costs associated with the provision of microdata sets are low.

Exhibit 2: Sources of DCS funding

	2017 (in LKR thousands)	2018 Revised Budget (in LKR thousands)	2019 Estimates (in LKR thousands)
Total Expenditure	981,129	997,627	1,007,000
Domestic Funds	958,221 (98%)	986,415 (99%)	988,200 (98%)
Foreign Grants	22,909 (2%)	11,212 (1%)	18,800 (2%)
Cost of conducting surveys	40,605 (4%)	30,000 (3%)	25,000 (2%)

Source: Budget estimates 2019, Volume 1[13]

13 Website of the Ministry of Finance, Sri Lanka, 'Budget Estimates 2019' at: http://www.treasury.gov.lk/web/guest/budget-estimates [last accessed]

Failure to successfully operationalize Sri Lanka's microdata dissemination policy impedes public access to the rich administrative data collected by the DCS that could otherwise be used by legislators, policy makers, businesses, researchers, and the public to make informed decisions.

To unlock the full potential of data, it is important that the data is widely accessible, affordable, and available in a timely manner. The following section describes how the sting DCS policy fails to deliver on this criteria.

2. KEY FINDINGS

2.1. Mismatch between DCS policy and practice

2.1.1 Public User Files are not accessible

PUFs are typically machine-readable files containing microdata that: (i) allow users to make sensible inferences on the phenomenon for which data were collected; and (ii) have been subject to statistical disclosure control methods that render the data non-confidential according to the legal and methodological standards applicable to the NSO that produced the statistics.[14] PUFs are commonly referred to as non-confidential microdata files and carry little to no risk from disclosure and can therefore be made available to virtually any user.[15]

According to the DCS data dissemination policy, access to PUFs is granted through the National Data Archive (NADA) housed on the DCS website. Data users must create an account, provide their name, and email address, and agree to abide by the terms and conditions applicable for PUFs as specified in the dissemination policy. Registering supposedly allows data users to download the PUFs.

In reality, the process is not this straightforward. Research carried out by Verité Research (VR) indicates that PUFs are not available for download on the NADA website. Additionally, communication with the DCS confirmed that PUFs are not available for the public to access and have not yet been housed on the NADA website.

2.1.2 The DCS fails to adhere to stated timelines

According to the DCS's dissemination policy, a data requester application (DRA) must be submitted to the Director General of the DCS by post/email to access licensed files (LFs). The NADA also provides a facility to electronically submit DRA forms, and where users can check the status of their requests. Accordingly, the policy stipulates that data requests will be evaluated by the DCS within a period of two weeks and if approved, data files will be released to the data requester upon payment.

However, VR's experience in procuring LFs is that the time taken to obtain approval and get access to LFs far exceeds the stipulated timelines. Additionally, this process is cumbersome and requires frequent follow-ups. To illustrate:

- The DCS took up to two months to process requests for LFs for the Labour Force Surveys and Household Income and Expenditure Surveys of 2012 and 2016. The DRA was submitted on 28 February 2018 via email and the data were received on 26 April 2018. Further, during this two-month period, VR had to follow up multiple times with the DCS.
- The DCS took approximately 1.5 years to provide microdata for the 2011 population Census. The DRA form to procure this dataset was submitted on 12 December 2017. A sample dataset was received only approximately six months later in June 2018. VR also had to make multiple rounds of inquiry to clarify queries with regard to the sample dataset, with the DCS sometimes taking as long as three months to respond to queries. Finally, the invoice for the dataset was sent on 29 April 2019 and the funds transferred on 24 May 2019. The final dataset was received on 1 June 2019.

Online requests submitted through the NADA portal also do not seem to be processed or picked up in a timely manner. VR submitted an online request on 9 June 2020. However, as at 18 August 2020, progress regarding the status of this request had not been updated on the portal (see Appendix 1).

14 OECD, 'Expert Group for international collaboration on microdata access Final Report', (2014), at: https://www.oecd.org/sdd/microdata.htm [last accessed 19th November 2020].

