



Verité Research is an independent think-tank based in Colombo that provides strategic analysis to high level decision-makers in economics, law, politics and media.

Salma Razick
Researcher

Nishan de Mel
Series Editor



Image courtesy of Unsplash

Making Public Data Public: Sri Lanka Misses its Own Targets While India Races Ahead

The Department of Census and Statistics aims to be the 'leader in the region' in producing and providing timely statistical information. This Insight shows, however, it has failed in providing public access to data in three respects: ease, speed and cost. By contrast, India has raced ahead on these metrics, to become the regional leader that Sri Lanka aspired to be.

In Sri Lanka, the Department of Census and Statistics (DCS) is the primary government agency responsible for collecting and providing access to data that can be used for statistical analysis. The stated vision of the DCS in Sri Lanka is to be the 'leader in the region' in producing timely statistical information to achieve the country's development goals.

This *Insight* finds that the DCS in Sri Lanka falls well short of its own stated commitments towards realising this vision. The analysis also finds that India, despite having a much later start, and having a more complex data collection environment, has raced ahead to become the regional leader that Sri Lanka aspires to be, in providing access to public data.

100
 Rupees

Charged per 50kbs of data

427,406
 Rupees

For a 5% sample of 2012 Census data

1.5
 Years

To obtain that dataset

The comparative analysis with India is made on three dimensions: (i) ease of access; (ii) speed of access; and (iii) cost of access.

The importance of timely data that is widely accessible

Data is one of the most important public assets of a country. High accessibility of data leads to multiple positive outcomes: it allows academics and professionals to generate better analysis, helps government to formulate better policies, facilitates better economic and investment decisions, and can also provide better visibility to society on actual outcomes and progress on government policies.

In order to understand the details and to make good decisions “to achieve the country’s development goals,” there needs to be open access to actual data

The term ‘data’ is often used to refer to summarised aggregate statistics. We use the term to mean actual data that is used to generate those statistics.

The availability of data, not just summarised statistics, is critical for decisions in relation to a country’s development goals. Take for example, the unemployment rate that is reported in the Annual Labour Force Survey. This estimate is obtained by taking the average number of unemployed persons in the total sample; it provides only an overarching macro picture of unemployment in Sri Lanka. But such summarised statistics do not reveal important details such as geographic and age dispersion of unemployment, and the extent of disparities in unemployment based on other demographic factors such as ethnicity, language or gender.

Data sets with units of observation at the individual, household, or firm level can provide information on aspects such as gender, age, education level, income level and district of residence etc., while being appropriately

anonymised to ensure privacy of the reporting unit. Such data paves the way to formulate better analysed policies that are more likely to be effective and durable. Therefore, in order to understand the details and to make good decisions “to achieve the country’s development goals” (as stated in the DCS policy), there need to be, first, open access to actual data.

Three dimensions of data accessibility: Sri Lanka misses and falls behind India

In order for data to be an effective tool for decision making, for both policymakers and the public, there are three dimensions to data accessibility that are important to assess : (1) Ease, (2) Speed, and (3) Cost. The analysis in this *Insight* shows Sri Lanka missing its own targets and falling behind India in all three of these important dimensions.

1. Ease of access: Sri Lanka fails to make any datasets available online

In October 2014, the DCS introduced a data dissemination policy. This policy addresses the need for the public to access data and provides guidelines on how it can be accessed. It specifies that ‘public’ datasets are available on their National Data Archive (NADA) for public use and can be downloaded from the DCS website.

However, more than six and half years later, DCS has failed to deliver on this key commitment under its own data dissemination policy. Even by June 2021, no datasets are available online on NADA. The research team made multiple attempts to obtain these ‘public’ datasets and found that they were not available to be downloaded as stated in the policy. DCS officials also confirmed this status quo.

India is ahead: In India, the function of collecting national level data and statistics is split between the Ministry of Statistics and Programme Implementation (MoSPI) and the Ministry of Home Affairs (MoHA). The MoSPI conducts periodic national economic censuses and large-scale national sample surveys, while MoHA conducts the decennial census.

India’s current data dissemination policy, which also states their commitment to make data available online, was adopted only in April 2019 (five years after Sri Lanka). Yet, in stark contrast to Sri Lanka, India has fully operationalised its data dissemination policy. Users, public or private, both within and outside India, can access public datasets through the National Repository housed on the MoSPI website.

The website has a step-by-step guide on how to access and download these datasets. The data can be downloaded and viewed on Nesstar, a free software that can be used for data analysis, or imported to other software such as STATA, and can also be converted to Excel format.

2. Speed of Access: Sri Lanka has a prolonged process

In addition to not making any public datasets available online, the DCS has also failed to adhere to its stated timelines in providing datasets offline. According to the DCS policy, once a request is made for access to certain datasets, the request must be evaluated within two weeks, and if approved, the datasets should be provided, upon receiving the stipulated payment.

India’s current data dissemination policy was adopted five years after Sri Lanka. Yet, India has fully operationalised its data dissemination policy.

Two case studies by Verité Research were integral to the assessment of speed. The case studies indicated that the DCS does not meet the timelines set out in its policy in providing access to datasets. The first case study relates to a request to obtain a 5% sample of the 2011 Census of Population and Housing. It took multiple follow-ups by the research team over approximately one and a half years before the data was received. The initial request was made on 12 December 2017; the final dataset was received only on 1 June 2019.

