

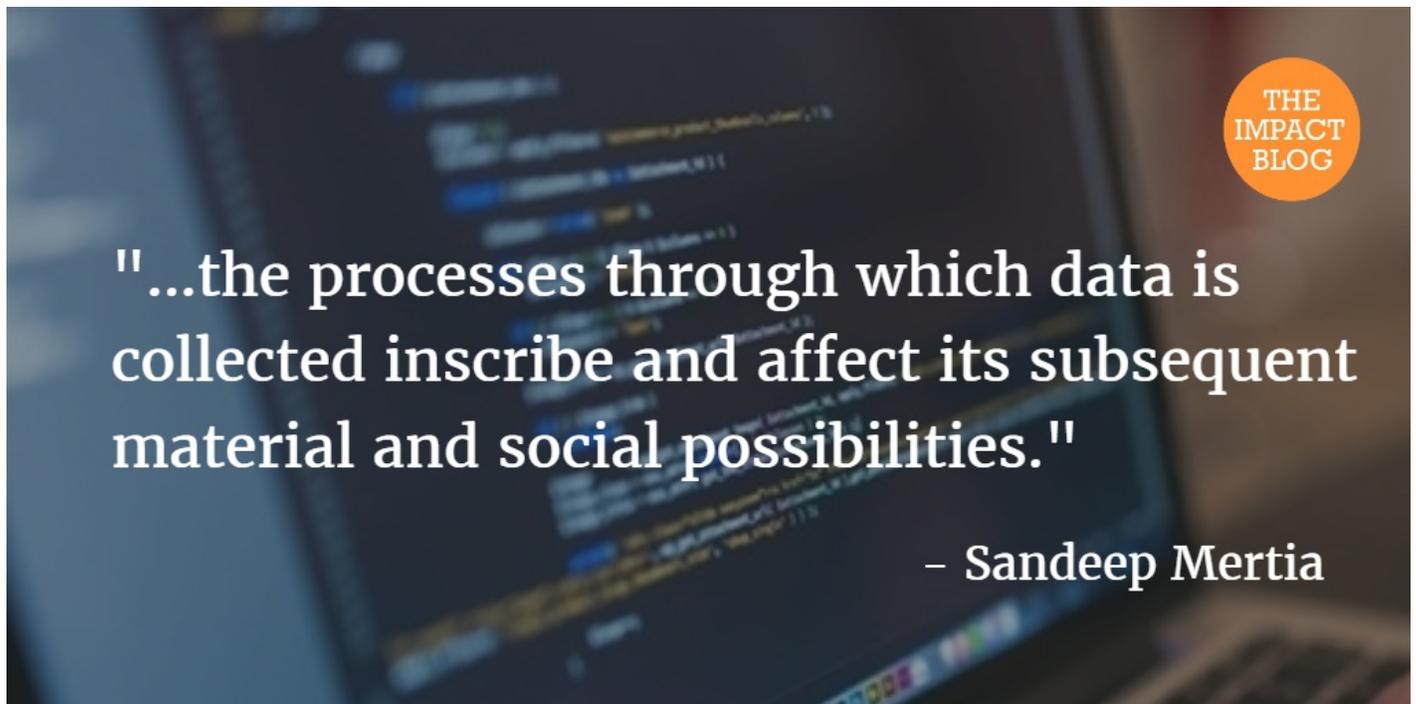
From Computing Clerks to Androids: Two bits on the material lives of social data in India.

*The social and material conditions of data collection have a significant bearing on how we think about and understand data. **Sandeep Mertia** looks at the history of data collection in India and how the conditions have changed over time. From the work of the eminent statistician and founder of the Indian Statistical Institute, Prasanta Chandra Mahalanobis, to the now large scale surveys conducted through tablets and Android apps, various mechanisms have shaped the material lives of survey data.*



In the world of Big Data, Cloud Computing and Software as a Service (SaaS), the material affordances and immanence of technology have reconfigured – if not transformed – earlier modes of collecting, storing, cleaning, processing, computing and visualising ‘data’. The nature of these reconfigurations is far from just ‘technical’ (if there exists any such thing). As we know ‘Raw Data’ is an oxymoron, and different disciplines cook and savour data as per their historically and culturally constructed epistemologies ([Gitelman 2013](#); [Bowker & Star 2000](#)).