15 Ibid.

2.1.3 Access to licensed files is costly

The DCS microdata policy states that the entity seeks to "encourage the broad use of its products by making them affordable for users".[16] As per this policy, LFs are priced according to the principles of cost recovery and the charges are intended to cover the cost of supplying microdata and not the cost of data collection. Since the DCS already prepares microdata sets for internal purposes and to publish their reports, the marginal cost of supplying a microdata set is likely to be extremely low.

However, VR's research finds the costs charged for data at present to be completely out of step with the policy's cost recovery principles. All users except Sri Lankan government institutes and students engaged in higher education must pay for the data. The standard charge for 50 KBs of data is LKR 100 for local users and USD 2 for foreign users. Since these are large data sets, the total cost can be quite significant (see Exhibit 3).

Exhibit 3: Actual costs incurred to procure various datasets by VR and associated researchers

Dataset	Size (in KBs)	Actual Cost Incurred by VR (in LKR)
Labour Force Survey 2012	16,907	33,800
Labour Force Survey 2016	22,222	44,400
Household Income and Expenditure Survey 2012	121,789	243,500
Household Income and Expenditure Survey 2016	73,503	147,000
5% sample of Census of Population and Housing 2001	226,787	453,574
5% sample of Census of Population and Housing 2012	213,703	427,406

16 Website of The Department of Census and Statistics, 'Dissemination Policy on Microdata', at: http://www.statistics.gov.lk/Datadessimination/DataDissaPolicy_2007Oct26 [last accessed 19th November 2020]

Far from "encouraging broader use",[17] the high cost of purchasing datasets excludes users who cannot afford these statistics and also discourages potential data users. This is highlighted in Exhibit 4, which compares the use of LF microdata files provided free of charge against those provided at a cost. The usage of microdata files disseminated by the DCS free is far higher than the number of files disseminated at a price. Over the four-year period shown in Exhibit 4, only 13 microdata files were disseminated to paid users. For example, the number of microdata files disseminated free ranged from 12.8 times to 36 times the number of microdata files disseminated at a cost for the years 2015-2018. Overall, 274 microdata files were provided free of charge during that period.

Exhibit 4: Microdata files disseminated by the DCS free vs. at a price

Year	Microdata files provided free (a)	Microdata files purchased (b)	(a)/(b)
2015	64	5	12.8
2016	72	2	36
2017	52	2	26
2018	86	4	21.5

Source: DCS performance reports 2015 to 2017[18]

17 Website of The Department of Census and Statistics, 'Dissemination Policy on Microdata', at: http://www.statistics.gov.lk/Datadessimination/DataDissaPolicy_2007Oct26 [last accessed 19th November 2020]

18 Website of the Parliament of Sri Lanka, 'Performance Report of the Department of Census and Statistics 2015, 2016 and 2017'. at: https://www.parliament.lk/uploads/documents/paperspresented/performance-report-department-of-census-and-statistics-2015.pdf https://www.parliament.lk/uploads/documents/paperspresented/performance-report-department-of-census-and-statistics-2016.pdf https://www.parliament.lk/uploads/documents/paperspresented/performance-report-department-of-census-and-statistics-2017.pdf [last accessed 3rd February 2021]

2.2. India: leading the way in microdata dissemination

In India, the Ministry of Statistics and Program Implementation (MoSPI) is responsible for conducting periodic, national economic censuses and large-scale national sample surveys. The responsibility for conducting decennial censuses lies with the Office of the Registrar General and Census Commissioner, which comes under the purview of the Ministry of Home Affairs.

To promote the wider use of data, the Government of India issued an official memorandum in April 2019, **which recognises official statistics as public assets** and as a key input for decision making and policy intervention. The memorandum stipulates that microdata access is to be provided **free of charge, with single point online access for conducting research for both public and private purposes** (see Appendix 2).[19] The rationale for this decision is highlighted in Box 1 below.

Box 1: Excerpts from the Office Memorandum issued by the Ministry of Statistics and Programme Implementation, Government of India, on the provision of free online access to microdata

"Official statistics are key inputs for decision making and policy intervention and become public assets for conducting research both in the public and private sphere. Recognizing the requirement and potential of data, Ministry of Statistics and Program Implementation(-MoSPI), Government of India, has decided to provide free of costs, single point access and support of micro-data of census and surveys conducted by the Ministry to Students/Researchers/Institutes in India and abroad through its National Repository from 1st April 2019"

Source: Office Memorandum on Online Microdata Dissemination of Census/ Surveys conducted by MoSPI[20]

The government has set up 12 workstations in universities around the country to provide free access to anonymized census microdata. These workstations are fully equipped and have multiple terminals, and researchers are permitted to use software such as SPSS and STATA for data analysis. The objective of the workstations is to permit qualified researchers optimum use of anonymized microdata from censuses. The facility is open to all, subject to prospective users receiving formal approval from a selected steering committee.

