Sandeep Mertia (Hg.)

Lives of Data. Essays on Computational Cultures from India

2020

https://doi.org/10.25969/mediarep/19259

Veröffentlichungsversion / published version

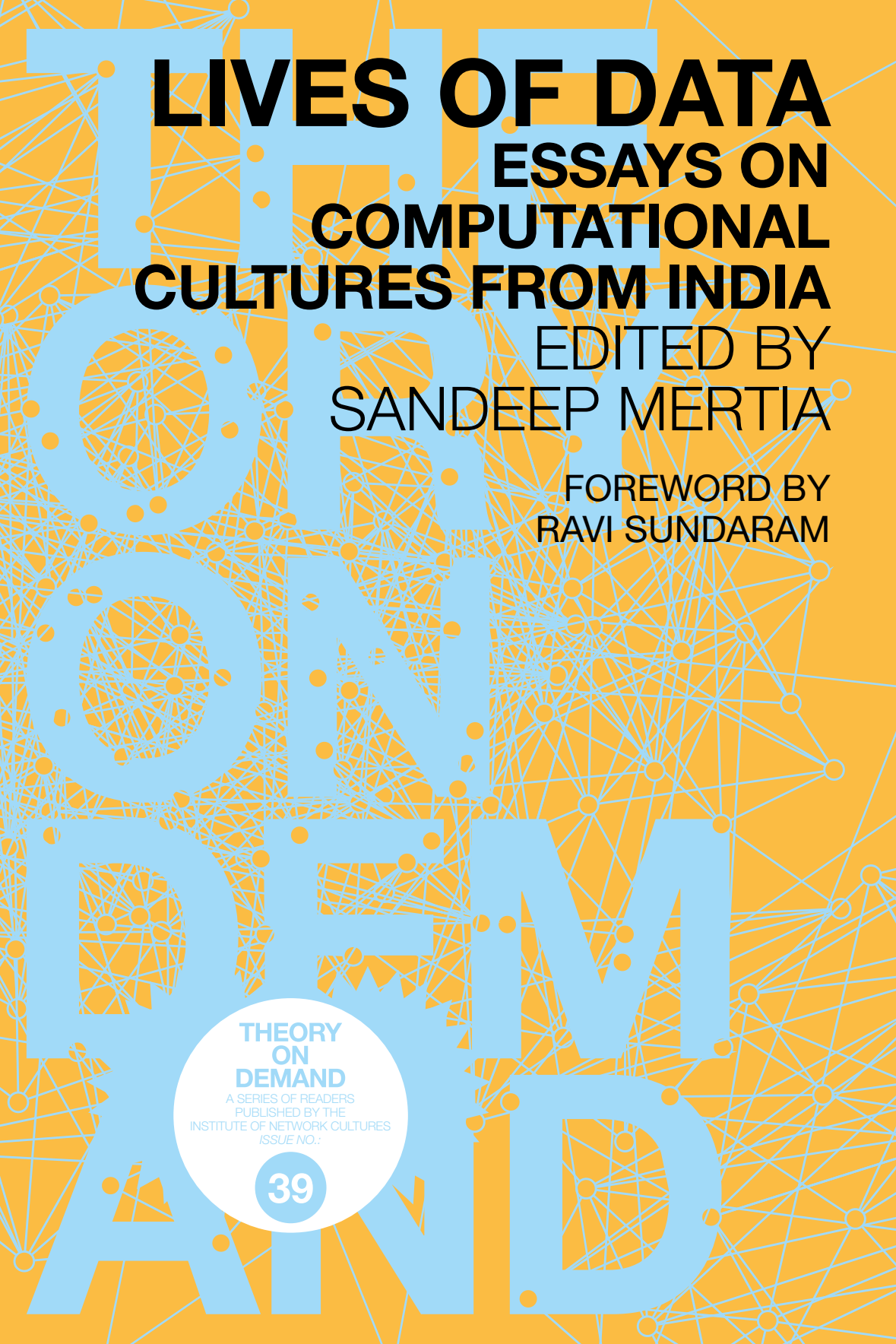
Buch / book

LIVES OF DATA

ESSAYS ON COMPUTATIONAL CULTURES FROM INDIA

EDITED BY
SANDEEP MERTIA

FOREWORD BY
RAVI SUNDARAM



LIVES OF DATA:

ESSAYS ON COMPUTATIONAL CULTURES FROM INDIA

EDITED BY SANDEEP MERTIA

FOREWORD BY RAVI SUNDARAM

Theory on Demand #39
Lives of Data: Essays on Computational Cultures from India

Edited by Sandeep Mertia
Foreword by Ravi Sundaram

Authors: Sandeep Mertia, Karl Mendonca, Sivakumar Arumugam, Ranjit Singh, Puthiya Purayil Sneha, Lilly Irani, Anumeha Yadav, Preeti Mudliar, Prerna Mukharya and Mahima Taneja, Guneet Narula, Gaurav Godhwani, Noopur Raval, Aakash Solanki, and Anirudh Raghavan

Copy-editing: Divya Chandok
Cover design: Katja van Stiphout
Design and production: Chloë Arkenbout
Publisher: Institute of Network Cultures, Amsterdam, 2020

ISBN print-on-demand: 9789492302717
ISBN EPUB: 9789492302700

Contact

Institute of Network Cultures
Email: info@networkcultures.org
Web: http://www.networkcultures.org

Order a copy or download this publication free of charge at
http://networkcultures.org/publications.

CONTENTS

ACKNOWLEDGMENTS

Lives of Data: Essays on Computational Cultures from India is an outcome of collaboration between a diverse group of researchers and practitioners interested in cross-disciplinary conversations and scholarship. The contributors to this volume, many of whom participated in the first 'Lives of Data' workshop at Sarai-CSDS, 05–07 January 2017, generously agreed to contribute short essays based on academic and practice-based research that could help initiate and refine debates on big data in India.

The 'Lives of Data' research project at Sarai-CSDS, from which the workshop series and this book have emerged, would not have been possible without the generous encouragement and support of Ravi Sundaram and Ravi Vasudevan. Their fortitude in giving an opportunity to a fresh engineering graduate to formulate a new research project on the ethnography of big data in India is emblematic of Sarai's commitment to interdisciplinary research and advancement of young scholars. My colleagues at Sarai: Ishita Tiwary, Siddharth Narrain, Ritika Kaushik, Satakshi Sinha, Riley Gold, Vebhuti Duggal, Mehak Sawhney, Ashish Mahajan, Vikas Chaurasia, and Sachin Kumar provided a stimulating work environment and lasting friendships. Aakash Solanki, who has been a co-conspirator in thinking about the 'social' of computing since our college days, helped refine my research and writing at all stages of this project. At CSDS, everyday conversations with Awadhendra Sharan (Dipu), Ravikant, Alankar, Aditya Nigam, Prathama Banerjee, Rakesh Pandey, Avinash Jha, Abhay Dubey, Rajeev Bhargava, Peter deSouza, Shail Mayaram, Sanjay Kumar, Ashis Nandy, Tarangini Sriraman, Melissa Myambo, Satendra Kumar, Sunalini Kumar, Saagar Tewari, Shireen Mirza, and Ashutosh Kumar provided a unique education with conviviality. The presence and collegiality of all faculty and staff at CSDS have enriched this project in many tangible and intangible ways for over three years.

The two 'Lives of Data' workshops were made possible with the support of special grants from the CSDS and the Indian Council of Social Science Research (ICSSR), respectively. I thank Divya Chandhok for meticulously copy-editing the manuscript. I want to offer a special thanks to Geert Lovink, Chloë Arkenbout and everyone at the Institute of Network Cultures for their support and collaborative efforts in bringing this publication out.

Sandeep Mertia

Jodhpur, September 2020

FOREWORD

RAVI SUNDARAM

It is now almost 25 years since the internet arrived in its early avatar in India in the mid-1990s. Following cycles of boom and bust, the digital economy in India has been expanding steadily for the past decade. The larger promise was the offer to reformat the infrastructure of governance and energize capitalist expansion. By 2020, it appears that this project has partially succeeded. The digital economy remains a significant part of India's future designs; not a day passes without new informational slogans emanating from state managers and regime planners. The catastrophe of COVID-19 suggests that these moves will, in fact, accelerate; it remains to be seen how the cheery optimism of the start-up era will play out in the context of authoritarian politics, socio-economic crisis, and pandemic melancholia.

The Sarai programme at the Centre for the Study of Developing Societies (CSDS), Delhi, was an intellectual response to the early years of digital culture in India. Conceptualized in the late 1990s, the waning years of the now-mythic early internet, Sarai began to address the radical, research, and practice implications for digital media in an unequal and non-Western country. The early years of Sarai fashioned a unique combination of experimental practice, rigorous fieldwork and writing, and regular publishing and exhibitions. This cross-disciplinary thrust stands out today as research into digital media has bifurcated into communication scholars, humanistic media studies, STS scholars, and information science.

Ian Hacking had famously referred to the 'avalanche of numbers' in the 19th century, which made populations, landscapes, and networks legible to governmental power.[1] Various documentary practices, like the cadastral map, census registers, and health records, were combined with enumerative/recording technologies like fingerprint and photography. Social security numbers and ration cards followed in the 20th century. Calculative strategies in colonial and postcolonial India were geared towards an orchestration of flows: of humans, technical artifacts, and species. In the West, the proliferation of statistical techniques and the rise of recording technologies, like typewriters, stencil duplicators, filing systems, allowed information to be indexed, retrieved, and transmitted. In India, manual writing technologies remained powerful, complemented by innovative statistical techniques for field surveys.

In 2006, at the cusp of the Web 2.0 transitions, Sarai organized the Sensor-Census-Censor conference, with scholars and artists.[2] It was set up as a series of cross-disciplinary and experimental encounters between historians, media theorists, activists, artists, curators, and researchers. The public call of Sensor-Census-Censor set up an ambitious informational map that included 'territorial surveys and census forms, public and private archives, documents

1 Ian Hacking, *The Taming of Chance*, Cambridge (UK): Cambridge University Press, 1990.

2 Sarai Media Lab, '*Sensor-Census-Censor: An International Colloquium on Information, Society, History and Politics*', New Delhi: Sarai-CSDS, 2007, http://archive.sarai.net/files/original/513f079389216ba9636dbe4e12ec8f17.pdf.

and databases, reports and records, surveillance cameras and electronic filters, informers and informants, fingerprints and biometrics, photographs and recordings, and a host of other technologies, methods and practices register the changes of state that occur in societies.' The catalog of questions in the Sensor-Census-Censor public call remarkably anticipated many of the debates to come in media scholarship. The material turn has opened up a range of questions for media studies. Witness the interest in the affective potential of objects, courtrooms are transformed into forensic theatres as the sanctity of human testimony has been blurred with media technologies. There is also an interest in the longer map of media infrastructure: ranging from photography in the 19th century to contemporary digital media. As Sudhir Mahadevan has shown us, we are dealing with a contemporary media, which is also a 'very old machine', with surprising jumps and returns.[3]

Sensor-Census-Censor had suggested that contemporary sovereignty relied on documentary artifacts, calculative technologies, and media storage systems. In fact, calculation has dramatically reinserted itself in contemporary debates, offering many things at the same time: a revitalization of policy through transparency indicators and real-time dashboards, a modulation of governance through device-led participation rather than contingent political speech.[4] Emerging calculative infrastructures have generated a volatile mix of actors: data intermediaries and server farms spread worldwide, managerial technocrats, older employees affected by audit culture. There are shifting interface zones for subaltern populations and migrants, along with para-legal networks of hardware, money-transfer, and ID documents. These changes offer us a diagnostic of the contemporary, its atmospheric shifts and technical affordances. These mixtures of the calculative and the sensory have raised all kinds of new questions, which could not have been anticipated by the Sensor-Census-Censor conference in 2006. Data infrastructures in 2006 were not manifest in the way they are in India in 2020. Data as both a category of infrastructure and as a philosophical provocation required a new generation of scholarship.

Edited by Sandeep Mertia, a next-generation Sarai researcher (now at New York University), *Lives of Data* begins the difficult yet pioneering task of engaging with India's informational turn in the past two decades. This was the significant unthought in the Sensor-Census-Censor conference, and this edited volume opens the way forward. Bringing together a new set of exciting researchers, this collection sets up encounters between STS, anthropology, information studies, and the history of science. Equally, *Lives of Data* brings us reports from data practitioners in India that are usually missing in collections of this kind.

In his framing introduction, Mertia points to the sociotechnical relationalities of data, which capture the complex overlaps between human and machine which mutate and shift through space and time. *Lives of Data* addresses both the power and limits of what Katherine Hayles has called the non-conscious cognition of computational systems, by filtering that concept

3 Sudhir Mahadevan, *A Very Old Machine: The Many Origins of the Cinema in India*, Albany: State University of New York Press, 2015.

4 Christopher M. Kelty, 'Too Much Democracy in All the Wrong Places: Toward a Grammar of Participation', *Current Anthropology* 58.S15 (2017): S77–S90.

through a non-Western lens.[5] At the same time, *Lives of Data* productively engages with the cross-disciplinary questions thrown up by the 2006 conference, bringing historical and conceptual debates to the fore. It does so in a map that stretches from the early statistical thinkers to contemporary biometrics and neoliberalism. In all ways, *Lives of Data: Essays on Computational Cultures from India* offers us the first collaborative steps towards understanding the informational present in India. *Lives of Data* shows us the narratives of populations imprisoned by digital platforms, or the 'death' of networks may be too easy; the task of research has just begun.

Delhi, September 2020

5 Katherine Hayles, *Unthought: The Power of the Cognitive Nonconscious*, Chicago: University of Chicago Press, 2017.

INTRODUCTION: RELATIONALITIES ABOUND

SANDEEP MERTIA

> It is not difficult to see what is wrong with official statistics in India. There is gap between theory and practice. There is gap between the means and the end in the absence of any clearly perceived purpose.
>
> - P. C. Mahalanobis, *Statistics as Key Technology*, 1965
>
> Data is its own means. It is an unlimited non-rivalrous resource. Yet, it isn't shared freely. What began as a differentiator is now the model itself.
>
> - Nandan Nilekani, *Why India needs to be a Data Democracy*, 2017

Data shadows our situation. Many believe it can determine our situation. There were enthusiastic claims that 'Big Data' would lead to a 'fourth industrial revolution' and the 'end of Theory', and that it will 'transform how we live, work, and think'.[1] Arguably, much of the early 2010s hype around the big data revolution has already been replugged into popular narratives of artificial intelligence (AI).[2] The media infrastructures that enliven digital data and the fast-moving claims of data revolution are now evidently more globalized and capitalized than ever before. If we look a little under the hood, techniques such as data mining have moved from the margins of techno-scientific practice to normative centers of global computing in less than two decades. How did data become so powerful, pervasive, and relatable in the first place? To understand the global momentum of the data revolution, it is crucial to inquire into the many lineages, affinities, and relations of data in context-sensitive ways.

The first step towards such an inquiry is to understand the relational nature of data in computational cultures. The actual and potential relations that various software, computational objects (e.g., biometric data) and techniques (e.g., micro-targeted advertising) have with our media, bodies, devices, and infrastructures are constituted by diverse kinds of production and processing of data. In broad terms, it is the cultivation of relationalities of data—for instance, mapping populations onto biometric databases that can be linked with bank accounts—that has emerged as a key feature of contemporary modes of governance and knowledge and value production. The stakes for understanding how data inscribes and mediates political power and socio-cultural difference are predictably high. Our bid here is to track the intricacy of the lives of data in theories and practices of human and natural sciences, technology,

1 Chris Anderson, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', *Wired*, 23 June 2008, http://www.wired.com/2008/06/pb-theory/; Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, 2013.

2 Geethika Bhavya Peddibhotla, 'Gartner 2015 Hype Cycle: Big Data Is Out, Machine Learning Is In', *KDnuggets* (blog), 28 August 2015, https://www.kdnuggets.com/gartner-2015-hype-cycle-big-data-is-out-machine-learning-is-in.html/.

media, governance, and politics to better understand emergent computational cultures in India and South Asia.

Dominant models of 20th-century information economy and governance, from cybernetics to notions of a post-industrial network society, universalized a certain context-free, mathematically representable view of information.[3] Actually, much of the world did not follow the pattern of first encountering digital information and computer networks in a military-industrial complex, followed by expansion into formal markets.[4] With an unprecedented number of people beginning to get proper access to the internet through smartphones in India and the Global South at large, both the state and private companies have been grappling with rapidly evolving conditions of governance, involving dynamic innovations and changes in media circulation and consumption. With the penetration of everyday infrastructures of computing and the emergence of new technological imaginaries in large parts of the world, 'context' is now subject to a sociotechnical production that demands fresh interdisciplinary inquiry. To explore the proliferating machinic and cultural ontologies of data, we need to rethink the relations between technological objects and their social lives. Rather than being unprocessed digital information, data needs to be approached as a constitutive technological object and cultural-economic commodity integral to infrastructures of media, governance, business, and life at large.[5]

Data is never produced in silos. Life of any kind of data is shaped by actual and potential relations with other existing data, classifications, paper and digital infrastructure, statistical techniques, data collection and cleaning practices, and possibilities of circulation.[6] Such a life of data is not entirely new and derives from the emergence of modern states and statistics over the past two centuries.[7] While there has obviously been a change in the proliferation of digital media-technologies in recent times, with global internet traffic reaching zettabytes (i.e., trillion gigabytes) of data per day, the event of a big data 'revolution' is not about data deluge. Even Abul Fazl's *Ain-i-Akbari*, a 16th-century administrative report composed under the Mughal Emperor Akbar, produced unprecedented numerical accounts of the army, agriculture, commerce, caste, and geography.[8] The big data 'revolution' is grounded in changing relationalities between data, techno-scientific practices, communication, and connectivity that can be thought of in computational terms and experienced as life in a world increasingly populated by composite human-machine networks.

3 Orit Halpern, *Beautiful Data: A History of Vision and Reason Since 1945*, Durham: Duke University Press, 2014.

4 Ravi Sundaram, *Pirate Modernity: Delhi's Media Urbanism*, London and New York: Routledge, 2009.

5 Geoffrey C. Bowker and Susan L. Star, *Sorting Things Out: Classification & Its Consequences*, Cambridge, Mass.: The MIT Press, 2000; Arjun Appadurai (ed.) *The Social Life of Things: Commodities in Cultural Perspective*, Cambridge: Cambridge University Press, 1986.

6 Lisa Gitelman (ed.) *"Raw Data" Is an Oxymoron*, Cambridge, Mass. and London, England: The MIT Press, 2013.

7 Alain Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning*, trans. Camille Naish, Cambridge, Mass.: Harvard University Press, 2002.

8 Norbert Peabody, 'Cents, Sense, Census: Human Inventories in Late Precolonial and Early Colonial India', *Comparative Studies in Society and History* 43.4 (2001): 819.

Lives of Data seeks to better understand the status of data objects, relationalities, and difference in computational cultures. A critical focus on India necessitates pluralistic vantage points for examining the contemporary global discourse of data revolution in relation to the enduring legacies of colonialism and 20th-century modernization programs. From state-supported technological boosterism of its 'digital superpower' status to everyday lives of over a billion people in one of the most diverse and unequal societies in the world, India's sociotechnical conditions assemble deeply contrasting lives of data. This collection of essays features a diverse group of interdisciplinary scholars and practitioners, engaging the emergence, limits, potentialities, politics, practices, and consequences of data-driven knowledge production and circulation. Encompassing history, anthropology, science and technology studies (STS), media studies, civic technology, data science, digital humanities, and journalism, the essays open up possibilities for a truly situated global and sociotechnically specific understanding of data, computing, and society. Thinking beyond India's storied emerging market and demographic size that draw data extractivist platforms, *Lives of Data* offers novel points of entry for critical inquiry into how computational cultures generate and modulate the global in context. In the rest of this essay, I introduce and contextualize the research questions and debates that have shaped this book.

Data Revolution(s) in Context

The contrast between the two epigraphs above is a good place to begin tracking lives of data. The first epigraph is from a lecture in 1965 at the 125th Annual Meeting of the American Statistical Association by P. C. Mahalanobis, founder of the Indian Statistical Institute (ISI) and a member of the Planning Commission, a powerful body at that time. In this lecture, he emphasized the need to establish a 'purposive' view of statistics as a 'fully developed technology of a multi-discipline character'.[9] This was especially so in the 'underdeveloped countries' where the 'principle of authority' of the government reigned supreme over 'independent' statistical analysis and interpretation.[10] Mahalanobis made these observations at a time when the ISI and India's official statistics and economic planning system were receiving global recognition for pioneering work in research, training, sample-survey methods, and economic planning (Chapter 1). He clearly placed statistical knowledge production in the service of postcolonial nation-building. The desire to perceive a clearly defined 'purpose' when the ISI was already at the cutting edge of large-scale data collection and processing stands in puzzling contrast to contemporary modes of data-driven governance which claim 'data is its own means'.

The second epigraph is from an opinion piece by Nandan Nilekani, co-founder of Infosys and founding chairman of Unique Identification Authority of India (UIDAI), the government body responsible for the world's largest biometric database, Aadhaar. In this article he argues for the value of big data and artificial intelligence for disrupting existing patterns of information management, and cautions against 'data colonization' by state and global platforms.[11] It is

9 P. C. Mahalanobis, 'Statistics as a Key Technology', *The American Statistician* 19.2 (1965): 43.

10 Ibid.

11 Nandan Nilekani, 'Why India Needs to Be a Data Democracy', *Livemint*, 27 July 2017, https://www.livemint.com/Opinion/gm1MNTytiT3zRqxt1dXbhK/Why-India-needs-to-be-a-data-democracy.html.

important to note that what we now know as Aadhaar actually began in 1999 as an identity card project for citizens living in border states.[12] The Rangarajan Commission, set up in January 2000 to look into the 'growing concern regarding the quality of data' in the entire statistical system, recommended the creation of a 'centralized database of citizens (population register)' in which every citizen would have a unique identification number.[13] Within a few years of the UIDAI being set up in 2009, Aadhaar became a primary key linking databases of bank accounts, mobile phones, income tax returns, payment apps, email IDs, and so on, even if such a linking is not mandated by the law.[14] Aadhaar has afforded development of application programming interfaces (APIs), and web and mobile applications with payment interfaces demanding Aadhaar verification for government and private services across domains.[15] Perhaps nobody in 2009 could have imagined connecting biometric data to mobile phone SIM cards. Anumeha Yadav (Chapter 7) draws on her detailed field reports to show how the project grew from select pilot implementation in 2011 to a national legal and policy imperative by 2017. She notes a growing public alertness to the importance of enrolling with Aadhaar to ensure the ratification of rights, irrespective of the unclear legal status and the widespread technological glitches in the everyday functioning of the project. The story of Aadhaar raises questions about *what* counts as data, *who* can design its purposes, and *how* its means and ends are discovered. It is a story that is at once expansionist and contingent: in India, the evolution of Aadhaar indicates that we need to reflect on computational culture without prefiguring the object of computation and its potential relationship to taxonomies of social control.

To understand the shift that has taken place between the data in the mid-20th-century statistical regime of economic planning and big data aggregation and prediction in the contemporary, we need to re-examine the history of computing in India, which has been largely tethered to the IT revolution.[16] We examine different techniques and affordances of computation in different media ecologies consisting of human computers and mass-media such as telecom in the decades before the emergence of the internet.[17] In Chapter 1, I explore the role of the 'first computers' of India—both human and electronic—from the 1930s to 1960s in generating official statistics. In Chapter 2, Karl Mendonca analyses the role of computerization in the 1980s at a major advertising company involved in the cinema business, and how the company later repurposed its cinema distribution network into a courier company. In different ways, both chapters challenge the notion of a clear and stable rationale for the evolution of computers and big data.

12 R. Ramakumar, 'What the UID Conceals', *The Hindu*, 21 October 2010, sec. Lead, https://www.thehindu.com/opinion/lead/What-the-UID-conceals/article15786909.ece.

13 Chakravarthi Rangarajan, 'Report of Dr. Rangarajan Commission', Ministry of Statistics and Program Implementation | Government of India, 2001, http://www.mospi.gov.in/report-dr-rangarajan-commission.

14 Reetika Khera, (ed.), *Dissent on Aadhaar: Big Data Meets Big Brother*, Hyderabad, Telangana: Orient BlackSwan, 2018.

15 'India Stack - The Bedrock of a Digital India', *IndiaStack* (blog), 17 November 2016, http://indiastack.org/india-stack-the-bedrock-of-a-digital-india/.

16 Dinesh C. Sharma, *The Outsourcer: The Story of India's IT Revolution*, History of Computing. Cambridge, Mass.: The MIT Press, 2015.

17 Paula Chakravartty, 'Telecom, National Development and the Indian State: A Postcolonial Critique', *Media, Culture & Society* 26.2 (2004): 227.

It was not until the early 2000s that database practitioners began to seriously look at data mining as a mode of knowledge production.[18] New concepts of scale and computational processing power emerged and developed through trade-offs and reconfigurations of statistical accuracy, localized data storage and retrievability, hardware and software load balancing, and electricity consumption. Of particular importance was the shift from 'relational' (structured design) to 'non-relational' (distributed design) database management systems.[19] Here, we must not forget the co-production of affordances, users, and publics.[20] After all, a computer database is only one specific instance of a wider set of relationalities made durable by the thoroughly material and well-constructed craft of software engineering—even if it is widely imagined to be abstract and mystical.[21] In the Indian context, while the IT industry has become symbolic of a new middle-class imaginary of technology and social mobility, the epistemic cultures of software engineering and their relations with global developments are yet to be adequately unpacked.[22] We do not know how India's political and infrastructural conditions affect Aadhaar's database design or the development of high energy-consuming data centers for 'data sovereignty', to name but two examples.[23]

In a post-colony like India, any critical engagement with data-driven knowledge production has to consider the persistent role of colonial biopolitics. It is well established that statistics—formerly termed 'political arithmetic'—have played a key role in the production of people, identity, and nation-states.[24] From the construction of enlightenment ideas such as the 'individual', national populations in Europe, and the 'citizen' in the USA, the intended and unintended consequences of counting and categorizing people run far and wide.[25] European colonies became sites for exotic and imperious enumerative and classificatory systems framed by orientalist pedagogies that displaced and serialized existing social orders. From the inventions of fingerprinting and the enumeration of complex traditions of faith and social difference into the fixities of religious identity and objectification of caste, such a biopolitics sought to make populations knowable and governable.[26]

---

18 Matthew L Jones, 'Querying the Archive: Data Mining from Apriori to PageRank', in Lorraine J. Daston (ed.) *Science in the Archives: Pasts, Presents, Futures*, Chicago: University of Chicago Press, 2017, pp. 311–328.

19 Paul Dourish, 'NO SQL: The Shifting Materialities of Database Technologies', *Computational Culture* 4 (2014), http://computationalculture.net/article/no-sql-the-shifting-materialities-of-database-technology.

20 Christopher M. Kelty, 'Preface: Crowds and Clouds', *Limn* 2 (April, 2012), https://limn.it/articles/preface-crowds-and-clouds/.

21 Matthew Fuller (ed.) *Software Studies: A Lexicon*, Leonardo Books, Cambridge, Mass: The MIT Press, 2008.

22 Carol Upadhya, *Reengineering India: Work, Capital, and Class in an Offshore Economy*, Oxford and New York: Oxford University Press, 2016.

23 Priyanka Sangani, 'Data Centres May Prove to Be the Next Big Opportunity in India', *The Economic Times*, 23 October 2019, https://economictimes.indiatimes.com/tech/internet/data-centres-may-prove-to-be-the-next-big-opportunity-in-india-/articleshow/71714171.cms?from=mdr.

24 Desrosières, *The Politics of Large Numbers.*

25 Ian Hacking, 'Biopower and the Avalanche of Printed Numbers', *Humanities in Society* 5.3–4 (1982): 279.

26 Arjun Appadurai, 'Number in the Colonial Imagination', in Carol Breckenridge and Peter Van Der Veer (eds) *Orientalism and the Postcolonial Predicament: Perspectives on South Asia*, Philadelphia: University of Pennsylvania Press, 1993, pp. 314–339; Information and Society Research Cluster

Post-independence India saw an expansion of bureaucracy, official statistics, and planning. Subsequently, government and transnational businesses used data modelling of the economy and populations to understand citizenship entitlements and consumer profiles. The intersections of state and market interests after economic liberalization in 1991 transformed the national political economy as well as the everyday cultural conditions of governance. In particular, the entry of private digital technology vendors and consultants in state and international development projects afforded new means and incentives for collecting and analyzing data. Supporters of the Aadhaar project often claim that the state is a much more benign collector of data than companies such as Google and Facebook. Putting questions of veracity aside, the implications of this distinction are suggestive. The purported commensurability between data imaginaries and practices of India's welfare state and those of big technology companies widens the scope of inquiry into the politics of data-driven governance and bureaucracy.[27] From state-owned biometrics to state-promoted transnational mobile apps, the contemporary (surveillance-friendly) road between the ideology of the state and that of popular digital media is punctuated by diverse and distributed data-driven pathways.

At one level, the shift from colonial fingerprinting to contemporary biometric technologies shows some continuity in terms of tactics of governance and subjectification of bodies. If we look closely though, the machinic-readability of fingerprints opens new analytical challenges for theorizing governmentality.[28] The contemporary modes of data-driven subjectification are deeply entangled with proliferation of digital technologies of identification in governance, finance, media, and consumer products across developmental and business models. How can we map this expansion and proliferation in sociotechnically specific ways? From navigating the nudge marketing of discount codes on mobile payment apps to facing new determinations of citizenship and identity through myriad paper-based and digital documents, among other things, the emergent mutations of power, subjectivity, and data demand a closer look into the design and material form of media. This is particularly challenging in conditions of fragmented digital infrastructures, where diverse intermedial forms emerge and coalesce in everyday practices for bypassing the lack of end-to-end connectivity and formal access.[29]

Sociotechnical Relationalities of Data

Perhaps the biggest irony about big data is that it has little to do with data per se. Rather, it has a lot to do with classifications, connections, and patterns that emerge or can be gen-

Sarai-CSDS (ed.) *Sensor-Census-Censor: An International Colloquium on Information, Society, History, and Politics*, New Delhi: The Sarai Programme, Centre for the Study of Developing Societies, 2007; U. Kalpagam, *Rule by Numbers: Governmentality in Colonial India*, Lexington Books, 2014.

27 Itty Abraham and Ashish Rajadhyaksha, 'State Power and Technological Citizenship in India: From the Postcolonial to the Digital Age', *East Asian Science, Technology and Society* 9.1 (2015): 65; Aakash Solanki, 'Management of Performance and Performance of Management: Getting to Work on Time in the Indian Bureaucracy', *South Asia: Journal of South Asian Studies* 42.3 (2019): 588.

28 Tarangini Sriraman, *In Pursuit of Proof: A History of Identification Documents in India*, New Delhi: Oxford University Press, 2018.

29 Aswin Punathambekar and Sriram Mohan (eds) *Global Digital Cultures: Perspectives from South Asia*, University of Michigan Press, 2019.

erated when large-scale, high-dimensional, real-time, and variably un/structured data are mashed up with other data. Individual Aadhaar card, credit card, internet history, or any other machine-readable data in itself does not mean much. It has to be positioned within the relationalities and infrastructures that demonstrate for example, the uniqueness of a fingerprint relative to the biometric data of 1.3 billion others, or the classification of one's online purchases in a cluster of other users who might be interested in buying a 'related' product. While invasive collection and monetization of data might feel like the infrastructural norm today, this is not always the case. The online ticket booking website of the Indian Railways (IRCTC) moved to a 'distributed in-memory database' (one of many big data architectures) in 2014 and is still trying to find ways to mine and monetize its treasure trove of user data.[30] Simply having large amounts of data does not afford analytics or intelligence. However, some of the IRCTC data was leaked in 2016 and the data dump was sold in gray markets online as well as in compact disks (CDs) for ten–fifteen thousand rupees.[31]

Intermediaries are key here: from contractual content moderators of the most industrialized platforms to entrepreneurial data brokers who market government-owned as well as private telecom and financial data, intermediaries of various kinds populate, innovate, pause, and punctuate data flows.[32] Neither proliferation nor circulation of data follows any universal law, architectural truth, or definitive model of intelligence. How then do certain actors, epistemes, platforms, and organizations emerge as dominant? There are important conceptual questions involved here, about how computational cultures are deployed to shape the circulation of power, knowledge, and capital in the contemporary, including the constitution of a distinct territoriality.[33] Many commentators, top businessmen, and government ministers have expressed concerns about 'data colonialization' by western technology companies in

30 'Indian Railways: Distributed In-Memory Data Management Solution Improves the Capacity and Availability of New E-Ticketing System', Case Study, Pivotal, 2014, https://content.pivotal.io/case-studies/indian-railways; 'Nothing to Fear about Artificial Intelligence (AI); Should Harness It for Organizational and Social Good: Piyush Goyal', *Press Information Bureau*, 24 March 2018, http://pib.nic.in/newsite/PrintRelease.aspx?relid=177990.

31 Gopal Sathe, 'The Indian Railways Is Sitting on a Giant Trove of Your Travel Data. The Modi Govt Wants to Sell It, and Experts Are Worried', *HuffPost India*, 18 July 2018, https://www.huffingtonpost.in/2018/07/18/the-indian-railways-is-sitting-on-a-giant-trove-of-your-travel-data-the-modi-govt-wants-to-sell-it-and-experts-are-worried-dek_a_23484279/.

32 The history of information is inseparable from that of intermediaries, see, Christopher A. Bayly, *Empire and Information: Intelligence Gathering and Social Communication in India, 1780–1870*, Cambridge: Cambridge University Press, 1996. Intermediaries continue to be essential to lives of data, albeit in new kinds of human-machine networks. See, Aritra Sarkhel and Neha Alawadhi, 'Data Brokerage: How Data Brokers Are Selling All Your Personal Info for Less than a Rupee to Whoever Wants It', *The Economic Times*, 28 February 2017, https://economictimes.indiatimes.com/tech/internet/how-data-brokers-are-selling-all-your-personal-info-for-less-than-a-rupee-to-whoever-wants-it/articleshow/57382192.cms; Ciaran Cassidy and Adrian Chen, *The Moderators*, Documentary, 2017, https://fieldofvision.org/the-moderators.

33 Ned Rossiter, 'Imperial Infrastructures and Asia beyond Asia: Data Centres, State Formation and the Territoriality of Logistical Media', *The Fibreculture Journal* 29 (2017): Computing the City, http://twentynine.fibreculturejournal.org/fcj-220-imperial-infrastructures-and-asia-beyond-asia-data-centres-state-formation-and-the-territoriality-of-logistical-media/.

India.[34] In response, so far, we have witnessed policies such as data localization and promotion of popular technological nationalism by companies such as Jio, which is rapidly monopolizing India's digital economy.[35] We should be careful to not conflate the physical locations of data (i.e., server farms and networked devices) with computational territories of extraction of value and accumulation of power wielded through data. The latter has become possible for companies such as Jio and Uber through assemblages of big data technologies such as Cisco's 'Network Automation Platform' and Apache's 'Kafka' (distributed stream platform), respectively, among many others.[36] In just a few years, these technologies of 'distributed computing' have apparently helped businesses and nation-states to engineer and optimize relationalities of data in the service of large-scale centralization of capital and control. Any critical engagement with such developments demands a robust sociotechnical understanding of data-driven knowledge production and circulation.

Thus, *sociotechnical relationalities* of data—the possible ways in which data generates and is generated by the relations amongst objects (digital and analog), people (collectives of users and non-users), and phenomena (social and mathematical)—are key to understanding the historical and emergent conditions of data-driven knowledge. It is best to approach the constitution of data in provisional and context-sensitive terms. For example, if we look at artificial neural networks (type of machine-learning algorithms) for a) speech recognition and b) facial recognition from datasets of comparable resolution, human-machine relations in the two applications will be qualitatively different from each other. Further, the socio-political preconditions and ramifications of the two applications may radically differ, depending on which nation-state or technology company conducts them and how. Unlike the prevalent notion that a 'full-stack' of tools and skills is required to build and maintain digital platforms, it is impossible to aggregate and predict the social relations through which different layers of digital technologies are constituted. We are thus compelled to cast a wide net to capture the sociotechnical relationalities that set up the plural lives of data.

One might wonder what the point is of focusing so closely on sociotechnical relationalities instead of *means* of production and circulation. Ownership and control are obviously important: it is not for purely stochastic reasons that the celebrated era of big data and related digital revolutions has emerged in conjunction with the global oligarchy of a few technology companies, gig economy, unprecedented circulation of hate speech and fake news, and unbridled surveillance. However, the many layers of technological abstraction through which

34 Osama Manzar, 'What Is Data Colonisation and Why It Matters to Us in India', *Business Standard*, 17 August 2017, https://www.business-standard.com/article/economy-policy/who-owns-your-data-india-needs-to-tackle-data-colonisation-soon-117081700234_1.html.

35 Data localization means storing a nation's citizens data on servers located within its sovereign boundaries. See, 'Notification: Storage of Payment System Data', Reserve Bank of India, 6 April 2018, https://www.rbi.org.in/scripts/NotificationUser.aspx?Id=11244.

36 'Reliance Jio Is Powering the World's Largest All-IP Network', *Cisco*, 2018, https://www.cisco.com/c/m/en_us/network-intelligence/service-provider/digital-transformation/rjio-powering-world-largest-ip-network.html; Chinmay Soman et al., 'UReplicator: Uber Engineering's Robust Apache Kafka Replicator', *Uber Engineering Blog*, 4 August 2016, https://eng.uber.com/ureplicator-apache-kafka-replicator/.

all these changes have become possible cannot be made visible simply by opening technological black-boxes (we will only find semi-conductors and electrons at the bottom of it all) or by accounting for the industrialization of software engineering. The rhizomatic nature of software in general and data analytics in particular has to be taken into consideration before trying to analyze these phenomena on normative political grounds that demand transparency. Unlike hardware-centered notions of value production, the speculative and material affordances of data may affect and even govern capital and the deployment of technologies in the contemporary. Critical here is wider research into a variety of sites to unsettle standard utopian and dystopian narratives of a developmentalist parochialism centered on Europe and North America.[37]

If I may paraphrase Mahalanobis, it is not difficult to see what is wrong with the study and practice of data-driven knowledge production. There are wide gaps in our understanding of global technological proliferation and socio-cultural conditions of access and circulation. Crucially, these gaps provide fresh ground for a politics of hope and alterity in computational cultures, challenging a status quo dominated by a few technologists and organizations. The struggles for privacy laws, open data, algorithmic fairness, inclusive access, and progressive technological governance unfold in intimate relation to infrastructural and cultural proliferation of apps, media content, and techniques of data collection, classification, aggregation, and re-purposing. For instance, the global movement for Open Government Data, with India as one its early adopters, has raised many important questions on privacy and access to government data in specific formats, and the role of experts and civil society institutions as intermediaries in making an open data community.[38] To understand the stakes involved in using various kinds of government and social sector data for advancing public accountability, we must engage with open data practitioners. The chapters by Gaurav Godhwani and Guneet Narula offer novel, hands-on insights on organizational and technological challenges involved in scraping and opening budget data from PDF files (Chapter 11), and practices of collecting and opening data in development sector organizations (Chapter 10). Both chapters help us better understand the value of 'openness' in relation to different computational and institutional choices involved in data collection and circulation.

To explore possible worlds beyond those projected by Silicon Valley and its mirror servers, how can we conceptualize research afresh? How can we bring together researchers and practitioners to think through and beyond computing, as it exists right now and in possible futures?[39] How do sociotechnical relationalities of data emerge and circulate in different contexts and parts of the world? How can we develop an expansive vocabulary—beyond West and the

37 Yuk Hui, 'Cosmotechnics as Cosmopolitics', *e-Flux* 86 (November, 2017), https://www.e-flux.com/journal/86/161887/cosmotechnics-as-cosmopolitics/; Kavita Philip, Lilly Irani, and Paul Dourish, 'Postcolonial Computing: A Tactical Survey', *Science, Technology, & Human Values* 37.1 (2012): 3.

38 Sumandro Chattapadhyay, 'Opening Government Data through Mediation: Exploring the Roles, Practices and Strategies of Data Intermediary Organisations in India', 2015, http://ajantriks.github.io/oddc/.

39 I use the term 'futures' in a broad political sense, inspired by Ashis Nandy's observation that 'no hegemony is complete unless the predictability of dissent is ensured'. Ashis Nandy, 'Bearing Witness to the Future', *Futures* 28.6–7 (1996): 636.

rest—to understand diverse systems such as China's social credit, Kenya's m-Pesa, e-Estonia, and Malta becoming a 'blockchain island', on their own terms? If data-driven technologies are key to contemporary global systems of knowledge and value production and circulation, what can we learn from following the lives of data in specific sites, as this book does with India?

Mapping Lives of Data in Digital India

What does big data in India want? What kinds of input and output of resources and ideas does it command and conjure? At the very least, big data techniques and projects rely upon the availability of apparently seamless digital infrastructures, techno-managerial expertise, large number of users, and tangible market or public policy outcomes. India has the world's largest number of software engineers, fastest growing mobile internet user base and market, and nation-wide government programs for building a 'Digital India', 'Startup India', and one hundred 'Smart Cities'. And yet it has highly fragmented infrastructural conditions of technological access, and nearly half of the population *still* does not have broadband internet access. However, looked at differently, the rapidly growing number of users of digital media—largely through mobile phones—display creativity in working out different approaches for imagining, accessing, and sharing media.[40] The co-existence of large-scale state sponsored technological projects and global technology ventures in India are opening new pathways for innovation and capital accumulation.[41] We must remember that even as the Indian economy grows and political shifts release new aspirational energies, the social context is defined by incredible hierarchies of religion, caste, gender, and class.[42] These exist alongside dizzying variations in formal literacies, languages, aesthetics, and media ecologies.[43] The context sets up many generative and disruptive encounters with long-standing hierarchies and techno-cultural orders.[44] In particular, digitally enabled forms of ethno-nationalism and majoritarian popu-

40 Robin Jeffrey and Assa Doron, *Cell Phone Nation: How Mobile Phones Have Revolutionized Business, Politics and Ordinary Life in India*, Gurgaon: Hachette India Local, 2013; Nimmi Rangaswamy and Payal Arora, 'The Mobile Internet in the Wild and Every Day: Digital Leisure in the Slums of Urban India', *International Journal of Cultural Studies* 19.6 (2016): 611; Aditi Deo and Vebhuti Duggal, 'Radios, Ringtones, and Memory Cards or, How the Mobile Phone Became Our Favourite Music Playback Device', *South Asian Popular Culture* 15.1 (2017): 41; Sandeep Mertia, ''Timepass' Development: Situating Social Media in Rural Rajasthan', *Economic and Political Weekly* 52.47 (2017): 69.

41 Jahnavi Phalkey and Sumandro Chattapadhyay, 'The Aakash Tablet and Technological Imaginaries of Mass Education in Contemporary India', *History and Technology* 31.4 (2015): 452; Lilly Irani, *Chasing Innovation: Making Entrepreneurial Citizens in Modern India*, Princeton: Princeton University Press, 2019; Ishita Tiwary, 'Amazon Prime Video: A Platform Ecosphere', in Adrian Athique and Vibodh Parthasarathi (eds) *Platform Capitalism in India*, Palgrave Macmillan, 2020, pp. 87–106.

42 Nivedita Menon and Aditya Nigam, *Power and Contestation: India Since 1989*, London and New York: Zed Books, 2007.

43 A critical understanding of media histories and ecologies in India and South Asia is essential for engaging with contemporary digital media. See, Ravi Sundaram (ed.) *No Limits: Media Studies from India*, Oxford and New York: Oxford University Press, 2013; Arvind Rajagopal (ed.) *The Indian Public Sphere: Readings in Media History*, Oxford and New York: Oxford University Press, 2009; Ravi S. Vasudevan et al., 'A Vision for Screen Studies in South Asia', *BioScope: South Asian Screen Studies* 1.1 (2010): 5; William Mazzarella, 'Beautiful Balloon: The Digital Divide and the Charisma of New Media in India', *American Ethnologist* 37.4 (2010): 783.

44 The tensions between digital infrastructures and 'social order' can be traced across diverse registers

lism—from propaganda-savvy uncles in WhatsApp family groups to the ruling party's IT cell that commands an army of bots and trolls—are unsettling conventional wisdom about India's democracy in deeply consequential ways.[45]

In the face of accelerating media circulation, the state strives to assert sovereignty by censorship, internet shutdowns, data localization, biometric identification, and digital payments. Meanwhile, an imagined collective of the aspirational 'next billion' users and excited publics capable of full-stack defining (and at times, defying) use of mobile phone-driven media, are participating in un/making infrastructures and subjectivities that few scholars have been able to anticipate and theorize.[46] The rapidly changing relationalities of computing in India imitate, adapt, and confuse ontologies of big data in ways that we have only started to explore.[47]

A dearth of critical scholarship on science, technology, and society allows for vague formulations about the transformative power of the digital.[48] Key here are the claims of technological 'leapfrogging', in which evolutionary stages of digital technologies are bypassed to arrive directly in a smartphone ecosystem.[49] However, to follow the lives of data we need to step back from the generalized hubris of the digital, and interrogate the specifics of the avowed epistemic and material stability projected onto computational imaginaries and practices.

The biopolitics and governmentality of big data operate in curious amalgamations of historical and emergent mathematical, technological, and political dynamics. While technological and political forces are prominent in big data discourse, the mathematics behind how our subjectivities interface with computing remains largely obscure. In Chapter 3, Sivakumar Arumugam illuminates the lives of data as it plays out between modelling and governmentality by examining the Duckworth-Lewis-Sterne (DLS) model to estimate revised targets in cricket. The semiotic activities of the DLS model, Arumugam argues, orients communities

---

of new media use and experience. See, Rahul Mukherjee, ''City Inside the Oven': Cell Tower Radiation Controversies and Mediated Technoscience Publics', *Television & New Media* 18.1 (2017): 19; Siddharth Narrain, 'Dangerous Speech in Real Time: Social Media, Policing, and Communal Violence', *Economic and Political Weekly* 52.34 (2017); Arvind Kumar Thakur, 'New Media and the Dalit Counter-Public Sphere', *Television & New Media* 21.4 (2019): 360; Sujatha Subramanian, 'Is Hindutva Masculinity on Social Media Producing a Culture of Violence Against Women and Muslims?', *Economic and Political Weekly* 54.15 (April, 2019).

45 Ravish Kumar, *Free Voice: On Democracy, Culture and the Nation*, revised edition, New Delhi: Speaking Tiger, 2019.

46 Ravi Sundaram, 'Post-Postcolonial Sensory Infrastructure', *e-Flux* 64 (April, 2015), https://www.e-flux.com/journal/64/60858/post-postcolonial-sensory-infrastructure/; Punathambekar and Mohan (eds), *Global Digital Cultures.*

47 Sandeep Mertia, 'Socio-Technical Imaginaries of a Data-Driven City: Ethnographic Vignettes from Delhi', *The Fibreculture Journal*. 29 (2017): Computing the City, http://twentynine.fibreculturejournal.org/fcj-217-socio-technical-imaginaries-of-a-data-driven-city-ethnographic-vignettes-from-delhi/.

48 Shiv Visvanathan, 'Democracy, Governance and Science: Strange Case of the Missing Discipline', *Economic and Political Weekly* 36.39 (2001): 3684.

49 Vinod Khosla, 'How to Win at Leapfrog', McKinsey & Company, December 2013, https://www.mckinsey.com/featured-insights/asia-pacific/how-to-win-at-leapfrog.

towards counterfactual, probabilistic, and algorithmic futures. This argument helps us understand how data modelling is deployed in a specific context, the everyday infrastructures of leisure and the governance of sport.