It also follows that, how data is framed and how it frames us is contingent upon the tools, techniques and infrastructures available to work with it. While we have a substantial body of work on what digital infrastructures and software and databases in particular, *do to and with data* ([Blanchette 2011](#); [Manovich 1999](#); [Mackenzie 2012](#); [Dourish 2015](#); [Castelle 2015](#)), we still have very little meta-data on the lives of data in the Global South.



To understand what the data revolution means in the Global South, it is imperative to formulate lateral, non-teleological entry points into contemporary debates on politics of data in societies which have had and continue to have fairly different encounters and experience with statistics, analytics, paper-files based bureaucracies, electrification, telecom, Internet and so on ([Mahalanobis 1949](#); [Mahalanobis 1958](#); [Gupta 2015](#); [Evans 1992](#); [Mazzarella 2010](#); [Jeffrey and Doron 2013](#)).

India's social data ecosystem with its rich history of (modern) data-driven knowledge production since at least the British colonial times and its contemporary multitudinous transitions from paper to digital systems, offers many relevant sites for meditating on the social of the computational and vice versa. I use the term 'social data ecosystem' in an indicative sense to include the central, state and local governments, Non-Governmental Organisations (NGOs), policy think tanks, development and philanthropic organizations, survey agencies and the technology companies which work directly or indirectly on collecting or analyzing data related to education, health, rural development and other social indicators.

We know that the processes through which data is collected inscribe and affect its subsequent material and social possibilities. In a system where data exists in multiple or mixed materialities – with several micro differences within macro categories like paper and digital – data collection becomes key for opening the debates on *how* data is produced.

Before Post-Humanism: When Computers were Humans

Before the first digital 'computer' was imported in 1955, *computers* in India in 1930-40s were the '*statistical staff*' members for calculation and tabulation work of large scale sample surveys conducted by the pioneer statistician and founder of the Indian Statistical Institute (ISI), Prasanta Chandra Mahalanobis. In his [reflections](#) on those surveys, he wrote:

In 1937 there was not a single trained field worker, and only about half a dozen computers ... The whole of the field staff was recruited for only three or four months, and continuity of employment could not be guaranteed ... On the statistical side, however, it became possible to train up and give more or less continuous employment to a good proportion of computers by employing them on other projects.

In 1953 the [first general report on the National Sample Survey \(NSS\)](#) conducted under his leadership, noted that:

*To make suitable arrangements for the work of tabulation and analysis of the primary data, **more than 100 additional computing clerks** were appointed and given training in the Indian Statistical Institute. Since much of the work was to be done by tabulating machines, training was also given to a large number of punchers and verifiers in the Institute both in Calcutta and its branch at Giridih in Bihar. Arrangements were made to hire the latest types of tabulating machines from the International Business Machine Corporation (IBM) of New York; and by latter part of 1951 the Institute had two new models of IBM tabulators, a new multiplier and several sorters, reproducers, etc. in addition to some of the machines of the British Tabulating Machine Co. which the institute had been using for the some time. An Electronic Statistical Machine (a high power combined sorter-tabulator) was also rented from the IBM. [Emphasis added]*

As we can see from just these two instances that in about a decade's time the infrastructure and language of statistics and computing went through a tectonic change (with even more disruptive changes in the political climate post the second world war and India's independence). Broadly these changes constituted the material basis for the post-independence Nehruvian era of centralised planning with meteoric rise of statistics as a discipline and creation of all the key institutions which manage official statistics till date.