Box 2: Excerpts from Government of India's workstations for research for sample microdata policy

"As information is power and as collection of information itself is very costly, many countries allow researchers the access to this huge micro dataset for individual research. After anonymizing the data of individual identity and sensitive information, the dataset is made available for in-depth research on the varying characteristics of the population of the country. The researchers, under secure environment are permitted to analyse the data and generate outputs of extract information at aggregate levels for use in their research"

Source: Workstations for Research on Sample Micro-Data from Census, 2011[21]

2.3. Practice better aligns with stated policy

India has successfully operationalized its data dissemination policy. Users are able to register with MoSPI's National Repository and obtain instant access to all available microdata files. Additionally, a step by step guide on how to register and access data is provided on the MoSPI website. Data can be downloaded and viewed on Nesstar, a free software that can also be used for data analyses, imported to other software such as STATA, and converted to Excel files.

The crucial difference between Sri Lanka's and India's microdata dissemination policies is that the latter provides free of cost, single point online access to all users. In contrast, Sri Lanka only provides free access to selected users. The rest of the public, which includes,

19 Website of the Ministry of Statistics and Programme Implementation, 'Online Microdata Dissemination of Census/ Survey conducted by MoSPI'. at: https://web.archive.org/web/20200110165726/https://mospi.gov.in/sites/default/files/data_disemination/OM%20_to_disseminate_data%20free_.pdf [last accessed 19th November 2020]

20 Website of the Ministry of Statistics and Programme Implementation, India, 'Office Memorandum on Online Microdata Dissemination of census/surveys conducted by Ministry of Statistics and Programme Implementation', at: http://www.mospi.gov.in/data-dissemination [last accessed 3rd February 2021]

21 Website of the Ministry of Home Affairs, India, 'Workstations for research on sample micro-data from Census,2011', at: https://censusindia.gov.in/2011census/workstation.html [last accessed 3rd February 2021]

private think tanks, private researchers, consultants, and the private sector are required to purchase these datasets. Additionally, where single point online access is not feasible, India has taken steps to set up remote access facilities. By comparison, Sri Lanka does not have such mechanisms in place. Finally, unlike in India, Sri Lanka has been deficient in operationalizing some of elements of its microdata policy, such as the provision of PUFs and adherence to timelines for granting data access.

3. Conclusion

The data dissemination policy introduced by the DCS is a step in the right direction in terms of making data more widely accessible to the public. However, there exists a significant mismatch between policy and practice, undermining the ability of the DCS to achieve its own objectives. Research conducted by VR finds three key areas where practice is at odds with the stated policy:

1) the DCS fails to provide online access to PUFs as stated in its policy;