In Chapter 4, Ranjit Singh draws from the literature in STS and Information Science to examine how large-scale datasets are constructed, managed, and processed. Singh uses the example of Aadhaar to show how information infrastructures are imagined as layers on top of existing organizational practices. Over time though, it becomes difficult to delineate these layers as information infrastructures imbricate more datasets and organizational practices. The relationships through which an enrollee enters Aadhaar's database using existing identity documents are not the same as the ones involved in authenticating his/her Aadhaar identity for a transaction at a later point in time. To follow the changes that happen in between, Singh offers a methodological maxim: 'study the imbrication'. In Chapter 5, Puthiya Purayil Sneha describes new forms and meanings of 'text as data' and databases of cultural material in the field of Digital Humanities. Examining online archives of poetry and cinema such as *Bichitra*, *Indiancine.ma* and *Pad.ma*, Sneha shows how affordances of database search, editing, transcription, and optical character recognition (OCR) for vernacular languages, among other things, are changing the very media objects of humanities research and practice. The tensions between the new and old data objects and related practices are not a product of any intrinsic quality of digital tools or infrastructures per se. Rather, shifting practices, scale and epistemology of data are intertwined with changes in material design of media-technologies and forms of computation. The key formal quality of digital medium lies in how big data methodologically challenges existing ways of studying media and media infrastructures by assembling abstract models and material objects in capriciously relational ways.

While critical sociotechnical research demands a careful look at technological inscriptions and imbrications, the stakes are different for practitioners. How do technology practitioners approach the 'social' of Digital India? This can't be speculated upon and analyzed purely by the growth in machine-readable data, since a vast majority of users have only very recently started using the internet. An Indian entrepreneur recently wrote in the MIT Technology Review that 'India's mess of complexity is just what AI needs', to help train the machine learning systems to solve more complex social problems.[50] This is not an example of vanilla technological optimism. India today has the third largest number of technology start-ups in the world. Both state and venture capital are deeply invested in nurturing a culture of aspiration and entrepreneurial drive. There is special investment in using cutting-edge technologies to solve social problems. India's national AI policy, for example, identifies five focus areas for AI technologies: healthcare, agriculture, education, smart cities, and transportation.[51] How are technology entrepreneurs engaged in the pursuit of such policies?

50 Varun Aggarwal, 'India's Mess of Complexity Is Just What AI Needs', *MIT Technology Review,* 27 June 2018, https://www.technologyreview.com/s/611478/indias-mess-of-complexity-is-just-what-ai-needs/.

51 'National Strategy for AI #AIFORALL', Discussion Paper, NITI Aayog, June 2018, https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf.

In Chapter 6, Lilly Irani makes visible the hidden pedagogies of 'bias to action', 'management of the political', and information infrastructures that shape this entrepreneurial ethos. Remarkably, various agencies, from design studios to the World Bank use hackathons to assemble entrepreneurial opportunities. These feature challenges to complete time-bound tasks involving extractive use of data infrastructures to arrive at software solutions and design prototypes. By connecting value speculation with social good they become vehicles of what Irani has called 'entrepreneurial citizenship'.[52] On the other hand, open source hackathons foreground politics of care, development, and maintenance of shared infrastructures. In comparing the two, Irani shows how hackathons gather and transform data labor. How are such data labors scaled? Who has to deal with the failures of projects that try to technologically hack social problems? In Chapter 8, Preeti Mudliar takes a closer look at the biometric authentication failures in the use of Aadhaar based on ethnographic research in Ajmer district in Rajasthan. Documenting the experiences of people denied access to food supplies because of Aadhaar authentication failures, Mudliar shows how the burden of repair is put on bodies of excluded citizens, as they become what she calls 'broken data' in the big data system of Aadhaar. The excluded are urged to understand this to be a failure of their bodies to match with stored biometric data. What advocates of Aadhaar refer to as 'teething problems' denies any responsibility to care for those excluded and marginalized by its technological failure.

If centralized data-driven misgovernance is common, there are also many on-going creative attempts at rethinking data analytics to solve grounded problems. In Chapter 9, Prerna Mukharya and Mahima Taneja discuss the work of their organization, Outline India. Foregrounding the 'fieldwork' component of survey research and the use of digital tablets and unmanned arial vehicles (UAVs) in rural areas, they describe wide-ranging efforts to collect good-quality data from the hinterland to drive social policy. Mukharya and Taneja describe steps such as cognitive testing of survey instruments, training modules for fieldworkers, and background studies that have to be conducted even before fieldwork. They note how geo-spatial data from the UAV (drone) needs to be complemented with transect walks, participatory resource mapping, and household surveys to map demographic and caste divisions in a village. It is here that we get a rich glimpse of the axiom, '"raw data" is an oxymoron', and also how to carefully 'cook' data in non-ideal infrastructural conditions.[53] Chapters 9, 10, and 11 by data practitioners provide unique first-hand accounts of the work that goes into building and implementing data-driven systems. We need to carefully follow practitioners in order to understand the technological complexities and potentials of computing in context. Such an engagement is crucial to open out inquiries and interventions in meaningful ways that do not re-produce disciplinary tunnels of thought and practice.

The last section of the book consists of three ethnographic accounts of data analytics in the Indian context that map the everyday life-worlds of different practitioners and technologies. Noopur Raval (Chapter 12) describes how drivers who work for ridesharing apps such as Uber and Ola make sense of the data provided to them by the app dashboard. Raval focuses on drivers' handwritten account notebooks (*Hisaab-Kitaab*, in Hindi) in which they record

52 Irani, *Chasing Innovation*.
53 Gitelman (ed.) *"Raw Data" Is an Oxymoron*.

numbers relating to rides, timing, distance, amount, and mode of payment. These are paperized information objects which drivers use to navigate, narrativize, and personalize the flattened data provided by the app. There are disjunctive lives of data here, in the counterpoint between personal account notebooks and algorithmic rationality of app-based data. Such complications of data driven operations help us think about what keeping a record of one's labor means when apps appear to be the primary medium of adjudication and governance of work. Aakash Solanki (Chapter 13) draws attention to a case of shifting materiality of governance within the bureaucracy. Solanki looks at the adoption and use of a management information system (MIS) in a state government education department to describe how practices of working with paper and digital formats collide and coalesce in unexpected ways. The MIS system was designed to do away with the need for phone calls for sharing data between different offices of the department. In practice, Solanki shows, the MIS becomes part of a messy interplay of computation and writing practices in Indian bureaucracy. Solanki traces the back and forth in and between various digital and paper versions of PDF format of spreadsheets, the filing and annotation practices of bureaucrats, and mobile-phone camera pictures of files 'WhatsApped' from one office to another. He gives us a grounded view of the enduring role of the paper file and its circulation, now reconfigured not against but in tandem with digital media. Anirudh Raghavan (Chapter 14) looks at data flows at the Integrated Disease Surveillance Program (IDSP) in Delhi. This system uses non-specific data from various paper and computerized forms—sourced from patients, nurses, doctors, and paramedics—to predict the emergence of an epidemic and trigger rapid response in 'real-time'. To manage false positives, they conduct on-field investigations for data cleaning that usually takes one to two weeks and involves the work of 'waiting'—as a modality of action as well as patience for 'data to come to life'. Raghavan argues it is this work of waiting that makes possible the algorithmic promise of immediacy. Dreams of total standardization and immediacy have many co-constitutive discontents. These three ethnographies clearly establish how data analytics emerge in contexts that entangle it with older infrastructures, materialities, and information practices, weaving complex relations amongst data objects, technologies, users, and the social.

...

Overall, the essays in this volume offer arguably the most comprehensive and interdisciplinary view of big data and computational cultures in India. Big data in India, or any other place for that matter, is neither a grand global technology paradigm nor a local or national invention. Data, big or small, always produces and is produced by contexts, shadows, and relationalities. The wide-ranging and rapidly evolving problems and problematics of data-driven knowledge production and circulation covered here may conjure very different devices, headlines, and buzzwords in a few years, perhaps even in the next few weeks. Unless a complete digital autopoiesis is around the corner, sociotechnical relationalities of data will continue to shadow our situation in myriad, context-sensitive ways. To rethink the status quo, for a potentially decent, sustainable, and open-ended sociality, alterity, and generativity in computational cultures, we will need more creative ways to think about lives of data and the growing world of things and beings related to them. Meanwhile, relationalities abound.

References

Abraham, Itty, and Ashish Rajadhyaksha. 'State Power and Technological Citizenship in India: From the Postcolonial to the Digital Age', *East Asian Science, Technology and Society* 9.1 (2015): 65–85.

Appadurai, Arjun. 'Number in the Colonial Imagination', in Carol Breckenridge and Peter Van Der Veer (eds) *Orientalism and the Postcolonial Predicament: Perspectives on South Asia*, Philadelphia: University of Pennsylvania Press, 1993, pp. 314–339.

_____(ed.). *The Social Life of Things: Commodities in Cultural Perspective*, Cambridge: Cambridge University Press, 1986.

Bayly, Christopher A. *Empire and Information: Intelligence Gathering and Social Communication in India, 1780–1870*, Cambridge: Cambridge University Press, 1996.

Bowker, Geoffrey C., and Susan L. Star. *Sorting Things Out: Classification & Its Consequences*, Cambridge, Mass.: The MIT Press, 2000.

Cassidy, Ciaran, and Adrian Chen. *The Moderators*, Documentary, 2017. https://fieldofvision.org/the-moderators.

Chakravartty, Paula. 'Telecom, National Development and the Indian State: A Postcolonial Critique', *Media, Culture & Society* 26.2 (2004): 227–249.

Chattapadhyay, Sumandro. 'Opening Government Data through Mediation: Exploring the Roles, Practices and Strategies of Data Intermediary Organisations in India', 2015, http://ajantriks.github.io/oddc/.

Deo, Aditi, and Vebhuti Duggal. 'Radios, Ringtones, and Memory Cards or, How the Mobile Phone Became Our Favourite Music Playback Device', *South Asian Popular Culture* 15.1 (2017): 41–56.

Desrosières, Alain. *The Politics of Large Numbers: A History of Statistical Reasoning*, trans. Camille Naish, Cambridge, Mass.: Harvard University Press, 2002.

Dourish, Paul. 'NO SQL: The Shifting Materialities of Database Technologies', *Computational Culture* 4 (2014), http://computationalculture.net/article/no-sql-the-shifting-materialities-of-database-technology.

Fuller, Matthew (ed.). *Software Studies: A Lexicon*, Leonardo Books, Cambridge, Mass.: The MIT Press, 2008.

Gitelman, Lisa (ed.). *"Raw Data" Is an Oxymoron*, Cambridge, Mass. and London, England: The MIT Press, 2013.

Hacking, Ian. 'Biopower and the Avalanche of Printed Numbers', *Humanities in Society* 5.3–4 (1982): 279–295.

Halpern, Orit. *Beautiful Data: A History of Vision and Reason Since 1945*, Durham: Duke University Press, 2014.

Hui, Yuk. 'Cosmotechnics as Cosmopolitics', *e-Flux* 86 (November, 2017), https://www.e-flux.com/journal/86/161887/cosmotechnics-as-cosmopolitics/.

Information and Society Research Cluster Sarai-CSDS (ed.). *Sensor-Census-Censor: An International Colloquium on Information, Society, History, and Politics*, New Delhi: The Sarai Programme, Centre for the Study of Developing Societies, 2007.

Irani, Lilly. *Chasing Innovation: Making Entrepreneurial Citizens in Modern India*, Princeton: Princeton University Press, 2019.

Jeffrey, Robin, and Assa Doron. *Cell Phone Nation: How Mobile Phones Have Revolutionized Business, Politics and Ordinary Life in India*, Gurgaon: Hachette India Local, 2013.

Jones, Matthew L. 'Querying the Archive: Data Mining from Apriori to PageRank', in Lorraine J. Daston (ed.) *Science in the Archives: Pasts, Presents, Futures*, Chicago: University of Chicago Press, 2017, pp. 311–328.

Kalpagam, U. *Rule by Numbers: Governmentality in Colonial India*, Lexington Books, 2014.

Kelty, Christopher M. 'Preface: Crowds and Clouds', *Limn* 2 (April 2012), https://limn.it/articles/preface-crowds-and-clouds/.

Khera, Reetika (ed.). *Dissent on Aadhaar: Big Data Meets Big Brother*, Hyderabad, Telangana: Orient BlackSwan, 2018.

Kumar, Ravish. *Free Voice: On Democracy, Culture and the Nation*, revised edition, New Delhi: Speaking Tiger, 2019.

Mahalanobis, P. C. 'Statistics as a Key Technology', *The American Statistician* 19.2 (1965): 43–46.

Mayer-Schönberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, 2013.

Mazzarella, William. 'Beautiful Balloon: The Digital Divide and the Charisma of New Media in India', *American Ethnologist* 37.4 (2010): 783–804.

Menon, Nivedita, and Aditya Nigam. *Power and Contestation: India Since 1989*, London and New York: Zed Books, 2007.

Mertia, Sandeep. 'Socio-Technical Imaginaries of a Data-Driven City: Ethnographic Vignettes from Delhi', *The Fibreculture Journal* 29 (2017): Computing the City, http://twentynine.fibreculturejournal.org/fcj-217-socio-technical-imaginaries-of-a-data-driven-city-ethnographic-vignettes-from-delhi/.

_____. ''Timepass' Development: Situating Social Media in Rural Rajasthan', *Economic and Political Weekly* 52.47 (2017): 69–76.

Mukherjee, Rahul. ''City Inside the Oven': Cell Tower Radiation Controversies and Mediated Techno-science Publics', *Television & New Media* 18.1 (2017): 19–36.

Nandy, Ashis. 'Bearing Witness to the Future', *Futures* 28.6–7 (1996): 636–639.

Narrain, Siddharth. 'Dangerous Speech in Real Time: Social Media, Policing, and Communal Violence', *Economic and Political Weekly* 52.34 (2017).

Peabody, Norbert. 'Cents, Sense, Census: Human Inventories in Late Precolonial and Early Colonial India', *Comparative Studies in Society and History* 43.4 (2001): 819–850.

Phalkey, Jahnavi, and Sumandro Chattapadhyay. 'The Aakash Tablet and Technological Imaginaries of Mass Education in Contemporary India', *History and Technology* 31.4 (2015): 452–481.

Philip, Kavita, Lilly Irani, and Paul Dourish. 'Postcolonial Computing: A Tactical Survey', *Science, Technology, & Human Values* 37.1 (2012): 3–29.

Punathambekar, Aswin, and Sriram Mohan (eds). *Global Digital Cultures: Perspectives from South Asia*, University of Michigan Press, 2019.

Rajagopal, Arvind (ed.). *The Indian Public Sphere: Readings in Media History*, Oxford and New York: Oxford University Press, 2009.

Rangaswamy, Nimmi, and Payal Arora. 'The Mobile Internet in the Wild and Every Day: Digital Leisure in the Slums of Urban India', *International Journal of Cultural Studies* 19.6 (2016): 611–626.

Rossiter, Ned. 'Imperial Infrastructures and Asia beyond Asia: Data Centres, State Formation and the Territoriality of Logistical Media', *The Fibreculture Journal* 29 (2017): Computing the City, http://twentynine.fibreculturejournal.org/fcj-220-imperial-infrastructures-and-asia-beyond-asia-data-centres-state-formation-and-the-territoriality-of-logistical-media/.

Sharma, Dinesh C. *The Outsourcer: The Story of India's IT Revolution*, History of Computing. Cambridge, Mass.: The MIT Press, 2015.

Solanki, Aakash. 'Management of Performance and Performance of Management: Getting to Work on Time in the Indian Bureaucracy', *South Asia: Journal of South Asian Studies* 42.3 (2019): 588–605.

Sriraman, Tarangini. *In Pursuit of Proof: A History of Identification Documents in India*, New Delhi: Oxford University Press, 2018.

Subramanian, Sujatha. 'Is Hindutva Masculinity on Social Media Producing a Culture of Violence Against Women and Muslims?', *Economic and Political Weekly* 54.15 (April, 2019).

Sundaram, Ravi. 'Post-Postcolonial Sensory Infrastructure', *e-Flux* 64 (April, 2015). https://www.e-flux.com/journal/64/60858/post-postcolonial-sensory-infrastructure/.

_____ (ed.). *No Limits: Media Studies from India*, Oxford and New York: Oxford University Press, 2013.

_____. *Pirate Modernity: Delhi's Media Urbanism*, London and New York: Routledge, 2009.

Thakur, Arvind Kumar. 'New Media and the Dalit Counter-Public Sphere', *Television & New Media* 21.4 (2019): 360–375.

Tiwary, Ishita. 'Amazon Prime Video: A Platform Ecosphere', in Adrian Athique and Vibodh Parthasarathi (eds) *Platform Capitalism in India*, Palgrave Macmillan, 2020, pp. 87–106.

Upadhya, Carol. *Reengineering India: Work, Capital, and Class in an Offshore Economy*, Oxford and New York: Oxford University Press, 2016.

Vasudevan, Ravi S., Rosie Thomas, Neepa Majumdar, and Moinak Biswas. 'A Vision for Screen Studies in South Asia', *BioScope: South Asian Screen Studies* 1.1 (2010): 5–9.

Visvanathan, Shiv. 'Democracy, Governance and Science: Strange Case of the Missing Discipline', *Economic and Political Weekly* 36.39 (2001): 3684–3688.

01. DID MAHALANOBIS DREAM OF ANDROIDS?

SANDEEP MERTIA

> Statistics is not a branch of mathematics but is a technology which is essentially concerned with the contingent world of reality [...] Mathematics and probability theory are only the means to promote the use of statistical methods in the world of reality.
>
> - P. C. Mahalanobis, 1946[1]

In October 2016, after several months of searching for 'big data' in the Indian government and social sector, I landed in a big government office near the Parliament of India to meet the Director General of National Sample Survey Organisation (NSSO). I began the conversation by asking him about the evolution of data analytics in official statistics and the growing number of technology start-ups conducting field surveys for the government and social sector organizations through digital tablets and customizable Android apps. He agreed that private players in this space are increasing, while noting that 'they only cover small pockets here and there, we [the NSSO] are the only nation-wide survey with scientific methodology'. He added that the processes for NSSO surveys, sample design, data validation, tabulation, reporting, etc. have evolved over six decades and provide highly accurate estimates of social indicators, labor, poverty, etc. They were the first organization to get *computers* in India for large-scale data processing for economic planning in the 1950s, and on the data collection side, about ten years ago, they had 'experimented with *palmtop-based* surveys and conducted pilots. It didn't work out. One of the problems was that the questionnaire was too long to be conveniently filled on the palmtop'. Recently they accepted the World Bank's recommendation for using computer-assisted personal interview (CAPI) app for data collection through digital tablets.[2] In a recent review report, the Parliamentary Standing Committee on Finance has recommended the NSSO to use management information system (MIS) for streamlining 'statistics collection machinery'.[3] Clearly, collecting and managing data is an evolving, non-trivial problem even for an organization that has been at the forefront of official statistics for several decades.[4]

---

1 As quoted in Ashok Rudra, *Prasanta Chandra Mahalanobis: A Biography*, Delhi: Oxford University Press, 1996, p. 176 (emphasis mine).

2 'Computer Assisted Personal Interviews (CAPI)', The World Bank, 2016, https://dimewiki.worldbank.org/wiki/Computer-Assisted_Personal_Interviews_(CAPI).

3 'Review of National Statistical Survey Office (NSSO) and Central Statistics Office (CSO)', PRS Legislative Research, 30 January 2018, http://www.prsindia.org/report-summaries/review-national-statistical-survey-office-nsso-and-central-statistics-office-cso.

4 The National Sample Survey was established in 1950 in the Indian Statistical Institute with its fieldwork component under a separate entity called Directorate of the NSS. It was reorganized as a single organisation, NSSO, in 1970.

Beyond 'Computerization': Towards a Historical Anthropology of Computing in India

The story of computerization of the National Sample Survey (NSS), from the days of India's first electronic digital computers to contemporary Android tablets and MISs, much like the overall history of computing in India, is often clamped between mid-20th-century 'technology transfer' from the West and the millennial IT revolution.[5] In this essay, I will explore a different kind of history of computing in India and the Global South, by examining the epistemic and material culture of computing under the leadership of Prasanta Chandra Mahalanobis (1893–1972), a world-renowned statistician (though a physicist by training), who founded the Indian Statistical Institute (ISI) and the NSS.[6] Mahalanobis is widely credited to be one of the first visionaries to realize the value of electronic computers for large-scale data processing for national planning.[7] A lesser-known genealogy of his celebrated vision for computing lies in the extensive work he did for training 'human computers' from the early 1930s to the late 1950s. In fact, the first 'staff' at ISI in 1932 was a 'part-time computer'.[8] In addition to its professional training programs for human computers that began in 1938, ISI made electronic computer training a core part of its curriculum after importing India's first electronic computer in 1956. It is globally acknowledged that under Mahalanobis, India emerged as one of the leading countries in statistics research, particularly in sample survey techniques, and attracted scientific and political interest in these techniques from many other countries, including China.[9]

The use of computers in official statistics and planning in India during 1950–60s, when electronic computing in the West was expanding from census and military applications to scientific and industrial uses, and cybernetics and operations research, opens possibilities to rethink national and global histories of computing.[10] The computational imaginaries and practices at work in India's official statistics system did not just affect the processes of economic planning, they forged new relations between data-driven knowledge production, governance, and postcolonial nation-building.

---

5 Dinesh C. Sharma, *The Outsourcer: The Story of India's IT Revolution*, History of Computing. Cambridge, Mass.: The MIT Press, 2015.

6 Rudra, *Prasanta Chandra Mahalanobis*.

7 Homi J. Bhabha (1909–1966), an eminent nuclear physicist, had a different vision for computing in India than Mahalanobis. See, R. K. Shyamasundar and M. A. Pai, *Homi Bhabha and the Computer Revolution*, New Delhi: Oxford University Press, 2011. The competing pursuits for importing and building the first electronic computers in India by Mahalanobis and Bhabha—both of whom were close to Nehru—sheds light on the intertwining of postcolonial nation-building, science, and computing. See, Nikhil Menon, ''Fancy Calculating Machine': Computers and Planning in Independent India', *Modern Asian Studies* 52.2 (2018): 421.

8 'Indian Statistical Institute: Twenty-Fifth Annual Report: April 1956-March 1957', *Sankhyā: The Indian Journal of Statistics (1933-1960)* 20.1 (September, 1958): 109.

9 W. Edwards Deming, 'In Memoriam: P. C. Mahalanobis (1893–1972)', *The American Statistician* 26.4 (October, 1972): 49; Arunabh Ghosh, 'Accepting Difference, Seeking Common Ground: Sino-Indian Statistical Exchanges 1951–1959', *BJHS Themes* 1 (2016).

10 Greg Adamson, 'Norbert Wiener and Prasanta Chandra Mahalanobis', in *2012 IEEE Conference on Technology and Society in Asia (T&SA)*, 1–5, Singapore: IEEE, 2012.

Computers—human, (electro/)mechanical and electronic (both analog and digital)—have more than a century-long history of data processing.[11] In relative terms, the desire to efficiently compute large numbers, at scale, is much older than the contemporary techniques and devices that we identify with 'big data'. If the history of computing is replete with the problem of ever-increasing volume of data for storage and processing, what is specifically new about the contemporary data revolution? The conventional answer to this question is variety, velocity, and the value of data today. An immediate limitation that this response runs into is that all of these features are essentially relative, and their historical precedents in human computers, punch-cards, and navigational or relational database management systems have as much ontological validity as that of 'big data' (or non-relational database systems).

How do we then begin to develop a historically informed view of the novelty, promise, and perils of 'big data'? Did Mahalanobis simply dream of large number-crunching machines? Is the story of India's first electronic computers yet another paradigmatic example of pursuit of efficiency by the postcolonial developmental state? Mahalanobis's body of work suggests otherwise. There are wide-ranging historical, mathematical, and technological connections between the development of computing and Mahalanobis's long career as a statistician. In this short essay, I would like to propose a crucial step for developing an expansive view of those connections: to begin to decenter 'computers', that is, electronic stored-program computational devices, particularly the first-generation machines, in the history of computing in India. This might appear to be too simple. India did not have too many of those fancy machines to begin with, and the history of computing prior to the millennial IT revolution is a marginal area of research in India and the Global South at large. One might ask, why engage in beating the dead machines when we already have the full inventory of 'technology transfer', and when there are more urgent and critical beasts to tame (or run away from) in contemporary computational cultures. What is at stake in dwelling on the relationalities and entanglements of computers in mid-20th-century India, particularly with reference to official statistics? How are Mahalanobis's human and first-generation electronic computers relevant in our cutting-edge software saturated here and now? A historical anthropology of computing would critically engage with these questions, to develop a context-sensitive understanding of human-machine relations, meanings, and practices of computing—its constitutive cultural and political pasts, limits, and possibilities.

Computing Centered Humans, c.1930–50s

> [N]o one should be considered to have qualified as a statistician without having gone through an apprenticeship as a computer.
>
> - P. C. Mahalanobis, 1946[12]

11 Martin Campbell-Kelly et al., *Computer: A History of the Information Machine*, Third Edition, Boulder, CO: Westview Press, 2014.

12 P. C. Mahalanobis, 'Recent Experiments in Statistical Sampling in the Indian Statistical Institute', *Journal of the Royal Statistical Society* 109.4 (1946). Reprinted in *Sankhyā: The Indian Journal of Statistics (1933-1960)* 20.3/4 (December, 1958): 392 (emphasis mine).

The first NSS, October 1950–March 1951, arguably modern India's first data revolution event, which was also the world's largest statistical and computational exercise of its kind, happened much before the country got its first electronic digital computers. Much of the then existing system of fieldworkers, statistical staff, human computers, and a range of computational devices, such as desk calculators, punch-cards, tabulators, sorters, etc., was set up by Mahalanobis over two decades of research and training carried out at the ISI as well as the official survey work that he did for the British colonial government. While this system was substantively upgraded to handle the unprecedented amount of data from the NSS, it survived and thrived for the entire decade of 1950s without an electronic digital computer suitable for large-scale data processing.

Why was Mahalanobis so driven, against all geopolitical, economic, and technological odds, to import and indigenously develop electronic computers in India? He started making efforts to import them even before the formal beginning of the NSS. Historian Nikhil Menon has meticulously described Mahalanobis's quest for electronic computers, beginning from the late 1940s to mid-1960s, as driven by the 'urgent' need for efficient large-scale computation of NSS data for National Planning.[13] No doubt electronic computers offered much faster computation and data modelling solutions than the existing system, but perhaps there is something more to the story of India's first computers. It is important to remember that computers in that era were not 'plug and play' type machines, as the NSS Review Committee headed by Sir R. A. Fisher noted in 1956–57:

> The actual and potential work-load appears to us to be sufficiently large to justify the installation of a large computer. High speed input and output and reasonably fast computing speed will be required. We would emphasise, however, that the adoption of electronic methods of computation is a considerable undertaking, and two or three years are likely to elapse before a computer, when installed, can be put to full use. Considerable specialised skill and experience is required to programme computers effectively for complicated jobs, and the planning and construction of programmes and their subsequent testing takes, at best, a good deal of time. Considerable technical skill is also required to keep a computer in good running order. The ISI has recently acquired a small electronic computer (the British Hollerith HEC 2M), and is expecting early delivery of a Russian machine (the URAL). Neither of these machines is suitable for full scale work of the NSS type, but they can provide a useful opportunity for testing out methods and will provide useful experience in the 'programming' (i.e., writing instructions for the machine) required for this type of work. We strongly recommend that they should be used for this purpose to the maximum extent possible. They may also prove of permanent value for research studies.[14]

Indeed, the first-generation electronic computers, both imported and indigenous, were primarily used for experiments and training. They were inadequate for large-scale data process-

13 Menon, 'Fancy Calculating Machine'.

14 P. C. Mahalanobis, 'Indian Statistical Institute: National Sample Survey Review Committee Report', *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 26.3/4 (1964): 301.

ing requirements of the NSS. Thus, it is worth looking at the computational challenges for the NSS and economic planning not so much as a quest for 'computers', but rather in terms of 'computing', techno-scientific expertise, and operations experience. The latter could not simply be imported or manufactured but only constructed with long-term practice, including the pedagogical use of electronic computers.

Michael S. Mahoney has argued that the difference between the first-generation electronic computers that occupied full rooms and a contemporary laptop is not 'evolution but social construction, a lot of it. The difference is not the result so much of working principles as of pursuing the possibilities of practice'.[15] This emphasis on 'possibilities of practice' in computing is crucial for developing a grounded view of how different mathematical and material relationalities are imagined and pursued in a given organizational context. Further, Jon Agar, in a fascinating study of history of computing in science and government in the USA and UK, has shown that 'computerization, using electronic stored-program computers has only been attempted in settings where there *already existed* material and theoretical computational practices and technologies'.[16] The computational practices at ISI too were already well evolved in terms of technological practices by the time of the first NSS. Consider the following note by Mahalanobis, in the NSS general report no. 1:

> 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. As 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 at 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 the latter part of 1951 the Institute had 2 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 some considerable time. An Electronic Statistical Machine (a high powered combined sorter-tabulator) was also rented from the IBM. This expansion in staff and machines called for a large increase in office and storage space and a new office building with a floor space of about 20,000 sq. feet was constructed by the Institute in 1951 mainly for the work of the National Sample Survey.[17]

The most intriguing part of the above assemblage of human 'punchers' and punch-card tabulating machines, in the context of the NSS, is the availability of such a large number of computing and statistical staff and training facilities. The maximum number of human computers working on Mahalanobis's jute sample surveys in Bengal in 1941 was ninety.[18] Anyone familiar

15 Michael S. Mahoney, 'The Histories of Computing(s)', *Interdisciplinary Science Reviews* 30.2 (2005): 131.

16 Jon Agar, 'What Difference Did Computers Make?', *Social Studies of Science* 36.6 (2006): 872.

17 P. C. Mahalanobis, 'The National Sample Survey: General Report No. 1. First Round: October 1950-March 1951', *Sankhyā: The Indian Journal of Statistics (1933-1960)* 13.1/2 (1953): 59.

18 P. C. Mahalanobis, 'On Large-Scale Sample Surveys', *Philosophical Transactions of the Royal Society of*

with literacy rates and the accuracy levels of official statistics in late colonial India would appreciate the institutional efforts that would have gone into assembling this community of practice over two decades, long before the invention of electronic stored-program computers.

In 1943, after conducting several large-scale sample surveys for estimating rice and jute production in Bengal—struck by the colonial state's genocidal famine—Mahalanobis (in line with the government enquiry committee on the famine) diagnosed the problem as a complete lack of accuracy and reliability in official statistics and articulated a detailed vision for statistics in the post-war period. He noted that for accurate data collection, 'it is essential to build up an efficient human organisation with carefully selected and trained staff. This takes time. And unless such time is allowed the results are often not only useless, but even harmful'.[19] In his view, the 'need of planning', at different scales, was seen as both a condition of possibility and an applied use of good-quality data. In his extended reflections on large-scale sample surveys, published by the Royal Society of London, he noted that, '[I]n 1937 there was not a single trained field worker, and only about half a dozen computers'.[20] While discussing the challenges in training the staff, he emphasized the problem of the seasonal availability of fieldworkers: 'a large number, especially the abler men, left after one season and did not come back, so that work had to be carried on with a large proportion of untrained men each year'.[21] Since it was possible to employ human computers on other projects, and not just surveys, he was able to build a somewhat stable community of practice of human computers.[22]

It is important to stress that computational practices, involving human and/or electronic computers, even with their working principles grounded in discrete mathematics and measurable outcomes, are never limited to the actual moments of calculation or data processing. Rather, they are co-constituted by imaginaries of what kinds of knowledge and labor are possible and desirable in relation with different techniques and machines for computation.[23] Mahalanobis's whole survey organization, even the mathematical methods for preparing sample units, optimization of (human) 'computer-hours', continuous tabulation and analysis of data, monitoring of error rates, and all other related steps were designed with an epistemic and material focus on scale and standardization of computational work in conditions of limited resources and staff, that too in India's large and linguistically and socio-culturally diverse geography. Even the work done by human computers was broken down into smaller tasks such as 'copying three-figure tables, adding four-figure quantities, squaring three-figure entries, [and] preparing frequency tables with not more than ten classes'.[24] Each step had an associated standard

London. Series B, Biological Sciences 231.584 (31 October, 1944): 329.

19 P. C. Mahalanobis, 'Organisation of Statistics in the Post-War Period', *Proceedings of the National Institute of Sciences of India* 10.1 (March, 1944): 69.

20 Mahalanobis, 'On Large-Scale Sample Surveys', p. 409.

21 Ibid.

22 This was, presumably, an entirely male community of practice. I have not come across any mention of women computers in Mahalanobis's writings. In contrast, women computers played a formative role in the development of computing in the West. See, Jennifer S. Light, 'When Computers Were Women', *Technology and Culture* 40.3 (1999): 455.

23 Matthew L. Jones, 'Calculating Devices and Computers', in Bernard Lightman (ed.) *A Companion to the History of Science*, Hoboken, NJ: John Wiley & Sons, 2016, pp. 472–487.

24 Mahalanobis, 'Recent Experiments in Statistical Sampling in the Indian Statistical Institute', p. 337.

rate of output as well as a rate of mistakes. The output of each of the human computers was punched on Hollerith cards and tabulated at the end of every month. Clearly, Mahalanobis's vision of 'statistics as key technology' for India's development and planning was far from a one-way application of statistical methods to understand social realities.[25]

Conclusion: Computing After Mahalanobis

The data revolution under Mahalanobis's leadership was a long-term endeavor that found a particular productive materialization in the NSS. The above examples are but a few samples of the computational imaginaries and practices in mid-20th-century India. A robust historical anthropology of computing would have to reckon with a wide range of questions concerning sociotechnical imaginaries of statistics in and beyond Mahalanobis's work, epistemic and material virtues of sample-surveys and data modelling for national planning, constitution of governmentality and possible subjectivities of/for the surveyed 'social', and the genealogies of our seemingly untamable computational present. 'Big data' or large-scale data collection and processing have transformed in copious ways since Mahalanobis's time, and the Planning Commission he helped set up has been replaced by a subtly named think tank, National Institute for Transforming India. Fortunately, technologies and epistemologies of computing do not tend to follow teleological transformations. After all, in as early as 1979, C. R. Rao, former Director of the ISI and a prodigy of Sir R. A. Fisher and P. C. Mahalanobis, noted in his presidential address to the International Statistical Institute that the 'enormous speed [of computers] appears to be both a boon and a hindrance to statistical research'.[26] For all their well-known benefits, computers had also 'encouraged uncritical use of statistical methods through the commercially available computer package programs [. . .] It is thought that what is lacking in sophistication of methodology can be made up by acquiring more data and processing by computers using less efficient procedures'.[27] Dr. Rao's forty-year old diagnosis of computational hubris of 'big data' should ring a few 'cutting-edge' bells. Still, history and past visions of computing do not repeat themselves in toto. Perhaps if we decenter computers in the histories of computing from Mahalanobis's human computers to NSSO's on-going adoption of Android tablets for conducting surveys and look at computational practices situated in specific contexts and imaginations of data and society, we can better survey the transformative potentials and actualities of computing and how they cohere different regimes of knowing and governing the social.

25 History of statistics in colonial India is dominated by the debate on caste classification and enumeration in the Census and its effects. See, Arjun Appadurai, 'Number in the Colonial Imagination', in *Modernity at Large: Cultural Dimensions of Globalization*, Minneapolis: University of Minnesota Press, 1996, pp. 114–135. Mahalanobis must have encountered some aspects of this politics of numbers after his return from Cambridge in 1915. However, I could not find any descriptions of caste in Mahalanobis's writings, except for numerical and anthropometric ones.

26 C. R. Rao, 'Perspectives in Statistics', *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 41.3 (1979): 136.

27 Ibid.

References

Adamson, Greg. 'Norbert Wiener and Prasanta Chandra Mahalanobis', in *2012 IEEE Conference on Technology and Society in Asia (T&SA)*, 1–5, Singapore: IEEE, 2012.

Agar, Jon. 'What Difference Did Computers Make?', *Social Studies of Science* 36.6 (2006): 869–907.

Appadurai, Arjun. 'Number in the Colonial Imagination', in *Modernity at Large: Cultural Dimensions of Globalization*, Minneapolis: University of Minnesota Press, 1996, pp. 114–135.

Campbell-Kelly, Martin, William Aspray, Nathan Ensmenger, and Jeffrey R. Yost. *Computer: A History of the Information Machine*, Third Edition, Boulder, CO: Westview Press, 2014.

Deming, W. Edwards. 'In Memoriam: P. C. Mahalanobis (1893–1972)', *The American Statistician* 26.4 (October, 1972): 49–50.

Ghosh, Arunabh. 'Accepting Difference, Seeking Common Ground: Sino-Indian Statistical Exchanges 1951–1959', *BJHS Themes* 1 (2016): 61–82.

'Indian Statistical Institute: Twenty-Fifth Annual Report: April 1956-March 1957'. *Sankhyā: The Indian Journal of Statistics (1933-1960)* 20.1 (September, 1958): 109.

Jones, Matthew L. 'Calculating Devices and Computers', in Bernard Lightman (ed.) *A Companion to the History of Science*, Hoboken, NJ: John Wiley & Sons, 2016, pp. 472–87.

Light, Jennifer S. 'When Computers Were Women', *Technology and Culture* 40.3 (1999): 455–483.

Mahalanobis, P. C. 'Indian Statistical Institute: National Sample Survey Review Committee Report', *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 26.3/4 (1964): 301–336.

_____. 'Recent Experiments in Statistical Sampling in the Indian Statistical Institute', *Journal of the Royal Statistical Society* 109.4 (1946). Reprinted in *Sankhyā: The Indian Journal of Statistics (1933-1960)* 20.3/4 (December, 1958): 329–398.

_____. 'The National Sample Survey: General Report No. 1. First Round: October 1950-March 1951', *Sankhyā: The Indian Journal of Statistics (1933-1960)* 13.1/2 (1953): 47–214.

_____. 'On Large-Scale Sample Surveys', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 231.584 (31 October 1944): 329–451.

_____. 'Organisation of Statistics in the Post-War Period', *Proceedings of the National Institute of Sciences of India* 10.1 (March, 1944): 69–78.

Mahoney, Michael S. 2005. 'The Histories of Computing(s)', *Interdisciplinary Science Reviews* 30.2: 119–135.

Menon, Nikhil. ''Fancy Calculating Machine': Computers and Planning in Independent India', *Modern Asian Studies* 52.2 (2018): 421–457.

Rao, C. R. 'Perspectives in Statistics', *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)* 41.3 (1979): 129–137.

Rudra, Ashok. *Prasanta Chandra Mahalanobis: A Biography*, Delhi: Oxford University Press, 1996.

Sharma, Dinesh C. *The Outsourcer: The Story of India's IT Revolution*, History of Computing. Cambridge, Mass.: The MIT Press, 2015.

Shyamasundar, R. K., and M.A. Pai. *Homi Bhabha and the Computer Revolution*, New Delhi: Oxford University Press, 2011.

02. PROGRAMMING THE INTERMISSION: 'BIG DATA', SOFTWARE, AND INDIAN CINEMA

KARL MENDONCA

Introduction

The historical shift from *human* to *electronic* computers to perform calculations on large data sets has been vividly traced by D. A. Grier in *When Computers Were Human*.[1] Highlighting the central but neglected role that women have played in computation, Grier's narrative is epic, spanning several centuries and contexts. Given the breadth of material covered, one cannot fault Grier for maintaining a sharp geographic focus on Europe and the US. But in doing so, the project reinforces an unconscious anchoring of the history of computation and associated labor in the West. Contemporary scholarship on 'big data' takes this attribution for granted and builds on the premise, even as it actively engages with other sites. On the subject of 'big data' especially, the ontology of the digital is closely scrutinized in literature, but there is a marked reticence to tackle the 'spatialization of time' on a more fundamental level.[2] But what is to be gained from an understanding of vernacular histories of computation and data? And how might we approach the production of such histories from both a methodological and epistemological standpoint? This paper attempts to respond to this provocation via a case study of the Blaze Advertising distribution network and a long history of 'big data' and the intermission in Indian cinema. Structurally, the paper is divided into four sections—in the first, I introduce the concept of the intermission in Indian cinema and the central role that Blaze Advertising played as a distributor; in the second, I briefly review a robust taxonomy of 'big data' and outline a framework to discuss the role of computation and software; in the third, I trace a material history of computation specific to the Blaze network; and finally, I conclude with insights from the case study.

A Brief History of Blaze Advertising

While the '*samosa* break' (as the intermission is referred to in the Bombay vernacular) has been phased out from cinemas in most parts of the world, it is an entrenched cinematic event for Indian audiences. For those unfamiliar with the concept, the mechanics are quite simple: about halfway through a film the house lights turn on and interstitial advertising is displayed on the screen for 10–15 minutes while patrons stretch their legs or visit the concession stand. What is perhaps less known is that *Blaze Advertising*, a 70-year old distribution agency set up in the 1950s, had an almost complete monopoly on delivering interstitial advertising to theaters across India for close to four decades. The company had its beginnings in Bombay (present-day Mumbai), when journalist Mohan Bijlani and his business partner Freni Variava

1 David Alan Grier, *When Computers Were Human*, Princeton, NJ: Princeton University Press, 2007.
2 Kavita Philip, 'Why Software? A Keynote Conversation', Computer History Museum: N.p., 2017.

founded *Blaze*, a print-based magazine that covered news and events related to the Indian cinema. Bijlani and Variava entered the business of distributing cinema advertising by accident, when a client gave the duo control of the intermission for a few theatres that he owned. Working out of a small office in Worli, Bombay, Variava and Bijlani systematically purchased the advertising rights for other theatres in the region and gradually across the country. By the mid-60s they had established a monopoly and were a centralized booking agency mediating between advertising agencies and cinema owners across India. The network was at its prime in the 70s, with several hundred employees across four national offices, subdivided into state and regional districts, based on taxation policies, language, and administrative efficacy.[3]

The organizational makeup of the company comprised three core functions—Client Management & Accounting (typically handled by senior executives who liaised with advertising agencies); Scheduling, which involved managing the distribution and exhibition of advertising programming across cinemas; and Operations, the on-the-ground network of warehouses and 'runners' that delivered the ad films to cinemas. The high cost of film prints made it unfeasible to strike a 35mm print for each ad and supply the growing number of cinemas in India at a 1:1 ratio. Instead, advertising agencies would prioritize specific cinemas for a *first run* and then circulate advertisement reels across other cinemas in the region. It was up to the Scheduling department to plan out the ad playlist for each cinema and manage a calendar to ensure that the reels were updated and ads re-circulated on a weekly basis. This involved negotiating the competing demands of orders from multiple offices for national brands with ad placements made by local businesses. As one might imagine, the process of scheduling was labor intensive and error prone, involving a vast body of junior clerks, assistants, schedule checkers, and typists. In 1982, Lalit Bijlani, the son of Mohan Bijlani, who had taken over operations of Blaze Advertising after his father's passing, decided to 'computerize' the scheduling and planning of the intermission ads. The process of reviewing work orders, manually updating ledgers, shuffling schedules, and typing out the final instructions was transformed into feeding data and COBOL-based instructions into an IBM 7044 and IBM 1401 on punch cards and waiting for a printout to appear on an 1403 Line Printer. Interestingly, the computational 'programming' of the intermission occurred at a time when projection technology was entirely analog, that is, the advertising showreel was projected on 35mm film projectors and glass plate slide projectors.

'Big Data' and the Sign of the Empty Archive

> In 1984 alone, Blaze moved approximately *9 million film prints* and slides to and from each of *11,000 cinema halls*, in over *3,000 cities and towns*. Controlling and coordinating this network…is Blaze's key strength. And the foundation on which Blaze have planned all further diversification.
>
> — Blaze Advertising Marketing Brochure

3 In-person interview with Ramdas Mundacheery, former Regional Manager at Blaze Advertising, March 2016.

Although the term 'big data' was coined fairly recently in response to extremely large, predominantly digital data sets, it serves as an apt descriptor of the immense amounts of data generated by the Blaze Advertising network. A fundamental theoretical challenge when working with a concept like 'big data' is to produce a definition that is capacious enough to accommodate the heterogeneous composition of varying data sets while also providing an optic for analysis. Attributing the etymological origins of the term 'big data' to the computer scientist John Mashey in the 1990s, Kitchin and McArdle provide a useful overview of several taxonomies that articulates the key traits of big data as 'volume, velocity and variety' and 'exhaustivity, resolution, indexicality, relationality, extensionality and scalability'.[4] Whether or not the Blaze data qualifies as 'big data' is perhaps beside the point as even in Kitchin and McArdle's analysis of twenty-six contemporary data sets, only a few check all the boxes. With this framework in place, a logical next step for this paper would be to conduct a close examination of how the Blaze data can be plotted along each axis of the framework. This exercise would no doubt yield insights about the ontology of the data produced within the Blaze network. Given the vast quantities of data hinted at in the Blaze brochure, the endeavor would also depend on the availability of the original data in ledger form. However, all the paper-based records, including manifests, logbooks, exhibition certificates, and receipts used to track and manage distribution, were destroyed by heavy flooding that completely submerged the company warehouse.

This is where the story of *Blaze Advertising* takes yet another peculiar turn. The 1980s saw a decline in the overall popularity of cinema in India, due in part to the aggressive growth of television programming and a boom in the number of households with television sets. The anxiety of this downturn was compounded by a decade-long case against Blaze brought by the Government of India to break up the distributive monopoly using the Monopolies and Restrictive Trade Practices Act, 1969. In 1986, fast dwindling profits compounded by legal pressure forced the company to pivot—Blaze repurposed the network into a domestic courier company (similar to FedEx) with a franchise-based business model that exponentially increased their presence across India.[5] To support its new function as a courier company, Blaze developed a website to allow its customers to track the status and progress of deliveries. It is not without a sense of irony that the Blazeflash web database became the last remaining trace of the cinema distribution network. And yet again, unfortunately, all the logs for the database were lost when the company was shuttered in 2012.

But perhaps the absence of data is not a bad thing. The negative space is an opportunity to shift our efforts away from categorization to interrogating the relationship between software and data. To undertake on this task, we must first answer a fundamental question: What is software? As media theorist Matthew Fuller succinctly articulates: 'while much has been said about the use of digital media, the material of software has often been left invisible'.[6]

4 Rob Kitchin and Gavin McArdle, 'The Diverse Nature of Big Data', Social Science Research Network (September 2015), *Big Data & Society* (June, 2016), p. 1, https://ssrn.com/abstract=2662462.

5 In-person interview with Lalit Bijlani, former owner of Blaze Advertising and Blazeflash Couriers, March 2016.

6 Matthew Fuller (ed.) *Software Studies: A Lexicon*, Cambridge, Mass: The MIT Press, 2008, p. 6.

Defining the 'Object' of Software

How might we engage with the 'materiality' of software in a manner that does not turn code and computation into purely linguistic categories or lapse into technological essentialism? Despite the relatively recent development of Software Studies as a field formation, there are several approaches that address this question, ranging from the *genealogical* to the *formalist* to the *literary* to *logic and hardware* and even *rule systems*, each with its own strengths and limitations.[7]

But it is media theorist Wendy Chun's conceptualization of software that best serves the focus of this project. For Chun, software is a 'notoriously difficult concept' that must be understood not as a 'given' social and technical object, but as a discursive concept that is both material and ideological.[8] Chun's line of inquiry moves between a material analysis of software and hardware (snippets of code, vacuum tubes, logic diagrams) and historical sites of computation, focusing primarily on the period after World War II, where software, through programming, emerges from a gendered system of 'command and control'. What results from this method is a series of contradictions— '[a]s our machines disappear, getting flatter and flatter, the density and opacity of their computation increases'.[9] Despite, or rather because of, this opacity, software perpetuates certain notions of 'seeing as knowing', by mimicking both 'ideology and ideology critique [...] conflating executable with execution, program with process, order with action'.[10] For Chun, the comprehension of software's 'materiality' is not only a matter of unearthing a computational trace in hardware or demonstrating how digital processes have an agency that act independently of the human, it is rather a question of understanding how the 'immateriality' of software is part of its operational logic as a discursive sign, work that is 'glossed over if we just accept the digital as operating through 1s and 0s'.[11] While one might quibble with some of Chun's technical arguments (the boundary between hardware and software is not as arbitrary as she makes it out to be), the most compelling and useful aspect of her project is the recursive dialog between materiality and metaphor. Software is both a *thing* and an *ideological construct* that must be constituted within a historical context. Organized as a series of jump cuts, the final section of this paper builds on Chun's ideas via the interconnected material histories of software, hardware, and labor that collectively constitute the history of computation at Blaze.