In the second case study, a similar request to obtain the Labour Force Survey and Household Income and Expenditure Surveys of 2012 and 2016 also took multiple follow-ups and almost two months before the data was received (see Exhibit 1).

India is ahead: As India allows for online accessibility and download of data, the time taken to access the data can be as short as a matter of hours rather than months (and sometimes over a year) as is the case in Sri Lanka.

3. Cost of access: Sri Lanka charges a fee for most datasets; India provides them free of charge

In addition to the lack of ease, and the lack of speed, accessing data in Sri Lanka faces the added hurdle of costs. Except for Sri Lankan government institutions and students engaged in higher education, all users are

expected to pay a fee to gain access to data. The standard charge for a local user to access 50KBs of data is LKR 100. Since these datasets are quite large, the full cost can be quite substantial. For example, in the first case study in 2019, it cost LKR 427,406 to gain access to the 5% sample of the Census of Population and Housing 2012 (see Exhibit 2).

The stipulated fees, once again, are not consistent with the stated policy. The data is collected using public funds. The DCS's policy states that the pricing is designed only to "cover the cost of supply of microdata and is not intended at all to cover the all [sic] activities including cost of collection of data". Therefore, apart from the fixed transaction costs relating to delivering the data to the requester, the marginal cost of an additional quantum of data is close to zero. The linear pricing structure of LKR 100 for each 50KB unit,

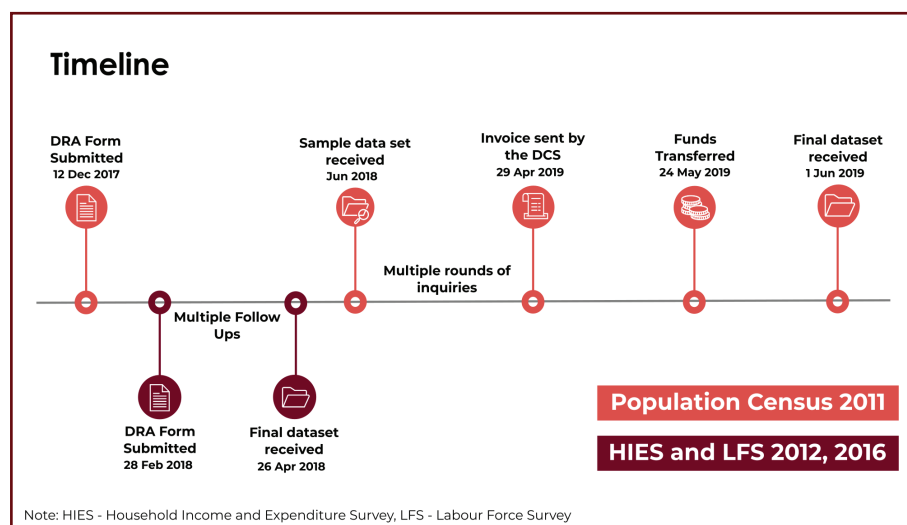
rather than a fixed transaction/delivery cost, is therefore not consistent with the purpose and principle of pricing set out in the stated policy.

In the first case study in 2019, it cost LKR 427,406 to gain access to the 5% Sample of the Census of Population and Housing 2012.

The data dissemination policy also states that it is designed 'to encourage broader use of its products by making them affordable to users'. Exhibit 2 shows the actual cost incurred by Verité Research in procuring selected datasets from the DCS, ranging from thirty thousand to four hundred and fifty thousand rupees. These prices, exceeding half the annual per-capita income of Sri Lanka for just 5% of the larger datasets, is more likely to discourage than encourage most users from using them by making it unaffordable rather than affordable. The pricing therefore is twice contrary to the stated policy. It is not limited to the transaction "cost of supply" and it is discouragingly costly (instead of being encouragingly affordable).

India is ahead: In India's case, the datasets are available for free. In April 2019, the Government of India recognised official statistics as a public asset and issued an official memorandum that specified that access to all datasets must be provided free of charge with single point online access for conducting research and for both public and private purposes.

Exhibit 1: Time taken by Verité Research to procure LFS following the DCS data request procedures.



The pricing of the datasets is twice contrary to the stated policy. It is not limited to the transaction "cost of supply" and it is discouragingly costly, instead of being encouragingly affordable.

Exhibit 2: Actual cost incurred in procuring datasets from DCS

Data Set	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

On data, India has moved ahead to become a role model for Sri Lanka

In providing access to data, India has taken on more progressive positions and has moved much faster to implement them. Sri Lanka, on the other hand, has taken on less progressive policies and has lagged behind in implementing them.

To quote the position of the Indian Bureaucracy (as stated in the Office Memorandum issued by the Ministry of Statistics and Programme Implementation, Government of India, on the provision of free online access to data):

"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 potential of data, the Ministry of Statistics and Programme Implementation, Government of India, has decided to provide free of costs, single point access and support to microdata of census and surveys conducted by the Ministry..."

This *Insight* assesses Sri Lanka's performance in terms of its own stated policies as well as its aspirations to be a regional leader providing access to timely statistical information.

The assessment against the stated policies of DCS suggests that Sri Lanka is currently failing to meet the goals that have been set out in its own policies. The comparison with India reveals that, despite a recognition in both countries of the importance of data for public users, India's recognition has translated into policies which are more progressive than Sri Lanka's, and have been better implemented in terms of improving the ease, speed and cost of access to data ♦