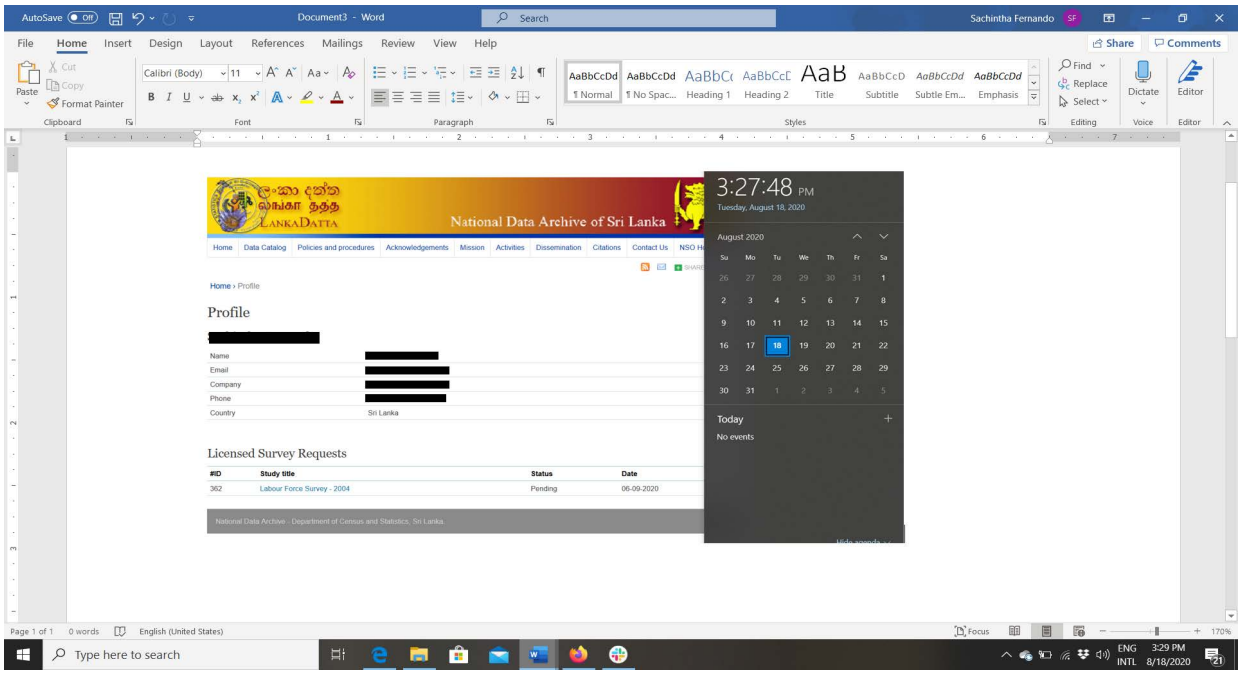
2) the DCS fails to adhere to the timelines stipulated in the policy when providing data; and

3) the cost of data does not align with the principles of costing specified in the policy.

Contrary to its objectives of encouraging wider use of data, these barriers serve to discourage users from accessing datasets from the DCS. This impedes the Sri Lankan public from reaping the full benefits from data collected using public funds.

Although the DCS aspires to be a regional leader in this field, India is far ahead of Sri Lanka in making data affordable and accessible to all users. The Indian policy is more inclusive because it provides the public and private sectors as well as foreign users equal, and free, access to national data. Sri Lanka, through the DCS, should move towards such a non-discriminatory, efficient, and low-cost approach. The DCS aims high in its vision statement. It is time to translate such "vision" into action. Such change would not be overly costly nor difficult, and would generate significant benefits for the country.

APPENDIX 1


Document3 - Word
Search
Sachintha Fernando
File Home Insert Design Layout References Mailings Review View Help
Share Comments
National Data Archive of Sri Lanka
Home Data Catalog Policies and procedures Acknowledgements Mission Activities Dissemination Citations Contact Us
Home › Profile
Profile
Name
Email
Company
Phone
Country
Sri Lanka
Licensed Survey Requests
#ID Study title Status Date
362 Labour Force Survey - 2004 Pending 06-09-2020
National Data Archive - Department of Census and Statistics, Sri Lanka
3:27:48 PM
Tuesday, August 18, 2020
August 2020
Today
No events
Page 1 of 1 0 words English (United States)
Type here to search
ENG INTL 3:29 PM 8/18/2020

APPENDIX 2

File No.O-11015/18/2015-ASI

F.No.O-11015/18/2015-ASI
Government of India
Ministry of Statistics & Programme Implementation
Data Storage & Dissemination Division

East Block-10, Sector-I,
R K Puram, New Delhi- 110066
Date: 22.03.2019

OFFICE MEMORANDUM

Subject: - **Online Microdata Dissemination of Census/ Surveys conducted by MoSPI**

Official statistics are key inputs for decision making and policy interventions and become public assets for conducting research both in public and private spheres. Recognizing the requirement and potential of data, Ministry of Statistics & Programme Implementation (MoSPI), Government of India has decided to **provide free of cost,** single point access and support of **Microdata of Census and Surveys conducted by the Ministry to Students/ Researchers/ Institutes in India and abroad through National repository from 1st April, 2019.**

(Hauliankap)
Under secretary to Government of India
e-mail: h.suantah@nic.in