---

7 Lev Manovich, *The Language of New Media*, revised edition, Cambridge, Mass.: The MIT Press, 2002; Friedrich A. Kittler, 'There Is No Software', in John Johnston (ed.) *Literature, Media, Information Systems: Essays*, Amsterdam: Overseas Publishers Association, 1997, pp. 147–155; Katherine N. Hayles, *My Mother Was a Computer: Digital Subjects and Literary Texts*, University of Chicago Press, 2010; Charles Petzold, *Code: The Hidden Language of Computer Hardware and Software*, 1st edition, Redmond, Wash.: Microsoft Press, 2000; Stephen Wolfram, *A New Kind of Science*, 1st edition, Champaign, Ill: Wolfram Media, 2002.

8 Wendy Hui Kyong Chun, *Programmed Visions: Software and Memory*, reprint edition, Cambridge, Mass.: The MIT Press, 2013.

9 Ibid., p. 2.

10 Ibid., p. 3.

11 Ibid., p. 139.

Programming the Intermission

The Invention of COBOL

Designed in 1959 as part of a U.S. Department of Defense initiative, common business oriented language (COBOL) has the distinction of being the most despised programming language in academic circles, while simultaneously thriving as one of the most popular languages to be used by businesses globally.[12] Unpacking this paradox, Ben Allen points out that COBOL was in fact preceded by FLOW-MATIC, a business-like data language, developed in 1955 by Grace Hopper and her team for use on UNIVAC (an acronym for a line of early digital computers). However, as one of the first, high-level programming languages, COBOL allowed programmers to call a list of over 300 reserved words in plain English, making the form of programing 'legible'. Allen charts an institutional history of the development of COBOL and the many decisions that informed the ultimate architecture of the language. He argues that although COBOL's syntax did not 'make programs written in it significantly easier to write or read, COBOL's resemblance to English-language business writing made *programmers themselves* more legible to the management figures responsible for purchasing machines and hiring programmers, and thus made programmers and also their machines seem more potentially trustworthy to these particular influential figures'.[13] COBOL was adopted and supported by IBM as a programming language that could be used for data processing on its early computers.

IBM at IIT, Kanpur

The tenuous history of IBM in India has been vividly charted by Dinesh C. Sharma, who provides a telling account in *The Outsourcer*, 2015, of the paradoxical role that the company played in India. In the early 1960s, IBM set up manufacturing plants in Bombay that assembled or 'reconditioned' old and discarded 1401 line computers from advanced markets. The company's business strategy involved leasing computers (rather than selling them outright) and charging maintenance fees (charged in US dollars, but paid in INR). Despite the exorbitant fees, they made huge profits from the circulation of computers that were unwanted and close to worthless in other parts of the world, while establishing a near monopoly of 80% of the Indian market. As the adoption of computers began to catch across various industries, the idea of computation was met with stiff opposition from labor unions led by George Fernandes. To counter this resistance, IBM launched a PR department that organized seminars, training, and outreach on the benefits of computers with the unions and in the popular media.[14] In this sense, it played a curious, paradoxical role—on the one hand, it profiteered from questionable business practices, while on the other, it was largely responsible for popularizing the idea of computation and conducting widespread training in India. By the time it was audited by the government in 1971, IBM had sold hundreds of the 1400 series computers and the 1620s

12 Ben Allen, 'Common Language: COBOL and the Legibility of Programming', Stanford University: N.p., 2016.

13 Ibid., p. 6.

14 Dinesh C. Sharma, *The Outsourcer: The Story of India's IT Revolution*, Cambridge, Mass.: The MIT Press, 2015.

to large government and private institutions in India, including IIT Kanpur. As Mehrotra and Shah point out, IIT Kanpur was the only university in India deemed fit for collaboration by an academic team of researchers from the Massachusetts Institute of Technology.[15] The computer science department, set up in 1965, was housed in the Department of Electrical Engineering that was headed by the V. Ramarajan. The two IBM 1401s and one 7044 were put to good use, thanks to the policy instated by Ramarajan that allowed other departments to also write and run programs for these machines. By the early 70s, these IBM machines were struggling to keep up with the workload and the complexity of calculations and were put up for an open auction. Indian Data Processes (IDP), the company hired by Blaze Advertising to computerize their network, purchased one of the 1401s and the 7044 for 15 lakh.

RANDOM( )

K. P. Kalyanam, one of the founding members of IDP, was a statistician trained by IBM as part of their larger effort to popularize computation in India. Along with K. S. Muthukrishnan, he developed a homegrown compiler that automated several subroutines and functions on the IBM 7044. When IDP was approached by *Blaze Advertising* to 'computerize' their network, the duo spent a month studying the existing process to understand the core functions they needed to support. The database structure was designed to keep track of three key units of information: a cinema code comprising 9 digits (the first two digits which were allocated to the state, the next three digits to the town, and the final four digits represented the cinema), an agency code (similar to the cinema code), and the product code (a four-digit code) to represent the actual advertising films and slides. Beyond the functions that one might expect (comparing cinema schedules, billing, etc.) the programmers were requested to create a unique 'randomizer' subroutine to compensate for overbooking. The demand to place advertising in theatres that served high-density populations in metros was extremely high. Rather than turn down advertisers, executives would 'overbook' the intermission slot. However, cinema owners did not want advertising to run longer than the allocated intermission time slot. This made it unfeasible to send an overbooked cinema all of the advertising, as this would incur the ire of the cinema owners and increase the odds of damaging the films themselves. The random function gave the management at Blaze Advertising a layer of opacity in the decision-making process, while simultaneously providing a sense of objectivity and efficiency. The scheduling data generated as part of the process of computerization, was 'always cooked', so to speak.[16]

Conclusion

What are some preliminary insights that can be gleaned from this curious narrative? Most obviously that data, software, and algorithms are socio-technical and material practices with ideological functions intimately tied to institutional practices. However, the implications of this insight increase the scope of work for the researcher and theorist. There is no doubt that much is to be gained from a close analysis of data in terms of classification and formats. But it is

15 S. P. Mehrotra and P. P. Shah, *The Fourth IIT: The Saga of IIT Kanpur*, Gurgaon, Haryana, India: Penguin Enterprise, 2015.

16 Lisa Gitelman (ed.) *"Raw Data" Is an Oxymoron*, Cambridge, Mass.: The MIT Press, 2013.

equally important to follow the action outside of the frame into a deeper understanding of the collection, codification, management, and interpretation of data sets. Such an effort requires an interdisciplinary approach and a broad methodological toolkit but will yield unexpected configurations and insights. As in the case of Blaze Advertising, a long history of computation and data reveals otherwise hidden entanglements between power, capital, and labor.

References

Allen, Ben. 'Common Language: COBOL and the Legibility of Programming', Stanford University: N.p., 2016.

Chun, Wendy Hui Kyong. *Programmed Visions: Software and Memory*, reprint edition, Cambridge, Mass.: The MIT Press, 2013.

Fuller, Matthew (ed.). *Software Studies: A Lexicon*, Cambridge, Mass: The MIT Press, 2008.

Gitelman, Lisa (ed.). *"Raw Data" Is an Oxymoron*, Cambridge, Mass: The MIT Press, 2013.

Grier, David Alan. *When Computers Were Human*, Princeton, NJ: Princeton University Press, 2007.

Hayles, N. Katherine. *My Mother Was a Computer: Digital Subjects and Literary Texts*, University of Chicago Press, 2010.

Kitchin, Rob, and Gavin McArdle. 'The Diverse Nature of Big Data', Social Science Research Network (September 2015), *Big Data & Society* (June, 2016), Available at SSRN: https://ssrn.com/abstract=2662462.

Kittler, Friedrich A. 'There Is No Software', in John Johnston (ed.) *Literature, Media, Information Systems*: *Essays*, Amsterdam: Overseas Publishers Association, 1997, pp. 147–155.

Manovich, Lev. *The Language of New Media*, revised edition, Cambridge, Mass.: The MIT Press, 2002.

Mehrotra, S. P., and P. P. Shah. *The Fourth IIT: The Saga of IIT Kanpur*, Gurgaon, Haryana, India: Penguin Enterprise, 2015.

Petzold, Charles. *Code: The Hidden Language of Computer Hardware and Software*, 1st edition, Redmond, Wash.: Microsoft Press, 2000.

Philip, Kavita. 'Why Software? A Keynote Conversation', Computer History Museum: N.p., 2017.

Sharma, Dinesh C. *The Outsourcer: The Story of India's IT Revolution*, Cambridge, Mass.: The MIT Press, 2015.

Wolfram, Stephen. *A New Kind of Science*, 1st edition, Champaign, Ill: Wolfram Media, 2002.

03. NUMBER, PROBABILITY, AND COMMUNITY: THE DUCKWORTH-LEWIS-STERN DATA MODEL AND COUNTERFACTUAL FUTURES IN CRICKET

SIVAKUMAR ARUMUGAM

The Duckworth-Lewis-Stern (DLS) model is an example of a data model that is drawn from and actively intervenes in a part of society, in this case rain-interrupted games of cricket. In this paper, I examine how the DLS model was itself put together and promoted, and some of the main issues that a consideration of the DLS model throws up. I suggest that such data models operate through a kind of data-based conduct of conduct—a kind of 'data governmentality'. The emphasis is on data as logic, intuition, and community. I will write about Brian Rotman on the compulsions of number, and Ian Hacking and C. S. Peirce on the relationship between feeling and probability. My argument, in brief, is that the DLS model has helped formulate what is acceptable in a data model to mass audiences for cricket around the world. I suggest that an important way to think about such data models is to examine how they work as semiotic activities, that is, not just attending to their inputs and outputs but also to how their inner workings help formulate new communities organized around thinking counterfactually and probabilistically about the future.

The DLS model is an algorithm optimized against a dataset of all recent one-day cricket games. The model is driven by relatively 'small data'. It uses only histories of the traditional scorekeeping of cricket games—balls bowled, runs scored, and so on—to formulate predictions of what might have happened in a game if rain had not prevented some play during the match. The model may nevertheless be a useful crucible in which to track developing ideas about data and algorithms in contemporary society. Unlike very recent developments with big data and algorithms, the DLS model has been in use in cricket for some 20 years. It has become a successful part of the infrastructure of thinking about cricket and is now rarely brought into question. This paper examines how and why it came to be accepted around the cricket playing and watching world.

The DLS model may seem complex, both as a mathematical formulation and in its application. Frank Duckworth, a statistician, and Tony Lewis, a mathematician, first met in January 1995. They had already been collaborating with each other at a distance since 1993. Duckworth had previously presented a short paper on a new rain-interruption rule at a conference. He had worked out an initial formulation and a computer program implementing it. Lewis, with the help of a student, had worked out some details using a small amount of data on cricket matches in England. By August 1995, they were meeting regularly to refine the model, pending a presentation to the Test and County Cricket Board.[1] Duckworth and Lewis set out

1 The board that has a de facto monopoly on organized cricket in England and is responsible for the national team. It is now the England and Wales Cricket Board. See, Frank Duckworth and Tony Lewis, *Duckworth Lewis: The Method and the Men Behind It*, Cheltenham: SportsBooks, 2011, pp. 31–34.

this story in their book, along with a first-hand account of the process by which their model came to transform how cricket is played. They suggest that Duckworth had in fact begun with the wrong kind of question, 'how many runs on average should one have made after *y* overs with *w* wickets down?' rather than 'how many runs can be made, on average, with *u* overs *remaining* and *w* wickets down?'[2] The former is prescriptive—what ought to have happened so far—and the latter is predictive—what will happen now. Duckworth, in other words, had shifted the task at hand from one of evaluating how well a team had done so far to evaluating its future. This, in fact, corresponds to the shift from theories of value in classical and 19th-century marginalist economics that predicated value on something—labor, corn, or some other substance—to a theory of value that rested on the future usefulness of a good.[3] The model also attempts to be fair. A rain-rule must not include the kinds of variables that someone gambling on the game, for example, will likely use. It would not do for a rain-rule to take into account that one team in a game has historically played much better than the other and is therefore likely to win, no matter what effect rain has had in curtailing the length of the game. Such a rule would award a rain-interrupted game to that team but not to another weaker team in exactly the same situation. With regard to individual and team differences in talent, the intuition is that, as they say in sports, 'anything can happen on the day'. There is a balance between fairness and prediction here. The more accurate you want the model to be, the more unfair it may become.

Duckworth and Lewis went through various possible formulas, all using a natural exponential function, to link together the idea of resources left to the batting team, from the number of overs and wickets left. The successive changes in the formula are interesting because they were entirely ungrounded in any explicit empirical considerations. They simply assume that a natural exponential function correctly describes the arc of the average, or rather the ideal, cricket game. The crucial innovation that they highlight is Duckworth's initial intuition that previous rain rules were deficient because they did not take into account both overs *and* wickets. Subsequent proposed models, different as they may be, all retain that core insight.[4]

The model is opaque in two different ways. First, it is a commercial and proprietary model. No one else but the regulators of the game has access to the Professional version of the model that calculates the changed target for the team to win a game. In application, the model is available only as a computer program—it is too complex a model to execute manually. The modelers have also kept secret, even from the regulators, how they used the dataset of previous games to calibrate their model. Second, the mathematics behind the model has been published but is largely incomprehensible to cricket players, coaches, and fans of the game. Certainly, there have been no journalistic attempts to explain how the model

2 Ibid., p. 31.

3 Philip Mirowski, *More Heat than Light: Economics as Social Physics, Physics as Nature's Economics*, Cambridge: Cambridge University Press, 1989; Philip Mirowski, 'Postmodernism and the Social Theory of Value', *Journal of Post Keynesian Economics* 13.4 (1991): 565.

4 Michael Carter and Graeme Guthrie, 'Cricket Interruptus: Fairness and Incentive in Limited Overs Cricket Matches', *The Journal of the Operational Research Society* 55.8 (August, 2004): 822; R. Bhattacharya, P. S. Gill, and T. B. Swartz, 'Duckworth–Lewis and Twenty20 Cricket', *The Journal of the Operational Research Society* 62.11 (November, 2011): 1951.

is constructed, nor have there been any other kinds of public discussions on the internal workings of the model.

In their first-hand account of the building of the model, Duckworth and Lewis note that in the first year, they 'had no problems with scorers, umpires, match managers or even the players', but that '[m]ost of the adverse comment came from the media'. They suggest that the media were 'critical but basically uncomprehending of the rationale and fairness of the method' and cite unfavorable media reports to the effect that their method is 'dreaded', 'much vaunted and complicated', and, simply, 'bizarre'.[5] In response to this media coverage, including a critical editorial piece in the 1998 edition of the Wisden Almanack, they decided that they would hold themselves accountable only to cricket regulators and not to the press or, even, the public.[6]

There is much to lament, I suppose, in this tale of proprietary and commercial data modelling. It is a tale that is likely familiar to you in other, and more serious, domains. But the main force of my paper lies elsewhere. I want to show how the DLS model is constructed, not in an attempt to dismiss it or demonstrate its artificiality, but rather, *pace* Bruno Latour, to show how data models both bring together and reshape a polity.[7] The various recent and useful critiques pointed at big data and algorithms apply to the DLS model. However, there is an underlying question worth asking: What is it about data and algorithm models that makes them so effective and consequently provides the impetus for critique in the first place? Given the initial reception of the DLS and its obvious failings in many respects, how did it establish itself so firmly in the cricketing imagination? I suggest that part of the answer lies in an underlying dynamic of logic and intuition, and compulsion and feeling, that drives forward a contemporary community that is oriented towards probabilistic and algorithmic futures.

Cricket is a game of numbers. Accurate scorekeeping and the laws of the game themselves date back to the mid-18th century. The principal driving force for both was gambling. It is hard to gamble on a game if the scoring is unreliable or if the laws of the game vary from one local match to another. Modern scorekeeping has a double entry book-keeping aspect to it. The batsmen's scores must total up to the bowler's figures, modulo some adjustments for different kinds of extras. We are all likely familiar with arguments about raw data—how data collection is itself a political act that carries its own consequences. I want to consider numbers themselves, however, and the things that can be proved with them, that is, I think it is worth examining model making itself as a semiotic activity.

Brian Rotman argues convincingly that the advent of the computer ought to re-formulate what counts as proof for mathematicians and computer scientists, and the overarching community of deductive proof-seekers they form. Rotman asks with regard to *natural* numbers—for him mathematics is essentially a practical, semiotic activity—'[i]sn't everything—everything cor-

5 Duckworth and Lewis, *Duckworth Lewis*, p. 68.

6 Matthew Engel (ed.) *Wisden Almanack,* London: John Wisden, 1998; Duckworth and Lewis, *Duckworth Lewis*, p. 69.

7 Bruno Latour, 'The Promises of Constructivism', in Don Ihde and Evan Selinger (eds) *Chasing Technoscience: Matrix for Materiality*, Bloomington & Indianapolis: Indiana University Press, 2003, pp. 27–46.

poreal—finite?'[8] Yet the set of natural numbers is infinite in extent. He goes on to suggest that writing and thinking about infinite numbers makes pressing the finite nature of human being and doing. For Rotman, the set of all natural numbers can only be legitimately reasoned about if it is taken as an actively constructed set. But if the emphasis is on the construction of the set, an immediate question follows: Who or what is it that can construct such an infinite set?

The orthodox formal position in mathematics treats its objects of study as timeless entities or forms—ones that are independent of human activity. Rotman emphasizes instead the practical, experiential, and semiotic process through which mathematics is conducted.[9] *Counting* is the fundamental mathematical act, and this is a counting that could only work through a repetition of signs. It is on this point that much of Rotman's argument rests. If mathematics is a practical activity, a construction of entities, rather than a discovery of them, it would follow an understanding of the work mathematics does can only be found in *how* it is undertaken as a practical activity of signification.

Rotman notes that written mathematical proofs are 'riddled with *imperatives*, with commands and exhortations such as "multiply items in *w*", "integrate *x*", "prove *y*", "enumerate *z*", detailing precise procedures and operations that are to be carried out'. In addition, such proofs are 'completely without *indexical* expressions' which raises the immediate questions: 'Who are the recipients of all these imperatives? What manner of agency obeys the various injunctions to multiply, prove, consider, add, count, integrate, and so on? How is the… lack of indexicality related to the impersonal, transcultural nature of mathematical knowledge?' Rotman argues that the implication of this for formal, classical, mathematics and its conceptualization of the infinite must be something like a 'disembodied Agent… —as near to God as makes no difference—[which] is a spirit, a ghost or angel required by classical mathematics to give meaning to 'endless' counting'.[10]

The solution for Rotman is the computer. He argues that computers are a kind of mathematical slave, working mathematical objects into being through the use of energy, time, and space.[11] Recognizing the computer in this way, however, can only be made by the mathematical community as a whole. Constructive mathematics is precisely this kind of practical and semiotic activity that a small section of contemporary mathematicians pursues, building on the intuitionism of L.E.J. Brouwer and work by Erret Bishop and others later. Computers, using energy to operate in space and time, are rule-following agents of this community of mathematicians. It is a commitment to value of those rules that forms the community and enables the computers to think mathematics.

One foundational argument for much science studies today can be traced, I would argue, to David Bloor's understanding of, precisely, rule-following. Bloor developed his theory of the

8 Brian Rotman, *Taking God out of Mathematics and Putting the Body Back in: An Essay in Corporeal Semiotics*, Stanford: Stanford University Press, 1993, p. xi.
9 Ibid., pp. 4–6.
10 Ibid., p. 10.
11 Ibid., p. 152.

social in his book *Wittgenstein: A Social Theory of Knowledge*. Taking Wittgenstein's example of completing the sequence '2, 4, 6, …' Bloor suggests that a 'Platonist has no trouble describing this example in terms of his theory' for 'the correct continuation of the sequence, the true embodiment of the rule and its intended application, already exists'.[12] All that is needed is to 'continue the sequence in the *same* way, and we can do this, and know what it means, by stating the rule of the sequence to ourselves'. But, suggests Bloor, the 'Platonist is actually presupposing the very competence that he is meant to be explaining' because the number sequence is itself the rule: 'its reality extends no further than our actual practice'.[13] There are after all an unbounded number of rules that fit the case '2, 4, 6, …', with each rule supplying possibly different subsequent numbers that 'fit' that beginning sequence. For Wittgenstein, an appeal to the simplicity of a rule is of little use because one would have to have some other principle that allowed one to order the possible rules by their degree of simplicity. What sort of grounds could there be for such a principle other than further social convention? Bloor explains that the force with which simple rule-following presents itself (and, more generally, mathematical proof), makes 'them appear fundamentally different from empirical happenings'. This force is due to the 'form taken in our consciousness by the social discipline imposed upon their use'.[14] It is that social discipline that grounds the correctness of some putative rule-following practice. And, I argue, it is this feeling of being forced into a practice—in this case, answering '8' in continuing the sequence '2, 4, 6, …'—that is a crucial part of how communities form and are reformed.

What these considerations suggest is that a mathematical model like the DLS brings together a cricket community through a rule-following compulsion it exerts on that community. This is a persuasive force built out of rules about number. Indeed, arguments about the DLS model explicitly take the form of social suasion. Duckworth and Lewis ask the reader to imagine a team scoring, say, 250 runs in 50 overs. Should a team chasing that target and on, say, 201 runs with 8 batsmen out after 40 overs when rain intervenes be deemed the winner or loser of the game? Much of the argument of their initial paper and rejoinders over the years by other statistics scholars depend on these hypothetical and counterfactual framings. Considerations of counterfactual futures ground the creation of a community spectatorship of, and reasoning about, cricket. And those grounds are built out of numbers.

Cricket depends on keeping accurate scores since it is the only means by which a game under dispute because it is incomplete can be settled. But, in addition to this basic dependency, following cricket comprises, to a large extent, an appreciation of numbers. Cricket fans can readily state the batting and bowling averages of their favorite players. Various other measures such as the average rate at which batsmen score or the rate at which bowlers get batsmen out are also readily available to fans. Discussions about cricket use these numbers as a way of grounding the conversation. They have the air of empirical fact. Yet, administrators, coaches, players, and fans alike never confuse the numbers for the quality of the performances they

12 David Bloor, *Wittgenstein: A Social Theory of Knowledge,* New York: Columbia University Press, 1983, p. 85.
13 Ibid., p. 85.
14 Ibid., p. 93.

index. The result is that numbers—batting averages and the like—undergird all conversations about cricket because they are always available, yet they never *determine* the outcome of those conversations. It is this ambivalent stance taken towards numbers in cricket that the DLS model successfully disturbed in quantifying the probability of winning and losing.

But what is probability? Ian Hacking argues that what he calls the avalanche of numbers depended on and helped create the idea of probability, starting in the late 17th century. Indeed, Hacking suggests that the identification of induction as a form of reasoning itself was not possible until the question of causes and effects could be dissociated from knowledge, that is, demonstrable knowledge, and placed firmly in the camp of opinion, read as the new notion of probability. The *analytic* problem of induction, Hacking argues, was already available. Here the problem is one of distinguishing between good and bad reasons to argue from induction.

Over the 19th century, this transformation from knowledge to opinion made room for the development of ideas about probability. For Hacking, Leibniz marked the beginnings of modern probability and Hume marked the setting in place of the possibility of inductive knowledge through reasoning about probabilities. The result was the bringing together of ideas about the physical world with a statistical concept of society. A society that had a population constructed out of *normal* people.

C. S. Peirce, in turn, stands in for an altogether different but equally transformative moment. By the end of the 19th century, it became possible to think of the world not as something known probabilistically—a form of understanding that used to be called mere opinion—but rather that the world itself might be probabilistic. It is in this sense that one can think both of normal people in a given population *and* of probabilistic laws of nature.

However, as Hacking argues, it is not that for Peirce an inductive inference could lend a probability to the conclusion of the inference. Rather, the inferential reasoning itself is probable to some possibly quantified degree. As Hacking puts it, deduction for Peirce is such that 'the conclusion of the argument is true whenever the premises are true', but for induction the 'conclusion is usually true when the premises are true'. When precise odds can be ascribed to the premises, Hacking suggests, 'the conclusion is reached by an argument that, with such and such probability, gives true conclusions from true premises'.[15]

Yet, what is often at stake is not the knowledge that a particular method of reasoning leads to truth more often than not, but rather that some particular given inference is reliable or not. Hacking quotes a passage from Peirce that makes the quandary that he had backed himself into quite clear:

> An individual inference must be either true or false, and can show no effect of probability; and, therefore, in reference to a single case considered in itself, prob-

15 Ian Hacking, *The Taming of Chance*, Cambridge: Cambridge University Press, 1990, p. 209.

> ability can have no meaning. Yet if a man had to choose between drawing a card from a pack containing twenty-five red cards and a black one, or from a pack containing twenty-five black cards and a red one, and if the drawing of a red card were destined to transport him to eternal felicity, and that of a black one to consign him to everlasting woe, it would be folly to deny that he ought to prefer the pack containing the larger proportion of red cards, although from the nature of the risk it could not be repeated. It is not easy to reconcile this with our analysis of the conception of chance.[16]

As Hacking notes, Peirce's solution to this problem is quite remarkable. Peirce finds solace in a notion of community:

> It seems to me that we are driven to this, that logicality inexorably requires that our interests shall *not* be limited. They must not stop at our own fate, but must embrace the whole community. This community, again, must not be limited, but must extend to all races of beings with whom we can come into immediate or mediate intellectual relation [. . .].[17]

Peirce is trying to resolve his quandary by focusing on the part of the hypothetical situation he set up that limits the drawing of the card to one instance (and then transporting one immediately to hell or heaven). If members of a community are individually drawing cards, it would be better for them collectively if they drew them from the pack containing more red cards. For Hacking, this is evidence that Peirce was committed to an ontology and metaphysics of chance through and through. Thus, Hacking writes that 'Peirce did not think that first all there is the truth, and then there is a method for reaching it. . . . His theory of probable inference is a way of producing stable estimates of relative frequencies. But on the other hand, the real world just *is* a set of stabilized relative frequencies whose formal properties are precisely those of Peirce's estimators'.[18] This is why Peirce needs to postulate a community of beings with a collective interest in order to resolve the problem he sets himself. The truth about the world just is the result of applying inductive methods because mind and matter evolve together. For Peirce, the world is made up of probabilities, and those quantifiable numbers are grounded in and depend on an expansive sense of community. I argue that data models, whether they draw us in to use them or to oppose them, are both undergirded by and help produce a rule-following, probabilistic, and contemporary 'data governmentality' based community.

Cricket has thrown up and is itself partly grounded in a statistical model of playing. The DLS model enumerates probabilities and predicts outcomes of a world it takes to be inherently probabilistic. Over the last 20 years, the success of the model within cricket has, I argue, silently shifted the cricket-watching public towards accepting probabilistic and predictive data models through a figuring of counterfactual futures. However, the model is constantly tested by cricket players, regulators, and spectators in the sense that its use requires thinking

16 Ibid., p. 211.
17 Quoted in Hacking, *The Taming of Chance*, p. 211.
18 Ibid., p. 213.

counterfactually about each rain-delayed game and balancing notions of fairness and prediction. The case of the DLS model stands in for more recent developments in algorithmic data modelling. These are all semiotic activities that draw together a particular kind of community. All data models, based on a computation of numbers, inscribe a community through logic and intuition, compulsion, and feeling. Not just in the sense that the inputs—the raw numbers fed into them—and the outputs—the effects of the application of the model—of such models are constructed by and for a polity, but also in the sense that the very working out of data models is itself produced through communities. Data models perform communities of feeling the future.

References

Bhattacharya, R., P. S. Gill, and T. B. Swartz. 'Duckworth–Lewis and Twenty20 Cricket', *The Journal of the Operational Research Society* 62.11 (November, 2011): 1951–1957.

Bloor, David. *Wittgenstein: A Social Theory of Knowledge,* New York: Columbia University Press, 1983.

Carter, Michael, and Graeme Guthrie. 'Cricket Interruptus: Fairness and Incentive in Limited Overs Cricket Matches', *The Journal of the Operational Research Society* 55.8 (August, 2004): 822–829.

Duckworth, Frank, and Tony Lewis. *Duckworth Lewis: The Method and the Men Behind It,* Cheltenham: SportsBooks, 2011.

Engel, Matthew (ed.). *Wisden Almanack,* London: John Wisden, 1998.

Hacking, Ian. *The Taming of Chance*, Cambridge: Cambridge University Press, 1990.

Latour, Bruno. 'The Promises of Constructivism', in Don Ihde and Evan Selinger (eds) *Chasing Techno-science: Matrix for Materiality*, Bloomington & Indianapolis: Indiana University Press, 2003, pp. 27–46.

Mirowski, Philip. *More Heat than Light: Economics as Social Physics, Physics as Nature's Economics*, Cambridge: Cambridge University Press, 1989

_____. 'Postmodernism and the Social Theory of Value', *Journal of Post Keynesian Economics* 13.4 (1991): 565–582.

Rotman, Brian. *Taking God out of Mathematics and Putting the Body Back in: An Essay in Corporeal Semiotics*, Stanford: Stanford University Press, 1993.

04. STUDY THE IMBRICATION: A METHODOLOGICAL MAXIM TO FOLLOW THE MULTIPLE LIVES OF DATA

RANJIT SINGH

> Data is the new oil!
>
> – Clive Humby, a Sheffield mathematician[1]
>
> India will go from data poor to data rich in five years as all of a sudden there is tsunami of data.
>
> – Nandan Nilekani, former chairman of Unique Identification Authority of India (UIDAI)[2]

Introduction

Data is the new currency. It is a condition and a resource for understanding knowledge production, dissemination, and consumption. It plays a crucial role in answering questions such as: How is knowledge created and represented? How does knowledge travel across contexts and circulate? How is knowledge understood and interpreted? These questions are certainly not new. They are asked and answered in unique ways in every academic discipline. However, the enthusiasm around big data is certainly new. I am using the term 'big data' here to colloquially address datasets, which often require computation for analysis. Algorithms designed to make sense of these datasets are resources to think statistically at a scale that was unimaginable even a decade ago. Situated in the organizational settings of not only the pursuit of producing more data about people but also the management of decisions that rely on this data, this essay conceptualizes *study the imbrication* as a maxim in researching the role of datasets in producing, distributing, and consuming knowledge.

This maxim is grounded in my research on Aadhaar (meaning Foundation), India's national biometrics-based identification infrastructure, to examine how data and knowledge about a particular resident/citizen/customer is put together and used to streamline bureaucratic and private services. As a topic of research, Aadhaar lends itself into thinking about questions of scale, precarity, materiality of biometric databases, and, most importantly, the politics of citizen data. However, given the length of the essay, I will restrict to research concerns around simplification and circulation in studies of large-scale data infrastructures such as Aadhaar.

1 Charles Arthur, 'Tech Giants May Be Huge, but Nothing Matches Big Data', *The Guardian*, 23 August 2013, https://www.theguardian.com/technology/2013/aug/23/tech-giants-data.

2 DHNS, 'India to Turn Data-Rich in 5yrs', *Deccan Herald*, 8 September 2015. http://www.deccanherald.com/content/499677/india-turn-data-rich-5yrs.html.

This article is divided into three sections. The first section explores how a data infrastructure is constituted. The second section investigates the ways in which data infrastructures operate as a layer on top of existing practices of organizing bureaucratic work (in the case of Aadhaar). The third concluding section presents the maxim of *study the imbrication* to analyze data infrastructures in terms of not only their design and appropriation but also their imagined, intended, and unintended consequences.

Constitution of a Data Infrastructure

Simply put, a data record is a simplified representation of a complex real-world phenomenon with a particular purpose in mind. It is an end as well as the means for the practice of counting. Martin and Lynch coined the word 'numero-politics' to highlight the political not only in the choice of methods for counting but also the consequences of counting practices on things/people that are counted.[3] 'Numero-politics implicates the work of assigning numbers to things and performing elementary arithmetical operations, but such work is embedded in disciplined fields, systems of registration and surveillance, technological checks and verifications, and fragile networks of trust'.[4] An investigation into numero-politics of Aadhaar lends itself into questions such as: who is counted, how they are counted, what the implications are of applying the chosen methods of counting to a resident identity, how residents resist or inspire a change in the methods of counting, what remains uncounted, and what the implications are for such uncounted residents/citizens.

Such concerns around the numero-politics of data infrastructures has inspired a range of scholarship in social studies of data.[5] Simplification has emerged as a salient critique of counting and, by extension, constituting data records within this scholarship. As Annemarie Mol argues:

> The point of asking what is being counted is not to argue that counting is doomed to do injustice to the complexity of life. This is certain. The point, instead, is to discover how and in what ways. For in that process something is foregrounded and something else turned into unimportant detail. Some changes are made irrelevant whereas others are celebrated as improvements or mourned as detrimental.[6]

3 Aryn Martin and Michael Lynch, 'Counting Things and People: The Practices and Politics of Counting', *Social Problems* 56.2 (2009): 243.

4 Ibid., p. 244.

5 See, for example, Lawrence Busch, 'Big Data, Big Questions | A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large Scale Data Sets', *International Journal of Communication* 8 (2014): 1727, http://ijoc.org/index.php/ijoc/article/view/2160/1160; danah boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon', *Information, Communication & Society* 15.5 (2012): 662, https://doi.org/10.1080/1369118X.2012.678878.

6 Annemarie Mol, 'Cutting Surgeons, Walking Patients: Some Complexities Involved in Comparing', in John Law and Annemarie Mol (eds) *Complexities: Social Studies of Knowledge Practices*, Durham: Duke University Press, 2002, p. 235.

Drawing on Theodore Porter, as biometric data gains dominance as stable description of an individual's identity in Aadhaar's appropriation, it simultaneously results in 'thinning' of the individual it describes.[7] To achieve greater precision, efficiency, accuracy, and objectivity, aspects of identity that do not fit in a neat formula for numerical or statistical analysis must be underplayed or removed from consideration. However, Porter also insists that 'We have [. . .] not intrinsic thinness, but thinning and thickening practices suited to diverse circumstances. [. . .] A faith in thinness [. . .] relieves [data] scientists of responsibility by implying that they are not engaged in subtle interpretation, but acting on evidence and in accordance with rules whose meaning is plain'.[8] Thinness is not just a characteristic of the description of a phenomenon under consideration, such as an individual's identity, it is also an instance of practically achieving simplification by following predefined rules of constituting a data record.

Concurrent with simplification is the representation of a phenomenon captured by the resulting data records. A critique of data analytics that limits itself to the challenge of simplification misses out on the amount of work that goes into producing and securing the validity of categories that are used to represent the phenomenon of interest. It obscures the creative ways in which categories establish qualities and make them accountable in a manner that does not simply reduce available information (though it is a common way such categories are justified). For example, Aadhaar captures four categories of demographic data (name, age, gender, and residential address) and three biometric modalities (ten fingerprints, two irises, and a facial photograph) in order to create a unique 12-digit identification number for every enrolled resident. While this certainly involves simplification of complex resident identities, it also involves production of a biometric identity that envisages a one-to-one correspondence between an Aadhaar number and an Aadhaar enrollee, thereby establishing 'uniqueness' of the enrollee. This biometric identity is then employed to resolve an Indian resident across multiple databases of public and private services. Analyzing production of data categories, Martin and Lynch have argued that 'Counting something as something is a condition for determining membership in the domain or field of things or persons counted. [. . .] 'Counting as' [. . .] is an epistemic achievement that involves categorical judgements'.[9] Focusing on these categorical judgements is essential to understand the work of producing a representation of resident identity through Aadhaar. Furthermore, these judgments also predicate the circulation of Aadhaar identity by making the identities of enrollees commensurable across databases.

Circulation of Data Records and Insights

This section traces the consequences of working with data, first, in terms of leveraging a data record to identify and represent a real-world entity (thing/person) and second, in terms of insights developed on a phenomenon under study through data analytics. Consider the example of Aadhaar again. At one level, Aadhaar creates 'reality' and 'uniqueness' of a person as an outcome of a data record that stores their demographic and biometric data. At another level, it

7 Theodore M. Porter, 'Thin Description: Surface and Depth in Science and Science Studies', *Osiris* 27.1 (2012): 209, https://doi.org/10.1086/667828.

8 Ibid., p. 222.

9 Martin and Lynch, 'Counting Things and People', p. 246.

becomes a resource to deduplicate records of below poverty line (BPL) beneficiaries of welfare programs to delineate 'real' and 'unique' beneficiaries from fake ones. In other words, the statistical category of 'uniqueness' must be created before it can be deployed in identifying BPL beneficiaries who fit into this category when they interact with the Indian state's welfare programs. This process lends itself into data analytic insights for accurately tracking real offtake of welfare entitlements.

Ian Hacking has conceptualized dynamic nominalism to describe the interplay of these multiple levels of reality.[10] 'The claim of dynamic nominalism is not that there was a kind of person who came increasingly to be recognized by bureaucrats or by students of human nature but rather that a kind of person came into at the same time as the kind itself was being invented'.[11] Hence, an analysis of the invention of a statistical category such as 'uniqueness' requires working through two interconnected vectors. First is the vector of labeling from above, that is, creation of a 'reality' (for example, unique beneficiaries) that identifies a certain human condition which is then appropriated by bureaucrats (in this case) for their own purposes. Second is the vector of human condition created by autonomous behavior of people (such as claiming uniqueness) that needs to be recognized by the bureaucrats. Hacking argues for a Foucauldian understanding of these two vectors to suggest that they are connected to each other by a whole series of intermediate relations.[12]

One way of approaching these intermediate relations is to investigate how data is managed and processed through data infrastructures. Specifically with respect to behavior of people (claiming welfare benefits, shopping, voting, and so on), dynamic nominalism operates at the intersection of how data about people with particular characteristics becomes constitutive of a dataset (in terms of tables of data categories, etc.) and how data analytics produces people with particular characteristics (inferred as patterns of behavior after analysis) within the dataset. In constituting big data, behavior of people is reflected in what is stored in the databases. This data after analysis informs judgements (such as suitability of methods to distribute welfare) to streamline targeting of people with particular characteristic patterns. The behavior of these people (influenced by such judgements in different ways) goes on to then reflect those characteristic patterns more firmly within the data stored in the databases. Thus, people and big data analytics become enmeshed in a circularity of mutually constituting each other.[13] Along similar lines of critique as observed with respect to simplification, many studies have pointed out the amplification of certain 'realities' and a simultaneous reduction, if not erasure, of other 'realities' in the circulation of data records and data analytic insights

10 Ian Hacking, 'Making Up People', in Margaret Lock and Judith Farquhar (eds) *Beyond the Body Proper: Reading the Anthropology of Material Life*, Durham and London: Duke University Press, 2007, pp. 150–163.

11 Ibid., pp. 155–156.

12 See, for example, Michel Foucault, *Security, Territory, Population: Lectures at the Collège de France 1977-1978*, ed. Michel Senellart, trans. Graham Burchell, New York: Palgrave Macmillan, 2007; Michel Foucault, *The Birth of Biopolitics: Lectures at the Collège de France, 1978-1979*, ed. Michel Senellart, trans. Graham Burchell, New York: Palgrave Macmillan, 2008.

13 Geoffrey C. Bowker, 'Data Flakes: An Afterword to "Raw Data" Is an Oxymoron', in Lisa Gitelman (ed.) *"Raw Data" Is an Oxymoron*, Cambridge, Mass.: MIT Press, 2013, pp. 167–171.

across different contexts.[14] This is also evident in the arguments for marginalization produced by Aadhaar in the distribution of welfare benefits.[15]

Collating these observations, data infrastructures become tools that draw on and fit into existing practices of accomplishing distributed work. They are not an end in and of themselves, rather they are a means to an end that can be (re)specified over time. They are imagined as a layer that operates on top of existing practices and remain relational in their ability to inform and influence these practices. Thus, the question to consider in exploring the relationship between a data infrastructure and existing practices is: When does a data infrastructure connect with such existing practices and when does it become an extension of them? Taking the example of Aadhaar again, when the Aadhaar number is used to deduplicate beneficiary records in social welfare databases, it instantiates a connection between Aadhaar and the social welfare databases. Concurrently, when the Aadhaar number is used to authenticate a beneficiary before they receive their entitlements, it becomes an extension of the process of managing welfare. This distinction outlines how the consequences of appropriating data infrastructures can change significantly when it becomes an extension of an organized practice when compared to when it simply connects as a layer on top of such practices. However, it becomes increasingly difficult to delineate boundaries of this layering over time. Data infrastructures get gradually imbricated into and extend the very nature of the organized practice that they draw on and fit into.

Conclusion: Study the Imbrication

This essay provides methodological indicators for any study that captures lives of data in terms of attention to processes involved in making up data categories and records, and the consequences of using them. Both are equally important in understanding the trajectory of the flow of data and the nature of emerging data analytics-based insights on any organized practice. Taking this idea of flow seriously, it becomes important to carefully choose the moments of time when the nature of this flow is investigated.[16] Star and Ruhleder frame this concern by asking—'when is an infrastructure'—rather than asking what a data infrastructure is.[17] Their focus on temporality is an analytical intervention to unpack the relationships that sustain appropriation of an infrastructure over time. Indeed, as they quote Gregory Bateson, 'What can be studied is always a relationship or an infinite regress of relationships. Never a "thing"'.[18] Data infrastructures are *thick things*—'a phrase meant to invoke the multiple meanings ascribed to particular material artifacts'.[19] However, their thickness unfolds over

14 See, for example, Busch, 'Big Data, Big Questions | A Dozen Ways to Get Lost in Translation'.

15 See, for example, Ursula Rao, 'Biometric Marginality'. *Economic and Political Weekly* 48.13 (2013): 72, http://www.epw.in/review-urban-affairs/biometric-marginality.html.

16 Steven Jackson et al., 'Collaborative Rhythm: Temporal Dissonance and Alignment in Collaborative Scientific Work', in *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, New York, NY, USA: ACM, 2011, pp. 245–254, https://doi.org/10.1145/1958824.1958861.

17 Susan Leigh Star and Karen Ruhleder, 'Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces', *Information Systems Research* 7.1 (1996): 111.

18 Ibid., p. 112.

19 Ken Alder, 'Focus: Thick Things, Introduction', *Isis* 98.1 (2007): 80, https://doi.org/10.1086/512832.

time and their sociomateriality is never a given at any particular moment or place.[20] It must be (re)specified as relationships between a data infrastructure and existing practices change with time.

Lampland and Star use the metaphor of a stone wall to illustrate this slow process of change.[21] A data infrastructure like a good stone wall is an uneven imbrication: an overlapping assemblage of uncemented solutions, 'including discourses, actions, architecture, work, and standards/quantifications/models'.[22] Contesting the static portrayal of infrastructure as layers of stacks, the metaphor of the stone wall highlights how the imbrication that constitutes an infrastructure changes slowly over time and across places. 'Some stone walls fall down; some survive for thousands of years. [. . .] A keystone at one time—a rigid standard, say—may become a minor interchangeable end stone at another, later time'.[23] An imbrication changes over time as new elements are added to it and older elements are partially changed or removed. A good example here is data drift, when data collected on an phenomena of interest changes over time. Different scholars have pointed out different moments of time to elaborate on this change. For example, Star and Ruhleder present one such moment in arguing that infrastructures become (functionally) visible upon breakdown.[24] In a moment of breakdown, the relationships that hold the infrastructure and the existing practices together experience tensions that make them analytically accessible for social science research. Another approach is Geoffrey Bowker's call for 'infrastructural inversion' as a tool to decenter technological solutions in discourses of modernity, progress, and infrastructural development.[25] The analyst 'take[s] a claim that has been made by advocates of a particular piece of science/technology, then look[s] at the infrastructural changes that preceded or accompanied the effects claimed and see[s] if they are sufficient to explain those effects - then ask[s] how the initial claim came a posteriori to be seen as reasonable'.[26] Infrastructural inversion requires the analyst to specify the moment of time when the inversion is brought to bear upon the study of existing practices. In both cases, deciding on the moment allows for analysis of the imbrication to unfold.

To conclude, I offer the maxim that has been a resource as well as an analytic lens in my research on the relationship of Aadhaar with Indian governance: *Study the Imbrication*.[27] This approach situates data infrastructures as extensions of existing practices and unpacks relationships that hold them together at specific times and places. The constitution of data as

20 Wanda J. Orlikowski, 'Sociomaterial Practices: Exploring Technology at Work', *Organization Studies* 28.9 (2007): 1435, https://doi.org/10.1177/0170840607081138.

21 Martha Lampland and Susan Leigh Star, *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca: Cornell University Press, 2009.

22 Ibid., p. 20.

23 Ibid., pp. 20–21.

24 Star and Ruhleder, 'Steps Toward an Ecology of Infrastructure'.

25 Geoffrey C. Bowker, 'Information Mythology: The World of/as Information', in Lisa Bud-Frierman (ed.) *Information Acumen: The Understanding and Use of Knowledge in Modern Business*, London: Routledge, 1994, pp. 231–247.

26 Ibid., p. 235.

27 Ranjit Singh and Steven J. Jackson, 'From Margins to Seams: Imbrication, Inclusion, and Torque in the Aadhaar Identification Project', in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, CO: ACM, 2017.

well as its consequences must be made practically accountable at each chosen moment. For example, in the moment of construction of an Aadhaar record on an enrollee, the imbrication is of Aadhaar with existing ID documents that are used by the Indian bureaucracy. In the moment of authenticating a welfare beneficiary with their Aadhaar number in the process of securing welfare, Aadhaar imbricates with the practices that manage the last mile delivery of welfare entitlements. These two moments provide different portraits of the imbrication that sustains the usability of Aadhaar and its consequences for existing practices. *Lives of data are trajectories of movement within the imbrication that holds their relevance together.* One way to study these trajectories is to follow them as they circulate within this imbrication.

References

Alder, Ken. 'Focus: Thick Things, Introduction', *Isis* 98.1 (2007): 80–83, https://doi.org/10.1086/512832.

Arthur, Charles. 'Tech Giants May Be Huge, but Nothing Matches Big Data', *The Guardian*, 23 August 2013, https://www.theguardian.com/technology/2013/aug/23/tech-giants-data.

Bowker, Geoffrey C. 'Information Mythology: The World of/as Information', in Lisa Bud-Frierman (ed.) *Information Acumen: The Understanding and Use of Knowledge in Modern Business*, London: Routledge, 1994, pp. 231–247.

_____. 'Data Flakes: An Afterword to "Raw Data" Is an Oxymoron', in Lisa Gitelman (ed.) *"Raw Data" Is an Oxymoron*, Cambridge, Mass.: MIT Press, 2013, pp. 167–171.

boyd, danah, and Kate Crawford. 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon', *Information, Communication & Society* 15.5 (2012): 662–679, https://doi.org/10.1080/1369118X.2012.678878.

Busch, Lawrence. 'Big Data, Big Questions | A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large Scale Data Sets', *International Journal of Communication* 8 (2014): 1727–1744, http://ijoc.org/index.php/ijoc/article/view/2160/1160.

DHNS. 'India to Turn Data-Rich in 5yrs', *Deccan Herald*, 8 September 2015. http://www.deccanherald.com/content/499677/india-turn-data-rich-5yrs.html.

Foucault, Michel. *Security, Territory, Population: Lectures at the Collège de France 1977-1978*, ed. Michel Senellart, trans. Graham Burchell, New York: Palgrave Macmillan, 2007.

_____. *The Birth of Biopolitics: Lectures at the Collège de France, 1978-1979*, ed. Michel Senellart, trans. Graham Burchell, New York: Palgrave Macmillan, 2008.

Hacking, Ian. 'Making Up People', in Margaret Lock and Judith Farquhar (eds) *Beyond the Body Proper: Reading the Anthropology of Material Life*, Durham and London: Duke University Press, 2007, pp. 150–163.

Jackson, Steven J, David Ribes, Ayse Buyuktur, and Geoffrey C Bowker. 'Collaborative Rhythm: Temporal Dissonance and Alignment in Collaborative Scientific Work', in *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, CSCW '11, New York, NY, USA: ACM, 2011, pp. 245–254, https://doi.org/10.1145/1958824.1958861.

Lampland, Martha, and Susan Leigh Star. *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Ithaca: Cornell University Press, 2009.

Martin, Aryn, and Michael Lynch. 'Counting Things and People: The Practices and Politics of Counting', *Social Problems* 56.2 (2009): 243–266.

Mol, Annemarie. 'Cutting Surgeons, Walking Patients: Some Complexities Involved in Comparing', in John Law and Annemarie Mol (eds) *Complexities: Social Studies of Knowledge Practices*, Durham: Duke University Press, 2002, pp. 218–257.

Orlikowski, Wanda J. 'Sociomaterial Practices: Exploring Technology at Work', *Organization Studies* 28.9 (2007): 1435–1448, https://doi.org/10.1177/0170840607081138.