A defining feature of the rise of statistics from the late 30s to the 50s, was the conceptualisation and evolution of the 'sample survey' methodology by Mahalanobis. His idea of 'data' was explicitly defined by what was considered *feasible* to collect by the 'field staff' with low literacy levels, over large populations and area sizes, and with minimum

expenses. In his notes he gives vivid details like the organisation of field staff's compensation depending upon the 'journey time' – different for day and night timings – between two sampling units. Such social and material conditions of data collection in Mahalanobis' times have a significant, but largely unexplored, bearing on how we continue to collect, and think about, 'data' in India.

Fast Forward to 2015: The Emergence of Post-Data-Entry World & Related Myths

Advancements in computing and proliferation of data over the years didn't quite disrupt the old model of 'field staff' and 'statistical staff'—now referred to as IT staff which includes data entry operators who digitise survey responses from paper. The use of mobile computing devices—tablets for data collection in large scale surveys is a very recent occurrence.

As part of my ongoing ethnography of emerging data ecosystems in India, I've been observing large scale surveys in remote, rural locations being conducted through an *Android* data collection application (or 'app') on tablet devices. Beyond the celebratory discourse of efficiency of digital over paper-based systems and performativity of software, the app based data collection mechanism seems to engender a new material whirlpool in the lives of survey data.



Figure 1 – Charging Tablets in a Fieldsite

The field staff, who are trained to use an android tablet and the data collection app for a week before they are sent back to their villages to conduct the surveys, have now subsumed the role of the data entry operator. While collecting data and syncing it with the company's servers is imagined and marketed to be a real time process, in practice it involves a series of infrastructural and cultural negotiations involving highly varied levels of expertise among the field staff, their physical mobility and distances from town and other staff, procedural accuracy and completion of the survey, network connectivity, SIM card activation, mobile internet pack balance, battery life of the tablets, availability of electricity (and chargers and extension cords) [see image], version of the android app and last but not the least, bugs!

This convoluted techno-cultural trail of data opens new zones of not just remediation of surveys and the surveyors, but it also reveals the epistemic and material investment in obtaining machine readable (digital) data which can then be sliced and diced and represented in myriad ways.

The move from paper to digital in such conditions thus, does not just fulfill a certain criteria of efficiency, more importantly, it redistributes the values associated with knowledge of the 'social' and representation of both the knowledge and the apparatus – especially software – involved in producing it.

Mining Data for (tentative) Archaeology of Knowledge

While it may be historically odd to attempt a juxtaposition of two data collection systems which seem to be temporally and technologically at a great distance from each other, such archaeological tactics are now vital to forge connections which allow us to revisit the always already new lives of data.

From an ethnographic perspective, when '[deep learning](#)' on [e-commerce websites](#) happens in the same data ecosystem where last mile internet and telecom connectivity is a mammoth challenge, it is safe to assume that the mixed materiality of systems and their webs of connections will continue to animate the lives of data. Perhaps it is also safe to assume that to think through such ecosystems – of computing clerks and androids without a clear sense of boundaries between their respective 'fields' and databases – we will need a few epistemological version updates. Software Studies 2.0 (*not* Ethnic-Software Studies), anyone?

Note: This article gives the views of the author, and not the position of the LSE Impact blog, nor of the London School of Economics. Please review our [Comments Policy](#) if you have any concerns on posting a comment below.

About the Author

Sandeep Mertia is a Research Associate at The Sarai Programme, Centre for the Study of Developing Societies, Delhi. He graduated from the Dhirubhai Ambani Institute of Information and Communication Technology with a Bachelors in Information and Communication Technology (BTech. (ICT)) in 2014. He received the Social Media Research Fellowship from Sarai-CSDS, May-November 2014, for ethnographic research on digital technologies and social media ecologies in rural Rajasthan. His research interests lie at the intersections of Science & Technology Studies, Software Studies and Anthropology. He Tweets at [@SandeepMertia](#).

- Copyright 2015 LSE Impact of Social Sciences - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.