

Finding 'buried' data on South Asia at LSE Library



*In response to a BBC article requesting for further information about other collections of 'missing' or not widely known data, **Inderbir Bhullar** looks at LSE's holding of South Asian statistical material (India and the subcontinent, pre- and post-Independence) revealing that many of the 9,577 titles may be unique to the Library at the London School of Economics.*

In January 2016 an article by Justin Rowlatt, the BBC's South Asia correspondent, appeared on the BBC news website called [Hunting for buried treasure with Thomas Piketty](#). It reported on the discovery of "a huge cache of unpublished data" which had been used as the basis for a report by Unicef looking into child and female malnourishment in India. This report had initially not been released by the Indian government until Rowlatt and *The Economist* (who first covered the story in an article called *India's malnourished infants*) shed light on its existence. This then forced their hand and led to the release of the publication called the [Rapid Survey on Children](#).

In Rowlatt's article, Thomas Piketty referred to the importance of data transparency, particularly relating to Income Tax data in India, and its importance in enabling researchers to trace and record levels of inequality. The article ended with a request for further information about other collections of 'missing' or not widely known data which led to us in the Library taking a closer look at our holdings of South Asian data. In this article I review LSE's collection and provide an overview of the variety and the uniqueness of the data.

The LSE has amassed a large collection of statistical material relating to India (both pre- and post- Independence) and the Subcontinent. A total of 9,577 titles were found on our catalogue, primarily located in three sections of the library: The Government Publications (Lower Ground Floor), The Historical Statistics (Lower Ground Floor), The Statistic Collection (Lower Ground Store)

Just under 80 percent of this material is to be found in the Government publications section. There are also nearly 550 pre-twentieth century publications in the various collections (though most can be found, again, in the Government Publications section). The majority of these are split between publications published in or about Mumbai, Bengal and Madras which align with three of the East India Company's main regions of operation. The vast majority of the material is English language, though many make use of Indian units of measurement.

The Government Publications relating to India are housed on 16 rows of (single sided) shelving. There is a huge variety of data here which spans a large series of time and forms a very useful and significant national collection.

It is somewhat difficult to summarise a collection of this size and variety, however, there are strong collections of material relating to different states, provinces and regions. Bombay and the Punjab (both 'East' and 'West') are best represented in terms of region specific information with over 700 titles relating to them. Bengal and Madras are also other areas which have strong collections of associated material. Given the larger scope of the government material, other countries and regions which made up the Subcontinent are better represented and make up 22 percent of the total. Pakistan and Sri Lanka have the largest share of titles, with Bangladesh comprising the next greatest share. I've included materials catalogued as 'East Pakistan' towards the Bangladeshi figure rather than Pakistan.

The variety of material is particularly striking, from early colonial records detailing the nascent efforts to understand legal and financial matters to later village and land tenure surveys that describe local socio-economic life. Industries (particularly agriculture), literacy (primary, secondary and tertiary education), prison and legal issues are all analysed through various reports. There is much less in the form of raw data than, for instance, the statistical and historic statistical materials and far more usage of secondary forms of data to create analytical interpretive forms of writing – although these are usually heavily informed by and contain a good deal of statistical data.



Image credit: An edited image of the LSE Library from [LSE Alumni](#), LSE.

The Historical Statistics contain a strong collection of Indian censuses which date back to the first to be carried out nationwide in 1881. There are collections from other parts of the Indian subcontinent but they are much smaller in number and space. Pakistani materials are limited to foreign trade statistics although there are also incomplete census derived publications from the 1950s and 60s. Bangladesh appears to have a greater number of items in the historical statistics, though these mainly date from the 1970s onwards. There are also holdings relating to Sri Lanka and Nepal although these are fewer still and quite patchy.

The statistics which can be found in the Library's store predominantly consists of Indian census data and statistical publications relating to economics, with a large number of monthly economic abstracts and census tables. There are numerous volumes of state based statistics again derived from the censuses of 1991 onwards. There are also hundreds of titles of other government publications produced and about labour relations, demography, agriculture, crime and other topics.

With regards to the original query raised by Piketty, data relating to income tax and other forms of taxation can be found in several titles. These include many mid-nineteenth century reports on the systems in place of obtaining taxation and revenue from land holders (whether tenant farmers or landlords). There are also the *All-India income-tax report and returns for the year* spanning 1922-1941 and 1945-50.

Uniqueness

As part of the survey I also chose a random sample of 116 titles in order to compare availability in other HE and Research libraries via the collective Copac catalogue. Of those 116, 44 were found to be unique to LSE Library and not found elsewhere i.e. almost exactly 38 percent.

I also surveyed 41 titles dated pre-1850 using the same method and here LSE had even stronger and more unique collections with around 48 percent (20 of 41 titles) not recorded anywhere else.

This means that if we were to broaden out the findings, it is likely that the Library holds not only comprehensive collections but also information that is, at the very least, hard to find elsewhere and in many cases likely to be the only example in the UK. While the data has officially been published, the fact that it is rare and/or undiscovered brings us back to the very reason for starting the project in the first place – uncovering buried treasure. Budding social science researchers are welcome to browse and search through our holdings to locate some of these materials and uncover some buried treasure of their own.

Conclusion

While the library holds many materials from all over the world, there is certainly a case to be made to suggest that the amount of data relating to India constitutes a strong and relatively comprehensive collection. This is a large amount of 'buried' statistical data which could be used to aide researchers interested in a wide range of queries given the breadth of the holdings. This is a largely untapped resource and while further work could be done to analyse how much duplication there is between the LSE and other collections of similar scope, it is probably fair to say that there is likely to be a significant amount only found here. Whether, and then how, this can be used to aide current or budding Pikettys are the questions to consider.

This article gives the views of the author, and not the position of the South Asia @ LSE blog, nor of the London School of Economics. Please read our comments policy before posting.

Researchers, students and academics familiar with LSE's Library and classification system can find the historical statistics and government publications on the open access shelves on the lower ground floor. The store statistics can be requested to be viewed in the Women's Library Reading Room and further information about access arrangement can be found [here](#). To request a a copy of the full report or a full list of the publications contact email: i.bhullar@lse.ac.uk.

About The Author



Inderbir Bhullar is curator of economics and social policy at LSE Library where he was worked since 2013, connecting students, staff and anyone interested with our rare and unique [special collections](#). He tweets [@EconomicCurator](#).