Porter, Theodore M. 'Thin Description: Surface and Depth in Science and Science Studies', *Osiris* 27.1 (2012): 209–226, https://doi.org/10.1086/667828.

Rao, Ursula. 'Biometric Marginality', *Economic and Political Weekly* 48.13 (2013): 72–77, http://www.epw.in/review-urban-affairs/biometric-marginality.html.

Singh, Ranjit, and Steven J. Jackson. 'From Margins to Seams: Imbrication, Inclusion, and Torque in

the Aadhaar Identification Project', in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, CO: ACM, 2017.

Star, Susan Leigh, and Karen Ruhleder. 'Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces', *Information Systems Research* 7.1 (1996): 111–134.

05. DATA LIVES OF HUMANITIES TEXT

PUTHIYA PURAYIL SNEHA

The 'computational turn' in the humanities has brought with it several questions and challenges for traditional ways of engaging with the 'text' as an object of enquiry.[1] In fields such as humanities computing, cultural analytics, and now digital humanities (DH), the use of computational methods is steadily becoming prevalent in working with and studying cultural artifacts today. This development is both necessitated by and adds to the availability of a large corpora of materials (digitized and born-digital) in an array of formats and across varied platforms. These cultural data sets have grown in abundance because of many factors, including better access to digital technologies, and the ubiquitous presence of the internet-facilitated new modes of documentation and circulation of information. The prevalence of data-driven scholarship in the humanities offers several challenges to traditional forms of work and practice, with regard to theory, tools, and methods. In the context of the digital, 'text' acquires new forms and meanings, especially with practices such as distant reading.[2]

This essay will explore how 'data' in the humanities has become a new object of enquiry as a result of several changes in the media landscape in the past few decades. The availability of a vast corpora of digital materials and the advent of new tools and methods are primary factors here, resulting from the large-scale digitization of cultural artifacts, creation of new online archival platforms, and the growth of processes such as curation, annotation, referencing, visualization, and abstraction in research and practice. Drawing upon excerpts from a recently completed study on DH in India, this essay will discuss how data in the humanities is not a new phenomenon; concerns about the 'datafication' of humanities, now seen prominently in DH and related fields is actually reflective of a longer conflict about the inherited separation between humanities and technology. Fields such as DH provide a space to illustrate these conflicts, and, in doing so, open up possibilities to trace a twinned history of humanities and technology. This essay will also discuss how reading 'text as data' helps understand the role of data in the making of humanities texts and redefines traditional ideas of textuality, reading, and the reader. Importantly, it seeks to understand the growth of such data-driven scholarship as informed by an 'archival turn' in the practice of humanities and arts, which remains imperative to advancing new forms of enquiry and in framing its concepts and methods. Through this, the essay will attempt to provide an insight into the data lives of humanities texts.

'Data' in the Humanities

The emergence of data-driven scholarship in the humanities appears be a relatively new phenomenon. The proliferation of gadgets and a culture of sharing fostered by the ubiquity of social media and other online spaces of collaborative knowledge production, such as Wiki-

1 David M. Berry, 'The Computational Turn: Thinking About the Digital Humanities', *Culture Machine* 12 (2011), https://sro.sussex.ac.uk/id/eprint/49813/1/BERRY_2011-THE_COMPUTATIONAL_TURN-_THINKING_ABOUT_THE_DIGITAL_HUMANITIES.pdf.

2 Franco Moretti, *Distant Reading*, London: Verso, 2013.

pedia, have contributed to this change. The growth of private online archival spaces, aided by access to the internet and availability of infrastructure in the form of tools, platforms, and new technologies for documentation, circulation, curation, and use of digital material also forms an important context to these developments. Computational methods, however, have been part of humanities study and practice for some time. Julia Flanders notes that 'the ubiquity of computing resources means that it's no longer remarkable for humanities scholars to work with computers'; so the idea of 'humanities data' is not of recent emergence, even though it has occasioned much debate and conflict.[3] This resistance to fields like DH is illustrative of this, where the use of computational methods is seen as taking away from traditional approaches to engaging with texts.[4]

This resistance to datafication of the literary or textual has been countered in early DH discourse, largely by locating a history of the field in humanities computing, in processes like concordance, stylometery, and lemmatization.[5] Text mining or text analytics using methods from natural language processing (NLP) are other examples.[6] Mathew Kirschenbaum states that 'after numeric input, text has been by far the most tractable data type for computers to manipulate. Unlike images, audio, video, and so on, there is a long tradition of text-based data processing that was within the capabilities of even some of the earliest computer systems and that has for decades fed research in fields like stylistics, linguistics, and author attribution studies.'[7] The re-textualization of 'literary objects' through digital media, such as Facebook or YouTube, also renders them as new objects of enquiry. The making of these digital objects involves a process of disaggregation, producing different kinds and large volumes of data, often as ancillary material. These digital objects demand a new form of engagement with the 'text', as the primary artifact itself has been rendered different through digitalization. One example is distant reading, and, more broadly, through questions about forms of textuality, materiality, and medium which emerge as pertinent, locating this notion of data within humanities, and why it evokes divided opinions. The problem of an abundance of data generated by making, sharing, and using these new digital objects has resulted in processes like curation, annotation, referencing, visualization, and abstraction becoming important methods of parsing and creatively making meaning of content. These processes also urge a rethinking of the concept of the reader and practices of reading, if indeed they may still be called reading. Kirschenbaum, in his paper on implications for data mining in literature, elaborates that its 'potential to "provoke" a human subject expert may yield insights not readily obtainable otherwise.' He adds that

3 Julia Flanders, 'The Productive Unease of 21st-century Digital Scholarship', *Digital Humanities Quarterly* 3.3 (2009), http://www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html.

4 For more on this see, Stanley Fish, 'Mind Your 'Ps' and 'Bs': The Digital Humanities and Interpretation', *New York Times*, 23 January 2012; Stephen Marche, 'Literature Is Not Data: Against Digital Humanities', *Los Angeles Review of Books*, 28 October 2012; and Adam Kirsch, 'Technology Is Taking Over English Departments', *New Republic*, 2 May 2014.

5 Susan Schreibman et al., *A Companion to Digital Humanitie*s, Oxford: Blackwell, 2008.

6 Anne Kao and Steve R. Poteet (eds) *Natural Language Processing and Text Mining*, New York: Springer, 2007.

7 Matthew Kirschenbaum, 'What Is Digital Humanities and What's It Doing in English Departments?', in Mathew K. Gold (ed.) *Debates in the Digital Humanities*, Minneapolis: University of Minnesota Press, 2012, http://dhdebates.gc.cuny.edu/debates/text/38.

'[r]eading is not so much "at risk" as in the process of being remade, both technologically and socially.'[8]

Gitelman and Bowker suggest that several preliminary concerns about 'data', however, remain to be addressed, including: 'What are the histories of data within and across disciplines? How are data variously "cooked" within the varied circumstances of their collection, storage, and transmission? What sorts of conflicts have occurred about the kinds of phenomena that can effectively — can ethically — be "reduced" to data.'[9] They propose that 'one productive way to think about data is to ask how different disciplines conceive their objects, or, better, how disciplines and their objects are mutually conceived.'[10] To extend this further, Christine Borgman asks: 'What constitute data in the humanities? What are data sources? How are they made, shared, valued, used, and reused?'[11] Pre-conceived notions often accompany the term 'data'; the imagination comes from the use of the term in the natural and social sciences, more often than not as quantitative, abstract, objective, and, maybe, inflexible, and before interpretation, so 'raw' as Gitelman points out, although there have been efforts to rethink these notions and redefine what data means in humanities.[12]

Reading Data as Text

While the use of data is central in natural and social sciences, its significance for a humanities scholar is contested. In adopting computational methods, are the disciplinary questions also changing? What is the difference or novelty in these questions for the humanities? Arguably these conflicts between text and data are a result of blurring boundaries, and a field like DH which seeks to be collaborative and interdisciplinary, and has consequently provoked much debate and even criticism about the (increased) role of technology in the humanities, could be a space to explore these conflicts.

Although the use of data-driven methodologies in humanities is not prevalent in India, there have been some recent digital initiatives, even if practical constraints have restricted their elaborate use. *Bichitra* is an online variorum of the works of the Indian writer and poet Rabindranath Tagore, developed by the School of Cultural Texts and Records (SCTR) at Jadavpur University, Kolkata.[13] It contains most versions of Tagore's works—poetry, drama, fiction, and nonfiction—but excludes letters, speeches, textbooks, and translations, except those done by Tagore himself. Digitization is a lengthy process of sourcing material, photographing/scan-

8 Matthew Kirschenbaum, 'The Remaking of Reading: Data Mining and the Digital Humanities', 2007, http://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf.

9 Lisa Gitelman and Virginia Jackson, 'Introduction', in Lisa Gitelman (ed.) "*Raw Data" Is an Oxymoron*, Cambridge, Mass.: MIT Press, 2013, p. 3.

10 Ibid., p. 7.

11 Christine L. Borgman, 'The Digital Future Is Now: A Call to Action for the Humanities', *Digital Humanities Quarterly*, 3.4 (2016), http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html.

12 Trevor Owens, 'Defining Data for Humanists: Text, Artifact, Information or Evidence?', *Journal of Digital Humanities* 1.1 (2017), http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/2011.

13 Bichitra, http://bichitra.jdvu.ac.in/index.php.

ning, making copies searchable with optical character recognition (OCR), uploading, and cross-referencing. The website has three unique functionalities—the bibliography, search engine, and a collation software named *Prabhed* (meaning *difference* in Bengali). The bibliography is linked to the scans and transcriptions of different versions of a text, and these are open to the data tracking resources of the website. Using the search engine to track a word or phrase leads to all its occurrences in the entire corpus. *Prabhed* collates the different versions of a work at three levels—chapter, paragraph, and word—and tracks all the migrations and variations across editions. The project has received an overwhelming response but not without some unique challenges, such as locating and acquiring content, lack of OCR for Bengali fonts, and problems of privacy and access, among others.[14] Using computational tools, it is now possible to search across such a large corpus of material and pose new questions related to their access (in digital form), context, and usage.[15] It also offers important provocations for understanding language and representation in the digital context, and our interaction with technology. What is the role of such a platform/resource for a literary studies/humanities scholar? How can we see traditional practices of reading and writing being reimagined in this context?

Indiancine.ma and *Pad.ma* are two online archives that are significant in terms of the archival questions and possibilities emerging with the transition of film from celluloid to digital.[16] The *Public Access Digital Media Archive*, or *Pad.ma*, is a collection of audio and video materials ranging from found footage, stills, sound clips to unfinished films. The database is searchable, and materials can be viewed/listened to and downloaded. Users can work with the material in multiple formats and can add transcripts, descriptions, events, keywords, and maps through annotations and referencing. Like *Pad.ma*, *Indiancine.ma* is an online archive of films that are out of copyright (released sixty years ago) and is built upon a free/libre and open source software (FLOSS) named Pan.do/ra, a web application that helps organize and manage large decentralized archives of video materials, and create metadata and time-based annotations in the forms of text, photographs, images, and posters. Users can edit and annotate a particular sequence in the film according to a time code, and search and organize content through different filters, such as colour and object recognition. This offers a different mode of engagement with the film, by creating a new kind of research object, structured though different forms of meaning–time, date, maps, and so on. The film object is layered by different kinds of data—texts, images, writing, tagging, and annotations—thus facilitating new ways of reading the primary text. This is possible precisely because of the digital, and it also illustrates the ways in which the primary object of enquiry, the film or archival object, as well of as the methods of study, have evolved or need to evolve in response to advancements in technology.[17]

14 For a more detailed description, see the interview with Prof. Sukanta Chaudhuri in, P. P. Sneha, *Mapping Digital Humanities in India,* Bangalore: Centre for Internet and Society, 2016, https://cis-india.org/papers/mapping-digital-humanities-in-india.

15 Sukanta Chaudhuri, *Bichitra: The Making of an Online Tagore Variorum*, New York: Springer, 2015.

16 Indiancine.ma, https://indiancine.ma/; Pad.ma, https://pad.ma/.

17 For more on this, see the interview with Ashish Rajadhyaksha in, Sneha, *Mapping Digital Humanities in India*.

Pad.ma, *Indiancine.ma*, and *Bichitra*, being essentially large databases of cultural material, offer several possibilities of working with computational tools. Color and object recognition is already a part of the filters on *Indiancine.ma*.[18] The affordances of these tools are also reshaping older analytical practices of studying film, like statistical style analysis, an approach predating digital technologies, as illustrated by Barry Salt, David Bordwell, and others, and further developed through computational methods.[19] The makers of *Bichitra* are exploring the possibilities offered by topic modeling.[20] While this is helpful in parsing data in a large corpus, the researchers working with these projects emphasize the need to understand how these efforts may add to the study of the primary object, which is the film or printed text. The motivation behind the development of some of these tools is also varied, like surveillance for example, where tools are used to gather data from social media, or CCTV footage, raising questions about privacy and data protection which often go unaddressed—hence the purpose of the technology and its limitations may also determine or restrict the scope of enquiry.[21] Importantly, such projects require expertise and skills spanning diverse domains, and DH, in encouraging collaborative and interdisciplinary work, helps articulate epistemological conflicts.

Data Lives of Humanities Texts

Fields like DH enable creation of spaces such as digital archives and labs, and methods wherein the making of cultural artifacts may be illustrated explicitly. The process of making humanities texts, especially in the digital, is as much about data as it is about acts of reading and writing. It is also one that is simultaneously changing and evolving. The digital medium is processual, where objects or images are constantly being made, and unmade.[22] The process of curation is important here—gathering, organization, remaking, and representation are demonstrative of many stages that an object goes through within the archive to become available as a text for further reading. Tracing several versions of a text, tracking minute changes across one edition to another involve specific decisions about its classification, metadata, search, retrieval, and use. Processes like transcription from image to text and developing

18 For an example of the scope of work that could be undertaken with such tools such as color and object recognition, see Selfiecity, http://selfiecity.net/.

19 Barry Salt, 'Statistical Style Analysis of Motion Pictures'*, Film Quarterly* 28.1 (1974): 13, https://online.ucpress.edu/fq/article/28/1/13/38835/Statistical-Style-Analysis-of-Motion-Pictures; David Bordwell, *The Way Hollywood Tells It: Story and Style in Modern Movies,* University of California Press, 2006. For more work on this using computational methods, see work by Yuri Tsivian and others at Cinemetrics, http://cinemetrics.lv/tsivian.php.

20 Chaudhuri, *Bichitra*, pp. 146–164, describes topic-modelling as 'a type of computer operation that examines the frequency of certain sets of words in a corpus of texts, with a view to determining the topics common to them. It thereby allows us to detect the subjects or concerns operating in a discourse'.

21 For more on this, see, Michael Widner, 'The Digital Humanists' (Lack of) Response to the Surveillance State,' Author's Blog, 20 August 2013, https://web.archive.org/web/20130820213839/https://people.stanford.edu/widner/content/digital-humanists-lack-response-surveillance-state; Stéfan Sinclair and Geoffrey Rockwell, 'Teaching Computer-Assisted Text Analysis: Approaches to Learning New Methodologies', in Brett Hirsch (ed.) *Digital Humanities: Practices, Principles and Politics*, Open Book Publishers, 2012, https://www.openbookpublishers.com/reader/161#page/1/mode/2up.

22 Mark Hansen, *New Philosophy for New Media*, Cambridge, Mass.: The MIT Press, 2004.

OCR for Indian languages need to draw from the database capabilities offered by assembling and processing a large corpus of cultural material. The juxtaposition of different types of data (image, text, video, audio, maps, etc.) within a single space, around a specific text also help create new versions and readings, a new mode of *compositing* of the digital image.[23] With every iteration of the object, the aggregation and disaggregation of data changes; the purpose of digital archives or variorum like *Bichitra* and *Indiancine.ma* is to render these texts in the form of manipulable data. This allows a different kind of access to both the archive, and the archival object itself. The film or text is now accessible not as a complete, finished work, but can be disaggregated and made available for reading and remaking through various modes, such as annotations, filters, and collation. While this helps redefine conventional notions of the archive as a space of preservation, it emphasizes the perpetuation and growth of the archival object through its circulation and diversified use. An open, accessible, and collaborative archive as a space that juxtaposes and often collapses different processes—of making and interpretation, of practice and analysis—allows for a more nuanced, affective engagement with the object itself. The role of data is important here, as it is no longer pre-object or analysis, as Gitelman observes, but an integral part of the process of creating the text or cultural artifact, or it, in fact, becomes the object of enquiry, for example in cases where the original text is missing. It therefore poses an important question about what methods are required for working with these digital objects. As mentioned earlier, the use of computational methods such as text mining, including topic modeling, etc., as processes of making meaning of cultural data evokes a certain anxiety about displacing the 'text' as the object of enquiry; it also challenges the human reader as the privileged subject of interpretative acts. This notion of a 'non-human reader' is already a significant aspect of work in artificial intelligence, specifically machine learning, which endeavors to constantly push the limits of the computer's capabilities to replicate human thought and learning.[24] The mass production of data and the development of data-centric approaches are thus important in tracing a 'technologized' history of humanities and for exploring unresolved questions in fields like AI as well.

'Data' in the humanities would be a useful trope, therefore, to illustrate the manner in which disciplines are changing with the advent of digital technologies. While texts have always been a form of data, the manner in which they are produced, circulated, and used can be illustrated more explicitly now, with increased access to computational methods in emerging fields like DH. With its aim of bringing together humanities and technology (if they are indeed separate domains), DH can provide a space where some of these questions may be explored in detail, through both practice and analysis.

Acknowledgments

I am grateful to Sarai-CSDS for the opportunity to present my work at the Lives of Data workshop and for the helpful feedback and suggestions. This essay draws upon interviews and observations from a report on mapping the field of digital humanities in India. I would like to thank Sumandro Chattapadhyay and Nishant Shah for their insights on key questions explored in it.

23 D. N. Rodowick, *The Virtual Life of Film*, Cambridge, Mass.: Harvard University Press, 2007.
24 Stuart C. Shapiro (ed.) *Encyclopedia of Artificial Intelligence*, New York: John Wiley, 1992.

References

Berry, David M. 'The Computational Turn: Thinking About the Digital Humanities', *Culture Machine* 12 (2011), https://sro.sussex.ac.uk/id/eprint/49813/1/BERRY_2011-THE_COMPUTATIONAL_TURN-_THINKING_ABOUT_THE_DIGITAL_HUMANITIES.pdf.

Borgman, Christine L. 'The Digital Future Is Now: A Call to Action for the Humanities', *Digital Humanities Quarterly* 3.4 (2009), http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html.

Chaudhuri, Sukanta. *Bichitra: The Making of an Online Tagore Variorum*, New York: Springer, 2015.

Fish, Stanley. 'Mind Your 'Ps' and 'Bs': The Digital Humanities and Interpretation', *The New York Times*, 23 January 2012, http://opinionator.blogs.nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and-interpretation/.

Flanders, Julia. 'The Productive Unease of 21st-century Digital Scholarship', *Digital Humanities Quarterly* 3.3 (2009), http://www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html.

Gitelman, Lisa and Virginia Jackson. 'Introduction', in Lisa Gitelman (ed.) "*Raw Data" Is an Oxymoron,* Cambridge, Mass.: MIT Press, 2013, pp. 1–14.

Kao, Anne and Steve R. Poteet (eds). 'Overview', *Natural Language Processing and Text Mining*, New York: Springer, 2007, pp. 1–7.

Kirsch, Adam. 'Technology Is Taking Over English Departments,' *New Republic*, 2 May 2014, https://newrepublic.com/article/117428/limits-digital-humanities-adam-kirsch.

Kirschenbaum, Matthew. 'What Is Digital Humanities and What's It Doing in English Departments?', in Mathew K. Gold (ed.) *Debates in the Digital Humanities*, Minneapolis: University of Minnesota Press, 2012, http://dhdebates.gc.cuny.edu/debates/text/38.

_____. 'The Remaking of Reading: Data Mining and the Digital Humanities', 2007, http://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf.

Marche, Stephen. 'Literature is Not Data: Against Digital Humanities', *Los Angeles Review of Books*, 28 October 2012, https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities/#!.

Hansen, Mark B. N. *New Philosophy for New Media*, Cambridge, Mass.: The MIT Press, 2004.

Moretti, Franco. *Distant Reading*. London: Verso, 2013.

Owens, Trevor. 'Defining Data for Humanists: Text, Artifact, Information or Evidence?', *Journal of Digital Humanities* 1.1 (2011), http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/.

Rodowick, D. N. *The Virtual Life of Film*. Cambridge, Mass.: Harvard University Press, 2007.

Schreibman, Susan, Ray Siemens, and John Unsworth (eds). *A Companion to Digital Humanities*, Oxford: Blackwell, 2004.

Sinclair, Stéfan, and Geoffrey Rockwell. 'Teaching Computer-Assisted Text Analysis: Approaches to Learning New Methodologies', in Brett. D. Hirsch (ed.) *Digital Humanities: Practices, Principles and Politics*, Open Book Publishers, 2012, DOI: 10.11647/OBP.0024.

Sneha, P. P. *Mapping Digital Humanities in India*, Bangalore: The Centre for Internet and Society, 2016, https://cis-india.org/papers/mapping-digital-humanities-in-india.

Shapiro, Stuart C. (ed.) *Encyclopedia of Artificial Intelligence*, New York: John Wiley, 1992.

Widner, Michael. 'The Digital Humanists' (Lack of) Response to the Surveillance State', Author's Blog, 20 August 2013, https://web.archive.org/web/20130820213839/https://people.stanford.edu/widner/content/digital-humanists-lack-response-surveillance-state.

06. HACKATHONS: LABOR, POLITICS, AND THE ORGANIZATION OF PUBLIC PASSIONS

LILLY IRANI

The lives of data, like affects, are uncertain, animated by public cultures and passions directed through organizing. People engage one another, animated by drives, duties, fears, and hopes. Among those vying to shape those affects are state and philanthropic institutions, the private sector, and activists. The passions provoked by 'open data', 'innovation', and 'nation-building' can prove potent resources for experiments in statecraft, private-sector research and development, or activist infrastructures. They can do so in ways that strengthen the adaptive capacities of investors and governments, or they can do so in ways that strengthen the reproduction of resistance and transformative efforts. This chapter focuses on hackathons and the ways they can extend infrastructures, systems, and interpretive practices through which data comes alive.

Hackathons are just one labor process that brings data to life. Hackathons are intense, multi-day events that gather people in intense, urgent, and collaborative digital labor—often the labor of designing demos or prototypes of software-to-come. The events are often structured as a scramble towards hope, allowing participants to engage in intense technological labors that can benefit distant masses through the mediation of technology. In India, as in the United States, technology as a vehicle of development is hardly new. The temples of modern India, however, have shifted in scale, from dams produced by technocratic state to apps produced by technocratic entrepreneurs. The civil engineer has given way to the computer engineer and designer as an ideal citizen.[1] The Government of India, the World Bank, venture capitalists, and non-profits invite citizens to imagine change in the idiom of software. This is one practice of what elsewhere I have called 'entrepreneurial citizenship' that posits design and social entrepreneurship as a way Indians can do nation-building, create financial value, and author 'authentic' selves at the same time.[2] These institutions employ hackathons to proliferate opportunity; they manufacture urgency, gather people to work, and attempt to capitalize on existing infrastructures and labors hidden elsewhere. As devices for organizing affects—as energy, and as interpersonal relationships—they stir public passions to generate potential financial value.[3] But hackathons need not only expand accumulation. I conclude the

1 Philip, Kavita, 'Telling Histories of the Future: The Imaginaries of Indian Technoscience', *Identities* 23.3 (2016): 276; Ajantha Subramanian, 'Making Merit: The Indian Institutes of Technology and the Social Life of Caste', *Comparative Studies in Society and History* 57.2 (2015.): 291.

2 Lilly Irani, 'Hackathons and the Making of Entrepreneurial Citizenship', *Science, Technology, & Human Values* 40.5 (2015): 799.

3 See, Sreela Sarkar, 'Passionate Producers: Corporate Interventions in Expanding the Promise of the Information Society', *Communication, Culture & Critique* 10.2 (2017): 241; Lilly Irani, *Chasing Innovation: Making Entrepreneurial Citizens in Modern India,* Princeton, NJ: Princeton University Press, 2019. Both Sarkar and Irani find middle-class Indians react to the alienations of global, corporate workplaces described by Aneesh, *Virtual Migration*, Durham: Duke University Press, 2006; Kalindi

chapter with a discussion of a different use of public passion—hackathons in which people gather to care for infrastructures and data that sustain the publics and their politics in the face of environmental extraction.

I conducted the fieldwork that informs this chapter over 14 months, between 2009 and 2014, primarily immersed in a design studio in Delhi, India, and the work of those who moved around the studio. I'll call the studio DevDesign.

Delhi at the time of my fieldwork seemed a development boomtown. Since before independence, Delhi has been a center of development planning and calculation to modernize Nehru's 'needy nation'.[4] Five Year Plans and import controls had given way after liberalization to facilitating the movement of capital investment and the growth of public-private partnerships.[5] By 2004, Goldman Sachs directed global investors to the potential of emerging markets in BRICs, and C. K. Prahalad directed business leaders to seek their fortunes 'at the bottom of the pyramid'.[6] DevDesign worked in the speculative 'dream zones',[7] doing user research to develop designs for products and services for the 'bottom of the pyramid'. They did fieldwork for London-startups working on hand sanitation. They coached Indian college students in dreaming up improvements to water distribution. They consulted with multinational corporate social responsibility initiatives. They even consulted with the Government of India's 'smart cities' project. Acknowledging that times were flush in the Delhi development scene, the director of DevDesign once quipped, 'There's nothing wrong with a bubble if you are in at the beginning'. These designers speculated at the nexus of nation-building and new product development, adopting the role of developmental mediators circulating among villagers and *basti* dwellers—potential users and targets of development—and the investors, philanthropies, government agencies, and consumer product firms that hoped to intervene.

Beyond products, the studio evangelized design as a model for making Indians into entrepreneurial citizens. They put on an annual festival celebrating 'interdisciplinary action' directed

---

Vora, *Life Support: Biocapital and the New History of Outsourced Labor*, Minneapolis, MN: University of Minnesota Press, 2015; Shehzad Nadeem, 'Macaulay's (Cyber) Children: The Cultural Politics of Outsourcing in India', *Cultural Sociology* 3.1 (2009): 102; and Sareeta Amrute, *Encoding Race, Encoding Class: Indian IT Workers in Berlin*, Durham: Duke University Press, 2016—by investing their passions into corporate social responsibility and uplift projects. These 'passionate producers', as Sarkar calls them, bring poorer Indians into the very global information economy that they themselves found so alienating. DevDesign members were aware of this irony but responded to the structures of philanthrocapitalist funding agendas.

4 Srirupa Roy, *Beyond Belief: India and the Politics of Postcolonial Nationalism,* Durham: Duke University Press, 2007, p. 110.

5 Stuart Corbridge and Jonathan Harriss, *Reinventing India: Liberalization, Hindu Nationalism and Popular Democracy*, Cambridge, UK: Polity, 2000, p. 120; Atul Kohli, 'Politics of Economic Growth in India, 1980-2005: Part I: The 1980s', *Economic and Political Weekly* 41.13 (2006): 1251; Arvind Rajagopal, 'The Emergency as Prehistory of the New Indian Middle Class', *Modern Asian Studies* 45.5 (2011): 1003.

6 Dominic Wilson and Roopa Purushothaman, 'Dreaming with BRICs: The Path to 2050', *Global Economics Paper* 99 (2003): 1; C. K. Prahalad, *The Fortune at the Bottom of the Pyramid*, Pearson Prentice Hall, 2006.

7 Jamie Cross, *Dream Zones: Anticipating Capitalism and Development in India*, London: Pluto Press, 2014.

at students, planners, engineers, artists, and development workers. They showed existence proofs of activism, social business models, and even literary production in Indian vernacular languages. They reached wide to elicit 'progressive' sentiment, banners at the festival one year listed off words that appeal to the English-fluent: 'Brand - Community - Enterprise - Crafts - Innovation - Habitat - Ideation - New Media'. 'Activism - Impact - Curation - Culture - Tradition - Heritage - Reform - Experience - Sustainability'. I want to point out that in the festival, 'new media' evokes a kind of hope, but it is part of a mosaic concerned more with modernity than with the digital itself. DevDesign's civic entrepreneurialism was just one example of many schools, conferences, and contests I came across over the course of my fieldwork teaching similar attunements.

The hackathon that I now turn to was part of the studio's design festival as just one example of multi-day workshops meant to immerse participants in 'hands-on, hearts-on, minds-on' development activity. Other workshops included designing craft programs for a Gandhian NGO in Ahmedabad and developing solar power initiatives in Auroville. What the workshops had in common was that they brought together people who did not know each other to spend a few days dreaming of development projects, and then making those dreams concrete as demos, plans, and presentations.

The hackathon I participated in was like a multi-day software production party. It was one of a genre of events drawn from open source cultures but adopted recently in the development and corporate sectors as a way of recruiting volunteers to do experimental labor for free or to build excitement around an agenda. Examples included Indian Planning Commission hackathons to work with government data, Silicon Valley venture capital-sponsored hackathons to pitch startups in Bangalore, and an Infosys–World Bank hackathon to develop 'solutions' to sanitation problems. Organizers typically provide space, take out dinners, electricity, Wi-Fi, and a roof for anywhere from a day to a week; software engineers and designers can come together to meet people, test their skills, and produce a demo—a piece of software that operates like a promise of technology to come.

Hackathons began as a way for participants in globally distributed open source projects to work together, face-to-face for short periods of time. These open source hackathons were a way for programmers already familiar with one another to take advantage of rare moments of geographic copresence. Face-to-face programmers, who usually only connected online, could quickly, collaboratively, and intensively care for and maintain code and related infrastructures. These hackathons allowed for intense collaborative labor among programmers with already deep ties to the open source community.[8]

In recent years, companies, NGOs, universities, and even government agencies have taken up hackathons as a means to recruit volunteer labor, generate interest in social or technological platforms, and use participants to explore possible futures for a host organization. The company Facebook regularly hosts hackathons to explore future projects and to inculcate in

8 Gabriella Coleman, 'The Hacker Conference: A Ritual Condensation and Celebration of a Lifeworld', *Anthropological Quarterly* 83.1 (2010): 47.

employees the ability to 'move fast and break things'.[9] Hackathons entered a global lexicon of public culture when MTV featured a Facebook hackathon in a documentary about the company.[10] The World Bank organized a Global Water Hackathon in 2011 with 500 'hackers' across nine cities to direct entrepreneurial programmers toward partner agendas.[11] In 2013, for example, non-profits and government bodies across the United States participated in a National Civic Day of Hacking, an intense Saturday of coordinated digital volunteerism.[12] These events invite people to experiment with possibilities for social ventures, tools for mapping water in crisis regions, or prototypes of future startup offerings. While early, open source hackathons often focused on improving, repairing, and maintaining shared infrastructures, the hackathons have also grown to include speculation about technological futures that rely precisely on those infrastructures cared for elsewhere.

The theme of this hackathon was 'open governance'. As we ambled into the studio at 9 a.m. the first morning, the cook handed us *chai* and we sat with laptops open at a long table. The convener had us introduce ourselves and describe our motivations. The seduction of tangible action—of making and doing something other than words—was on many of our minds. A young Bangalore software consultant wanted to quit cribbing about governmental inefficacy to 'see if we can make a difference'. An Indian Institute of Technology-trained designer wanted to see if design could actually save the world instead of just 'making posters' for clients about it. I was there to see what would happen if I brought anthropological sensibilities critical of development and my coding skills together to attempt technology as a critical practice. Prem, a legal anthropologist, came because in his words, 'anthropologists sit and critique things, but they never get around to doing anything'. All the speech act theory in the world left him still wanting to experiment with other forms of intervention. In different ways, what was at stake for all of us was performing the promise of agency—of action which promises to make a difference, and promise is key here—in a messy, complex world through some kind of building.

We began by familiarizing ourselves with the domain. Vipin, the convener, had recruited a friend at Parliamentary Research Service who guided us towards Parliamentary standing committees as a site where we could inform legal deliberation through the software we would design. Most of us had experience making software, but few of us had knowledge of the legal process. We read through and critiqued a recent Road Safety Bill draft to put ourselves in the shoes of possible law-reading users. We learned about parliamentary procedures. Vipin, trained at IIT and Indian Institute of Management, kept up on business and computing trends. He pushed a stack of books on 'Open Government' and e-Government, exclusively based on American case studies, to me and told me to skim for anything 'that interested' me.

9 Alex Fattal, 'Facebook: Corporate Hackers, a Billion Users, and the Geo-Politics of the 'Social Graph'', *Anthropological Quarterly* 85.3 (2012): 927.

10 Andrew Huang, *Diary of Facebook*, Documentary, Biography, 2011, http://www.imdb.com/title/tt1882342/.

11 World Bank, 'Water Hackathons: Lessons Learned', Water Papers, Washington, D.C.: World Bank, May 2012.

12 Melissa Gregg, 'FCJ-186 Hack for Good: Speculative Labour, App Development and the Burden of Austerity', *The Fibreculture Journal* 25 (2015), http://twentyfive.fibreculturejournal.org/fcj-186-hack-for-good-speculative-labour-app-development-and-the-burden-of-austerity/.

These activities were interwoven with expressions of time anxiety. Someone, most often one of the software engineers, would ask us to sketch a production schedule. How long could we talk about the law? Could we scope the time of debate to assure ourselves that we could produce 'the demo'? As we negotiated milestone deadlines, Vipin pushed post-it notes around the board, representing the timeline leading up to the festival. This collective visualization of time forced us to work backward from the demo, bounding the time to build components, preceded by negotiating what we *could* do that we wanted to do, preceded by where we were now—understanding anything about the problem, to begin with.

Fairly quickly, major differences emerged in how Prem and Vipin understood politics to work. Vipin expressed technocratic fantasies of a website that could link dispersed Indian experts with state planners and politicians—a kind of 'Innocentive' for the development state, as he described it.[13] Vipin saw the law as a kind of code that sets incentives through punishment; fix the law, fix the nation. Prem, on the other hand, had studied the implementation of the Forest Rights Act and told stories of how the law moved through activists, district officials, and landless adivasis on the ground. The law as text was little match for the contingencies and power plays in which it was invoked. Prem, and many of us with him, did not share Vipin's faith in elite experts in substituting for the politics of the poor.

Prem and Vipin got into a heated debate and many of us sided with Prem. Working with and through Prem's ethnographic cases, our interactions that followed were peppered with the subjunctive: 'What you *could* do' and 'what if we'. Vipin left for a few hours, and taking advantage of his absence, we developed a concept called Jan Sabha, inspired by the Jan Lokpal, that would allow organizers to document face-to-face deliberations of poorer constituencies around central government issues. The hackathon seemed to accommodate more leftist politics. But, Prem warned us, it would require 'some REAL footwork' to get 'on the street' and work with existing organizations thinking in terms of political participation. As the sun sank deeper in the sky, we realized we had little time to reach out to NGOs or activist networks. We had little time to understand their information practices or to build trust with them. We could not even promise maintenance of any demo that came out of a potential collaboration.

That week, we weren't on the street. We were in the studio. The time, tools, and skills in the room were geared towards prototype work, not 'footwork'. Even the kinds of prototype work we could undertake was limited by the political economies of internet production in a country where few had direct access to the internet. Krish, a software engineer, explained to us that in the long term, the project could get into rural areas through interactive voice response phone systems, rural kiosks, or SMS-based systems. 'In Andhra, there's a women's radio station', he told us. 'The scope of what we want to envision is THAT. What we implement in five days is probably a website.' The skills in the room were of the web; web tools were those most at hand for urgent hacking. He continued, 'So we're going to go to a conversation where we'll chop

13 Shortly after Narendra Modi took office as Prime Minister, GoI announced a very similar website called mygov.in. The site called on citizen volunteers to offer 'expert advice' through design competitions and discussion forums. See, 'MyGov: A Platform for Citizen Engagement', https://web.archive.org/web/20141218060431/http://mygov.in/.

off everything. Cut. Cut. Cut. Cut. But if there's a master document that accompanying this chopped up little thing', he trailed off. The hackathon was an experiment in making prototypes of promising projects than dealing with the actual implementation of development work itself.

The next morning, Prem did not come back. While he liked the Jan Sabha idea, he did not trust Vipin to carry it forward faithfully. Vipin hoped to seek funding to carry the project forward from Ford Foundation and World Bank acquaintances. Whatever the politics we read into our demo, the demo would become a vehicle for generating more projects and funding to enrich the design studio, or perhaps the engineer-consultants who were at the hackathon. Jumping forward to today, I can tell you that we showed the demo at the festival and nothing in particular happened with it, but every year or so one of the engineers has written to ask me for mockups so he might build something finally. Hope springs eternal.

The hackathon carried with it a hidden pedagogy that I argue is in common with social enterprise and much design practice. I focus on three here in brief: 'a bias to action', the management of the political, and the elision of infrastructural labors.

The hackathon celebrated 'a bias to action'. This is not just my description, but an actor's category originating in McKinsey consultants Peters and Waterman's work on how to manage corporations in the face of the failures of rational, predictive, linear models.[14] The world, they argued, was one of complexity and rapid change. They advised that managers ought to quickly research, implement, experiment, and learn rather than run into 'analysis paralysis'. The 'bias to action', they advised, made it into job postings not only for the Delhi design studio but even for Google.[15]

To achieve a 'bias to action', politics and conflict had to be managed. Conflict could be useful for generating feedback about risks and opportunities to the project, but it ought not to stop action. Designers often discussed this problem as one of curbing 'talk'. After a particularly long debate, one designer told me, 'Give them lots of water. Lock the doors. They can't leave until they decide how to move forward'. Champions of 'the bias to action' contrasted it with stereotypes of other kinds of Indians: overly intellectual Malayali men who could find 'six sides to a cube', Bengali men in *adda* satisfied to talk deeply, or academics who attuned to political dilemmas over action. Collaborative design meant getting feedback from many kinds of people but not letting the project run aground over the political. The 'bias to action' celebrated by design works because of the kinds of networks, labor configurations, tools, and systems designers can mobilize quickly, extending their agencies out into the world.

This was the third hidden pedagogy: one of relying on hidden infrastructures—the building and maintenance labor of unseen others. The efficacy of hackathons required other labors—24/7 servers, code libraries written and maintained by others, Foxconn workers, and metal mining,

14 Thomas J. Peters and Robert H. Waterman, *In Search of Excellence*, New York: Harper Collins, 2004.

15 Eric Schmidt, Jonathan Rosenberg, and Alan Eagle, 'How Google Attracts the World's Best Talent', *Fortune*, 4 September 2014, http://fortune.com/2014/09/04/how-google-attracts-the-worlds-best-talent/.

for example. These infrastructures were ready to hand but maintained out of sight. As we prototyped a future system, we celebrated the design and the plan—the products of proper 'technological authorship' valorized in regimes the privilege of intellectual property and the creation of new forms.[16] These regimes vilify software pirates and while celebrating patent creators.[17]

At DevDesign, and in cultures of entrepreneurialism, these hidden pedagogies aligned with entrepreneurial citizenship in media beyond the digital.[18] At DevDesign and studios like it, designers similarly developed product design plans at a great distance from the extractive, factory, distributional labors that enabled an idea to actually matter to the masses. Design patents and design labor processes circumscribed moments of intention and form giving as creative. Such regimes took for granted and devalued the labors that make those forms available en masse—the labor of manufacturing workers and craftspeople that reproduce the design.[19] The maintenance and repair, or care, of these systems became an afterthought to the moment of innovation.[20] In the studios I worked, the labor of others—those other than designers—came to matter only when concerns of manufacturability threatened the authorial intentions of the designers and engineers.

These hidden pedagogies added up to an entrepreneurial ethos—one funders, philanthropists, and high-tech managers evangelized to transform civil society's relationship to capitalist development. The World Bank, for example, organized global hackathons to attract programmers—'non-traditional partners'—towards its water and sanitation partners and programs.[21] A bank's white paper on hackathons argued the events could 'orient non-subject matter experts to focus on the low-hanging opportunities'—opportunities for projects that aligned easily with the infrastructures, cultural practices, and institutional agendas of the bank and its allies.[22] Hackathons proliferate in the non-profit sector as a labor process to encourage experimental, digital labor. Participants bring their tacit knowledge, their desires, and even their existing working relations into a space where investors can evaluate and harvest emerging ideas and teams. They draw on the sociability, technical craft, and playfulness of the hacker to speculate in value.[23]

Other hackathons are possible.

16 Kavita Philip, 'What Is a Technological Author? The Pirate Function and Intellectual Property', *Postcolonial Studies* 8.2 (2005): 199.
17 Ibid.; Philip, 'Telling Histories of the Future'.
18 Irani, 'Hackathons and the Making of Entrepreneurial Citizenship'.
19 Arindam Dutta, 'Design: On the Global (R)Uses of a Word', *Design and Culture* 1.2 (2009): 163; Adrian Forty, *Objects of Desire: Design and Society from Wedgewood to IBM*, New York: Pantheon Books, 1986.
20 Steven Jackson, 'Rethinking Repair', in Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (eds) *Media Technologies: Essays on Communication, Materiality, and Society*, Cambridge, Mass.: The MIT Press, 2013, pp. 221–239.
21 World Bank, 'Water Hackathons', p. 7.
22 Ibid., p. 15.
23 Gabriella Coleman, 'Hacker', in Benjamin Peters (ed.) *Digital Keywords: A Vocabulary of Information Society and Culture*, Princeton: Princeton University Press, 2016, pp. 158–172.

Activists have employed hackathons not to proliferate potential, but to sustain and extend collective resources and infrastructures. Open source hackathons operated according to this logic. Programmers came together to care for and extend the platforms and open source libraries that made their relations as a public possible.[24] In 2017, the Aam Aadmi Party proposed a hackathon as a way of testing democracy's infrastructures. They called on the Election Commission to allow experts to hack electronic voting machines in search of vote-tampering vulnerabilities.[25] The party publicized the Commission's refusal to allow machine tampering to generate publicity around election security.[26] In North America, activists also convened hackathons as public provocation and ad hoc labor formation. As the Trump administration took office in the US, North American researchers feared the administration would remove publicly available climate science data. Information activists convened hackathons to scrape and save endangered data through 'guerilla archiving'.[27] Anthropologist Andrea Muehlenbach, describing the event, asked, 'How then do we think of this event not only as a technical meet-up but as a possibility for building a larger and durable transnational public around the anticipation and protection of vulnerable data? We have the technical capacities, but what of the collective energies captured and engendered by this event?'[28] Like the entrepreneurial hackathon, this hackathon gathered people in urgent labor. Yet rather than demos—the promise of technology to come—the gathered people worked to produce archives in the present for common use by others in the future. Through this work, organizers also extended a public and attempted to inculcate in them a 'collective habitus around vigilance'.[29]

Both the Aam Aadmi Party hackathon and climate change hackathons cultivated an anticipatory sociality; they called on people to act on the future by caring for and extending complex, layered networks of digital technologies.[30] They made issues public, whether through party-based social life or work with the press. The hackathon allows organizers to gather and

24 Christopher Kelty, 'Geeks, Social Imaginaries, and Recursive Publics', *Cultural Anthropology* 20.2 (2005): 185.

25 Pankaj Gupta, 'Reply to Dr. Zaidi', May 26, 2017, eci.nic.in/eci_main1/current/ReplyAAP_27052017.pdf.

26 'Aam Aadmi Party to Hold EVM Hackathon on Same Day as Election Commission's Challenge', *The Indian* Express, 1 June 2017, http://indianexpress.com/article/india/aam-aadmi-party-to-hold-evm-hackathon-on-same-day-as-election-commissions-all-party-challenge-4684180/.

27 Andrea Muehlenbach, 'Building an Archive of Vulnerability: #GuerrillaArchiving at #UofT', *EDGI*, 2 January 2017, http://flolab.org/wp19/building-an-archive-of-vulnerability-guerrillaarchiving-at-uoft/.

28 Ibid.

29 Ibid.

30 See, Vincanne Adams, Michelle Murphy, and Adele E. Clarke, 'Anticipation: Technoscience, Life, Affect, Temporality', *Subjectivity* 28.1 (2009): 246; Geeta Patel, 'Risky Subjects: Insurance, Sexuality, and Capital', *Social Text* 24.4 (2006): 25. Adams, Murphy, and Clarke, building on Patel and others, argue that anticipation is a future-oriented 'regime of being in time' equally part of Marxism, decolonization, feminism, but also insurance companies, population management campaigns, and immunization. Institutions attempt to manage futures through techniques of techniques of calculation, socialization, and representation, as well as through hegemony. People might contest and struggle over these futures. Michelle Murphy, co-author of 'Anticipation', also co-organized the climate data archiving hackathon. The hackathon organizers took a technique for proliferating futures under the gaze of corporate sponsorship and venture capital and transformed it into a way of galvanizing people's vigilance in the struggle to fight for land, air, and life.

condense people's labors of care around those infrastructures held more publicly, more in common.

Entrepreneurial value speculation and eventful public care can overlap in regimes of private-public partnership. Civic hacking in the United States builds on a histories of data transparency as activism.[31] And yet, US government agencies also call on citizens' civic sense to hail 'free labor' under regimes of neoliberal fiscal austerity.[32]

Hackathons gather labor—technical, imaginative, communicative. As a vehicle for entrepreneurial citizenship, hackathons transform craft, sociality, and even hope into investable, managed futures. They extract from data and data labors performed and promised elsewhere. As a vehicle of care, however, hackathons might attract people to the often invisible labor of protecting data, expanding access, and sustaining resources that expand the field of political contestation.

31 Andrew R. Schrock, 'Civic Hacking as Data Activism and Advocacy: A History from Publicity to Open Government Data', *New Media & Society* 18.4 (2016): 581.

32 Gregg, 'FCJ-186 Hack for Good'.

References

Adams, Vincanne, Michelle Murphy, and Adele E. Clarke. 'Anticipation: Technoscience, Life, Affect, Temporality', *Subjectivity* 28.1 (2009): 246–265.

'Aam Aadmi Party to Hold EVM Hackathon on Same Day as Election Commission's Challenge', *The Indian* Express, 1 June 2017, http://indianexpress.com/article/india/aam-aadmi-party-to-hold-evm-hackathon-on-same-day-as-election-commissions-all-party-challenge-4684180/.

Amrute, Sareeta Bipin. *Encoding Race, Encoding Class: Indian IT Workers in Berlin*, Durham: Duke University Press, 2016.

Aneesh, Aneesh. *Virtual Migration*, Durham: Duke University Press, 2006.

Coleman, Gabriella. 'The Hacker Conference: A Ritual Condensation and Celebration of a Lifeworld', *Anthropological Quarterly* 83.1 (2010): 47–72.

_____. 'Hacker', in Benjamin Peters (ed.) *Digital Keywords: A Vocabulary of Information Society and Culture*, Princeton: Princeton University Press, 2016, pp. 158–172.

Corbridge, Stuart, and Jonathan Harriss. *Reinventing India: Liberalization, Hindu Nationalism and Popular Democracy*, Cambridge, UK: Polity, 2000.

Cross, Jamie. *Dream Zones: Anticipating Capitalism and Development in India*, London: Pluto Press, 2014.

Dutta, Arindam. 'Design: On the Global (R)uses of a Word', *Design and Culture* 1.2 (2009): 163–186.

Fattal, Alex. 'Facebook: Corporate Hackers, a Billion Users, and the Geo-Politics of the 'Social Graph'', *Anthropological Quarterly* 85.3 (2012): 927–956.

Forty, Adrian. *Objects of Desire: Design and Society from Wedgewood to IBM*, New York: Pantheon Books, 1986.

Gregg, Melissa. 'FCJ-186 Hack for Good: Speculative Labour, App Development and the Burden of Austerity'. *The Fibreculture Journal* 25 (2015): Apps and Affect, http://twentyfive.fibreculturejournal.org/fcj-186-hack-for-good-speculative-labour-app-development-and-the-burden-of-austerity/.

Gupta, Pankaj. 'Reply to Dr. Zaidi', 26 May 2017, eci.nic.in/eci_main1/current/ReplyAAP_27052017.pdf.

Huang, Andrew. *Diary of Facebook*, Documentary, Biography, 2011, http://www.imdb.com/title/tt1882342/.

Irani, Lilly. 'Hackathons and the Making of Entrepreneurial Citizenship', *Science, Technology, & Human Values* 40.5 (2015): 799–824.

_____. *Chasing Innovation: Making Entrepreneurial Citizens in Modern India*, Princeton: Princeton University Press, 2019.

Jackson, Steven. 'Rethinking Repair', in Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot (eds) *Media Technologies: Essays on Communication, Materiality, and Society*, Cambridge, Mass.: The MIT Press, 2013, pp. 221–239.

Kelty, Christopher. 'Geeks, Social Imaginaries, and Recursive Publics', *Cultural Anthropology* 20.2 (2005): 185–214.

Kohli, Atul. 'Politics of Economic Growth in India, 1980-2005: Part I: The 1980s', *Economic and Political Weekly* 41.13 (2006): 1251–1259.

Muehlenbach, Andrea. 'Building an Archive of Vulnerability: #GuerrillaArchiving at #UofT', *EDGI*, 2 January 2017, http://flolab.org/wp19/building-an-archive-of-vulnerability-guerrillaarchiving-at-uoft/.

Nadeem, Shehzad. 'Macaulay's (Cyber) Children: The Cultural Politics of Outsourcing in India', *Cultural Sociology* 3.1 (2009): 102–122.

Patel, Geeta. 'Risky Subjects: Insurance, Sexuality, and Capital', *Social Text* 24.4 (2006): 25–65.

Peters, Thomas J., and Robert H. Waterman. *In Search of Excellence*, New York: Harper Collins, 2004.

Philip, Kavita. 'What is a Technological Author? The Pirate Function and Intellectual Property', *Postcolonial Studies* 8.2 (2005), pp. 199–218.

_____. 'Telling Histories of the Future: The Imaginaries of Indian Technoscience', *Identities* 23.3 (2016): 276–293.

Prahalad, C. K. *The Fortune at the Bottom of the Pyramid*, Pearson Prentice Hall, 2006.

Rajagopal, Arvind. 'The Emergency as Prehistory of the New Indian Middle Class', *Modern Asian Studies* 45.5 (2011): 1003–1049.

Roy, Srirupa. *Beyond Belief: India and the Politics of Postcolonial Nationalism*, Durham: Duke University Press, 2007.

Sarkar, Sreela. 'Passionate Producers: Corporate Interventions in Expanding the Promise of the Information Society', *Communication, Culture & Critique* 10.2 (2017): 241–260.

Schrock, Andrew R. 'Civic Hacking as Data Activism and Advocacy: A History from Publicity to Open Government Data', *New Media & Society* 18.4 (2016): 581–599.

Schmidt, Eric, Jonathan Rosenberg, and Alan Eagle. 'How Google Attracts the World's Best Talent', *Fortune*, 4 September 2014, http://fortune.com/2014/09/04/how-google-attracts-the-worlds-best-talent/.

Subramanian, Ajantha. 'Making Merit: The Indian Institutes of Technology and the Social Life of Caste', *Comparative Studies in Society and History* 57.2 (2015): 291–322.

Vora, Kalindi. *Life Support: Biocapital and the New History of Outsourced Labor*, Minneapolis: University of Minnesota Press, 2015.

Wilson, Dominic, and Roopa Purushothaman. 'Dreaming with BRICs: The Path to 2050', *Global Economics Paper* 99 (2003): 1–24.

World Bank. 'Water Hackathons: Lessons Learned', Water Papers, Washington, D.C.: World Bank, May 2012.

07. REPORTING THE WORLD'S LARGEST BIOMETRIC PROJECT

ANUMEHA YADAV

Rajesh Kumar, his head bent over the screen of his laptop, tried again to connect to the internet. Behind him, a long queue of people waited at the panchayat office in Dohakatu in Jharkhand's Ramgarh district. Kumar had to submit bank account forms online for rural workers that week, but he ran into many interruptions. 'The line [power] failed and came back only in the afternoon', he said. 'Last week, there was a power cut for two days. When the electricity line works, the server line disappears'.

Kumar, the banking correspondent in Dohakatu, was a private agent contracted by a public bank to deliver financial as well as internet services in the village. Two months earlier, local officials had asked him to pay rural workers through a hand-held device, a micro ATM, after verifying their details in a new system using Aadhaar, a biometrics-based identity number.

That afternoon, seven workers waited to receive their wage payments. While three workers successfully placed their fingertips on the machine and collected their wages, the machine did not recognize four of them.

Dashay Bediya, an elderly Adivasi farmworker in a white shirt and dhoti, was among those whose fingerprints were repeatedly rejected by the machine. Bediya tried eight times, placing different fingers on the small screen, hoping that one would work, and then went outside the office and scrubbed his weathered hands. He came back in and made five more attempts, getting more anxious and disappointed each time.

Kumar examined the machine. When that did not work, he advised Bediya: 'Put Vaseline or Boroplus and rub your fingers before you go to sleep', he said. 'Come after three-four days, and try again'. The elderly man went back without collecting his wages that day.

Since its launch, Aadhaar has been presented by the government as a scheme for the benefit of India's 'indigent and the marginalized'. Bureaucrats heading the Unique Identification Authority of India (UIDAI), the agency that issues Aadhaar identity numbers and manages the database, said a biometrics-linked number would allow payments to happen at 'door-steps' of beneficiaries through a network of banking correspondents like Dohakatu's Rajesh Kumar. It would especially help migrants and farmworkers by providing them an identification document they could access anytime.

In 2011, Ramgarh became one of the fifty-one districts of a total of over six hundred districts selected as pilot districts for use of Aadhaar in social schemes. As the Jharkhand state reporter for a national newspaper, I traveled to several districts, starting with Ramgarh, to document the effects. In Dohakatu, that December afternoon, watching Bediya trying to prove who he

was through his weathered hands but failing again and again, it was a first inkling that in reality, this shift was not going to be as smooth as the government was making it seem.[1]

Over the next five years, I interviewed several residents, first in Jharkhand and Rajasthan, and then in other states where farmworkers, quarry and construction workers, and the elderly described the coercion on them and the hardship of linking their existing government benefits to the biometrics scheme.

Aadhaar authentication relies on the use of biometrics, a sophisticated technology that requires several other technologies—computer, the internet, agencies servers' capacities, electricity, biometric devices, and even one's physical features—to work at the same time. When even one failed, it disrupted some of the poorest citizens' access to essential services.

Senior officials in Jharkhand and other states and even in Delhi described the difficulties stemming from the absence of infrastructure and biometric rejection faced by manual laborers as 'teething problems' as the Aadhaar project was still in its infancy. They argued that any public service delivery program would have some errors and exceptions, but to question biometrics technology would be to 'throw the baby out with the bathwater'.

In 2012, in Ramgarh, the bank agent could be seen prescribing skin-softening creams. Years on, more of this continued. In 2015, in Latehar, another district in Jharkhand, I interviewed officials who prescribed cleaning hands with flour and lemon juice to rural residents to pass the test of the biometric machines.

The government had stated repeatedly to the legislature and in courts that Aadhaar enrollment was completely voluntary. At the district level, however, citizens were presented with the choice of either enrolling or going without essential services. Some of the poorest citizens were cut off from welfare schemes and pensioners were even wrongly declared dead if they failed to enroll in the scheme.[2] In just four years, in an inversion of the initial promise of Aadhaar as an enabler, the quality of one's biometrics and enrolling in the database seemed to have in practice become the new basis to be able to access any public services, with Aadhaar becoming one more proof to be provided to qualify for any scheme.

The Promise of 'Welfare Delivery'

To get an Aadhaar number, residents have to submit biometrics and demographic information to private enrolling agencies hired by the government. A resident who does not possess any proof of identity or proof of address can enroll and get an Aadhaar by being introduced by a designated 'introducer'

1 Anumeha Yadav, 'To Pass Biometric Identification, Apply Vaseline or Boroplus on Fingers Overnight', *The Hindu*, 15 December 2012, https://www.thehindu.com/todays-paper/tp-opinion/to-pass-biometric-identification-apply-vaseline-or-boroplus-on-fingers-overnight/article4202157.ece.

2 Anumeha Yadav, 'Rajasthan's Living Dead: Thousands of Pensioners Without Aadhaar or Bank Accounts Struck Off Lists', *Scroll.in*, 6 August 2016, https://scroll.in/article/813132/rajasthans-living-dead-thousands-of-pensioners-without-aadhaar-or-bank-accounts-struck-off-lists.

.The government claimed Aadhaar would streamline administration in three ways. One, every time a person enrolls in Aadhaar, his/her biometrics and demographic details are compared with the other data previously recorded in the Aadhaar database. This is meant to identify duplicate entries. The second use is that a unique Aadhaar number is added to—or 'seeded in'—an existing database. Banks also link the Aadhaar number to the bank account of account holders and report this to the National Payments Corporation of India (NPCI).[3] The third proposed use was 'doorstep delivery' of benefits.

The promise that Aadhaar will provide identification to those without any identification document through the 'introducer system' turned out to be inaccurate. More than 99 percent of residents obtained Aadhaar after showing an existing proof of identity and address.[4] Having an Aadhaar did not automatically enable eligibility for social schemes as government welfare programs continued to have additional requirements.[5]

The promise of 'doorstep delivery' with authentication done by local banks agents ran into many problems—poor internet connectivity, non-upgradation of the banks to the new technology, insufficient numbers of new banking agents or tablets, and other 'technical glitches' and logistical difficulties. When doorstep delivery threw up multiple issues, the government focused in the initial years on the first two processes, enrolling residents, and comparing welfare databases against the Aadhaar database.[6]

Local authorities instructed beneficiaries to mandatorily submit their Aadhaar numbers to access welfare services. As officials compared the demographic details in two databases—the names and residences in the welfare scheme list and the demographic information collected by the UIDAI—to check if the two matched, the discrepancies that became apparent were described as 'ghosts' or 'fake claimants', people who did not exist or who had duplicate cards. But those who missed enrollment, or those who did not know about new requirements, or those simply not interested in enrolling into the database were simply struck off social registries.[7]

Early Signs

The United Progressive Alliance (UPA) under Manmohan Singh selected Jharkhand to introduce Aadhaar-linked cash transfers to pay workers in the Mahatma Gandhi National Rural

3 Anumeha Yadav, 'No Benefits for Beneficiaries', The Hindu, 6 March 2014, http://www.thehindu.com/opinion/lead/no-benefits-for-beneficiaries/article5753965.ece.

4 The Wire Staff, 'Most Aadhaar Cards Issued to Those Who Already Have IDs', *The Wire*, 3 June 2015, https://thewire.in/law/most-aadhar-cards-issued-to-those-who-already-have-ids.

5 Reetika Khera, 'UID: From Inclusion to Exclusion', *Seminar* 672 (August 2015), http://india-seminar.com/2015/672/672_reetika_khera.htm.

6 In-person interview with UIDAI officials in Ranchi and New Delhi.

7 Yadav, 'No Benefits for Beneficiaries'; Anumeha Yadav, 'No Aadhaar, No Scholarship to Jharkhand SC, ST Students', *The Hindu*, 8 October 2013, https://www.thehindu.com/news/national/no-aadhaar-no-scholarship-to-jharkhand-sc-st-students/article5213382.ece; Jean Drèze, 'Following the Grain Trail: On India's Public Distribution System', *The Hindu*, 16 January 2018, https://www.thehindu.com/opinion/lead/following-the-grain-trail/article22451645.ece.

Employment Guarantee Act (MGNREGA) scheme, which assures 100 days of work in a year to any rural household willing to do manual labor. In December 2011, officials said that over the next one year, they planned to pay 174,000 workers through bank accounts newly linked with Aadhaar. A year on, when I interviewed officials in the UIDAI regional office in Ranchi, they had paid a little over 5,000 workers through Aadhaar-linked bank accounts, less than 3 percent of their target.[8]

The pilot had not scaled as planned because of a range of reasons. The district collector of Ramgarh, Amitabh Kaushal, who had won the 'National Aadhaar Governance Award' in 2012 told me that the district administration's capacity was under strain. In many villages, people had not shown any interest in enrolling for Aadhaar, and in some places where they had, there were not enough bank branches and agents, said Kaushal. He also expressed concerns about whether even the existing banking correspondents would be safe carrying large amounts of cash to pay to workers while Jharkhand was witnessing an armed conflict between the paramilitaries and Maoists insurgents in its forested districts. Banking correspondents in turn told me that while they had been telecast making payments to MGNREGA workers when the prime minister Manmohan Singh had inaugurated the Aadhaar project on October 20, 2012, they were not paid wages for six months after the inauguration.[9] The state-level bankers' committee officials in Ranchi told me there were 'technical errors' because of which transactions did not reflect, so invoices could not be prepared, so the agents' salaries were kept pending.

When the pilots for Aadhaar payments had failed to scale, it was not clear what to expect in terms of benefits for the beneficiaries. But in 2013, despite no clear evidence of its benefits to people, there was a renewed push in the state to enroll people in Aadhaar and expand the experiment in social schemes. This coincided with the transfer of Ram Sewak Sharma, the former director-general of UIDAI and the second-highest-ranking official after the founder Nandan Nilekani in the agency till then, as the new Chief Secretary of Jharkhand in April 2013. Now, as the highest-ranking bureaucrat in the state, Sharma closely monitored Aadhaar linking in social schemes. He also started a new application of Aadhaar, launching the first Aadhaar-enabled back-end attendance system for state secretariat employees.

During the pilots earlier, the district officials had spoken of missing Aadhaar enrollment and bank account-linking 'targets' because of uneven infrastructure, patchy bank network, and irregular payments to banking agents. In 2013 and 2014, district officials described tremendous pressure from the top to show 100 percent Aadhaar 'seeding', and this was coupled with the fear of administrative action if they failed to do so. In some instances, to showcase Aadhaar enrollment and linking under their jurisdictions as 100 percent, local officials even resorted to removing beneficiaries from welfare schemes that were their legal entitlements.

8 Anumeha Yadav, 'Direct Benefits Transfer: Why Direct Transfer May Not Put Money in People's Pockets', *The Hindu*, 15 December 2012, https://www.thehindu.com/news/national/direct-benefits-transfer-why-direct-transfer-may-not-put-money-in-peoples-pockets/article4200661.ece.

9 Yadav, 'No Benefits for Beneficiaries'.

In Khunti, a predominantly Adivasi district 40 kilometers from the state capital Ranchi, social activists expressed concerns that the most vulnerable Scheduled Tribe beneficiaries in the interior villages with less information and poor connectivity were the first to fall through the cracks.

One of the primary aims of MGNREGA is to reduce distress migration by rural poor in non-farming months, by providing them the choice to work in their villages for around three months in a year. Khunti is wrecked by the conflict between Maoist insurgent groups and the government, and witnesses a regular stream of distress migration.

When residents of all villages had still not enrolled in Aadhaar more than a year after the first Aadhaar pilots had ended, the Khunti administration began 'deleting' their MGNREGA job-cards. On January 25, 2014, Khunti's district collector Mukesh Kumar wrote to all local officials that they would be asked to explain if they failed to show 100 percent Aadhaar 'seeding' of all those who held MGNREGA job-cards. Junior officials, in turn, stopped salary payments to panchayat *sewak* and *rozgar sewak*, the village and scheme-level functionaries, till they showed 100 percent adoption of Aadhaar.

Asked why he had given these instructions, Kumar described Aadhaar seeding as *pavitra karya* (sacred work), as the administration had 'no ulterior motive' in it. He described how he had set up a district 'control room' especially for Aadhaar and hired local private computer operators to 'seed' Aadhaar in all databases when the banks acted tardily.[10] On paper, who was a 'real', or genuine, beneficiary in the job schemes was to be determined after holding *gram sabhas* (public hearings). But this was seldom done.

In one instance, the staff deleted 2,211 workers' job-cards, while 11,234 workers' were 'tagged as deleted'—that is, these job-cards were marked as deleted but could be used if the worker applied afresh for work with proof of Aadhaar enrollment. In another case in Tirla village in Khunti, twenty-two workers had done land-leveling work, and all but three workers who had not enrolled in Aadhaar received their wages.

The administration had not laid down any formal processes for those whose payments and benefits were disrupted in the hasty transition to Aadhaar. It was only after these three workers from Khunti, with the help of activists, submitted an affidavit in the Supreme Court on the non-payment of their wages because of not enrolling in Aadhaar did the state machinery spring into action, with the chief secretary Sharma personally clarifying that there was no instruction to mandate rural workers seeking work in MGNREGA to enroll in or produce Aadhaar.

Following this incident, R Subramanyam, the then Joint Secretary, Ministry of Rural Development, which administers the rural employment guarantee scheme, issued formal instructions from Delhi that no worker should be deprived of the legal entitlement to work for not having an Aadhaar number.

10 Ibid.

As the Jharkhand state reporter of a national newspaper, I was covering a range of subjects such as political economy, resource use, and the ongoing Maoist conflict. Though I was not in a position to say whether central agencies were especially sensitive to any ground critique of the Aadhaar project or whether it was the Chief Secretary's personal enthusiasm for the biometrics project, I was surprised when I received direct clarifications twice from the Chief Secretary, the highest bureaucrat in the state, after I reported on beneficiaries getting cut off in Aadhaar linking and the glitches that were appearing. This happened immediately after I reported on the exclusion of workers in MGNREGA scheme and how Adivasi children were losing scholarships as they had not enrolled in the biometrics database.[11] On the ground, the pressure from local authorities on beneficiaries to enroll in Aadhaar continued, but in interviews, the senior officials responded that Aadhaar was not mandatory in any scheme.

The Legal Framework

While the Aadhaar project's promise was to enable the poor's legal rights, the project was run without a legal framework for six years. The UPA government introduced the National Identification Authority of India Bill in 2010. The parliamentary standing committee on finance under the Bharatiya Janta Party member of parliament Yashwant Sinha rejected the Bill in December 2011. It stated the 'collection of biometric information and its linkage with personal information of individuals' without amending the Citizenship Act appeared 'to be beyond the scope of subordinate legislation, which needs to be examined in detail by Parliament'. The committee also referred to the experience of the identity project for a national biometrics ID that the British government had dismantled, citing potential risk to public interest and the legal rights of its citizens.[12]

The UPA government, however, had continued to enroll residents into the Aadhaar database without reframing the Bill or initiating any further debate on it in the parliament. In 2012, Justice KS Puttaswamy challenged the Aadhaar project in the Supreme Court of India. The Supreme Court passed three orders between 2013 and 2015 that the state cannot make Aadhaar a pre-condition for accessing any public services (this also explained why officials in Jharkhand did not wish to be seen as if they were compelling people to enroll in Aadhaar).

In 2015, I moved to New Delhi for reporting. While writing an explainer on the legal status and policy around Aadhaar, I interviewed several central government officials who justified the haste to increase enrollment in Aadhaar and link more and more things to it. On paper, Aadhaar was still voluntary, but an advisor to government who worked on implementing Aadhaar explained that 'the system would work best if everyone is enrolled, with their details seeded at the time of enrollment, even if this took

11 Yadav, 'No Aadhaar, No Scholarship to Jharkhand SC, ST Students'.

12 Standing Committee on Finance (2011–12), Ministry of Planning, *The National Identification Authority of India Bill, 2010*, Report no. 42, New Delhi: Lok Sabha Secretariat, December 2011.

a few years'. His reason was that, over time, maintaining two lists of beneficiaries one with Aadhaar and other without it would be an administrative hassle. Also, it was necessary to get every resident on the database to improve the 'deduplication' efficiency. 'In the future, if more and more people without Aadhaar started appearing in the database, it will make it harder to authenticate who they are', he added.

There is a vast difference between a voluntary identity scheme and a compulsory national ID. But by presenting the scheme as a voluntary facility, the UPA government had skirted serious debate and questions over the creation and regulation of a vast infrastructure of social control and the rights of those for whom it was making it compulsory in practice to enroll in it. UIDAI officials' position was that the agency would 'not mandate Aadhaar', that they 'provided just a number, and it is up to various government agencies what they do with it'.

However, in its policy documents, UIDAI was explicitly arguing that linking Aadhaar to welfare schemes would help increase enrollment numbers. 'Since de-duplication in the UIDAI system ensures that residents have only one chance to be in the database, individuals are made to provide accurate data. This incentive will become especially powerful as benefits and entitlements are linked to Aadhaar'.[13] As more essential schemes were linked to Aadhaar, these schemes would serve as a 'killer application' to boost enrollment, the UIDAI argued.[14]

By July 2015, 87 crore, that is, 72 percent of India's population, and over 90 percent of adults, had been enrolled in the biometrics database. By asking the beneficiaries of just two schemes, the public distribution system, which provides food subsidies to 85 crore people, and the MGNREGA workers, to produce Aadhaar to continue getting their legal benefits, the Aadhaar project had managed to cover two-thirds of the country's population.

In May 2014, the National Democratic Alliance (NDA) government under Prime Minister Narendra Modi came to power at the center. In a court hearing in October 2015, the new government also defended the Aadhaar project like its predecessor. But it no longer claimed Aadhaar was 'voluntary'. Instead the NDA government asked the Supreme Court to allow it to make Aadhaar mandatory in around 80 social schemes. It even claimed that Aadhaar had become indispensable to welfare delivery and if the court restricted the project at this stage, it would disrupt wages of one crore workers under MGNREGA, besides pensioners' payments, when this was simply not true. Ministry of Finance data showed that though crores of workers had been enrolled in the database, more than 98 percent of payments were still happening as simple bank transfers, which did not require Aadhaar per se.

After this hearing, the court allowed the government the voluntary use of Aadhaar in MGNREGA and pension payments, but refused to allow the Aadhaar number's extended use as a

13 UIDAI Strategy Overview, 'Creating a Unique Identity Number for Every Resident in India', *UAIDAI*, 2010, http://www.prsindia.org/uploads/media/UID/UIDAI%20STRATEGY%20OVERVIEW.pdf.

14 Ibid. See also, Usha Ramanathan, 'Enrolment Saga', *Frontline* 28.24 (2011), https://www.frontline.in/static/html/fl2824/stories/20111202282402200.htm; Mohan Rao, 'False Promises', *Frontline* 28.24 (2011), https://www.frontline.in/static/html/fl2824/stories/20111202282401900.htm.

mandatory identity document in a range of schemes and asked a larger bench to define the right to privacy in India

.A senior official in the Direct Benefit Transfer department, then under the Ministry of Finance, expressed dissatisfaction at the court orders: 'If you are a public scheme beneficiary, I have the right to ask for your digital identity', he argued. He was dismissive of the Constitutional challenge to the project: 'If people can give thumbprints when they cannot sign a document, what is wrong if the government asks for the same thumbprint digitally?' he argued. 'Those who have emotional problems with submitting biometrics should be willing to forgo any digital government facilities then'.[15]

Six months after the Supreme Court restricted the use of Aadhaar as voluntary and limited to six schemes, the NDA government introduced the Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits, and Services) Bill in the parliament in March 2016. Section 7 of the Bill gave the government sweeping powers to require Aadhaar for a wide range of services, birth and death registrations, railways, telecommunication, and digital payments.

By introducing it as a money bill, the government managed to avoid debate in the Upper House of the parliament, where it lacked a majority. It rushed the Aadhaar law through the parliament in less than two weeks.

Lack of Transparency

One of the main claims of Aadhaar was that Aadhaar-based authentication and Management Information Systems would bring transparency to the opaque systems in existence. An initial blueprint on Aadhaar in welfare delivery on the public distribution system stated: 'Clear accountability through Aadhaar authentication, as well as the use of electronic records, would make data more available for community monitoring, and would strengthen the use of right to information in the public distribution system'. It added that an Aadhaar-enabled information technology grievance system 'would ensure that complaints are visible publicly and across different levels of government'.

But there was no way to tell if this was true. The UIDAI refused to make authentication failures rates public, make the macro-data available online, or even share these in the right to information (RTI) requests

.To an RTI request I filed on February 8, 2017, the authority stated it had received 3,310 million authentication requests between September 2012 and October 2016, but it refused to share how many of these requests had failed or succeeded, stating that this data 'is not readily available'.[16]

15 Interview in the Ministry of Finance.

16 Anumeha Yadav, 'How Efficient is Aadhaar? There's No Way to Know Since the Government Won't Tell', *Scroll.in*, 5 April 2017, https://scroll.in/article/833060/how-efficient-is-aadhaar-theres-no-way-to-know-as-the-government-wont-tell.

In interviews, UIDAI officials posted in state capitals in Rajasthan and Jharkhand stated they could not share information on how many transactions were failing or how many had required multiple attempts since there were 'technical issues' in deriving the data.

Experimentation on More Social Schemes

The claims of efficiency and convenience through the use of Aadhaar did not reflect on the ground. Across states, Aadhaar was disrupting welfare delivery and causing distress to the poor.

Its major claims on welfare delivery and efficiency had not come true across schemes. In Rajasthan, Jharkhand, Chhattisgarh, and Gujarat, there was little evidence that centralized biometrics schemes served local needs.[17]

Coercion and bureaucratic procedure reduced the need to communicate with people. States continued to evade questions around accountability. The technology did not present a significant challenge to local power structures or social inequalities. First in the absence of a legal framework on Aadhaar, and later, even after the Aadhaar Act was passed, residents were left without any effective grievances redress mechanisms.

In New Delhi, the NDA government, however, announced that it was ready to start linking Aadhaar to more welfare facilities, with health records being the next major social scheme.

References

17 Anumeha Yadav, 'Chhattisgarh's Way of Dealing with Aadhaar: When Fingerprints Fail, Take Photos', *Scroll.in*, 21 December 2016, https://scroll.in/article/822764/chhattisgarhs-way-of-dealing-with-aadhaar-when-fingerprints-fail-take-photos; Anumeha Yadav, 'Gujarat Matches the Fingerprints of the Poor Before Giving Them Food. Does the System Work?', *Scroll.in*, 16 December 2016, https://scroll.in/article/821651/gujarat-matches-the-fingerprints-of-the-poor-before-giving-them-food-does-the-system-work.

Drèze, Jean. 'Following the Grain Trail: On India's Public Distribution System', *The Hindu*, 16 January 2018, https://www.thehindu.com/opinion/lead/following-the-grain-trail/article22451645.ece.

Khera, Reetika. 'UID: From Inclusion to Exclusion', *Seminar* 672 (August 2015), http://india-seminar.com/2015/672/672_reetika_khera.htm.

Ramanathan, Usha. 'Enrolment Saga', *Frontline* 28.24 (2011), https://www.frontline.in/static/html/fl2824/stories/20111202282402200.htm.

Rao, Mohan. 'False Promises', *Frontline* 28.24 (2011), https://www.frontline.in/static/html/fl2824/stories/20111202282401900.htm.

The Wire Staff. 'Most Aadhaar Cards Issued to Those Who Already Have IDs', *The Wire*, 3 June 2015, https://thewire.in/law/most-aadhar-cards-issued-to-those-who-already-have-ids.

UIDAI Strategy Overview. 'Creating a Unique Identity Number for Every Resident in India', *UAIDAI*, 2010, http://www.prsindia.org/uploads/media/UID/UIDAI%20STRATEGY%20OVERVIEW.pdf.

Yadav, Anumeha. 'Chhattisgarh's Way of Dealing with Aadhaar: When Fingerprints Fail, Take Photos', *Scroll.in*, 21 December 2016, https://scroll.in/article/822764/chhattisgarhs-way-of-dealing-with-aadhaar-when-fingerprints-fail-take-photos.

_____. 'Direct Benefits Transfer: Why Direct Transfer May Not Put Money in People's Pockets', *The Hindu*, 15 December 2012, https://www.thehindu.com/news/national/direct-benefits-transfer-why-direct-transfer-may-not-put-money-in-peoples-pockets/article4200661.ece.

_____. 'Gujarat Matches the Fingerprints of the Poor Before Giving Them Food. Does the System Work?', *Scroll.in*, 16 December 2016, https://scroll.in/article/821651/gujarat-matches-the-fingerprints-of-the-poor-before-giving-them-food-does-the-system-work.

_____. 'How Efficient is Aadhaar? There's No Way to Know Since the Government Won't Tell', *Scroll.in*, 5 April 2017, https://scroll.in/article/833060/how-efficient-is-aadhaar-theres-no-way-to-know-as-the-government-wont-tell.

_____. 'No Benefits for Beneficiaries', *The Hindu*, 6 March 2014, http://www.thehindu.com/opinion/lead/no-benefits-for-beneficiaries/article5753965.ece.

_____. 'No Aadhaar, No Scholarship to Jharkhand SC, ST Students', *The Hindu*, 8 October 2013, https://www.thehindu.com/news/national/no-aadhaar-no-scholarship-to-jharkhand-sc-st-students/article5213382.ece.

_____. 'Rajasthan's Living Dead: Thousands of Pensioners Without Aadhaar or Bank Accounts Struck Off Lists', *Scroll.in*, 6 August 2016, https://scroll.in/article/813132/rajasthans-living-dead-thousands-of-pensioners-without-aadhaar-or-bank-accounts-struck-off-lists.

_____. 'To Pass Biometric Identification, Apply Vaseline or Boroplus on Fingers Overnight', *The Hindu*, 15 December 2012, https://www.thehindu.com/todays-paper/tp-opinion/to-pass-biometric-identification-apply-vaseline-or-boroplus-on-fingers-overnight/article4202157.ece.

08. BROKEN DATA: REPAIRS IN THE PRODUCTION OF BIOMETRIC BODIES

PREETI MUDLIAR

The nature of data lends itself to many actions. Among other things, it is created, recorded, protected, circulated, deleted, duplicated, and shared. When digitized, data is endowed with even more attributes. For governments, it becomes a way of claiming efficiency, streamlining functions, and logging and tracking transactions. Data in its digitized form is thus a means to enforce a seamless, stable, and electronic infrastructural discipline and capture information about people, their identities, and their transactions. However, even as digitized data is valued for its precise tidiness and orderliness, it is also equally susceptible to errors and omissions. This makes it essential to interrogate the disruption that surrounds the repair and maintenance of data infrastructures.

Paul Edwards et al. argue that over the past couple of decades, digital or e-infrastructures have fast scaled to a point where they have started resembling 'genuine infrastructures' such as railroads and telephone networks for their robust reliability as providers of essential services.[1] They note that these infrastructures are often built to order for governments or firms and span a wide range of services, national contexts, and information environments. Aadhaar, the largest biometric database in the world, is a similar case in point. It aids the Indian state's quest to eliminate corruption in the delivery of social welfare programs, which increasingly finds its solutions in the creation of digital data infrastructures. It is claimed that Aadhaar enables the administration to authenticate beneficiaries' identities and their transactions, thus weeding out phantom claimants to entitlements. The growing list of authentication failures in Aadhaar-linked schemes, however, demand a closer understanding of breakdowns in the lives of data.

Through a month-long fieldwork in March 2017, conducted a year after Aadhaar-linked public distribution system (PDS) commenced in Ajmer district in Rajasthan, I interrogate what it means for beneficiaries to experience breakdowns in biometric authentication and thus their food security supplies. Although these breakdowns occur for various reasons such as poor internet connectivity, database servers experiencing downtime, and malfunctioning of point of sale (PoS) machines, the scope of this essay is limited to reflect on what happens to those who fail in authenticating their biometric data.[2] Steven Jackson urges attention to the moral and ethical nature of repair as one

1 Paul N. Edwards, Geoffrey C. Bowker, Steven J. Jackson, and Robin Williams, 'Introduction: An Agenda for Infrastructure Studies', *Journal of the Association for Information Systems* 10.5 (2009): 364.

2 For instance, seeding Aadhaar numbers in the PDS project led to various entry errors, resulting in duplication of ration cards. The duplication when discovered at the time of collecting supplies from

that offers care, solidarity, and responsibility for restoring order.[3] He argues that the reductive functionalism that is used to address matters of technology takes on an ethical and moral dimension of care through which repair is accomplished. Therefore, in this essay, I ask how biometric failures for PDS are received and who administers care and healing. I contribute to ways of thinking about data infrastructures by drawing attention to how people encounter their bodies' failures in finding a biometric match with their stored data and what they do to repair these disruptions. While some forms of data are recognized as big data, how do we begin acknowledging people and their actions when databases tell them that they are broken data?

Broken World Thinking

Breakdowns in biometric systems are routine as observed by Shoshana Magnet, who strongly contests the notion that biometric data are a reliable indicator of identity.[4] She argues that real world deployment of biometric data for authentication is contingent on practices that are assumed to be transparent and reliable but are actually ambiguous and dependent upon inscription and interpretation. We can see Magnet's claims play out in the case of how the Aadhaar database was built. For instance, Johri and Srinivasan illustrate that while the quality of the data was of primary importance to the Aadhaar design team, this did not always find resonance in the way enrollment agents were collecting biometric data, given that the agents' remuneration was linked to the number of enrollments they secured.[5] In contrast to the motives of the design team, the enrolling agents were prioritizing the quantity of the data they collected over its quality. The differing motivations of the design team and the enrollment agents is just one instance of how errors and situations for breakdowns were being introduced into the system, courtesy fraudulent, duplicate, incomplete, or incorrect entries. Not

ration shops not only led to a deletion of all the ration cards from the database but also a denial of supplies to the beneficiary. Matching the ration card details in the PDS database to the biometric details in the Aadhaar database were often unsuccessful owing to different conventions that were adopted while writing names on ration cards that varied from the way the name was reported in the Aadhaar database. Further, finding matches between the two databases was also sometimes rendered problematic since the PDS database was primarily a record of family units, while Aadhaar enumerates individual biometric records. In addition, infrastructural challenges such as poor real time internet connectivity stall the seamless functioning of point of sale (PoS) machines. See, Jean Drèze, 'Dark Clouds Over the PDS', *The Hindu,* 10 September 2016, http://www.thehindu.com/opinion/lead/Dark-clouds-over-the-PDS/article14631030.ece. Drèze observes that the success of using the Aadhaar database as the sole authenticating factor for the distribution of food grains is heavily dependent on 'multiple fragile technologies working at the same time' such as the PoS machine, the biometrics, the internet connection, and remote servers that allow databases to authenticate identities.

3 Steven J. Jackson, 'Rethinking Repair', in Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot. *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, Mass.: The MIT Press, 2014, pp. 221–239.

4 Shoshana Magnet, *When Biometrics Fail: Gender, Race, and the Technology of Identity*, Durham: Duke University Press, 2011.

5 Aditya Johri and Janaki Srinivasan, 'The Role of Data in Aligning the 'Unique Identity' Infrastructure in India', in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, Baltimore, Maryland, USA: ACM Press, 2014, pp. 697–709.

only did this contribute to compromising the efficacy of the database itself, but it also set the stage for fault lines, gaps, and breakdowns that would return to confront people dependent on the database working as promised.

Even as the Indian state persists with its proclivity to create digital data infrastructures as a way to efficiently tame unruly processes of governance, it exhibits a curious indifference in confronting and accounting for eventualities of errors and the fractures that surround digital data. Here, I engage with what Jackson terms as 'broken world thinking' to think through processes and acts of doing and coping by people who find themselves confronted with their new-found status as pieces of broken data.[6] In the words of Susan Leigh Star, they are akin to 'orphans of infrastructure' who are rendered residual by a system with little recourse or assurance about how best to array their body's biometrics back as authentic matches to restore their disrupted social order.[7]

Jackson and Kang note that acts of repair call upon people to change, learn, and adjust to dysfunction.[8] These actions in turn unearth hidden features of social life that were hitherto unnoticeable when functioning. Although undertheorized and less visible than technology innovation, engagement with repair, maintenance, breakdowns, reuse, and repurposing of technology artefacts has been addressed in human computer interaction (HCI) literature.[9] However, lesser known is how people choose

6 Jackson, 'Rethinking Repair'.

7 Susan Leigh Star, 'Orphans of Infrastructure: A New Point of Departure', in Ann Light (ed.) *The Future of Computing: Visions and Reflections,* Oxford: UK, Oxford Internet Institute, 2007, https://www.oii.ox.ac.uk/archive/downloads/publications/FD11.pdf.

8 Steven Jackson and Laewoo Kang, 'Breakdown, Obsolescence and Reuse: HCI and the Art of Repair', in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*, Toronto, Ontario, Canada: ACM Press, 2014, pp. 449–458.

9 Work on repair and maintenance is found in the context of physical artefacts such as mobile phone repairs and people engaging in repair work. See, Julian E. Orr, *Talking about Machines: An Ethnography of a Modern Job*, Ithaca, N.Y: ILR Press, 1996; Steven J. Jackson et al., 'Repair Worlds: Maintenance, Repair, and ICT for Development in Rural Namibia', in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, Seattle, Washington, USA: ACM Press, 2012, pp. 107–116; Steven J. Jackson et al., 'Learning, Innovation, and Sustainability Among Mobile Phone Repairers in Dhaka, Bangladesh', in *Proceedings of the 2014 Conference on Designing Interactive Systems - DIS '14*, Vancouver, BC, Canada: ACM Press, 2014, pp. 905–914; Syed Ishtiaque Ahmed et al., 'Learning to Fix: Knowledge, Collaboration and Mobile Phone Repair in Dhaka, Bangladesh', in *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development - ICTD '15*, Singapore: ACM Press, 2015, pp. 1–10; Susan Wyche et al., ''If God Gives Me the Chance I Will Design My Own Phone': Exploring Mobile Phone Repair and Postcolonial Approaches to Design in Rural Kenya', in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15*, Osaka, Japan: ACM Press, 2015, pp. 463–473. On computational and software error, see, Mark Nunes (ed.) *Error: Glitch, Noise, and Jam in New Media Cultures*, New York: Continuum, 2011; Matthew Bellinger, 'The Rhetoric of Error in Digital Media', *Computational Culture* 5 (2016), http://computationalculture.net/the-rhetoric-of-error-in-digital-media-2/. More recently, Forlano writes on what it means to live with a cyborg body held up by technologies to control and manage bodily functions. See, Forlano, Laura. 'Maintaining, Repairing and Caring for the Multiple Subject', *Continent* 6.1 (2017): 30.

to repair their own bodies to array themselves back as data when their biometrics fail in instances where essential social welfare schemes are dependent on successful authentication.

PDS in India

The PDS is the bulwark of India's food security program in ensuring a steady supply of food grains to the poor. At the same time, it has been susceptible to errors of exclusion and inclusion determined by categorization into above or below poverty lines and corruption and leakages in its delivery systems. Khera notes that computerization of records by some states in the latter half of the 2000 decade was a welcome move towards increasing transparency in the PDS by streamlining the distribution chain, regularly updating records, and weeding out duplicates.[10] Digitization was therefore seen as a step in the right direction when adopted in tandem with other measures. These included ration cards for beneficiaries to track supplies, transparency of BPL lists by painting names on panchayat office walls or color-coding households, and effective grievance redressal systems adopted by states like Chhattisgarh and Tamil Nadu through helpline phone numbers.
In the recent past, Andhra Pradesh from 2014 and Rajasthan from December 2015 have adopted the Aadhaar biometric authentication system as the sole way of authenticating identity to distribute ration in all districts. News reports in the aftermath of this policy move, especially in the state of Rajasthan, suggest that making Aadhaar mandatory has resulted in thousands losing their food entitlements with only 45% of over 98 lakh ration beneficiaries successfully receiving their supplies after authentication.[11] In particular, authentication of data was especially problematic for manual laborers when machines refused to recognize their fingerprints leading to multiple failed attempts to secure their rations.

Even as Aadhaar relies upon its formidable database for various governance functions, it bears remembering that digital records and data are at best mythical in their ability to produce systems with unimpeachable design and implementation. For instance, work around medical health records show that patient care information systems tended to foster errors in entering and retrieving information. Rather than reduce inaccuracies, digital records were found to be disrupting the very communication and coordination processes that they were brought in to support.[12] Similarly, linking delivery of food grains conditional to the database finding a match for the Aadhaar numbers of beneficiaries has proved inimical to the promise of food security in India.

---

10 Reetika Khera, 'Revival of the Public Distribution System: Evidence and Explanations', *Economic and Political Weekly* 46.44–45 (2011): 36.

11 Anumeha Yadav, 'In Rajasthan, There is 'Unrest at the Ration Shop' Because of Error-ridden Aadhaar', *Scroll.in*, 2 April 2016, https://scroll.in/article/805909/in-rajasthan-there-is-unrest-at-the-ration-shop-because-of-error-ridden-aadhaar.

12 Joan S. Ash et al., 'Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-Related Errors', *Journal of the American Medical Informatics Association* 11.2 (2004): 104.

Breakdowns: Between Infrastructures and Bodies

The infrastructural landscape in the Global South has long been habituated to negotiations with breakdowns, repairs, and reuse. The quality of instability that has come to mark the functioning of infrastructures takes a life of its own and is imbricated as a familiar and everyday part in the life of communities.[13] Acts of repair, reuse, and repurpose of things are a commonly accepted practice when confronted by breakdowns and unstable infrastructures. In India specifically, the notion of *jugaad* described as a 'mend and make do work ethic' by Birtchnell has been both celebrated for its disruptive inventiveness and resilience under conditions of scarcity and criticized for the dangerous and unsafe practices that it sometimes symbolizes.[14] It is hard not to encounter different forms of *jugaad* cutting across materialities and various use contexts as people go about the business of everyday living in the Global South. From strategizing for daily life essentials such as water and electricity as observed by Schnitzer in South Africa to coping with relatively more casual slips such as darners to repair fabric tears, cobblers to patch up footwear, and repair and resell markets for electronic goods—the practice of *jugaad* is omnipresent in the Global South.[15] But what happens when bodies marked as data experience errors? What happens when an infrastructure like Aadhaar eschews alternatives in favor of biometric authentication being the only accepted gateway to a welfare scheme?

In Rajasthan, the official rules allow for alternative authentication via a one-time password (OTP) sent to a registered mobile number if the system returns three failed biometric matches for a beneficiary. However, as I witnessed during fieldwork, implementing the OTP alternative is contentious. First, not all beneficiaries have access to mobile phones and not everybody with a mobile phone has linked their mobile numbers to their Aadhaar number. Second, dealers are reluctant to make use of the OTP alternative and report being unfairly penalized if OTP transactions figure in their monthly records. Some dealers even claim that they don't know how to use the OTP option as an excuse for not implementing it. They, thus, turn away beneficiaries whose biometrics fail without providing them with their food supply entitlements, even if their mobile numbers are linked to Aadhaar.

Countering the dealers' claims on unfair penalizations are the inspectors from the food security department who contend that not all OTP transactions arise out of genuine biometric failures. They point out that dealers often use OTPs as a way to make multiple fraudulent entries on a single Aadhaar number to divert food supplies and sell them in the open market. Dealers with high frequencies of OTP transactions are then served a notice, calling for a written explanation along with a suspension of their dealership license.

13 Kathryn Furlong, 'STS Beyond the 'Modern Infrastructure Ideal: Extending Theory by Engaging with Infrastructure Challenges in the South', *Technology in Society* 38 (2014): 139.

14 Thomas Birtchnell, 'Jugaad as Systemic Risk and Disruptive Innovation in India', *Contemporary South Asia* 19.4 (2011): 357.

15 Nikhil Anand, 'Leaky States: Water Audits, Ignorance, and the Politics of Infrastructure', *Public Culture* 27.2 76 (2015): 305; Antina Schnitzler, 'Traveling Technologies: Infrastructure, Ethical Regimes, and the Materiality of Politics in South Africa', *Cultural Anthropology* 28.4 (2013): 670.

The OTP alternative that has been formally approved by the administration is thus not an option that the PDS dealers are always willing to implement in case of authentication failures. Nor are attempts at other informal means like *jugaad* explored, given the threat of suspension. This essay is by no means an endorsement of *jugaad* or a suggestion that it should have space to exist in the PDS, but only an underscoring of the lack of feasible alternatives that beneficiaries could negotiate with for their food entitlements if their Aadhaar authentications fail.

In such a scenario, beneficiaries whose biometrics fail are subject to great anxiety and make repeated trips until they can get authenticated. While they can also send other family members whose names are linked to the ration card to authenticate for the month's supplies, this is often problematic for a variety of reasons. People with infirmities, advanced age, ill health, and disabilities find themselves unable to physically visit the dealer for authentication. The challenges are particularly acute for people who are the sole surviving members of their immediate families and do not have relatives attached to their ration card. Migration for work is another common occurrence and families do not always have someone who can be physically present for authentication every month.
Thus, the insistence on biometric authentication as the sole authenticating factor for food supplies can render beneficiaries 'infrastructural orphans' when confronted with failure. Jackson and Kang write that to be human is to experience embeddedness and completion in a world of things as a fundamental part of our nature.[16] Therefore, experiencing a sense of exclusion as missing data can arguably also deepen the orphaning not only from an infrastructural constitution, but also alienation from a larger scheme of collective belonging. This makes it imperative to pay attention to the kind of attempts people make to repair their break from the biometric data ordering process that holds their data.

Repair Responsibility

Since the dealers are usually the first point of contact where people learn about their breakdown, I found that beneficiaries engage in a series of actions on the advice of the dealers in a bid to authenticate their biometric identity. If the internet connectivity on the SIM and the Aadhaar database are working without interruptions, then taking center stage is the PoS machine that receives top billing as the star of the show. It records, authenticates, and informs: '*aapka Aadhaar sahi hain*' or '*nahin*' (your Aadhaar is correct or not) to the ration dealer working the machine. The machine's every word is breathlessly anticipated by the many people anxiously bent over it and awaiting its verdict. For beneficiaries, it is akin to a public test of what I term as their 'fitness for food grains' that they must undertake every month even as they watch others take theirs. To perform well on this, they must ensure that their fingerprints do not betray them.

For some, the betrayal is a matter of routine. It may take many tries and a systematic trial by elimination to find a fingerprint that will match their Aadhaar. For some, there is no saying when their fingerprints will be returned without a match. They recall days when

16 Jackson and Kang, 'Breakdown, Obsolescence and Reuse'.

authentication has been a breeze, days when they have had several trials with the machine, and days when they had to make several trips on different days before successfully returning with the grain allotment due to them. Sometimes, seemingly unrelated happenings lead to authentication failures. My fieldwork in Rajasthan coincided with the festival of *Holi*, which meant that a lot of womenfolk had colored their hands with henna as part of celebratory rituals. They only remembered the importance of keeping their fingers unblemished when they presented themselves for biometric authentication and none of their fingers were recognized by the Aadhaar server. Many sighed ruefully when they realized that they had forgotten about their monthly technological ritual, 'but, coloring hands is our tradition. Does that mean that I won't get rations until the henna fades?'
And then there are some who are fortunate to never have had an authentication failure but approach their monthly tryst with Aadhaar authentication with a fair amount of anxiety and unease all the same.

For the elderly and the manual laborers whose fingers are callused, hardened, and cut up in the daily grind that characterizes their trade, preparation for the Aadhaar test begins a couple of nights before they actually present themselves at the ration shop. It usually takes the form of diligently scrubbing their fingers with salt, soap, and water and then slathering it with oil before they go to bed. A few nights of this routine are known to lessen the chances of the machine rejecting the fingerprints. Some carry 1 sachets of Bajaj Almond Oil with them to the ration shop. They continue massaging their fingers with oil even as they await their turn. Their time in the queue is often spent bantering and examining each other's fingers for any tell-tale sign of treachery that could result in a negative verdict. Many lament how the source of their livelihood or advancing age have rendered their fingerprints hazy. When their turn arrives, they place their fingers on the machine's sensor. The very earnest among them place their other hand on their finger and press down with all their might as added security. If the machine was programmed to identify and match forceful intentions with fervent sincerity, then authentication would have been instantaneous.

But sometimes, their worst fears are confirmed, and their authentication fails. The dealer allows them their fair share of trials to get at least one of their fingers to match even as the people watching and waiting in the queue behind begin to grow restless at each announcement of rejection. The beneficiaries sometimes give up and walk away. Their next destination is usually the nearest water source. Here, they squat and rub their hands in frustration against mud, stone, and concrete several times before washing them. The harvest season especially leaves many farm laborers to contend with cut fingers, '*Katai kar rahein hain. Dararein pad jaati hain. Ragadna padta hain. Phir shaayad ungli khulegi*' (We are harvesting crops right now and our fingers are all cut up because of it. We have to scrub them hard. Perhaps they will then be authenticated), they inform. Hands washed, they then go in search of oil. Some rub their fingers into their own oiled hair, some catch hold of the nearest person with oil applied to their hair. Some approach neighboring homes for a few drops of oil to smoothen their fingers. Far removed from the contentious din on the ever-widening ambit of Aadhaar, what they confront in the wake of their failed

authentications is not a debate on state surveillance or privacy rights but a routine and public erosion of their dignity in their bid to secure food grains.

Concluding Thoughts

Broken world thinking brings to bear an appreciation of the fragility and the limits of the natural, social, and technological worlds that we inhabit, as one where breakdowns and things falling apart are inevitable, leading to reconstitution and repair.[17] For a system that aimed to strengthen the welfare delivery system and provide more robust ways of inclusion, adopting the Aadhaar biometric system as the only valid way of authentication without recognizing and strengthening alternatives or efficient ways to repair breakdowns weakens any claims for improving the social welfare system. More importantly, it renders the people dependent on the system vulnerable and exposed without a safety net. They are left to grapple with ways to attain recognition for themselves and their needs in the form of complete, valid, and authenticated data that the Aadhaar database would acknowledge. The process of repairing the cracks that render valid beneficiaries invalid for want of a matching biometric authentication is a journey fraught with dependencies and carries significant costs in terms of time, money, and dignity.

17 Jackson, 'Rethinking Repair'.

References

Ahmed, Syed Ishtiaque, Steven J. Jackson, and Md. Rashidujjaman Rifat. 'Learning to Fix: Knowledge, Collaboration and Mobile Phone Repair in Dhaka, Bangladesh', in *Proceedings of the Seventh International Conference on Information and Communication Technologies and Development - ICTD '15*, Singapore: ACM Press, 2015, pp. 1–10.

Anand, Nikhil. 'Leaky States: Water Audits, Ignorance, and the Politics of Infrastructure', *Public Culture* 27.2 76 (2015): 305–330.

Ash, Joan S., Marc Berg, and Enrico Coiera. 'Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-Related Errors', *Journal of the American Medical Informatics Association* 11.2 (2004): 104–112.

Bellinger, Matthew. 'The Rhetoric of Error in Digital Media', *Computational Culture* 5 (2016), http://computationalculture.net/the-rhetoric-of-error-in-digital-media-2/.

Birtchnell, Thomas. 'Jugaad as Systemic Risk and Disruptive Innovation in India', *Contemporary South Asia* 19.4 (2011): 357–372.

Drèze, Jean. 'Dark Clouds Over the PDS', *The Hindu,* 10 September 2016, http://www.thehindu.com/opinion/lead/Dark-clouds-over-the-PDS/article14631030.ece.

Edwards, Paul N., Geoffrey C. Bowker, Steven J. Jackson, and Robin Williams. 'Introduction: An Agenda for Infrastructure Studies', *Journal of the Association for Information Systems* 10.5 (2009): 364–374.

Forlano, Laura. 'Maintaining, Repairing and Caring for the Multiple Subject', *Continent* 6.1 (2017): 30–35.

Furlong, Kathryn. 'STS Beyond the 'Modern Infrastructure Ideal': Extending Theory by Engaging with Infrastructure Challenges in the South', *Technology in Society* 38 (2014): 139–147.

Jackson, Steven J. 'Rethinking Repair', in Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot. *Media Technologies: Essays on Communication, Materiality, and Society*, Cambridge, Mass.: The MIT Press, 2014, pp. 221–239.

Jackson, Steven J., and Laewoo Kang. 'Breakdown, Obsolescence and Reuse: HCI and the Art of Repair', in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14*, Toronto, Ontario, Canada: ACM Press, 2014, pp. 449–458.

Jackson, Steven J., Syed Ishtiaque Ahmed, and Md. Rashidujjaman Rifat. 'Learning, Innovation, and Sustainability Among Mobile Phone Repairers in Dhaka, Bangladesh', in *Proceedings of the 2014 Conference on Designing Interactive Systems - DIS '14*, Vancouver, BC, Canada: ACM Press, 2014, pp. 905–914.

Jackson, Steven J., Alex Pompe, and Gabriel Krieshok. 'Repair Worlds: Maintenance, Repair, and ICT for Development in Rural Namibia', in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, Seattle, Washington, USA: ACM Press, 2012, pp. 107–116.

Johri, Aditya, and Janaki Srinivasan. 'The Role of Data in Aligning the 'Unique Identity' Infrastructure in India', in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '14*, Baltimore, Maryland, USA: ACM Press, 2014, pp. 697–709.

Khera, Reetika. 'Revival of the Public Distribution System: Evidence and Explanations', *Economic and Political Weekly* 46.44–45 (2011): 36–50.

Magnet, Shoshana. *When Biometrics Fail: Gender, Race, and the Technology of Identity*, Durham: Duke University Press, 2011.

Nunes, Mark (ed.). *Error: Glitch, Noise, and Jam in New Media Cultures*, New York: Continuum, 2011.

Orr, Julian E. *Talking about Machines: An Ethnography of a Modern Job*, Ithaca, NY: ILR Press, 1996.

Schnitzler, Antina. 'Traveling Technologies: Infrastructure, Ethical Regimes, and the Materiality of Politics in South Africa', *Cultural Anthropology* 28.4 (2013): 670–693.

Star, Susan Leigh. 'Orphans of Infrastructure: A New Point of Departure', in Ann Light (ed.) *The Future of Computing: Visions and Reflections,* Oxford: UK, Oxford Internet Institute, 2007, https://www.oii.ox.ac.uk/archive/downloads/publications/FD11.pdf.

Wyche, Susan, Tawanna R. Dillahunt, Nightingale Simiyu, and Sharon Alaka. ''If God Gives Me the Chance I Will Design My Own Phone: Exploring Mobile Phone Repair and Postcolonial Approaches to Design in Rural Kenya', in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing – UbiComp '15*, Osaka, Japan: ACM Press, 2015, pp. 463–473.

Yadav, Anumeha. 'In Rajasthan, There is 'Unrest at the Ration Shop' Because of Error-ridden Aadhaar', *Scroll*, 2 April 2016, https://scroll.in/article/805909/in-rajasthan-there-is-unrest-at-the-ration-shop-because-of-error-ridden-aadhaar.

09. OUTLINE INDIA: FIELD NOTES ON DATA PRACTICES AND INNOVATIONS

PRERNA MUKHARYA AND MAHIMA TANEJA

Introduction

Government and administrative bodies have used statistics, or 'science of the state', since the 19th century, to govern, plan, execute, and manage populations. The data that is collected and collated by government agencies is extensive in reach; nonetheless, it has undeniable limitations. Often, recording and reproduction of data are not timely, as there is a lag between collection and release of information; and certain datasets, such as the census, are collected only once in a few years. This is compounded by the fact that the existing records are often incomplete or inaccurate, and bureaucratic hurdles can hinder access to information.

These inefficiencies have immense implications for policymakers, as they hinder their ability to demonstrate progress and impact of policy changes in a rigorous manner. Evidence-based policy-making mandates conducting formative studies, building monitoring frameworks, evaluating impact, and making periodic revisions to plans and implementation strategies. Alongside the statistical information typically captured by governments, procuring qualitative information through case studies, qualitative tools, ethnographies, and documenting processes is essential to develop a comprehensive understanding of the policy context. Moreover, in the past four decades, with the discourse on decentralized planning and participatory governance assuming center-stage, it has become crucial that all stakeholders are involved in the process of identifying needs, setting priorities, and rolling out development interventions.

Despite the fact that ground-level data is foundational to policymaking, in the development sector, the emphasis is on analysis and consulting, with little importance attributed to fieldwork. This paper is an attempt to fill this gap by discussing the challenges that researchers face in large-scale data collection and documenting processes and learnings from the field. It will also underline the importance of a thoroughly qualitative phase of pre-testing of survey tools in relevant settings to develop a contextually germane study design and identify anticipated and unanticipated inconsistencies. In doing so, the essay argues that for research and policy-making to be informed, an integral part of collecting data is maintaining quality and relevance, ensuring standardization at all levels for reliability and validity, and minimizing non-sampling errors and biases. The paper will also attempt to critically engage with the use of technology in social sector research and fieldwork and how the confluence of human capital and technology interacts with the processes of knowledge production.

Pre-Data Collection Phase

Pre-testing survey instruments is an indispensable necessity to ensure adherence to objectives of the study, identify gaps in comprehension between the respondents and the enumerators, determine optimal length and order of the questions, and tailor the tool, given the linguistic and regional variations in cognition and context across India. Traditionally, researchers have focused primarily on standardizing question-wording in the survey questionnaire and ensuring that enumerators adhere to it.[1] This does not take into account the cognition levels of the respondents, dialectal variations, communication gaps between the enumerator and the respondent, or the reasons thereof. Consequently, in recent years, attention has shifted to the importance of cognitive testing of survey instruments, which combines empirical research with cognitive psychology, enabling researchers to develop more robust survey instruments.

Cognitive testing includes task-related pre-testing methods to identify sources of measurement error, which aids in identifying the reasons for non-responses or the so-called 'satisficial' responses.[2] Satisficial responses refer to the phenomenon wherein respondents provide seemingly legitimate answers to survey questions to 'satisfy' the enumerator even when they do not comprehend their meaning or intent, or find it hard to retrieve the required information from their memory.[3] Therefore, the researcher must accord significant attention to check for misunderstandings in the intent or the concept of a question, inconsistent interpretations, colloquial references, and gaps in study instruments during the pre-test.

Let us understand this through a recent survey that Outline India conducted on evaluating hand hygiene behaviors and attitudes in rural India. Before the survey instrument was deployed for a large-scale data collection, an extensive pre-test was conducted in an area that was demographically similar to the target population, but it was drawn from outside it. At the onset, an interesting trend was observed when all the respondents said that they diligently clean their hands by washing them with water at all 'critical times', as was defined by the study. After a few surveys, it was realized that this was because of the way the survey instrument was translated into Hindi. The question 'How do you clean your hands after defecating?', for example, was translated as '*shauch ke baad aap haath kaise dhote ho*?', which back-translates to 'how do you *wash* your hand after defecating?'. Note here that the question in Hindi is nudging the respondent in the direction of 'washing hands using water', because of its literal implications. As a result, hand-cleaning habits which involve the use of mud, dried leaves, or a piece of cloth ran the risk of going unaccounted for. Consequently, the translation of *haath dhona* (washing hands) was changed to *haath saaf karna* (cleaning hands) throughout the tool.

Another key observation during the pre-test of this tool was that the respondents shied away from discussing an intimate habit, such as washing hands after defecation, with an enumerator

1 Debbie Collins, 'Pre-testing Survey Instruments: An Overview of Cognitive Methods', *Quality of Life Research* 12 (2003): 229.

2 Lois Oksenberg, Charles Cannell and Graham Kalton, 'New Strategies for Pretesting Survey Questions', *Journal of Official Statistic*s 7.3 (1991): 349.

3 Collins, 'Pre-testing Survey Instruments'.

because of notions of shame, disgust, and, most importantly, privacy associated with it. As a result, they often said that '*subah hum fresh hokar hi kuch khaate hain*' (in the morning, we eat something only after freshening up). The definition of 'freshening up' here has contextual as well as gendered variations. While for some it meant defecating, for others it meant the entire order of defecating, washing hands, and brushing teeth. For women, particularly, it also included cleaning the house and washing hands and feet before entering the kitchen/cooking space. Identifying such innuendos and colloquial phrases and defining them is also an important task during the pre-test exercise.

Survey Tool Length and Order of Questions

Pre-testing also examines the length and order of questions in the survey tool. This is important because the respondent is not a passive subject but an active participant, who consciously perceives questions, retrieves information from memory, and provides answers based on the temporary and dichotomous relationship established between himself/herself and the enumerator.[4] For example, it is important to acknowledge that during a survey, the respondent's perceptions of practice and knowledge may tend to overlap, resulting in misguided answers. This is because the respondents might want to alter their responses to 'conform to notions of social desirability and self-representation', depending on the survey's context and the surveyor's attitude.[5]

Our experience suggests that survey tools that take from half an hour to forty minutes to conduct, based on need, are optimal. Longer survey tools need additional skills on the part of the enumerators, multiple visits, incentives, or creative methods of engaging with the respondent, retaining their attention, and maintaining the quality and relevance of the responses. This necessitates optimal ordering and prioritizing of questions, specifically in longer survey tools.[6]

The time of the day that a survey is conducted also plays an important role. For instance, surveys conducted during working hours for shop owners or during harvesting season for farmers may be met with hostility. On the other hand, variables such as migratory patterns in the case of panel data must be factored in. When interviewing children, schoolteachers, *anganwadi sevikas*, ASHA[7] workers, or NGO workers, their workday schedule must be considered, in the interest of eliciting relevant responses.

4 Ibid.

5 Ibid., p. 234.

6 The ordering of questions must include collecting identifiers information early on, followed by a focus on priority questions at the onset, and including questions that are intrusive or sensitive towards the end or in the middle. This is to ensure that all the necessary information about the respondents, their background, literacy, socio-economic status is captured; in the event the survey is stopped midway, this information will allow the field staff to return to the respondent at a later point, if needed. Sensitive questions usually includes asking about religion, sexuality, health, or finances. Including them in the middle or at later stages ensures that the subject topic does not discourage or discomfort the respondent from participation and/or from disclosing their 'true' responses.

7 ASHA stands for accredited social health activists, instituted under the National Rural Health Mission.

Handling of Subjects

It is important to maintain neutrality and adopt the right probing techniques when communicating with respondents. There is abundant literature on seeking consent, ethical treatment of subjects, and working with children, sensitive groups, and women, and we will refrain from discussing them in this essay. The field staff must be cognizant of the objectives of the study and the goals of the exercise. Often, respondents discuss subjects that are off-topic, unrelated, or offensive. It is the enumerator's responsibility to bring the discussion back to the subject while ensuring a smooth transition. Again, in rural areas, it is commonplace to be surrounded by passers-by, neighbors, other family members, and children, among others, while conducting surveys. However, it is paramount to ensure that the respondent is at ease while giving responses, does not feel judged while expressing an opinion or sharing views, and trusts that nothing said in the interview will be misconstrued by the community, leading to a backlash. Within the ambit of ethical research and safety procedures, the interviews must be conducted away from large crowds and in an open, quiet place, in the presence of necessary guardians, family members, or alone, as the case may be.[8] Further, irrespective of the responses they receive, the researchers and field staff must refrain from offering their personal opinion or making their opinions about the subject obvious, whether in their tone or body language. All respondents must be communicated to with a neutral, non-emotional demeanor.

Going back to our hand hygiene study, when respondents were asked to demonstrate how they cleaned their hands, it was noticed that people tend to alter their behavior when someone is observing and documenting their activity. Consequently, in most interviews, it was recorded that the respondents correctly demonstrated the steps of handwashing, that is, scrubbing both hands with a cleaning agent and water, but this was not the case in actual practice. This was the result of the enumerator–respondent dichotomy compounded by the limitation of digital data collection where the space to capture subjective and substantial inputs is severely curtailed. In paper-based data collection, enumerators often record their observations in writing, especially so when the response does not adhere neatly to any of the coded options. This changes in digital data collection, which does not account for additional space for the enumerator's observations and scribbled notes. This is compounded by enumerators' lack of familiarity with typing on digital platforms. The tablet itself then adds another layer of gap between the enumerator and the respondent.

Training of Enumerators

For large-scale data collection where multiple enumerators and interviewers are involved, the processes to ensure objectivity and standardization do not stop after pre-testing and review of survey instruments. Team structure, logistics, deadlines, ethi-

8 It is also vital that daily life is not interrupted, and neither is the typical course of events within the subjects' surroundings.

cal issues, and safety precautions need to be discussed at the onset. A detailed field movement plan must be discussed with the entire team.

The quality and reliability of data from the field is dependent on the complete and uniform understanding of the (i) objective, (ii) rationale and intent, and (iii) survey instruments used by the field enumerators. Training followed by mock surveys and field de-briefs provides the surveyors an opportunity to understand the study objectives and survey tools, translate theory into practice through mock interviews, raise doubts, seek clarifications, and resolve ambiguities in the meaning and intent of questions to prevent data mishaps. It also provides the space for researchers to further define and identify ambiguous terms in the survey instrument and acquaint the field staff with the concerns and debates of the development sector and methodologies of social science research. Further, it underlines the need to maintain privacy and research ethics and serves as a space for identifying and resolving biases of the field enumerators themselves.

When enumerators are drawn from the community to ensure regional familiarity, they often come with their biases and assumptions intact, risking the objectivity of the study. For example, in the tool for the above study on hand hygiene behaviors, the question, 'what do you generally use for cleaning hands before feeding the child?' was reworded by the enumerators when interviewing male respondents as: 'What does your wife use for cleaning hands before feeding the child?' This brought forth the deeply rooted gender-based assumptions and biases of the field worker and had to be resolved early-on during the debriefing sessions.

In another study undertaken by Outline India to evaluate the status of WASH infrastructure in government schools across four states in India, the following questions were included:

- Are students given regular training on menstrual hygiene?
- How many common toilets are there in the school?

Several concerns and ambiguities arise in these questions. What counts as 'regular'—weekly/monthly/periodic? What counts as 'training'—formal/informal/external/in-house/morning assembly? How does one count toilets? Does one include only functional units or dysfunctional, abandoned, and broken units as well? How do we standardize and hence define 'common' here?

Let's take another example. In a question on the number of male and female members in the household during a household-level survey on girl child education in rural Rajasthan and Bihar, it was observed that because of dialectical variations, the word *purush*, which is the Hindi translation of male, was not understood by the respondents. As an alternative, when the word *aadmi* was used to ask the number of male members, it was colloquially understood to mean 'persons', making the simple process of recording number of household members a challenge.

Thus, a robust survey instrument also has the potential for immense ambiguities and contextual, regional, and linguistic variations, and it is important to identify, adapt, and resolve them during training and monitoring, through a dialogue with the field team.

Challenges and Learning from the Field

With donor-driven goals and focus on certain areas and locations, it is fair to expect multiple surveys and, hence, interventions being rolled out within the same community. In this age of data saturation, distrust, resentment, or indifference among the communities may be expected. At times, communities themselves make efforts to subvert the processes of data collection by refusing to participate or falsifying information, either in the hope of availing benefits or because of distrust or disengagement with the government. Today, communities also conduct surveys to mobilize knowledge about themselves in aiding local governance. According to Arjun Appadurai, it is important for communities to undertake their own research to advance their rights and claims to resources.[9]

In our own experience, we were once refused and turned away from conducting any surveys in a village in Rajasthan because some unknown NGO had collected data there a few months ago, with an unfulfilled promise of transferring money to the villagers' bank accounts. Given such increasing distrust, it is worthwhile to reiterate that one should go to the field through appropriate channels and permissions, seek informed consent, and value the opinions and cultural or social differences of the respondents. The idea is to work with the local bodies and the communities, and not against the stakeholders.

Further, the surveyor–respondent dichotomy, while important to maintain for objectivity, should not alienate the respondents to the extent that they provide only socially acceptable responses. One way to address such challenges is at the level of tool design as well as training while making sure that the respondents do not feel that they are being evaluated when asked about their behaviors or attitudes. Having said so, this is a difficult feat to achieve.

Thinking Through Technological Innovations

Tablet-based surveys help us better capture, transport, monitor, and process the data collected during personal interviews and surveys. It allows the enumerator to click multiple pictures, prevent data loss, record locations, and deploy the same tool in various languages. It expedites and streamlines the process of large-scale data collection and allows researchers to monitor the data on a real-time basis, in addition to maintaining the authenticity of the collected data.

However, conducting long surveys using electronic devices may not be feasible, given that the devices get discharged, or the enumerators face difficulty in referring to previous questions. Further, digital data collection runs the risk of losing verbatim details and extraneous observations of the enumerators because of the encoding process. In addition, field enumerators often have limited technological exposure and lose confidence even though they understand the subject matter and have contextual familiarity. The specific materiality of digital data collection in the form of a haptic handheld device then adds a layer of gap between the enumerator, the process of knowledge production, and the respondent. Additionally, what are the implications of using tablets for taking written consent through digital signatures or oral

9 Arjun Appadurai, 'Why Enumeration Counts', *Environment & Urbanization* 24.2 (2012): 639.

recordings? Does digital data collection have an adverse impact on building trust with the respondent or on the nature of responses because of the lack of familiarity with the medium, or is it perceived as a welcome change because of its departure from bureaucratic paper trails and assurance of data security? Furthermore, should an increasing shift towards digital data collection and proliferation in digital platforms to create one's own surveys be seen as a move towards democratization of data collection (and production) or a step back? These are questions that remain unexplored.

Outline India, with its focus on disrupting the way we engage in research and execute ground-level work, has been working on two innovative ideas. One of these entails the usage of drones or unmanned aerial vehicles (UAVs) to add a third layer of data through a geographic information system (GIS) and maps in combination with quantitative and qualitative information. In our quests across rural parts of India, we observed that often no maps are available, or they are available in minuscule detail, or they have outdated information. To address this issue, Outline India undertook a pilot study in Haryana, using UAVs to map infrastructure resources in a rural village and aid evidence-based decentralized planning and development initiatives. While the government has undertaken similar initiatives, such as the Bhuvan project, to collate geospatial information and push for decentralized planning through asset mapping and area profile reports, operationalizing them remains a challenge with outdated geospatial information, excessive reliance on satellite data on one hand and administrative data on the other, together with other technological gaps.

In a bid to explore the potential of refining, collating, and using geospatial information for social sector research and development initiatives, Outline India conducted this pilot study using a bottom-up mixed complementary approach. After a thorough review of various policies regarding the infrastructural provisions in rural areas, including the Minimum Needs Programme and Five Year Plans, the study mapped a rural village in Haryana using UAVs and complemented it through transect walks, participatory resource mapping, and household-level surveys. The emergent data was collated to spatially visualize and establish the demographic and caste-based distribution of the village and explore its co-relations with access to community assets and infrastructural resources.[10]

While the study was successful in assisting the local village representative extract information to feed into the Village Gram Panchayat Development Plan, as well as in identifying exclusions, several questions arose for a researcher: How is a physically distant and unfamiliar technological device like UAV perceived by the subjects of research? One can perhaps provisionally argue that some perceived the use of UAV by the Outline India Team, after liaising with the *Sarpanch*, as a sign—a physical and visual proof—of advancement and development. This was different from how a tablet is often perceived—merely a tool held by the enumerator, as

10 Outline India, 'Integrating UAVs in Social Research: Summary of a Successful Pilot Study', in Sonal Bahuguna, Sumeet Gupta, Gaurav Gaur and Maneesh Prasad (eds) *Geospatial Technologies in India: Select Success Stories*, Delhi: FICCI-Geospatial Today, 2017, pp. 75–80, http://ficci.in/spdocument/20873/Geospatial%20Technologies%20in%20India%20-%20Success%20Stories.pdf.

opposed to an *unmanned* aerial vehicle. However, for similar reasons in a different context, UAV can also be perceived as a threat and lead to further distrust in the community, which underlines the need for building participatory and responsible approaches to integrating technologies like this in development sector research. One must also ask—what are the perceptions of privacy and research that are ruffled by the introduction of drones in social sector research? How does one ensure that appropriate protocols are developed, established, and followed before scaling it up in the face of policy challenges? Undoubtedly, when incorporating new technologies in social sector research, one needs to be mindful of such ethical and privacy concerns. These are issues that Outline India is seeking to address and systematize.

Conclusion

This paper underlines the importance of standardizing data collection through various procedures and argues that the *processes*, *learnings*, and *challenges* from the field should be documented with equal rigor as attested to methodologies and data analysis. Using insights from examples of Outline India's projects across rural India in the field of sanitation, education, and infrastructure mapping, to name a few, the paper emphasizes on thoroughly testing survey instruments, using various pre-test methods and cognitive psychology tools to minimize satisficial responses and to identify and resolve sources of measurement errors during the testing. To reiterate, it is vital to predict and collate potential and actual errors that arise because of comprehension and cognitive, regional, linguistic, and contextual variations; identify innuendoes and colloquial references; and optimize the length of the survey instruments during the pre-data collection phase. This is a crucial step in developing a study design. Additionally, it is also pertinent to conduct extensive field training and monitoring of enumerators in relevant settings to resolve ambiguities and redundancies in survey tools in a bid to ensure standardization and order in data collection processes. The paper also discusses the learnings and challenges of using technology such as tablet assisted personal interview (TAPI) platforms and UAVs, arguing that while such technologies aid in expediting and monitoring data collection, maintaining quality, and adding a layer of geospatial information to assist in evidence-based policy-making, they face the challenge of deepening the subject–enumerator gap and losing qualitative insights. This further goes to indicate the importance of a qualitative research stage in data collection and a thorough pre-testing to determine the optimal methodologies, tools, and processes.

References

Appadurai, Arjun. 'Why Enumeration Counts', *Environment & Urbanization* 24.2 (2012): 639–641.

Collins, Debbie. 'Pre-testing Survey Instruments: An Overview of Cognitive Methods', *Quality of Life Research* 12 (2003): 229–238.

Oksenberg, Lois, Charles Cannell, and Graham Kalton. 'New Strategies for Pretesting Survey Questions', *Journal of Official Statistics* 7.3 (1991): 349–365.

Outline India. 'Integrating UAVs in Social Research: Summary of a Successful Pilot Study', in Sonal Bahuguna, Sumeet Gupta, Gaurav Gaur and Maneesh Prasad (eds) *Geospatial Technologies in India: Select Success Stories*, Delhi: FICCI-Geospatial Today, 2017, pp. 75–80, http://ficci.in/spdocument/20873/Geospatial%20Technologies%20in%20India%20-%20Success%20Stories.pdf.

Outline India. 'Understanding the Journey of Data from the Eyes of a Practitioner', Presentation, *Lives of Data Workshop*, Sarai-CSDS, New Delhi, 5–7 January 2017, http://outlineindia.com/downloadFolder/CSDS%20Paper.pdf.

10. COLLECTING OPEN DATA: DATA PRACTICES, TOOLS, LIMITATIONS, AND POLITICS

GUNEET NARULA

This essay looks at development sector organizations and their projects and programs—their data practices, needs, and/or uses—through the lens of data. The insights shared in this essay are based on the observations and experiences of the author, an information technology professional working in the data space and the development sector for the past few years. Significant questions that the essay attempts to answer include how far the data of such organizations is from being open, and what and how much more needs to be done to promote open data practices in this sector. The essay unpacks the tools, limitations, and the politics of collecting development sector data that can be released with an open license.

Open Data for the Uninitiated

Before I dive into this topic, it is important to define what makes data open—or what 'open data' is—especially since the word 'open' is used rather freely by otherwise closed (or just 'not open') systems and projects. Essentially, data that can be freely used, re-used, and redistributed by anyone is called open data. What this means in practice is that the data needs to be made available:

- through an easily accessible method
- at a reasonable cost
- in a convenient, machine-readable format
- with appropriate licenses to allow distribution and use, subject only, at most, to the requirement of attribution and sharing in the original form

Here, 'machine readable' means that the information inside a file can be 'understood' by a computer. For instance, the computer will not be able to spell check a scanned document, but the same content in a Word document can be spell checked easily and automatically. For data, this means sharing in formats like Excel.

Data and the Development Sector

It does not take long for an organization that works in the development sector to understand the importance of data. Data is both a conversation starter and killer; it is the currency that begets more currency. It is a necessity for monitoring and evaluation, as well as for innovation.

These remarks are from the perspective (read bias) of an information technologist and an open data enthusiast/evangelist working with a variety of development sector organizations—generally non-governmental in nature—on their challenges of data collection, management, and use. This disclaimer stands for the rest of the essay. I got involved with the open data community in India through DataMeet in around 2012, and in 2015 I began consulting with

the Akvo Foundation, which has several international partners. It creates open source tools to collect, manage, and use data in the development sector. At Akvo I had the opportunity to engage with many international and Indian organizations in South and Central Asia, such as the World Bank, Aziz Premji Philanthropic Initiative, Welthungerhilfe, Helvetas, Splash, Aga Khan Foundation, Mars Foods, and Innovative Change Collaboration, among others. This engagement helped me understand the data practices of these organizations.

Before we proceed further, let's understand what this 'development sector' is.

Development sector is an umbrella term that covers a wide range of work that deals with infrastructure, living conditions, and livelihoods, among other significant spheres of socio-economically under-privileged regions and communities, both urban and rural. So, from drinking water to sustainable farming to maternal health, school education, and nutrition, this sector deals with a variety of projects. Collecting and managing data related to such development projects and schemes comes with its own set of challenges.

Before we dive into these data challenges, to understand them better, it is necessary to establish what development sector organizations do with data. If we are to examine this from a distance, using data as the lens, it is essentially two things. Most organizations:

1. track or monitor public resources (say water points in a village), and

2. play the role that the state and/or the market economy is supposed to play (ensure the water points are functional throughout the year).

Is this work inconsequential? Not entirely, but it is beyond the scope of this essay to figure out why they do what they do. However, we can definitely ask the question: 'Why they collect data' to do what they do. They do this:

1. to show that a problem exists (for example, most public hand pumps are defunct), and

2. to show that a particular solution works or does not work (for example, rainwater catchments are good substitutes, while building 'capacities' of the local bodies to lobby the government and getting the broken hand pumps fixed).

The problem and the solution are, of course, limited to the scope and geography of an organization's work and so are the sustainability and the 'repeatability' of the solution. Again, these issues are beyond the scope of this essay.

Having helped several organizations collect and examine the required data, I will attempt to answer the question: What does data really show? For one, it shows how much work an organization has done. This is important for both the implementation team (local partner) and the funding team (the donor). Usually the dimensions of an Excel sheet give this away, but mostly it is the report that is derived from analyzing this data that gets the point across. In other words, the amount of data collected by an organization (read dimensions of the Excel

sheet) is an indicator itself. But it is only when this data is crunched, dissected, and sometimes tortured that one gets a report of how much work has been done, how much money has been spent, where it has been spent, for how many 'beneficiaries', and so on.

Secondly, and more importantly, if you dig deep enough, the data reveals several frightening truths about the living conditions of vast swathes of lands and populations. It reveals the severe lack of reach of public and private services and modern technological advancements. It reveals the wealth and comfort enjoyed by the ruling class that makes up our governments, corporations, institutions, and even our development sector organizations, and the marginal conditions in which a vast majority of people live in. These realities may not feature in such plain and straightforward words in the reports, but the evidence is found and consumed by those who analyze the Excel sheets.

Note that the question I raised earlier was 'why they *collect* data' and not 'why they *need* data'. This is an important distinction. It is understandable that an organization needs data to prove that a problem exists or to make sense of the problem, but what is the reason behind the emphasis on collecting this data? Has this data never been collected before? Sure, there are times that the required data does not exist in an easy-to-use format, but certainly this is not true every time. The development sector collects and produces vast amounts of data, but it mostly exists in silos. Only the reports derived from the collected data are shared and not the datasets themselves, nor are these datasets easy to find or access. To put it simply, development sector data is not open.

So Close Yet So Far

For someone involved in the open data community and its growing movement in South Asia, this is obviously frustrating, not only because the data (shared in the reports) is not machine-readable, in closed formats, or is improperly licensed but, also because numerous resources are spent on collecting new versions of already-existing datasets. Furthermore, it should not be a surprise that the standards for collecting data in different fields of development are rarely available. And when they are, there is little incentive to follow them.

At this point, there is one important question: how *far* is this data from being *open*?

Not very.

Yes, the datasets are (potentially) pretty close to the desired standard of openness. To elaborate, let's look at a common data pipeline of development sector organizations I work with:

- They have been using digital data collection tools for the past few years. The debate on paper versus digital (mobile) does not exist anymore, and almost all 'stakeholders' are convinced—except where privacy is paramount and digital solutions will only create new problems.
- There are tools to produce machine-readable formats. So, when Excel sheets, CSVs (comma separated values), JSONs (JavaScript object notation) became more accessible (meaning one did not have to be a computer engineer to publish a dataset in a format such as CSV), it became easier for the computer to consume the data.

- Data cleaning has been able to make significant space for itself in the pipeline. Almost all organizations ensure quality by following basic to high-level cleaning techniques.

This marks progress towards opening a dataset. At this point, the dataset is not far from being shared publicly at all, at least in terms of the effort needed. There are two very important steps that must be followed diligently before releasing data with an open license:

Step 1: 'Anonymization' and/or aggregation

Step 2: Publishing with appropriate licenses so that a lot more people can use the dataset

For any organization with basic Excel capacities, this does not require much effort. We may consider that ideally there ought to exist a standard catalog or repository to publish such datasets to, but that is not a necessity, to begin with. Despite the above-stated technological affordances, we still have far and few open datasets from the development sector.

The key question is: How do we then place 'open data' at the heart of a development sector project or program?

Collecting Open Data

Of course, we cannot just collect open data. Collected data needs to be 'made' open. So, the intention here is to suggest the significance of opening the collected data. Not only does it increase the 'shelf life' of the dataset, which can be used for a time period longer than the collecting organization's project or program under which it was collected, but it also increases collaboration, cooperation, and efficiency of both the organization and the program.

All development sector programs understand that data plays a central and indispensable role in their work. If the objective of a program, for instance, is to build toilets for all households in a village, and the role of the data is to only prove that the objective was achieved, then this is usually done by publishing a report based on the collected data. This serves the immediate purpose, but the dataset created in the process itself is a useful resource. Another organization can use it to push for behavioral change around toilet use, a journalist could use this data to challenge a local government narrative, or a researcher could use this to find best practices and suitable toilet models in different geographies.

All it takes is to treat the role of data with more importance. It is not a background filler in the scene, but it is the supporting cast. Essentially, the program's objectives should not only be about what needs to be achieved, by whom, by when and how, but also to carefully look at the role of data: what to do with it, how to collect or manage it, and then how to publish and share it. Until the development sector organization does not value the hundreds of rows of data it has collected, the Excel sheets will just be archived or trashed after the final report is made.

One of the reasons for not handling development sector data with due diligence is that the organizations themselves feel intimidated or incapacitated when data challenges arise. Technology has become much more advanced and accessible now. Organizations do not need dedicated, expensive IT departments to tackle data challenges anymore. First, there are ample tools and services available to make sure efficiency is not lost. Second, managing data is not rocket science anymore. Most of the time we can get by with well-organized folder structures containing just spreadsheets, while using vertical lookups (usually applied to connect multiple sheets with common columns), and data filter functions freely in Microsoft Excel. And third, big data is not the only form of data that we should aspire for. In fact, most development sector programs deal only with data that can be smoothly read by Excel since the scale of work (and therefore the data collected) is limited.

Finally, if we do develop an ecosystem of open data in this sector, none of us will be alone to face the challenges anyway. What remains to be answered though is how a donor organization can be convinced that open data is worth their money. It is here, in my experience, that technological advances and the merits of open data meet or rather clash with the long-standing issues of politics of development and information access. This clash is obviously not unidirectional and has many historical and cultural dimensions to it. To conclude, I think the interesting possibilities for the formation of new collectives and communities around open data will lead to more grounded theory of data for development practices and vice versa.

11. MAKING INDIA'S BUDGETS MACHINABLE

GAURAV GODHWANI

Background

In August 2017, a shocking number of 290 children died in the Baba Raghav Das Medical College in Gorakhpur, Uttar Pradesh, and in 2019, 940 infants died in the JK Lon government hospital in Kota, Rajasthan. Several media reports suggested poor hospital infrastructure and services being the key reasons behind these deaths of children. These include lack of functional hospital equipment such as oxygen cylinders, ventilators, nebulizers, heaters, etc., as well as several disease-prone conditions and contaminated surfaces found inside our public hospitals. Several researchers have attributed these causes to a significant lack of adequate government budget allocations and timely fund transfers to public health services in the country.[1] In such circumstances, one wonders if it would ever be possible to track an accurate and timely flow of more than 30 lakh crores rupees (0.4 trillion USD) of government budgets being spent across India.

Government budgets are globally considered as 'moral documents', reflecting the priorities and values of the state and its people.[2] They determine the government's take on their promises and past decisions, detail prioritizations across sectors, and explain the allocation of a significant percentage of the state's economy. Budgets are leveraged as a tool for enabling trust in the government's financial activities by providing transparency on public funds and can act to support impactful and equitable public policies. But to do so, governments need to publish their entire budgets in a timely manner and an easy-to-use format, as well as disclose the complete picture of their financial activities in the public domain. They are also responsible for creating appropriate channels for facilitating sustained public participation in budgeting processes.

Budget transparency can lead to efficient use of resources and less corruption, but to sustain it, governments need to invest in creating open systems.[3] These open systems will enable collaboration with citizens by giving them the right to access timely information, government budgeting documents and data, and opportunities to get more involved in various legislative processes through multiple channels. Open Data is a core component of such open systems. It

1 Abhay Shukla, Ravi Duggal, and Richa Chintan, 'How Gorakhpur Was Choked', *The Indian Express*, 1 September 2017, https://indianexpress.com/article/opinion/columns/gorakhpur-hospital-tragedy-gorakhpur-hospital-deaths-brd-hospital-uttar-pradesh-how-gorakhpur-was-choked-4823005/.

2 Dylan Matthews, 'Budgets Are Moral Documents, and Trump Is a Moral Failure', *Vox*, 16 March 2017, https://www.vox.com/policy-and-politics/2017/3/16/14943748/trump-budget-outline-moral.

3 Darshana Patel, Martin Luis Alton, and Sanjay Agarwal, 'Budget Transparency: What, Why, and How?', *Budget Transparency Initiative, World Bank*, 21 September 2011, https://siteresources.worldbank.org/EXTSOCIALDEVELOPMENT/Resources/244362-1193949504055/4348035-1352736698664/BT_What_Why_How.pdf.

is publicly available data that can be universally and readily accessed, used, and redistributed free of charge. It is structured for both usability and computability for humans and machines alike.[4] Open government data is now becoming an essential foundation to establish accountable infrastructure for governments, engage civic actors, and enable trust among citizens. It is believed to have high economic value along with the capacity to boost economic innovation and social transformation. Implementation of open data policies can boost cumulative G20 gross domestic product (GDP) by around 1.1 percentage points, almost 55% of the G20's five-year growth target. Combining all G20 economies, the output could possibly increase by USD 13 trillion cumulatively over the next five years.[5] Despite these clear benefits, we still have very few open budget data initiatives across the globe, and those that exist are yet in their nascent phase.

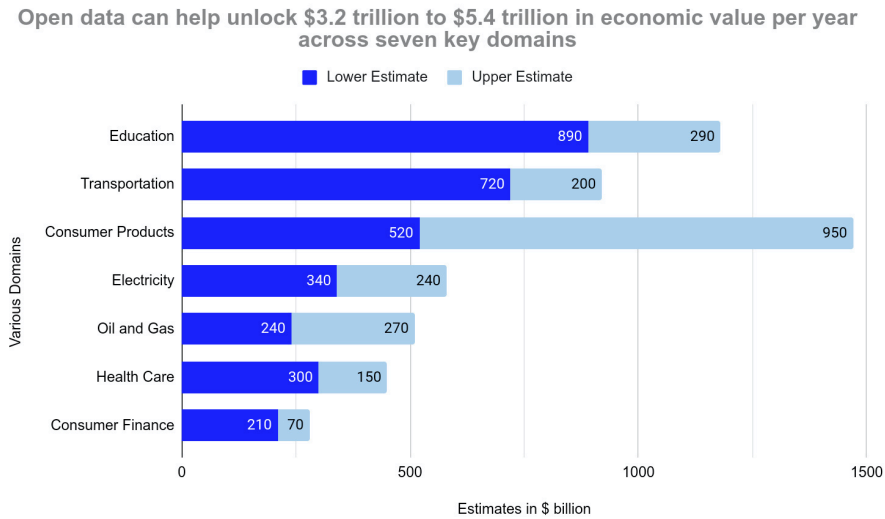


Fig. 1: Author's analysis based on McKinsey Global Institute's estimates in Open Data—Unlocking innovation and performance with liquid information[6]

India follows a federal fiscal architecture that allows for the provision of public goods and services through multiple tiers of government, with each level being assigned to provide a

4 Stefaan Verhulst and Andrew Young, 'The Global Impact of Open Data', *O'Reilly Media, Inc.*, September 2016, https://www.oreilly.com/library/view/the-global-impact/9781492042785/.

5 Nicholas Gruen, John Houghton, and Richard Tooth, 'Open for Business: How Open Data Can Help Achieve the G20 Growth Target', *Omidyar Network*, June 2014, https://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf.

6 James Manyika, Michael Chui, Peter Groves, Diana Farrell, Steve Van Kuiken, and Elizabeth Almasi Doshi, *Open Data: Unlocking Innovation and Performance with Liquid Information*, McKinsey Global Institute, October 2013, https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_data_FullReport_Oct2013.ashx.

fixed set of goods and services. But public access to government budgets data diminishes significantly as we go deeper from the union (central) government to local governments, particularly at the district and subdistrict levels. This gap has constrained public engagement with locally relevant budget information and processes. The union government has started publishing most of its budget documents in XLS (Microsoft Excel file format) format since 2011–12. But at the level of states, budget data is still not available in an easily accessible manner as some of the state governments still do not publish the complete sets of their budget documents online, and those who do, publish the budgets only for recent years as PDFs (Portable Document Format). The only exception is the Sikkim government, which has been publishing its budget documents in XLS format. As we move further to municipal corporations, the availability of budget documents online significantly reduces and the variation in data representation drastically increases. Only close to 100 out of over 200 municipal corporations that have a website publish their budgets in the form of PDFs and scanned copies. As of now, municipal corporations of only Pune, Nagpur, Surat, and Mira Bhaindar publish their budgets in XLS format.

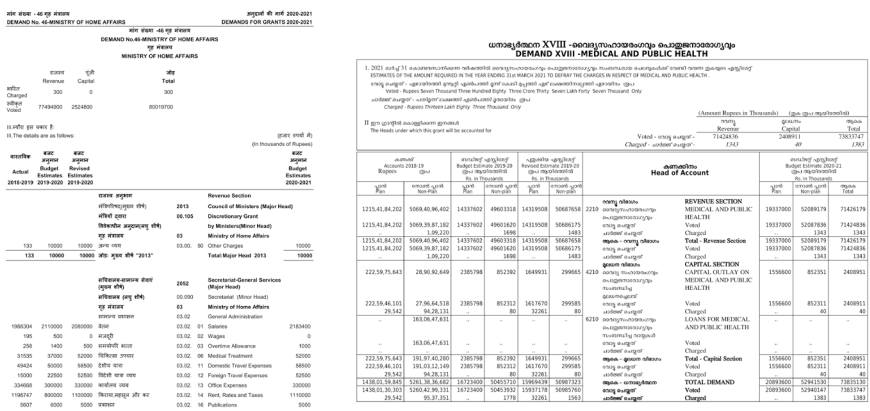


Fig. 2: Budget Documents 2020–21 from Finance Departments of Ministry of Home Affairs and Govt. of Kerala

In such a scenario, the Centre for Budget and Governance Accountability (CBGA), in collaboration with several other organizations and individuals, has developed Open Budgets India—an open data initiative to help make India's budgets more open, timely accessible, usable, and easy to comprehend. CBGA is an independent, non-profit policy research organization working towards enhancing transparency and accountability, and fostering people's participation in governance by demystifying government budgets.[7] Increasingly, people across the country are keen to understand and participate meaningfully in discussions on government budgets. But understanding government budgets is a multi-stage complex process and making budget documents available in a timely manner and machine-readable is just the first step towards this exercise of demystification. This chapter attempts to describe our journey so far to develop an open data collaborative for government budgets in India. This

7 See, https://openbudgetsindia.org/; http://www.cbgaindia.org/.

initiative has been made possible because of the generous financial support and guidance from a number of institutions, including Bill and Melinda Gates Foundation (BMGF), Omidyar Network, International Development Research Centre (IDRC)—Think Tank Initiative—and National Foundation for India (NFI).

Sowing the Seeds of Co-Creation

The civic-tech and data-for-good ecosystem is still in its nascent stage in India. The idea of using technology and data to improve the quality of public delivery systems and lives for millions of people is still struggling to find a stronghold. One of the common trends that I have observed while working with various non-profits in India is that most of them appreciate the use of technology in their work but don't invest enough to grow their own tech and data capacity. They still hugely rely on outsourcing their technological requirements, thus often missing the beat on how they can automate some of the human-intensive methods in their day-to-day research and advocacy. They end up paying high costs and suffer too steep a learning curve by using proprietary software, painstakingly dealing with vendor lock-ins, having dependencies on outdated frameworks, facing data security and privacy issues, thus overall hampering their future development and innovation. Moreover, this practice of relying on proprietary software severely restricts the possibility to co-create and engage with experts from various disciplines.

One of the key aspects of this initiative is to leverage the power of communities. It was in early May 2015 when Omidyar Network, a popular philanthropic investment foundation, reached out to DataKind Bangalore to help CBGA with this work. DataKind Bangalore is a local chapter of an international community that helps other non-profit organizations start their data science journeys by leveraging a pro-bono group of volunteers working over the weekend. Along with my colleagues, I got involved in the initial discussions to understand the state of budgets in the country. We assured CBGA that some of the difficult work of generating machine-readable data could be automated with the help of technology. We realized the need to conduct a series of consultations to facilitate brainstorming among budget researchers, social scientists, technologists, policy advocates, and other open data activists.

Next, Centre for Internet and Society, DataMeet—a community of open data enthusiasts—and DataKind Bangalore came together to brief CBGA about open data standards, the need for developing metadata, and how to harness free and open source software (FOSS).[8] We drew commonalities from how other open data initiatives across the globe have incorporated FOSS in their work for rapid and agile development.[9] This was followed by a community event where volunteers from DataKind Bangalore and researchers from CBGA came together to explore how a data pipeline could be developed to generate machine-readable data and how documents could be arranged on the platform. We explored various ways to visualize complex

8 Kenneth Wong and Phet Sayo, *FOSS – A General Introduction*, International Open Source Network and UNDP Asia-Pacific Development Information Programme, 2004, https://en.wikibooks.org/wiki/FOSS_A_General_Introduction/Introduction.

9 'About Federal Spending Transparency – Agile Development Methodology', *Data Act Collaboration Space*, n.d., https://fedspendingtransparency.github.io/about/.

budget data.[10] The enthusiasm and expertise of the volunteers made researchers at CBGA more confident about the potential of opening up their process, tech, and design along with their data, leading to avenues for regular community feedback. It was around this time in October 2015, when I was brought in full-time as the Technical Lead for the project to facilitate the platform development in-house and in open. Moreover, we set up an Advisory Committee consisting of experts from diverse backgrounds, including budget research, public finance, accounting and audits, policy research, and open data, to provide continuous inspiration, guidance, and suggestions to our work.

To facilitate collaboration in a multidisciplinary and geographically distributed in-house team and volunteer groups, we relied heavily on tools and techniques used in the software development world. We used Slack for our daily active communications; Github for version control and publication of our code and designs; Trello for team-wise tasks management; cloud-based servers to run our platform; and a shared Google Drive for storing documentation, metadata, and datasets, which were reviewed and uploaded on the platform. We followed Agile methodology that facilitates iterative and incremental development. It advocates adaptive planning, evolutionary development, early delivery, and continuous improvement, and it encourages rapid and flexible response to change.[11] We conducted weekly check-in calls to explain our progress on our individual tasks, bridge the communication gaps, and plan for future development.

We partnered with a few other organizations that have already been working with budget data. Macromoney Research Initiatives helped us make available the budget data of a large number of municipal corporations.[12] Budget Analysis Rajasthan Centre (BARC), Jaipur; National Centre for Advocacy Studies (NCAS), Pune; and Pathey, Ahmedabad contributed their efforts to collecting, collating, and translating budget data of a number of municipal corporations. These collaborations helped us secure key datasets on the platform, which otherwise would have been quite difficult to obtain. It was this ensemble of the right set of people, organizations, and technologies that made this data initiative possible.

Route to Machinability

To make government budget data more accessible and actionable, it's essential to understand the concept of machinability. Being machinable means having the ability to be consumed and processed by machines—in the case of data, it refers to computer programs. Not all digital materials are machinable. As described by Open Data Handbook, PDF documents containing tables of data are definitely digital but not machine-readable because a computer would struggle to access the tabular information; they are human-readable, though.[13] The

10 Gaurav Godhwani and Rohith Jyothish, 'Opening Up the Discussion on Public Finance in India: A Tool for Budget Analysis', *International Budget Partnership*, 30 March 2016, https://www.internationalbudget.org/2016/03/public-finance-in-india-tool-for-budget-analysis/.

11 'What is Agile Software Development?', *Agile Alliance*, n.d., http://www.agilealliance.org/the-alliance/what-is-agile/.

12 http://www.publicfinance.in/.

13 Open Knowledge Foundation, 'Machine Readable', in *Open Data Handbook*, June 2016, http://

equivalent tables in a format such as a spreadsheet are machine-readable. Machinability is key to facilitate the use of open data, as it enables users to perform timely analysis and comparisons on all digital platforms. Machinable data is utilized across various countries as a tool for advocacy on government spending, evidence-based research, and policy recommendations. The impact of machinable data can be globally observed in various key social sectors, from tackling corruption and transparency to social mobilization and informed decision-making.

Unfortunately, most of the budget documents published across various levels of government in India are in the form of PDFs, scoring just one star out of five as per Tim Berners-Lee open data standards, and this still remains one of the biggest challenges for us.[14] Thus, we decided to develop an automated data pipeline that could enable the acquisition of budget documents from various websites, facilitate tabular data extraction, perform data cleaning, and generate clean machine-readable datasets. To avoid reinventing the wheel, we started working with existing popular open-source software available in the open data ecosystem. We went ahead with the Comprehensive Knowledge Archive Network (CKAN), a powerful open-source data publishing platform that makes data accessible by providing tools to streamline publishing, sharing, finding, and using datasets.[15] We positioned ourselves to develop the codebase, process, design, and visualizations open-by-default, including our experiments and prototypes.[16]

Extracting tabular data from budget documents was the core element of this data pipeline. PDF was never designed to be a data format; instead it was developed as a print-friendly 'electronic format', positioning all text by placing each character at minutely precise coordinates in relation to the bottom-left corner of the page. PDF was invented in 1993 by Adobe, which acknowledges its shortcomings when it comes to data. 'For the person who wants raw data, PDF isn't the right choice!'.[17] Thus, algorithms need to rely on computer vision techniques to detect tabular information.[18] We used Tabula as the base for our PDF parsing. It detected boundaries of the table rows, and if the table contained ruling lines, it used their position to generate the boundaries (top and bottom) of each row.[19]

For our use case, we had to add some more intelligence specific to each budget document as input to Tabula. We developed a more nuanced way to detect the most prominent tabular boundary for the data, as most of the budget documents have multiple boundaries,

opendatahandbook.org/glossary/en/terms/machine-readable/.

14 Tim Berners-Lee, 'Linked Data', 27 July 2006, *w3.org*, https://www.w3.org/DesignIssues/LinkedData.html.

15 https://ckan.org.

16 https://github.com/cbgaindia.

17 Jim King, 'Inside PDF - My PDF Hammer (revision)', *Adobe*, 19 October 2011, https://web.archive.org/web/20170810234419/https://blogs.adobe.com/insidepdf/2011/10/my-pdf-hammer-revision.html.

18 Computer vision is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do. See, Wikipedia contributors, 'Computer Vision', https://en.wikipedia.org/wiki/Computer_vision, accessed 1 September 2020.

19 http://tabula.technology/

which makes them difficult to parse. We added heuristics to detect the number of columns and their coordinates. Further, we added detection of page dimensions, alignment, and layout for each document, which in turn enabled detection support for portrait and landscape layouts, various page sizes, etc. With these additions, our efficacy of detecting data tables in budget documents increased significantly. Moving forward, we planned to explore a couple of deep learning techniques such as Convolutional Neural Networks to make PDF parsing more efficient.[20]

Enhancing Data Usability and Budget Literacy

For each budget dataset, we worked to create an intuitive metadata vocabulary. Metadata describes a particular dataset and performs a function similar to that of a 'catalog card' in a library. It provides explanations for both the functional and administrative classification of budgetary information and fund flows in India and enhances the searchability of the data. In the absence of a common budget metadata standard, it becomes difficult to arrange, classify, search, or even compare budget data across tiers of government and even across years. This exercise of metadata preparation was spearheaded by a group of researchers at CBGA, bringing in their experience of studying and analyzing a variety of budget documents across India. All content on the platform developed by us is under the licensing agreement of Creative Commons Attribution 4.0 (CC-BY), which allows users to copy, distribute, display, and arrive at analyses with only one request of giving appropriate credit and attribution to the platform. Moreover, all datasets can be searched, accessed, and downloaded via robust APIs.[21]

We worked to create a couple of dynamic data visualizations for machine-readable datasets, which allow users to compare and analyze time-series datasets. To develop each data visualization, we followed an iterative process and gathered continuous feedback from researchers, ensuring easy delivery of key insights for the users. One can directly embed these visualizations in their blogs, case studies, and other forms of digital content.

20 Deep learning (also known as deep structured learning or hierarchical learning) is the application to learning tasks of artificial neural networks (ANNs) that contain more than one hidden layer. Simpler ANNs contain zero or one hidden layer. Deep learning is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms. See, Wikipedia contributors, 'Deep Learning', https://en.wikipedia.org/wiki/Deep_learning, accessed 1 September 2020; A convolutional neural network (CNN) is made of one or more convolutional layers (often with a subsampling step) and followed by one or more fully connected layers as in a standard multilayer neural network. See, http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/.

21 In computer programming, an application programming interface (API) is a set of subroutine definitions, protocols, and tools for building application software. In general terms, it is a set of clearly defined methods of communication between various software components. See, https://www.hcltech.com/sites/default/files/apis_for_dsi.pdf.

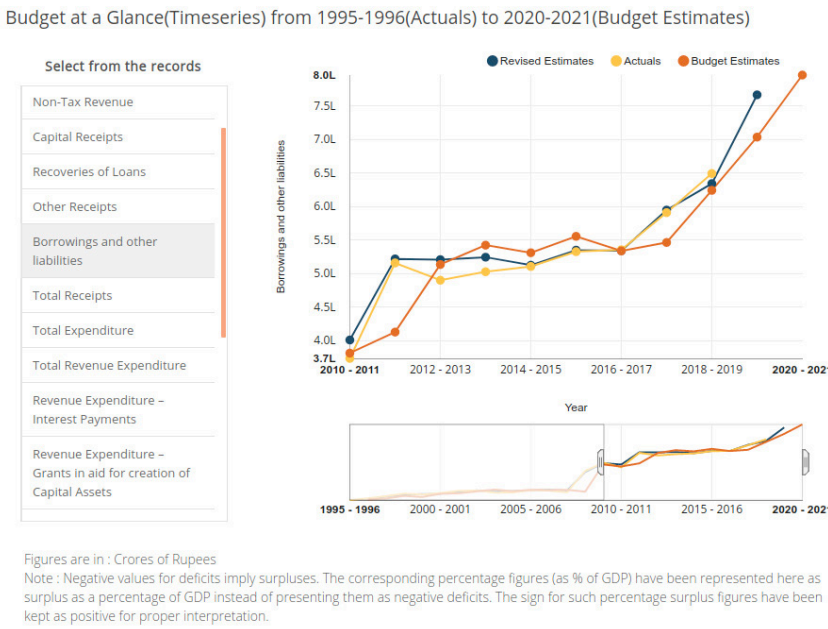


Fig. 3: Open Budgets India—Union Budget at a Glance (Time-series) [22]

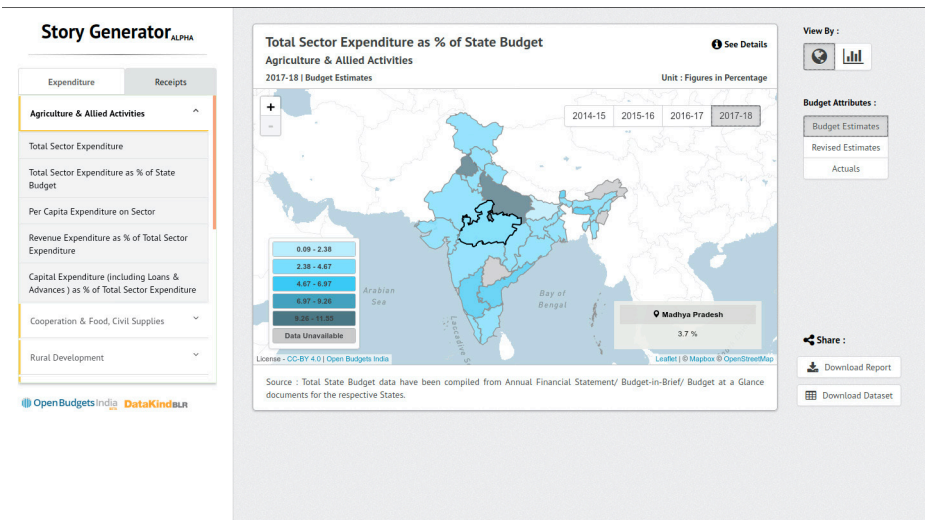


Fig. 4: Open Budgets India—Story Generator [23]

22 https://openbudgetsindia.org/dataset/budget-at-a-glance-timeseries
23 https://cbgaindia.github.io/story-generator

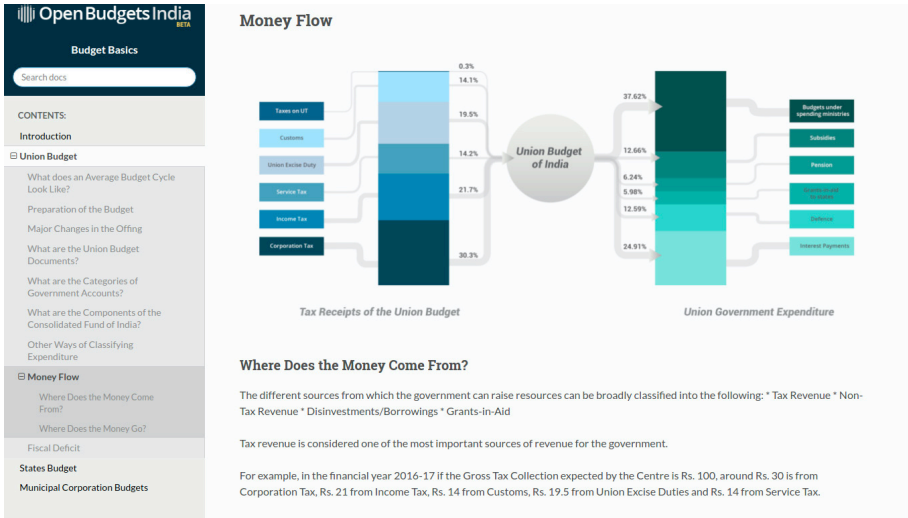


Fig. 5: Open Budgets India—Budget Basics[24]

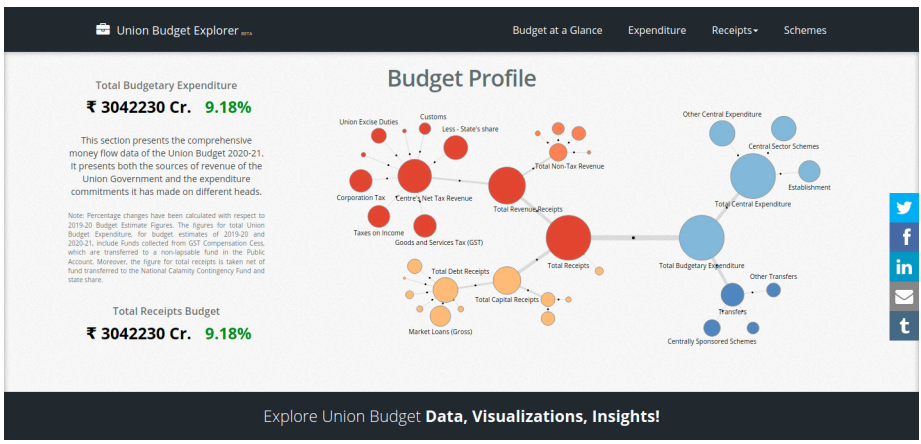


Fig. 6: Open Budgets India—Union Budget Explorer 2020–21[25]

One of the other key objectives of our open data collaborative is to educate users about how government budgeting works in the country. We strive to simplify the information related to budgeting processes, the flow of public funds, the format of budget documents and codes, etc. Thus, we created Budget Basics, a guide to understanding India's Budgets, which explains various fiscal terminologies and gives insights about various pro-

24 https://openbudgetsindia.org/budget-basics/
25 https://union2020.openbudgetsindia.org

cesses involved in the union government, state government, and municipal corporation budgets.[26]

We continued working with DataKind, engaging in a long-term DataCorps collaboration to build Story Generator, an open-source tool enabling comparison of key fiscal indicators across the states and financial years.[27] A dedicated team of volunteers from DataKind worked with our in-house design and visualization experts for more than ten months to shape this project. The data for this tool comprises various receipts and expenditure indicators across twelve key development sectors for twenty-six states (including Delhi). This extensive data preparation exercise was led by a team of researchers at CBGA.

Even after making this data machinable, we realized, with time, that citizens still face difficulty analyzing budget data and struggle to timely participate in crucial budget discussions. It's cumbersome to go through over 150 documents and find budget trends across years on such things as allocations for important centrally sponsored schemes like National Rural Employment Guarantee Act (MGNREGA), National Health Mission, etc. Moreover, one needs to search and sum up data from multiple files to get an accurate picture of sectoral allocations. Seeing these issues, we decided to develop a 'Union Budget Explorer' for each budget cycle to make it simpler for citizens to visualize and explore union budget data time-series, expenditure, receipts, schemes, and more in one place.[28]

Strengthening Fiscal Transparency for States and Districts

As part of our advocacy efforts, we created a list of recommendations for state finance departments, which included best practices to make their budget data more open, accessible, and citizen-friendly.[29] These recommendations also detailed various steps that state finance departments should take to become compliant with the National Data Sharing and Accessibility Policy (NDSAP) and its implementation guidelines.[30] We started sending these recommendations to each state, seeking appointments to meet the respective finance secretaries responsible for budget preparation and publication. Assam was the first state to respond to us, and after a series of discussions, CBGA signed a memorandum of understanding (MoU) with Assam Society for Comprehensive Financial Management System (AS-CFMS) to become a knowledge partner in institutional strengthening, finan-

26 https://openbudgetsindia.org/budget-basics/.

27 http://www.datakind.org/datacorps; https://cbgaindia.github.io/story-generator/. See, Akshay Verma, 'A Look into State Budget Analysis — Story Generator', *Open Budgets India*, 29 June 2018, https://blog.openbudgetsindia.org/a-look-into-state-budget-analysis-story-generator-67a4e015e6b9.

28 https://union2020.openbudgetsindia.org/.

29 'Best Practices for Publishing State Budget Documents Online', *Open Budgets India,* June 2017, https://openbudgetsindia.org/pages/best-practices-for-publishing-state-budget-documents-online.

30 Ministry of Science and Technology, Government of India, 'National Data Sharing and Accessibility Policy (NDSAP) 2012 (Gazette Notification)', *The Gazette of India*, March 2012, https://data.gov.in/sites/default/files/NDSAP.pdf; 'Implementation Guidelines for National Data Sharing and Accessibility Policy (NDSAP) Ver. 2.4', *Open Government Data Division - National Informatics Centre*, November 2015, https://data.gov.in/sites/default/files/NDSAP%20Implementation%20Guidelines%202.4.pdf.

cial reforms, and capacity building.[31]

We are focused on helping the Assam government publish more open budgets data, and citizen and sectoral budgets, ensuring transparency of financial and procurement information. We have worked with them to launch Assam Budget Explorer, a data platform for citizens to visualize and analyze budget highlights, grant-wise detailed expenditures, receipts, etc. We have also been doing regular workshops for the Assam government staff to promote more use of FOSS to analyze their data sources, design thinking, participatory design, and more.

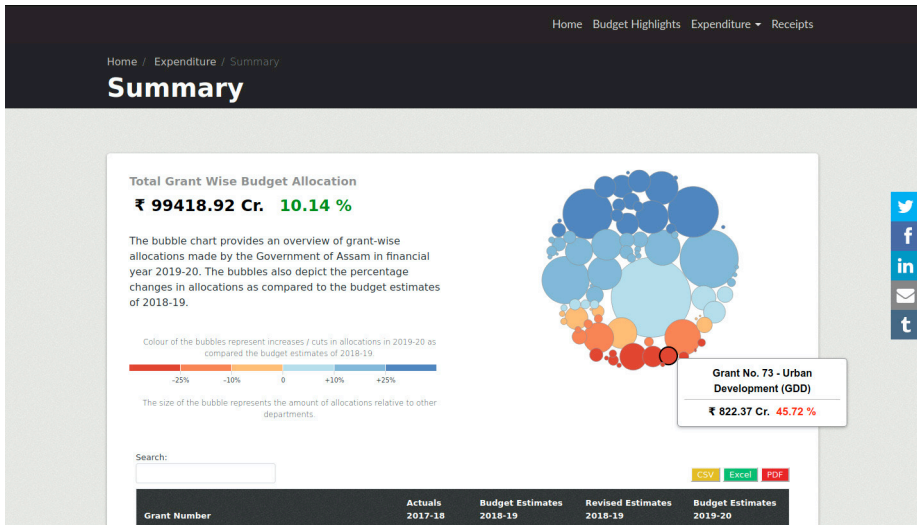


Fig. 7: Open Budgets India—Assam Budget Explorer 2019–20 [32]

31 Simonti Chakraborty, 'Open Budgets India's workshop with Assam's Finance Department', *Open Budgets India*, 17 August 2018, https://blog.openbudgetsindia.org/open-budgets-indias-workshop-with-assam-s-finance-department-6f796b5f683b.

32 https://assam2019.openbudgetsindia.org

Fig. 8: Open Budgets India's workshop with Assam's Finance Department[33]

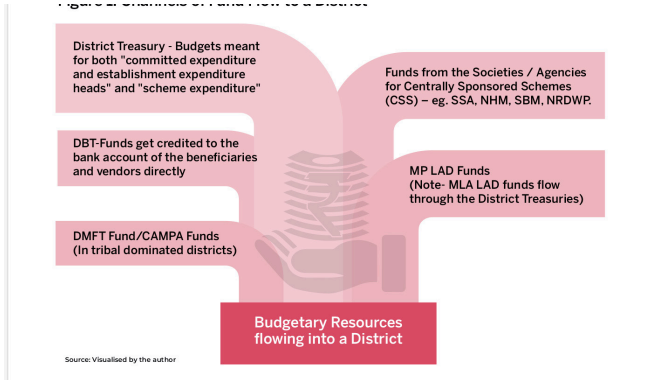
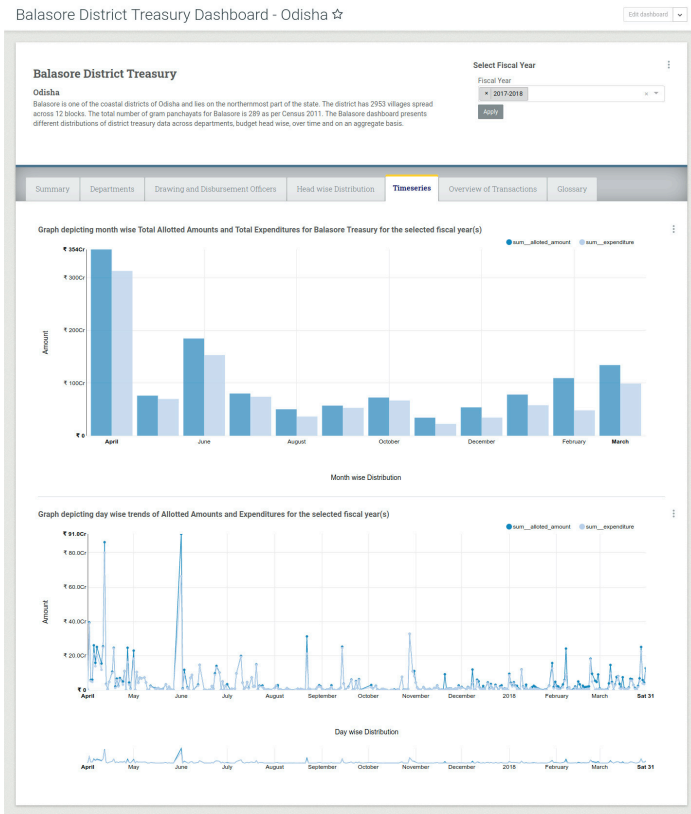


Fig. 9: Channels of Fund Flow to a District, Strengthening Budget Information Architecture at the District Level[34]



34 https://dash.openbudgetsindia.org/superset/dashboard/odisha_balasore_treasury_dashboard/?standalone=true

Fig. 10: Balasore District Treasury Dashboard—Open Budgets India

After making the budget data of a few states more machinable, our next target was to make budgets and spending information on the district-level more accessible and usable. A considerable proportion of money flows into a district through district treasuries, where drawing and disbursing officers (DDOs) procure money for specific service delivery in their designated subdistrict. Most Indian states have adopted the integrated financial management system (IFMS) to closely monitor budget preparation and distribution, real-time expenditure, accounting and reconciliation, bill preparation and disbursements, and other fund management services on a district treasury level.35

As a pilot, we started mining month-wise spending data for ten districts from Andhra Pradesh and Odisha. We also developed dashboards for Balasore district of Odisha and Krishna district of Andhra Pradesh, making it easier for citizens to drill down years of data and draw their own insights on how these treasuries have been spending across various departments and schemes.

The Road Ahead

In early 2018, I, along with some other colleagues, started CivicDataLab, a research lab that harnesses data, tech, design, and social science to strengthen the course of civic-engagements.[36] We work to harness the potential of the open-source movement to enable citizens to engage better with public reforms. We aim to grow the data and tech capacity of governments, non-profits, think-tanks, media houses, universities, etc. to enable data-driven decision-making at scale. We continue to work in the public finance space to strengthen the Open Budgets India initiative. We are also expanding our work in the space of law and justice, Indic languages, and urban development.

In terms of new developments, we are working closely with the Assam government to help them with participatory budgeting and co-creation of engaging citizen budgets. For Himachal Pradesh, we are co-creating fiscal data explorer, a unique tool where citizens can explore both budgets and granular day-wise spending data of state governments in an easy-to-comprehend and simple-to-use manner.[37] With CBGA, we are working to scope out the next phase of Open Budgets India, to expand our data coverage, analyze, and publish open budget and spending data for key sectors and schemes for various parliamentary constituencies in India. We plan to do more public consultations and consensus building with various stakeholders to evangelize data standardization and increased publication

35 Nilachala Acharya and Vijayta Mahendru, 'Strengthening Budget Information Architecture at the District Level', *Centre for Budget and Governance Accountability and Tata Trusts*, January 2020, http://www.cbgaindia.org/wp-content/uploads/2020/02/Budget-and-Expenditure-Information-at-District-Level-Policy-Brief.pdf.

36 https://www.civicdatalab.in/.

37 Gaurav Godhwani, Shreya Agrawal, Simonti Chakraborty, and Thomson Muriyadan, 'Roadmap for Co-Creating the Fiscal Data Explorer', 18 July 2019, https://medium.com/civicdatalab/roadmap-for-co-creating-the-fiscal-data-explorer-79818a53728f.

and accessibility.

Lastly, we are working to seek more contributions and support from diverse communities. We plan to engage with more budget researchers, policymakers, civil society organizations, and journalists to understand their needs and increase the uptake of Open Budgets India. We also aim to evolve our processes, documentation, and communications so that we can facilitate easier onboarding for volunteers. We hope to reach out to more people and communities to continue our adventurous journey to explore possibilities to track how government budgets are being spent across India. Together we aim to continue our efforts in making India's budgets more open, usable, and easy to comprehend.

References

Acharya, Nilachala, and Vijayta Mahendru. 'Strengthening Budget Information Architecture at the District Level', *Centre for Budget and Governance Accountability and Tata Trusts*, January 2020, http://www.cbgaindia.org/wp-content/uploads/2020/02/Budget-and-Expenditure-Information-at-District-Level-Policy-Brief.pdf.

Berners-Lee, Tim. 'Linked Data', 27 July 2006, *w3.org*, https://www.w3.org/DesignIssues/LinkedData.html.

'Best Practices for Publishing State Budget Documents Online', *Open Budgets India,* June 2017, https://openbudgetsindia.org/pages/best-practices-for-publishing-state-budget-documents-online.

Chakraborty, Simonti. 'Open Budgets India's Workshop with Assam's Finance Department', *Open Budgets India*, 17 August 2018, https://blog.openbudgetsindia.org/open-budgets-indias-workshop-with-assam-s-finance-department-6f796b5f683b.

Gruen, Nicholas, John Houghton, and Richard Tooth. 'Open for Business: How Open Data Can Help Achieve the G20 Growth Target', *Omidyar Network*, June 2014, https://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf.

Godhwani, Gaurav, Shreya Agrawal, Simonti Chakraborty, and Thomson Muriyadan. 'Roadmap for Co-Creating the Fiscal Data Explorer', 18 July 2019, https://medium.com/civicdatalab/roadmap-for-co-creating-the-fiscal-data-explorer-79818a53728f.

Godhwani, Gaurav, and Rohith Jyothish. 'Opening Up the Discussion on Public Finance in India: A Tool for Budget Analysis', *International Budget Partnership*, 30 March 2016, https://www.internationalbudget.org/2016/03/public-finance-in-india-tool-for-budget-analysis/.

Open Knowledge Foundation. 'Machine Readable', in *Open Data Handbook*, June 2016, http://opendatahandbook.org/glossary/en/terms/machine-readable/.

Patel, Darshana, Martin Luis Alton, and Sanjay Agarwal. 'Budget Transparency: What, Why, and How?', *Budget Transparency Initiative, World Bank*, 21 September 2011, https://siteresources.worldbank.org/EXTSOCIALDEVELOPMENT/Resources/244362-1193949504055/4348035-1352736698664/BT_What_Why_How.pdf.

Verma, Akshay. 'A Look into State Budget Analysis — Story Generator', *Open Budgets India*, 29 June 2018, https://blog.openbudgetsindia.org/a-look-into-state-budget-analysis-story-generator-67a4e015e6b9.

Verhulst, Stefaan, and Andrew Young. 'The Global Impact of Open Data' *O'Reilly Media, Inc.*, September 2016, https://www.oreilly.com/library/view/the-global-impact/9781492042785/.

Wong, Kenneth, and Phet Sayo. *FOSS – A General Introduction*, International Open Source Network and UNDP Asia-Pacific Development Information Programme, 2004, https://en.wikibooks.org/wiki/FOSS_A_General_Introduction/Introduction.

12. HISAAB KITAAB IN BIG DATA: FINDING RELIEF FROM CALCULATIVE LOGICS

NOOPUR RAVAL

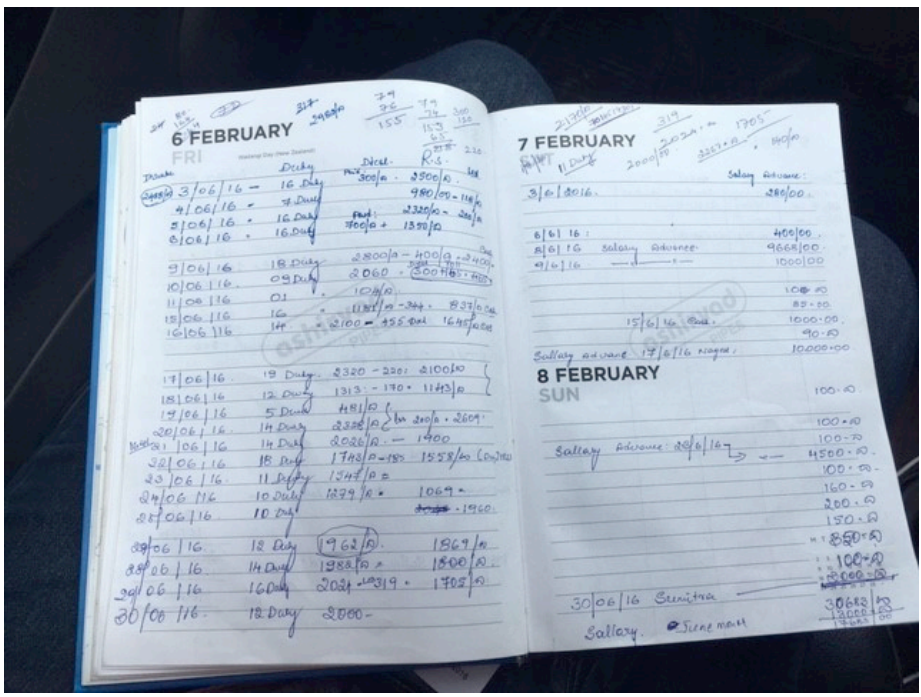


Fig. 1: A Bangalore ridesharing driver's account book

Much like other Indian ridesharing cab drivers, Jagdheesh, an Uber driver in Bangalore, is keenly aware of words like 'incentives', 'earnings', 'duty', and 'device' that dominate daily conversations among drivers and passengers within the Indian ridesharing space. While he drove me to my destination, we talked about Uber and its rival Ola cabs, work before the arrival of these apps, how passengers behave, the work hours that drivers put in, their monthly earnings, and so on. As soon as we reached my destination and I proceeded to get out of the car, I noticed that Jagdheesh reached for a small notebook atop his dashboard. I saw him write the exact amount of my trip to a list. The notebook was half-filled with several such lists, each page containing the date, number of trips, and earnings from each trip on that day. When I asked him why he maintained an actual physical account book when the app already displayed his daily and weekly earnings, he told me it was for his 'own record'.

After that trip, I started noticing that almost all drivers had a similar notebook stashed away under the wheel or kept on the dashboard. The persistent presence of the physical account

book made me curious because ridesharing apps such as Uber, Ola, and others already display daily and weekly trip earnings as well as incentives. Upon further probing, it became clear that drivers were well capable of reading the numbers and text—they knew these numbers represented their earnings. However, they continued to meticulously and habitually jot down the same numbers in their notebooks, too. Often, they would rearrange and even split earning numbers into smaller figures to retain the differences (such as cash vs. digital wallet payments, for instance) to make the app analytics 'consumable' in a way that the numbers made the most sense for their daily-life calculations. This motivated me to dig deeper and look into the *hisaab kitaab* (account) notebooks to grasp why drivers were constantly reordering and reinterpreting numbers provided to them by the app dashboard, what was being gained in such moments of intimate reorientation of analytics, and what that could tell us about living along with Big Data.

Account Books as Communicative Genres

Before going to my conversations with the drivers about account books, let's dwell a bit on the historical material life of the *hisaab* book. Physical account books are a common information artifact in the South Asian public sphere. Similar-looking rugged notebooks or foolscap books containing a pen stuck in between are an integral component of various kinds of informal work—found stowed away in auto-rickshaws, kept in *kirana* stores (local grocery shops), placed at the table of the local *dhobi* (washerman) shops, carried in personal bags by domestic helps and cooks, and often found in the kitchen usually maintained by the woman homemaker of the family. It's useful to ask what these books do and how they are used as communicative and calculative devices. For instance, the auto-rickshaw driver's book is not only for record-keeping of daily earnings for himself but also often an account produced for the owner of the vehicle who leases rickshaws out to drivers. Similarly, the grocery store account ledger (there are multiple ledgers) or the washerman's book there not to just record transactions but produce collaborative accounts of the engagement between customers and service providers, hinting at the importance of *co*-producing an account as an exercise in building and maintaining trust. Going a step further, the account book at the grocery store often doubles up as a credit register using which shopkeepers extend credit facilities allowing their regular clients as well as poorer customers to purchase essential goods without paying immediately. This helps homemakers budget their expenses without completely relying on the 'men of the family' for every daily expense.[1] It also allows customers to plan and reflect upon their earnings or expenditures, thus producing affordances of time and money that are housed within the rhetorical and discursive world of the account book.

Two things immediately surface from this brief exposition on the 'account book'. First, as we situate the account book within its local contexts of use, it materializes as an object of knowledge and power beyond its information content, compelling us to look at the relationships it mediates and how these relationships, in turn, shape the form a particular book of accounts takes. But, second, to re-emphasize the inherently communicative nature of an account book,

1 Fiona Leach and Shashikala Sitaram, 'Microfinance and Women's Empowerment: A Lesson from India', *Development in Practice* 12.5 (2002): 575.

despite the presence of so many varying account books in different contexts such as those I described, the book *as a form* performs a limited uniform function.

It is then worth asking how we all know what to do with account books, why they must be maintained, what to expect from them, and how they might be changed or broken. As Yates and Orlikowski explain in their work on 'communicative genres' within organizational action, the memo, the meeting, or, in this case, the account book as *generic* and pervasive communicative forms embody a set of rules and expectations that are very much shaped by negotiations among social actors over time but also simultaneously, much like infrastructure, gradually acquire the 'moral and ontological status of taken-for-granted'.[2]

While acknowledging the rich histories of book-keeping and audit cultures globally as well as the many colonial origins of enumerating practices, for this article, I want to draw attention to the role of established communicative genres within communities, organizations, and cities in meaning-making. To elaborate, especially within the informal sector in India,[3] where finding work, job referrals, credit practices, and more happen through family, friends, and local community networks, establishing and maintaining trust and familiarity are crucial to all transactions and interactions. Extending these 'mental models' to ridesharing apps, I noticed that Indian ridesharing drivers also make conscious choices to drive for one company or another (or both) and gave specific reasons for why they stopped driving for a company. A common refrain was, 'Ola is a fraud company. They don't pay me the correct amount and when I call their customer care, they promise to look into it but then nothing happens.'

Drivers reported similar experiences with Uber, too, citing cheating as a reason to leave. During my interviews, drivers explained that it was both inefficient and unfair that the app did not give them passengers' phone numbers. Companies also repeatedly reminded drivers (through text messages and in training) to 'not disturb' the passenger by calling them. There are several points in the ridesharing workflow where drivers felt a lack of trust and prepared to lose the passenger or get low ratings, but that is beyond the scope of this article. Returning to the earnings dashboard, there as well, drivers expressed a sort of 'gap', a feeling of uncertainty and skepticism, an associated need to check and *own* the numbers to ensure that their weekly payments corroborated with their daily calculations. In that sense, the sheer presence of a number (or numbers) is not enough to institute or replicate familiarity and trust in transactions. There is something more, a social and an aesthetic life to how numbers are communicated and where they appear (in books, apps, screens) that is crucial to what numbers can do as governing devices.

---

2 Wanda J. Orlikowski and Joanne Yates, 'Genre Repertoire: The Structuring of Communicative Practices in Organizations', *Administrative Science Quarterly* 39.4 (1994): 541; Joanne Yates and Wanda J. Orlikowski, 'Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media', *Academy of Management Review* 17.2 (1992): 305.

3 Useful to note that despite there being disputing figures on the size of the Indian informal sector, it is widely accepted that a majority of those employed in the country work in the informal sector. While it is beyond the scope of this article, the national push and celebration of the transition from informality to formality (including small, individual, and personal changes such as using digital technologies (over paper books, cash, etc.) must then also be located within developmental and electoral politics.

Personal Accounts in Time of Apps

Fig. 2-3: How earnings are displayed in Ola, India's leading ridesharing app / A foolscap register, typically found on ridesharing cars' dashboards in India, used to jot daily accounts

Building further on drivers' claims of being 'cheated' by ridesharing companies, I noticed their experiences weren't limited to discrepancies in weekly payments. The pervasive anxiety around their *datafication* and uncertainty regarding its implications on their 'real lives' (outside of the app) was reinforced at multiple levels. For instance, drivers reported that Ola and Uber would arbitrarily change their incentive models. As a business strategy, ridesharing companies are known to initially offer humongous 'incentives' (monetary amounts to match actual per ride earnings and guaranteed income just for staying online for a fixed period) to build a reserve force of drivers and to instill confidence among their customers in new markets. In India, too, drivers signed up in droves after seeing the incentives their peers received through the apps. However, as more drivers signed up and incentives decreased, the following trends emerged as consequences of the two dominant incentive models.

In the 'number of trips' model, drivers are incentivized to do as many trips as they can, leading to them canceling on passengers who are far away or driving past them without picking them up. In the 'earnings model', drivers are motivated to do longer trips to make each trip count. Both models are, of course, pitched against Bangalore traffic that fundamentally shapes and limits how much value drivers can extract from an hour's worth of being on the road at a certain time of the day in a certain area. Reduced incentives also exacerbated the effects of 'soft control' by design (drivers get limited passenger information, no phone numbers in advance to prevent them from planning or canceling the less profitable rides). This led to a series of tactical actions where drivers learned to switch off their apps, make customers cancel rides, use multiple SIM cards, and leverage deserted spots in a city to book 'fake rides'. In response,

companies revised their interface designs to maintain unpredictability by providing 'locked' devices (smartphones with limited capabilities) as well as 'OTP' confirmations to ensure that drivers behaved in anticipated ways. It is worth mentioning that across major Indian cities, ridesharing drivers have been protesting and logging-off en masse in the face of decreased incentives and their inability to pay off leases on financed cars. Against this backdrop, for those who continue to drive, *keeping an account* of the viability of work, assessing the claims that ridesharing companies make, and sharing information and mobilizing against unfair practices have become key to surviving as a ridesharing driver in India.

Questions of profitability and viability don't have straightforward answers, which, I argue, can only be answered by rendering information intimate and situating it contextually. To be able to work for ridesharing apps, drivers have had to make investments in smartphones, data packs, buying cars, and, of course, the daily and less visible expenses of fuel, maintenance, tolls, and so on. They also must endure a fair amount of risk as they come to unwittingly represent ridesharing companies in their fight against auto-rickshaw and taxi unions. While the app dashboard gives them a solid earnings figure with a breakdown of earnings and incentives, clearly these numbers in isolation cannot convey the feasibility of ridesharing work without *accounting for* the expenses I mentioned above.

This, in turn, highlights the gap between the rhetoric of new (algorithmic and datafied) technologies and the way their promises are realized by those working with and through them. Questions regarding the viability of ridesharing work can then be answered only by reckoning with the 'intimate space', where the person encounters algorithmic management, where one's own body marked by class, caste, gender, age, and other affordances and restraints configures the relative *personal* investment and earnings through ridesharing.[4]

For instance, another Bangalore ridesharing driver told me that he preferred working for apps rather than working for fleets that provided drivers to tech company employees because of the 'relative flexibility'. What he meant is that he did not mind working as much as twelve to fourteen hours a day or even forgoing his Sundays as long as he could 'take-off' any time his pregnant wife wanted his help. Driving for ridesharing companies also allowed him to use his vehicle in his own time for when family members visited or if he wanted to take up an outstation duty. Given that the per-hour (surge) rates change in ridesharing and so do the expenses (depending on the 'cost' of plying to busier or distant areas), there are no easy answers to determine fair compensation or work timings for ridesharing drivers, especially in India where informal workers such as drivers operate at the intersection of loose enforcement of labor protections, their necessity to earn for subsistence, and the materialities they operate in. In that sense, how ridesharing drivers' collective fates might be determined within a city depends on numerous factors such as the historical presence of transport unions, the influx of high-skilled, affording immigrant workers as well as the social conditions produced because of urban development and governance. And this is where the personal account book steps in. To be able to do the math that ridesharing apps don't do *for* the drivers, to be able to recon-

4 Alexander R. Galloway, *Gaming: Essays on Algorithmic Culture,* Minneapolis: University of Minnesota Press, 2006.

figure the 'flattened' numbers of earnings, trips, and incentives presented within ridesharing apps, drivers maintain their own record book. It allows them to not only have a more enduring record of their history of earnings (versus the relatively ephemeral and constantly updated app accounts) but also 're-gather' their long-terms and short-term earnings as well as make notes of things they might want to dispute or clarifications they might want from ridesharing companies at the end of the day.[5] The physical account book's situated temporality—whereby drivers get to annotate time spent at work with details that matter *to them* is also crucial to the utility of a personally produced record versus the app-produced record.

Social Life of Quantification

Then, what does the symbolic persistence of the personal account book reveal to us about the life of data? As demonstrated earlier in the section on the account book, multiple things are worth noting. Firstly, quantification or enumeration of public life is not merely an exercise in producing information. Simultaneously, it is in the narrativization of information that it comes to gain a social life—data by itself (digital or otherwise) does not mean anything or could literally mean anything. Secondly, while information objects have a social life, their production and deployment as infrastructure mediating public life are crucial to shaping, encouraging, hiding, and producing specific kinds of affective exchanges. As Thrift would say of infrastructure, so is true of the 'qualculative' logics of devices such as earnings, ratings, waiting times, and fares.[6] Unless we unpack them, count against them, and most importantly render visible the interplay of algorithmic logics and physical infrastructure, we may never understand the constant frictions, lapses, and patchwork underneath neat datafied categories of ratings, efficiency, and earnings. The maintenance of a personal book of *hisaab* as I've shown here is one such kind of possibility and a provocation to think about both the social life of datafied categories and the negotiations that escape our analysis of data when we take datafied categories for granted or as transparent. We miss the 'qualculative life' of datafication as well as the ongoing negotiations such as those undertaken by drivers to make data intimate and to renarrativize it through situatedness.

To conclude, my aim behind foregrounding the continued life of a physical information object such as the account book in the ridesharing space was to push back against the popular and totalizing rhetoric of datafication that is increasingly recasting all forms of public activity to produce data as the *doer* or the driving force of social action. In a controversial article in the *Wired* magazine that received a fair deal of attention, Chris Anderson made a provocative argument about the relevance of theory in the age of 'data deluge'.[7] Much against what Gitelman sought to *counter* in her 'raw data is an oxymoron' argument, Anderson wondered if theorizing, broadly read as inferring and modeling had any place in a world of increasing

5 See Amoore and Piotukh's analysis of data analytics, ingestion, and their flattening effect on information, using Henri Bergson's work on forms of perception. Louise Amoore and Volha Piotukh, 'Life beyond Big Data: Governing with Little Analytics', *Economy and Society* 44.3 (2015): 341.

6 Nigel Thrift, 'Movement-Space: The Changing Domain of Thinking Resulting from the Development of New Kinds of Spatial Awareness', *Economy and Society* 33.4 (2004): 582.

7 Chris Anderson, 'The End of Theory: The Data Deluge Makes the Scientific Method Obsolete', *Wired*, 23 June 2008, https://www.wired.com/2008/06/pb-theory/.

real-time data available at our fingertips.[8] His provocation is, in a sense, about the self-evident facticity of data, especially of numerical information as its own truth and at times, a motivator for specific kinds of socio-technical solutions. Through this short essay, by illuminating the actions that happen around data and by dwelling on the material and affective life of a data form in use, I have attempted to skewer the mythology of datafied interactions as naturally transparent and efficient. In response, I have located the personal account book as one such intimate 'other'—a tactical way of reshaping big data to posit what user-centric data practices may look like.

8 Lisa Gitelman, "*Raw Data" Is an Oxymoron*, Cambridge: Mass.: The MIT Press, 2013.

References

Amoore, Louise and Volha Piotukh. 'Life beyond Big Data: Governing with Little Analytics', *Economy and Society* 44.3 (2015): 341–366.

Galloway, Alexander R. *Gaming: Essays on Algorithmic Culture,* Minneapolis: University of Minnesota Press, 2006.

Gitelman, Lisa. "*Raw Data" Is an Oxymoron*, Cambridge: Mass.: The MIT Press, 2013.

Leach, Fiona, and Shashikala Sitaram. 'Microfinance and Women's Empowerment: A Lesson from India', *Development in Practice* 12.5 (2002): 575–588.

Orlikowski, Wanda J., and Joanne Yates. 'Genre Repertoire: The Structuring of Communicative Practices in Organizations', *Administrative Science Quarterly* 39.4 (1994): 541–574.

Thrift, Nigel. 'Movement-Space: The Changing Domain of Thinking Resulting from the Development of New Kinds of Spatial Awareness', *Economy and Society* 33.4 (2004): 582–604.

Yates, Joanne, and Wanda J. Orlikowski. 'Genres of Organizational Communication: A Structurational Approach to Studying Communication and Media', *Academy of Management Review* 17.2 (1992): 299–326.

13. UNTIDY DATA: SPREADSHEET PRACTICES IN THE INDIAN BUREAUCRACY

AAKASH SOLANKI

Introduction

'…I want my MIS to be *sacrosanct*…', demanded a senior bureaucrat, in charge of an e-governance program being executed by the education department of a north Indian state, during his first meeting with me in 2015. The program was being run from the department headquarters in Datanagar.[1] I had joined the program for a year as a volunteer-researcher to help improve the utilization and effectiveness of a management information system (MIS) for tracking school resources, personnel records, and student performance, among other parameters. A multinational consulting firm and a transnational funding agency had already been helping the government build the MIS.

The department wants the MIS to become the 'single source' for all information required by its staff. This MIS is expected to reduce the number of data requests sent from Datanagar to schools so that teachers would be able to focus more on teaching.[2] Once fully built, it will purportedly have accurate data about schools, students, teachers, and other staff of the department. It will also allow staff to generate predefined reports and, when needed, customized ones as well. It is also supposed to enable the department to redesign some of its critical administrative processes such as staff appointments, transfers, and promotions to make them more efficient.[3]

Hitherto, the department did not maintain a centralized repository of data about schools, students, and employees in the state. It has digitized information since the 1990s through a series of e-governance projects. However, neither do these databases interoperate with one another (albeit except via printed artifacts on paper) nor are they accessible online, maintaining a single data interface for all stakeholders. They are often not up to date.[4] The MIS's principal

1 A pseudonym.

2 India's Right to Education Act mandates that the teachers be only asked to teach and not do administrative work. Nevertheless, not only does the Indian state routinely engages teachers in administrative work, but it also involves them in elections, census activities, and much more, making teachers rather important actors in the political processes of Indian democracy. A reader familiar with South Asian life may not need further explanation of the consequences of such an arrangement within electoral politics.

3 However, please note that as of 2015 and much of 2016, major organizational restructuring projects proposed by the consultants had been eventually shut down by the senior bureaucrats saying that the government staff was being lazy and did not need organizational structuring but some reprimanding. The saying goes, '*Sarkar me sab dande se chalta hai*' (Everything in the government runs by the stick [not carrots]).

4 Except for the Transfers database, which has been in use since the early 2000s and continues to have near 100% accuracy given the stakes involved at both ends. The department wants to make sure that it

goal is to digitize administrative data in an accessible manner such that it becomes the only information source for every decision. Before the MIS, to make a decision, Datanagar staff would send requests to individual schools and teachers to put together reports regarding the numbers of students in various categories.[5] According to the management consultants and senior bureaucracy whom I interviewed, such a workflow was vulnerable to manipulation and introduced inordinate delays in decision-making. They would get different numbers from different sources, and on many occasions, they were unable to contact the concerned school representative on time. Despite having made significant strides in developing the MIS and numerous impositions by the higher bureaucracy to compulsorily use MIS data for decision-making, a large number of staff members continue to fall back on the older modus operandi owing to a lack of trust in the MIS's data. The administrative burden exceeded the academic work of the teachers, and the consultants and senior officers speculated it to be a significant reason for low learning level outcomes in this state. The MIS project emerged in this broader context to make the department a data-driven organization.[6]

Nevertheless, what does it mean to become a 'data-driven' organization? Does digitization—the transition from paper-based files and documents to digital—entail a move to 'data-drivenness'? How may we study the many ways in which the ideology and the practice of new media technologies constitute and reconstitute the materialities of Indian bureaucracies? What artifacts do these technology projects introduce in the everyday working of the bureaucracy? How do they affect and are affected by the micro-practices of the bureaucracies? In this essay, I provide brief vignettes of Datanagar bureaucracy's efforts at becoming 'data-driven' in nature and to reflect upon some possible ways to engage with the emergent changes in technology and governance.

I focus on how the newer regimes of data literacy and numeracy are changing bureaucracies, which are coming to terms with the newer forms of data collection, analysis, and dissemination being introduced to their workflows. Studying how these projects are changing the micro-practices of bureaucracies is helpful for understanding the lives of data at large.

is paying only those who are still eligible to work and the right amount based on requisite seniority and other parameters. The teachers, on the other hand, want to make sure that the data about their position in a school is always up to date so that they keep receiving their salary and are considered legitimate staff.

5 Categories here imply students belonging to various affirmative action groups such as scheduled castes, scheduled tribes, physically disabled, and religious minorities.

6 One of the prime goals of building the MIS is to avoid leakage in the public distribution system, so much so that avoiding leakage is far more critical than improving learning level outcomes. It is just that the two goals seem to have gotten conflated in public debate. However, the senior officers I interacted with seem to have a very nuanced understanding of the goals of the MIS vis-à-vis improving education outcomes in the state as against improving leakage of funds in the public delivery system.

Data Work

One set of the department's everyday activities comprises responding to requests for transfer made by teachers. For every request, the Transfers branch in Datanagar evaluates how they should respond. The staff works with data on students, schools, teachers, and the transfer policy to model how the transfer may affect the distribution of teachers across the state. The concerned branch works out whether the student-teacher ratio will be maintained, whether the teacher is eligible for a transfer given their seniority, past transfer requests, time to retirement, and other factors such as the influence of political networks.

Hitherto, they did this work in the following manner. To know student and teacher numbers, Datanagar staff consulted school principals by calling them over the phone. They then used a separate database (offline, accessible only on one computer in the office) to pull the teacher's employment history. However, they often relied on data acquired over phone calls to concerned schools and subordinate officers at the district and block level offices. The calls would percolate from headquarters in Datanagar to districts and from the districts to blocks and from blocks to schools, or at times, the staff in Datanagar would break the rank hierarchy of information flow and directly consult teachers in a particular school.

The MIS purports to replace the 'phone call' by providing data in the form of PDFs, spreadsheets, and HTML pages. The consultants, bureaucrats, and the international funding agency believed that the need for calling over the phone would go away, as such data would be readily available to the staff at the click of a few buttons. They believed that the adoption of MIS is bound to save a considerable amount of time, paper, energy, and money in basic decision-making and welfare distribution. In the next section, I illustrate what happens when the definition of what counts as a trusted data point shifts from the interpersonal aural data acquired over phone to tabular numerical data directly accessed via the MIS.

Avoiding Phone Calls, Processing Spreadsheets, Printing PDFs

The department has made some datasets public on the MIS portal, which can be downloaded as spreadsheets accessible in Microsoft Excel.[7]

7 And other open-source spreadsheet processing software. However, given Microsoft's market share in office computing, original and pirated copies of Microsoft Office suites are more commonly used.

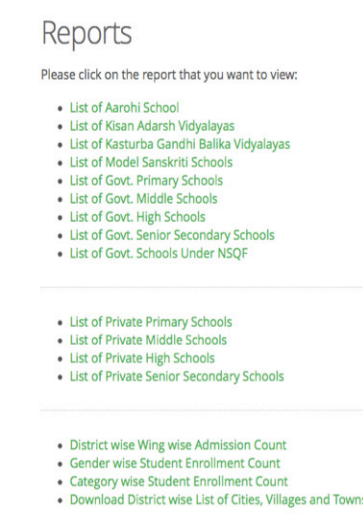


Fig. 1: List of reports made publicly available

Let us look at the dataset, 'District-wise, Wing-wise Admission Count', for one district. It is formatted in the following manner.

Count of Admitted Students(SRN)										
Classes 1-5 in Primary Schools	Classes 6-8 in Independent Middle Schools	Classes 9-10 in High Schools	Classes 11-12 in Senior Secondary Schools	Teacher Education Institutes	Classes 6-8 in Middle wing of High Schools	Classes 6-8 in Middle wing of Senior Secondary Schools	Classes 9-10 in High wing of Senior Secondary Schools	Total	Classes 1-5 in Primary Schools	Classes 6-8 in Independent Middle Schools
29122	7230	5834	6558	0	5882	9595	12170	76391	113	5
29122	7230	5834	6558	0	5882	9595	12170	76391	113	5

Visitor on 17 Jun, 2017 04:44:03 PM 1 of 1

Fig. 2: Default view of data in 'District-wise, Wing-wise Admission Count' file

To anyone familiar with spreadsheet processing, it will be apparent that such a presentation of data—even if born-digital—is not amenable to even fundamental Pivot Table analysis, one of the initial steps in many data analysis workflows. A Pivot Table is a data summarization tool found in the spreadsheet processing software. It is used to sort, count, sum, or average data stored in a spreadsheet and make some basic graphs. One can have several data summaries

by dragging and dropping fields graphically, as shown in the Pivot Table Builder for MS Excel 2011 for Mac in the figure below. Let us try Pivoting the data in Fig. 2.

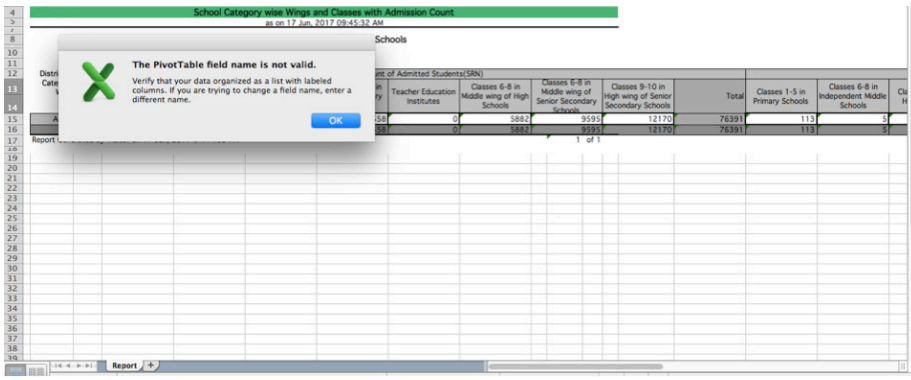


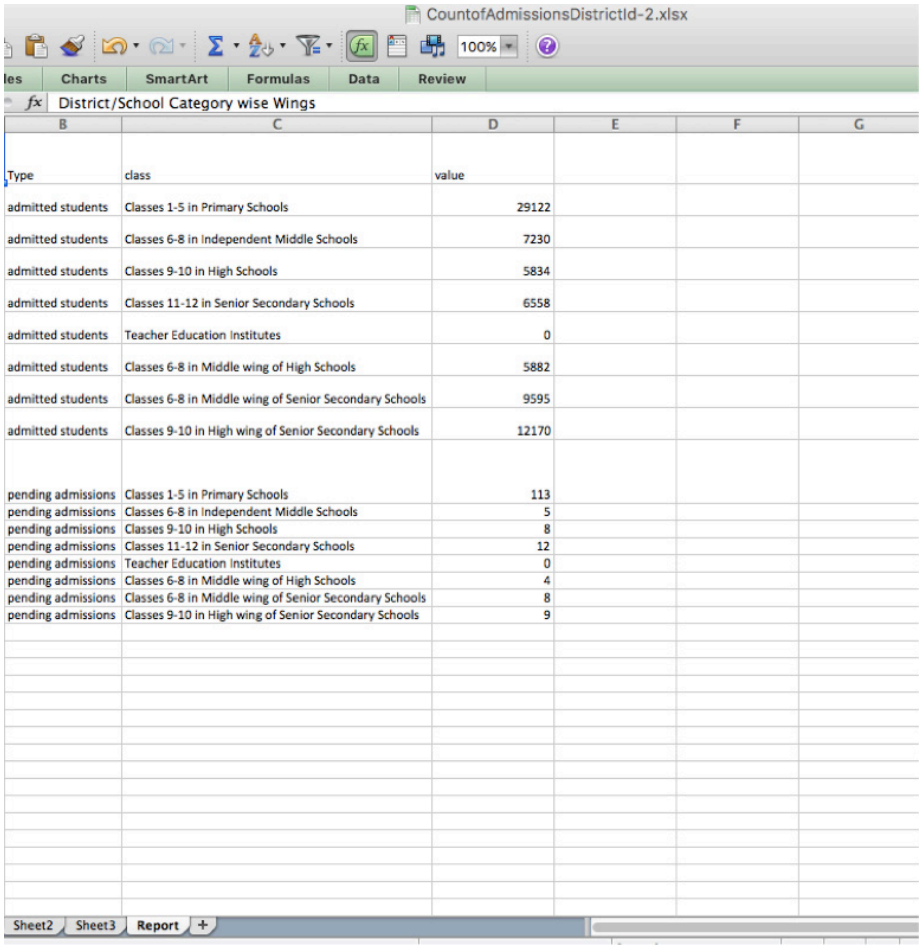
Fig. 3: Error prompt activated upon trying to run Pivot Table analysis on data organized as in Fig. 2

Upon trying to carry out a Pivot Table analysis, we run into the error shown in Fig. 3. Excel prompts the user to make sure that the user organizes the 'data' as a list with labeled columns. This error alludes to the notion of 'tidy data', which is common among those with a background in statistics, computer science, and allied fields. Hadley Wickham, an influential statistician in the data science community, defines 'tidy data' as:

> …a standard way of mapping the meaning of a dataset to its structure. A dataset is messy or tidy depending on how rows, columns, and tables are matched up with observations, variables, and types. In tidy data:
>
> 1. Each variable forms a column.
>
> 2. Each observation forms a row.
>
> 3. Each type of observational unit forms a table.[8]

8 Hadley Wickham, 'Tidy Data', *Journal of Statistical Software* 59.10 (2014): 14, https://doi.org/10.18637/jss.v059.i10.

Applying Wickham's exposition and transforming the spreadsheet in Fig. 2 yields Fig. 4.


CountofAdmissionsDistrictId-2.xlsx

Charts SmartArt Formulas Data Review

District/School Category wise Wings

	B	C	D	E	F	G
	Type	class	value			
	admitted students	Classes 1-5 in Primary Schools	29122			
	admitted students	Classes 6-8 in Independent Middle Schools	7230			
	admitted students	Classes 9-10 in High Schools	5834			
	admitted students	Classes 11-12 in Senior Secondary Schools	6558			
	admitted students	Teacher Education Institutes	0			
	admitted students	Classes 6-8 in Middle wing of High Schools	5882			
	admitted students	Classes 6-8 in Middle wing of Senior Secondary Schools	9595			
	admitted students	Classes 9-10 in High wing of Senior Secondary Schools	12170			
	pending admissions	Classes 1-5 in Primary Schools	113			
	pending admissions	Classes 6-8 in Independent Middle Schools	5			
	pending admissions	Classes 9-10 in High Schools	8			
	pending admissions	Classes 11-12 in Senior Secondary Schools	12			
	pending admissions	Teacher Education Institutes	0			
	pending admissions	Classes 6-8 in Middle wing of High Schools	4			
	pending admissions	Classes 6-8 in Middle wing of Senior Secondary Schools	8			
	pending admissions	Classes 9-10 in High wing of Senior Secondary Schools	9			

Sheet2 Sheet3 Report +

Fig. 4: Data in Fig. 2 after being organized according to 'tidy data' principles of Hadley Wickham (2014)

Now let us attempt Pivot analysis once again. Only by organizing the data, as shown in Fig. 4, can one engage in any Pivot Table analysis.

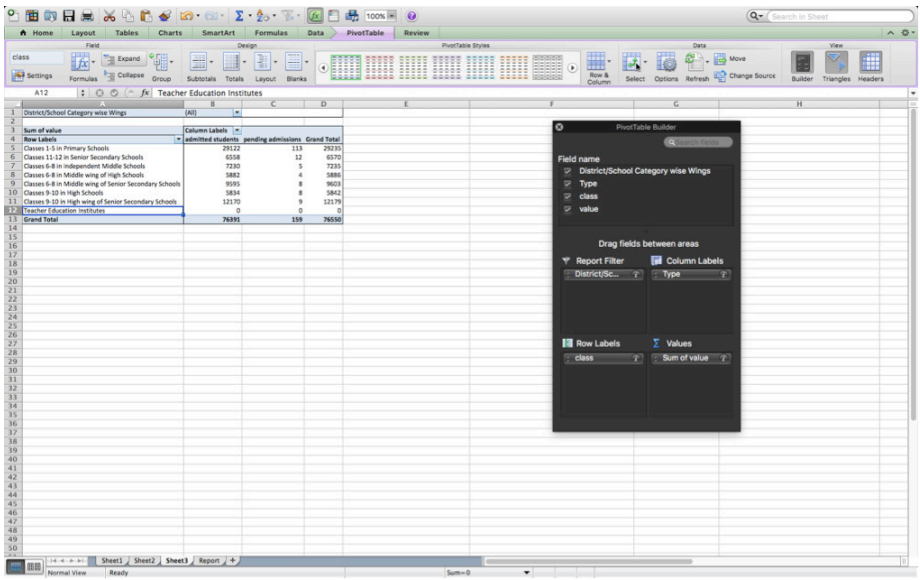


Fig. 5: Now that the principles have been followed, the Pivot Table view is available without error

Upon looking at these and many other datasets available on the portal, I came to wonder why Wickham's 'tidy data' principles do not find increased usage in the department despite its emphasis on data-drivenness. Based on interviews with the MIS developers, it seems that they understand 'tidy data', but they could not explain why the MIS reports were formatted as shown in Fig. 2. They pointed to the 'government's requirements' as the reason for this. A management consultant working on this project told me that senior officers in Datanagar realize that there is little digital numeracy among most staff members; most of them can barely operate a computer, let alone do intermediate and advanced spreadsheet work. They want the developers to pre-build all possible data analyses into the MIS portal such that the government staff finds the learning curve of using the MIS to be low. However, they made it to be almost flat. The only skill that the staff needs to develop is the ability to log in to the MIS, identify the required report, download a PDF, print it, and put it in a government file, which then takes on a life of its own. And with that, in the garb of the PDF as an electronic document, the form of paper's materiality returns, albeit with new digital entanglements.[9]

9 Matthew S. Hull, *Government of Paper the Materiality of Bureaucracy in Urban Pakistan*, Berkeley: University of California Press, 2012.

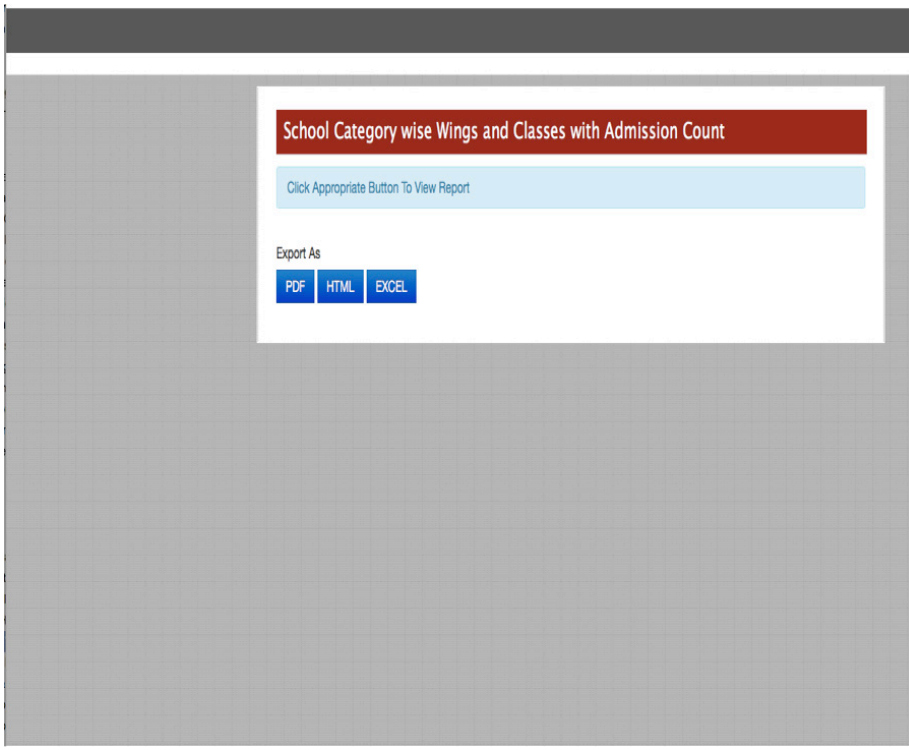


Fig. 6: Notice how the PDF option is presented first, (left to right eye movement) to the user

Endurance of the PDF

Once a staff member prints out and places a digitally generated document in a paper file, it becomes an 'application'. This 'application' then goes around the requisite hierarchy of concerned officials, whose comments have to be first incorporated into the document before it can become a 'noting' or a 'government order' (GO). Usually, senior officials do not have the typing speed to create documents on their own, so a staff of 'computer operators' is in place to type the document, print a copy, and hand it to the office to make corrections.

At times, computer operators do this on a spare desktop computer available at the officer's desk or the officer annotates the printed copy with a pen (usually a red ink pen). The operator is then sent off to edit the document based on the corrections and send an updated version to the officer.[10] If satisfied with the grammar, structure, and aesthetics of the document, the officer gives it a go-ahead, signs it off with other instructions that are given using pen on paper only and the application, making a file (with a designated file indexing number), which

10 I cannot develop it further here, but I am using the descriptor 'operator' in connection with the figure of the 'scribe' to imply a more extended connection between the contemporary digital labor and the scribal labor worlds articulated by Bhavani Raman. See, Bhavani Raman, *Document Raj: Writing and Scribes in Early Colonial South India*, Chicago: The University of Chicago Press, 2012.

becomes a 'noting' and is passed onto a different officer for further evaluation. Once an officer makes the decision, the noting becomes a GO by the power of a designated officer (usually from the Indian Administrative Service). It is then 'WhatsApped' or emailed to the concerned officer. Emailing or WhatsApp sharing usually entails taking a picture of the GO from a mobile phone camera and sending it as an email attachment or WhatsApp media to the concerned staff members.[11] The concerned staff then prints a picture of the order, makes annotations on paper, and again takes a photograph and shares it via WhatsApp or email to the staff members on the other end. Thus, nowhere in the entire communication process in the bureaucracy is any work purely paper-based or digital. The staff marshals various media to get the work done.

It appears to me that there is a certain finality that the PDF, and even more, a printed copy of the PDF, affords which the spreadsheet does not. The PDF allows for 'structuring authorship' in a manner that perhaps the spreadsheet processors do not.[12] The PDF author(s) is separated from the reader(s), and this relative non-editability of the PDF touches upon a certain valence in bureaucracies that are anxious about one or the other staff member changing details of a document for personal gains at the expense of public loss. When a PDF generated from the MIS in a format as shown in Fig. 2 is printed on paper and looked at by government officers and other staff, it is easier to access compared to the PivotTable analysis of 'tidy data' of MS Excel and other spreadsheet processing software. This points to a certain enduring capacity of paper and how it conditions the experience of engaging with the information presented on a piece of paper despite the concentrated efforts to digitize, almost globally.

Conclusion

The transition from paper to digital is not a simple one of permanently letting go of paper-based bureaucratic record-keeping and adopting a wholly digitized workflow. While their digital counterparts are replacing paper files inscribed by age-old writing practices, bureaucratic work continues to be governed by the media technology of paper as bureaucracies develop newer, hybrid media practices. Hardly any of the government files are purely paper-based and handwritten, or even typed on a typewriter anymore. Increasingly, files are first generated as a Microsoft Word document, and the .doc or .docx file is then printed out on what is usually an A4 sheet of paper. It is then made part of an official government 'file'.[13] However, depending on the requirements, it could be a government letterhead or the many kinds of legal or non-legal paper. Instead of focusing on digitization or the digital, it is helpful to think of the current epoch in which the analog and the digital are being used continuously in tandem.

The brief ethnographic vignettes presented in this chapter raise some crucial questions for further research. How do technology projects challenge our understanding of what develop-

11 Such micro-practices of new media usage in everyday life of the Indian bureaucracies need further scrutiny. How staff, bureaucrats, and others make sense of and engage with new media may help us understand the culture, which is leading to various Aadhaar related data leaks in India.

12 Lisa Gitelman, *Paper Knowledge: Toward a Media History of Documents*, Durham: Duke University Press Books, 2014, pp. 111–135.

13 This situation persists even though government offices are gradually adopting the use of 'e-Office', a digitized file workflow system developed by the National Informatics Center.

ment projects do, when 'improvement' is delegated to the vagaries of data analysis contingent upon experts comprised of human-algorithm assemblages?[14] Geoffrey Bowker has argued how we record knowledge inevitably affects the knowledge that we record.[15] If databases record knowledge about knowledge imparted and received in schools, how do the paper and digital interfaces that capture it affect that knowledge? How do they limit, obscure, or overdetermine what is measurable and improvable about education? What kinds of publics do such information infrastructures mobilize? How do the specific materialities of information systems allow for the kind of information processing they allow as in the case of the MIS?[16] While it has been a practice at least since independence, more and more Indian government agencies now hire a large number of external consultants for projects small and large. How are these sets of expertise being mobilized, and how do they affect and are affected by the micro-practices of the bureaucracies they intend to influence? A lot more work needs to happen in this direction, but looking at the micro-practices of spreadsheet management in state bureaucracies provides a crucial foray into the lives of data from India.

14 Tania Murray Li, *The Will to Improve: Governmentality, Development, and the Practice of Politics*, Durham: Duke University Press Books, 2007.

15 Geoffrey C. Bowker, *Memory Practices in the Sciences*, Cambridge, Mass.: The MIT Press, 2008.

16 Paul Dourish, 'No SQL: The Shifting Materialities of Database Technology', *Computational Culture* 4 (November, 2014), http://computationalculture.net/article/no-sql-the-shifting-materialities-of-database-technology; Paul Dourish, 'Spreadsheets and Spreadsheet Events in Organizational Life', in *The Stuff of Bits: An Essay on the Materialities of Information*, Cambridge, Mass.: The MIT Press, 2017, pp. 81–104.

References

Bowker, Geoffrey C. *Memory Practices in the Sciences*, Cambridge, Mass.: The MIT Press, 2008.

Dourish, Paul. 'No SQL: The Shifting Materialities of Database Technology', *Computational Culture* 4 (November, 2014), http://computationalculture.net/article/no-sql-the-shifting-materialities-of-database-technology.

_____. 'Spreadsheets and Spreadsheet Events in Organizational Life', in *The Stuff of Bits: An Essay on the Materialities of Information*, Cambridge, Mass.: The MIT Press, 2017, pp. 81–104.

Gitelman, Lisa. *Paper Knowledge: Toward a Media History of Documents*, Durham: Duke University Press, 2014.

Hull, Matthew S. *Government of Paper the Materiality of Bureaucracy in Urban Pakistan*, Berkeley: University of California Press, 2012.

Li, Tania Murray. *The Will to Improve: Governmentality, Development, and the Practice of Politics*, Durham: Duke University Press, 2007.

Raman, Bhavani. *Document Raj: Writing and Scribes in Early Colonial South India*, Chicago: The University of Chicago Press, 2012.

Wickham, Hadley. 'Tidy Data'. *Journal of Statistical Software* 59.10 (2014): 1–23, https://doi.org/10.18637/jss.v059.i10.

14. THE WORK OF WAITING: SYNDROMIC SURVEILLANCE AND THE PARADOX OF IMMEDIACY

ANIRUDH RAGHAVAN

India is witnessing what some have called a 'revolution in epidemic intelligence' with the state investing in a disease surveillance network stretching through the district, state, and national levels, supported by a digital network for the rapid transmission of data between experts and health workers across the country.[1] The large number of institutions that hitherto managed disease intelligence have been aggregated into a single body, the Integrated Disease Surveillance Programme (IDSP). The IDSP, established in 2004, marks a shift in the mode of epidemiological governance, as the management of disease is seen as a problem of information.[2] Hospital capacity, availability of medicines, and medical training were the anchors for disease management in the 20th century. Increasingly, the management of epidemics in a population is as much a problem of containing infection as it is of modulating the flow of information about the disease-event.[3]

Further, the epistemic value of information is distinct in this mode of action. The surveillance techniques of the 19th century focused on the collection of information after the disease-event. Thus, laboratory reports confirming the existence of an epidemic in a region were the primary source of data. Syndromic surveillance is focused on detecting a disease-event in its emergence. The attempt is not to prevent the event but to prepare for its eventuality. Non-specific data sources such as media reports, reports by para-medical workers, and pharmaceutical sales data are aggregated and analyzed for the clustering of syndromes such as cough or cold in specific regions. This generates warning signals, alerting authorities to the possibility of an epidemic.[4]

---

1 Vivek Singh and Biranchi Jena, 'Syndromic Surveillance in the Integrate Disease Surveillance Programme and Pre-Hospital Emergency Care in India', *International Society for Disease Surveillance*, lecture, 26 April 2012, https://knowledgerepository.syndromicsurveillance.org/syndromic-surveillance-integrated-disease-surveillance-project-isdp-and-pre-hospital-emergency-care.

2 Rajeev Sharma et al., 'Communicable Disease Outbreak Detection by Using Supplementary Tools to Conventional Surveillance Methods under Integrated Disease Surveillance Project (IDSP), India', *The Journal of Communicable Diseases*, 41.3 (September, 2009): 149.

3 See, Donna Haraway, 'The Biopolitics of Postmodern Bodies: Determinations of the Self in Immune System Discourse', in Janet Price and Margrit Shildrick (eds.), *Feminist Theory and the Body: A Reader*, New York: Routledge, 1999, pp. 203–214; for a defining position on the diminishing role of the physical body and the clinic with the rise of information as the foundation for governmental techniques. For a critique of this position, see, Roma Chatterji et al., 'Death of the Clinic? Normality and Pathology in Recrafting Aging Bodies' in Janet Price and Margrit Shildrick (eds) *Vital Signs: Feminist Reconfigurations of the Bio/logical Body*, Edinburgh: Edinburgh University Press, 1998, pp. 171–196.

4 Lyle Fearnley, 'Signals Come and Go: Syndromic Surveillance and Styles of Biosecurity', *Environment and Planning A: Economy and Space* 40.7 (2008): 1615.

The IDSP may be located within the larger rubric of global health security, a paradigm of governmentality that merges the concerns of public health and national security.[5] Disease no longer emerges when a diseased body is identified in the clinic or hospital, or made visible to the technologies of the state, but always already exists in potential. The mutability of the microbial genome, the rapidity of zoonotic microbial transfers between animals and humans, the global network of human and animal transport, and the acquisition of antibiotic resistance genes in bacterium become the multiple sites at which a disease emerges.[6] The disease-object cannot be subject to fixity and this ruptures any possibility of fully quelling it.[7] Governmental techniques must work to pre-empt disease-events to prevent a cascading catastrophe. To this end, surveillance moves from a panoptic location and isolation of bodies to an archival mode—the detailed, ever-expanding collection of data about the population and an algorithm-based visualization of data patterns that become the grounds for political action.[8] Big Data-driven disease surveillance is focused on acting on predictive assumptions rather than the confirmation of disease.[9] 'Real-time' coverage and 'immediacy' are the key terms in this global political technology geared towards pre-empting catastrophic disease-events.

This essay will investigate the contradictory framings of immediacy as a technological and political project, through the constitution of 'real-time' as necessarily supplemented by its unwanted other, waiting. Through two ethnographic scenes of the emergence of an H1N1 epidemic in Delhi in 2015, I will investigate 'waiting' as a modality of mediation between two sets of actors—data analysts and health workers, and the surveillance institution and the media. Waiting as a temporal space, I will argue, is the site within which the meanings and value of immediacy and real time are negotiated.

A New Object

The history of syndromic surveillance may be written as a convergence of three distinct yet interrelated genealogies in public health and security. The first is a shift in the object of medical surveillance from individual bodies to populations. Before the 1950s, surveillance proceeded along the disciplinary lines outlined by Foucault for the 18th century—the constitution of panoptic structures to make abnormal bodies visible and the segmentation of these

5 For a comprehensive survey of global health security literature, refer to Carlo Caduff, *The Pandemic Perhaps: Dramatic Events in a Public Culture of Danger,* Oakland: California University Press, 2015.

6 Ibid.

7 Nicholas B. King, 'The Scale Politics of Emerging Diseases', *Osiris* 19 (2004): 62–76.

8 I draw this concept of the archive as a political technology from Feldman's work on the loss of a structured enemy in post-Cold War US and, more pointedly, from Cohen's work on the Aadhaar biometric identification project as the imagination of the nation as archive itself. Cohen argues that with Aadhaar one may see the emergence of a political technology wherein the archive does not serve external objectives but is its own object. See, Allen Feldman, *Archives of the Insensible: Of War, Photopolitics and Dead Memory*, Chicago: University of Chicago Press, 2015; Lawrence Cohen, 'Duplicate, Leak, Deity', *Limn* 6, (2016), https://limn.it/articles/duplicate-leak-deity/.

9 The term Big Data is used, here, to refer to the reliance on large data sets for governance. Big Data is also used in a more specific sense to non-relational databases which are used by social media platforms like Facebook to aggregate vast amounts of varied data.

bodies in space.[10] However, in 1955, faced with a resurgence of polio in the US, the Centre for Disease Control (CDC) instituted a system of continuous surveillance of mortality data and epidemiological reports from sentinel laboratories across the country, thus shifting its focus from the management of the body to the generation and classification of information on diseases in a population, across space and time.[11]

The second genealogy is a shift within microbiology in the ontological status of the infectious entity—the microbe. By the 1980s, the disease-causing entity was understood to have been conquered by means of antibiotics and vaccination. But beginning in the 1990s, microbiologists Joshua Lederberg and Stephen Morse challenged this narrative. In a definitive report titled *Emerging Infections*, the microbe was defined as an ever-changing entity subject to constant mutation aided by the greater density of interaction between humans and animals and the smooth flow of transmission made possible by modern transportation.[12] The AIDS epidemic of the '80s and '90s marked the age of the 'new' emerging disease, constantly rebuking efforts at complete control.

The third is a shift in the conception of risk within surveillance at large, in the aftermath of 9/11. A global security apparatus came into being which also transformed disease management. Firstly, the object of risk became virtual as security was conceptualized as always and constantly at threat from events in potential. Secondly, the surveillance network was required to be both global in the scope of its coverage and decentralized in the ability to mobilize action.[13] Eugene Thacker describes this as the becoming-virus of the state and the surveillance apparatus.[14]

With the convergence of these developments in the 21st century, the possibility of pandemic events became a catastrophic threat to the nation's (specifically the US) security and global health. Pandemic events were conceived as not just natural occurrences but as fundamentally social—the result of a deliberate terror plot or an inadvertent result of an infection spreading through air transport networks. It was in this milieu of urgency that the first syndromic surveillance systems were developed to enable rapid action upon the possible emergence of disease.

In India, the shift in the political technology of epidemic management can be traced to the aftermath of the 1994 Surat plague. The epidemic was the first to receive global coverage as both CNN and BBC dedicated prime broadcast time to the resurgence of a 'medieval disease'

10 Michel Foucault, *Security, Territory, Population: Lectures at the College De France, 1977-78*, Springer, 2004.

11 Lorna Wier and Eric Mykhalovskiy, 'The Geopolitics of Global Public Health Surveillance in the Twenty-First Century', in Alison Bashford (ed.) *Medicine at the Border: Disease, Globalization and Security, 1850 to the Present*, London: Palgrave Macmillan, 2007, pp. 240–263.

12 Joshua Lederberg, Robert Shope, and Stanley Oaks (eds) *Emerging Infections: Microbial Threats to Health in the US*, National Academic Press, 2004.

13 See, Stephen Collier, Andrew Lakoff, and Paul Rabinow, 'Biosecurity: Towards an Anthropology of the Contemporary' in *Anthropology Today* 20.5 (2004): 3. Also see, Melinda Cooper, 'Pre-empting Emergence: The Biological Turn in the War on Terror', *Theory, Culture and Society* 23.4 (2006): 113.

14 Eugene Thacker, 'Living Dead Networks', *The Fibreculture Journal* 4 (2005), http://four.fibreculturejournal.org/fcj-018-living-dead-networks/.

in the far corner of the Third World. The embarrassed Congress government immediately initiated experiments in modernizing the country's limited and ailing surveillance infrastructure. This resulted in the National Surveillance Program for Communicable Diseases, which was replaced by the World Bank-funded Integrated Disease Surveillance Project (IDSP) in 2004, under which an extensive digital network of epidemiological intelligence was established by linking district-level health centers with state and national level surveillance and coordination centers. The IDSP was integrated into the landscape of state-funded disease management in 2010, with a specialized allocation in the Centre's budget.[15]

The IDSP's mandate is twofold: to integrate a surveillance network across three tiers—district, state, and national—and to provide a platform for decentralized action on warning signals through the training and maintenance of Rapid Response Teams (RRTs) to follow-up on the early warning provided by data visualization. Surveillance is carried out through three paper-based and computerized forms used for reporting. Form L is collected from sentinel labs across the country and provides the traditional data confirming the existence of disease in a given region. Forms P and S constitute the syndromic components. Form P is to be filled by nurses and doctors to reflect the initial diagnosis after a check-up and Form S is filled by paramedical workers to reflect the number of persons with symptoms such as cough and cold in a given community or ward. The syndromic data is analyzed by an algorithm that detects any unusually high clustering of symptoms in a region, alerting the RRT to perform a ground-level investigation.

The IDSP functions within a larger ideological framework of Big Data and algorithmic decision-making, wherein data as a semiotic entity is seen to possess 'immediacy'.[16] Immediacy has two resonances—the first is im-mediation. That data can reveal its meaning without mediation or translation. The second is immediate in the temporal connotation of the term.[5] That Big Data aggregation and algorithm-based visualization of patterns make possible an instantaneous capture of an event as it occurs. This is the imagination of a total archive—one whose purpose does not extend beyond archiving, fed by a desire to collate any and all information about a population. Thus, the IDSP website lists the increasing number of suspected epidemics reported across the country as a sign not of the failure to successfully prevent disease-events but as a sign of its coverage and the ease of communication. The increasing number of disease-events across the country reflects the efficacy of the surveillance archive as a record of emergent viral activity.

15 See, Lalit Kant and Sampath K. Krishnan, 'Information and Communication Technology in Surveillance in India: A Case Study', *BMC Public Health* 10.1 (2010): S11. Also see, K. Suresh, 'Integrated Disease Surveillance Project through a Consultant's Lens', *Indian Journal of Public Health* 52.3 (2008): 136.

16 I have drawn the concept of 'immediacy' from Mazzarella's work on e-governance in the context of the Tehelka sting operation in 2001. He argues that the promise of transparency in e-governance is the result of a hidden, back-stage mediation. What caused the public outrage, was not the revelation of corruption amongst politicians but the revelation of this inevitable structure of mediation framing the hyper-modern fetish of digital communication. See, William Mazzarella, 'Internet X-Ray: e-Governance, Transparency and the Politics of Immediation in India', *Public Culture* 18.3 (2006): 473.

However, the everyday functioning of the surveillance apparatus does not bear out the certainty with which expert literature makes claims for immediation, and neither does immediacy evoke unequivocal value.

Noise and Data Cleaning

Syndromic surveillance depends upon non-specific data sources (the ones which do not confirm the existence of a disease); consequently, the information it generates is structurally incomplete. This performative infelicity of data ensures the rapidity of perception and action.[17] But, this also results in an abundance of false positives or noise, that is, those signals that do not lead to any emergent epidemic on the ground. These false positives or noise may emerge for two reasons—either the algorithm detects a random cluster of syndromes rather than an index of an emergent epidemic or the data fed into the system for analysis is suspect. Syndromic data may also be wrongly recorded, misreported, reported with some delay, or symptoms might have been misidentified. Thus, both the algorithm itself and the workers who generate the data on the ground could generate noise. Noise, as we know from Michel Serres, functions as a parasite forcing the entire system to wrap around itself to accommodate it.[18] The false positives in the IDSP system perform a similarly parasitic role by placing enormous pressure on scarce manpower.[19]

Each positive signal from the system must be supplemented by an on-ground investigation for the surveillance mechanism to work. However, the quantum of signals generated by the system far outstrips the ability to conduct a follow-up. The system constantly operates within a lag between archival perception made possible by algorithmic processing and manual on-ground action. This lag is not simply the result of poor resource allocation for public health in India but is structured into the system. The algorithmic processing is designed to be rapid, processed as it is by computer-driven analytic power, while the on-ground follow-up by the team of microbiologists, epidemiologists, and pathologists is far slower. A typical on-ground investigation can range from one to two weeks, while corresponding warning signals are generated several times in one day. Thus, unsurprisingly, one study found that only 40% of signals resulted in a follow-up in the states of Delhi and UP.[20]

Thus, the analysts must conduct an extensive data-cleaning operation to ensure that they weed out false signals. A cleaning operation is a modality of action through waiting. The head of the Delhi state IDSP often referred to these operations as the 'true' job of

17 Carlo Caduff, 'On the Verge of Death: Visions of Biological Vulnerability', *Annual Review of Anthropology* 43.1 (2014): 105.

18 Michel Serres, *The Parasite*, Minneapolis: University of Minnesota Press, 2007.

19 Refer to Fearnley for an extensive analysis of the problem of the signal-noise ratio in the functioning of BioSense, a US-based syndromic surveillance platform developed by the CDC. See, Fearnley, 'Signals Come and Go'.

20 Manish Kakkar et al., 'Acute Encephalitis Syndrome Surveillance, Kushinagar District, Uttar Pradesh, India, 2011–2012' in *Emerging Infectious Diseases* 19.9 (2013): 1361.

the expert in a surveillance system, testing his ingenuity and innovation in an otherwise computerized environment.[21] In a cleaning operation, as one data analyst described to me—'life is invoked in the data'. This 'life' refers to the elaborate process by which data is made worthy of trust, a trust possible by means of negotiating the relationship between the analyst and the producer of data (the nurse or the para-medical worker). Alberto Corsin Jiménez notes that the trust of numbers and documents is, in fact, a mask for the establishment of relations between persons.[22] Waiting, then, becomes the supplement that makes surveillance possible through the establishment of trust between analyst and data-producer. As one analyst told me: 'Without waiting to know the data, we would all be flying blind. Acting like monkeys on every signal we receive. Our effectiveness depends on our waiting to act'.

I observed one such data cleaning operation, which took place in the Delhi IDSP in November 2015 when Delhi was gripped by an H1N1 outbreak. In a tense atmosphere, a data analyst reported a spike in fevers detected in Jehangir Puri, a low-income neighborhood in north Delhi. The unit was more strained for manpower than usual given the outbreak. Several frantic phone calls were made to different primary health centers in Jehangir Puri, with inquiries ranging from verifying the regularity with which the nurses and primary health workers reported for work to the rumors of a doctor's absence.

In the end, the signal was forgone, and the action was avoided. Upon asking the Unit Chief for the logic informing this decision, I was promptly told that Jehangir Puri was a 'dirty' area inhabited by poor, unhygienic migrants and laborers: 'Doctors are irregular there and there is no guarantee of finding a nurse in health centers. We conduct investigations there, but you know, signals from there are mostly false. They don't know how to report properly'. Another microbiologist extended this reasoning: 'It's not just that doctors there are bad but that area is full of disease anyway. How do we know that those fevers are swine flu? Given it's Jehangir Puri it must be cholera or something else. We focus on more certain signals'.

This extract reveals the double vulnerability of low-income poor neighborhoods like Jehangir Puri. They require the greatest intervention of surveillance systems given their unhygienic, infested surroundings, but also suffer the greatest discrimination as their data is perpetually subject to doubt. While waiting emerges as an essential supplement to data analytics, it also emerges as a site of exhaustion, frustration, and the denial of state services. One health worker in Jehangir Puri compared their situation within this system to being stuck in quicksand: 'We act and respond as fast as we can. We were trained to do that. But we keep waiting for the officials to respond to our warnings. They think we are not good enough'. Waiting reveals two potentials—as the site of mediation which makes data actionable and as the site of what Javier Auyero calls the 'temporal geography of state abandonment'.[23] In both cases, waiting operates in the gap that the surveillance modality opens between perception and action, between archival knowledge and its translation into a decision.

21 Personal interview, 2015.

22 Alberto Corsin Jiménez, 'Trust in Anthropology', *Anthropological Theory* 11.2 (June, 2011): 177.

23 Javier Auyero, *Patients of the State: The Politics of Waiting in Argentina*, Durham: Duke University Press, 2013.

Noise Leaks

Waiting as a modality of action also mediates the interface between the surveillance system and non-state actors including the media, professional unions, and NGOs. This became apparent to me in a controversy that engulfed the Delhi IDSP in October 2015. On the fifth of that month, two individuals (a man and a woman) succumbed to swine flu in Delhi's Safdarjung Hospital. This data was registered in the mortality records of the program, but no outbreak warning was sent. The Unit Chief reasoned that these individuals were migrant laborers from UP and Chhattisgarh, and hence the disease would have been picked up in these states. On 12 October, DNA ran a story titled 'Amid Dengue Crisis, Delhi Prepares for Swine Flu'.[24] The article announced the arrival of swine flu, citing the mortality data that was relegated as noise by the analysts.

Soon after, the Delhi Medical Association petitioned the state government to immediately provide vaccines for medical workers who were expected to be at the frontline of combating the expected epidemic. In its letter to the Chief Medical Officer (CMO) of the state, the government was chastised for undue delay in preparation in the face of an imminent threat. The officials at the IDSP vehemently opposed this petition. In a counter letter to the CMO, the Unit Chief wrote, 'This [controversy] has arisen due to imperfect analysis of data by media sources and others. Data is dangerous when used without caution and patience'.[25]

Later, the Unit Chief elaborated on his position more clearly: 'The problem always is this urgency with data. If you use data without waiting, you get panic, confusion. Data becomes counter-productive'.[26] What one can observe here is that immediacy, the much-feted objective of syndromic surveillance reliant on Big Data analytics, emerges only as a promise, constantly deferred to a future horizon. Action, negotiation, and mediated politics take place in a zone of waiting, where actors jostle to ascribe differing valuations to immediacy. The media and public health professionals find waiting to be an implication of the laxity of the state, while this laxity transforms into a necessary but hidden supplement to real-time surveillance for the analysts.

Conclusion

Carlo Caduff, analyzing biosecurity regimes in the US, drew attention to the structural infelicity of the security paradigm which operates only through the institution and maintenance of a permanent state of insecurity.[27] I take seriously his invitation to draw out the implications of this constitutive paradox. What we have here, in the context of Big Data-driven surveillance systems, is a promise of immediation, both in the senses of speed and abstract algorithmic decision-making, which is made possible through its hidden, obscene supplement—waiting.

24 'Amid Dengue Crisis, Govt. Prepares for Swine Flu', *Daily News and Analysis*, 12 October 2015, http://www.dnaindia.com/india/report-amid-dengue-crisis-delhi-govt-prepares-for-swine-flu-2132249.

25 The letter itself is private and was made available through personal communication.

26 Personal interview, 2015.

27 Caduff, 'On the Verge of Death'.

Waiting, I have attempted to demonstrate, has two potentials within it. It is a modality of action that cannot be avoided within the realm of expertise. To wait is to be patient with the data and to allow it to 'come to life'. Most crucial is the recognition that it is mediation, between health workers and experts or media and the data analysts that makes any action possible. However, when this structure of waiting becomes apparent outside of the circle of expertise, it takes on the semantic burden of revealing the ineptness and apathy of the state and its officials. Derrida puts it succinctly when he argues that the structure of waiting is a 'temporal aporia'—the object of waiting is permanently deferred but, in this deferral, it becomes intelligible.[28] Waiting is both the site where the failure of surveillance is known and where surveillance is made possible.

Recent literature on security and catastrophic imaginaries such as the Anthropocene has emphasized acceleration and speed as the markers of late modernity—an age rapidly moving towards apocalypse.[29] However, it may be argued that catastrophic imaginations of pandemic threat and terror attacks which call forth immediation as a political necessity also institute waiting as the temporal geography within which action must take place and a politics may emerge.

28 Jacques Derrida, *Aporias*, Stanford: Stanford University Press, 1993.

29 See Cultural Anthropology's curated section on Speed, especially, Vincent Duclos, Tomás Sánchez Criado and Vinh-Kim Nguyen, 'Speed: An Introduction', *Cultural Anthropology* 32.1 (2017): 1.

References

Auyero, Javier. *Patients of the State: The Politics of Waiting in Argentina*, Durham: Duke University Press, 2012.

Caduff, Carlo. *The Pandemic Perhaps: Dramatic Events in a Public Culture of Danger*, Oakland: University of California Press, 2015.

_____. 'On the Verge of Death: Visions of Biological Vulnerability', *Annual Review of Anthropology* 43.1 (2014): 105–121, https://doi.org/10.1146/annurev-anthro-102313-030341.

Chatterji, Roma, Sangeeta Chattoo, and Veena Das. 'Death of the Clinic? Normality and Pathology in Recrafting Aging Bodies', in Janet Price and Margrit Shildrick (eds) *Vital Signs: Feminist Reconfigurations of the Bio/logical Body*, Edinburgh: Edinburgh University Press, 1998, pp. 171–196.

Cohen, Lawrence. 'Duplicate, Leak, Deity', *Limn* 6 (2016), https://limn.it/articles/duplicate-leak-deity/.

Collier, Stephen J., Andrew Lakoff, and Paul Rabinow. 'Biosecurity: Towards an Anthropology of the Contemporary', *Anthropology Today* 205 (2004): 3–7.

Cooper, Melinda. 'Pre-Empting Emergence: The Biological Turn in the War on Terror', *Theory, Culture & Society* 23.4 (2006): 113–135.

Corsin Jiménez, Alberto. 'Trust in Anthropology', *Anthropological Theory* 11.2 (June, 2011): 177–196.

Derrida, Jacques. *Aporias*, Stanford: Stanford University Press, 1993.

Duclos, Vincent, Tomás Sánchez Criado, and Vinh-Kim Nguyen. 'Speed: An Introduction', *Cultural Anthropology* 32.1 (2017): 1–11.

Fearnley, Lyle. 'Signals Come and Go: Syndromic Surveillance and Styles of Biosecurity', *Environment and Planning A: Economy and Space* 40.7 (2008): 1615–1632.

Feldman, Allen. *Archives of the Insensible: Of War, Photopolitics and Dead Memory,* Chicago: University of Chicago Press, 2015.

Foucault, Michel. *Security, Territory, Population: Lectures at the College De France, 1977-78.* Springer, 2004.

Haraway, Donna. 'The Biopolitics of Postmodern Bodies: Determinations of the Self in Immune System Discourse', in Janet Price and Margrit Shildrick (eds), *Feminist Theory and the Body: A Reader*, New York: Routledge, 1999, pp. 203–214.

Kakkar, Manish, Elizabeth T. Rogawski, Syed Shahid Abbas, Sanjay Chaturvedi, Tapan N. Dhole, Sheikh Shah Hossain, and Sampath K. Krishnan. 'Acute Encephalitis Syndrome Surveillance, Kushinagar District, Uttar Pradesh, India, 2011–2012', *Emerging Infectious Diseases* 19.9 (2013): 1361–1367.

Kant, Lalit, and Sampath Krishnan. 'Information and Communication Technology in Surveillance in India: A Case Study', *BMC Public Health* 10.1 (2010): S11.

King, Nicholas B. 'The Scale Politics of Emerging Diseases', *Osiris* 19 (2004): 62–76.

Mazzarella, William. 'Internet X-Ray: e-Governance, Transparency and the Politics of Immediation in India', *Public Culture* 18.3 (2006): 473–505.

Serres, Michel. *The Parasite*, trans. Lawrence R. Schehr, Minneapolis: University of Minnesota Press, 2007.

Sharma, Rajeev, L. Ratnesh, Amit B. Karad, Himanshu Kandpal, A. C. Dhariwal, and R. L. Ichhupujani. 'Communicable Disease Outbreak Detection by Using Supplementary Tools to Conventional Surveillance Methods under Integrated Disease Surveillance Project (IDSP), India', *The Journal of Communi-*

cable Diseases 41.3 (September, 2009): 149–159.

Singh, Vivek, and Biranchi Jena. 'Syndromic Surveillance in the Integrate Disease Surveillance Programme and Pre-Hospital Emergency Care in India', *International Society for Disease Surveillance*, lecture, 26 April 2012, https://knowledgerepository.syndromicsurveillance.org/syndromic-surveillance-integrated-disease-surveillance-project-isdp-and-pre-hospital-emergency-care.

Suresh, K. 'Integrated Disease Surveillance Project Through a Consultant's Lens', *Indian Journal of Public Health* 52.3 (2008): 136–143.

Thacker, Eugene. 'Living Dead Networks', *The Fibreculture Journal* 4 (2005), http://four.fibreculture-journal.org/fcj-018-living-dead-networks/.

Wier, Lorna, and Eric Mykhalovskiy. 'The Geopolitics of Global Public Health Surveillance in the Twenty-First Century', in Alison Bashford (ed.) *Medicine at the Border: Disease, Globalization and Security, 1850 to the Present*, London: Palgrave Macmillan, 2007, pp. 240–263.

CONTRIBUTOR BIOGRAPHIES

Aakash Solanki is a PhD candidate in Anthropology and South Asian Studies at the University of Toronto. He is broadly interested in the genealogical study of states, statistics, and computing. His dissertation project is an historical anthropology of the Indian state's myriad efforts to address the challenges of 'uneven development', from the 1960s to the present-day as an entry point into studying the interrelationships between data, development, bureaucracy, and technical expertise. In the past, he has worked on the collection, classification, management, and politics of information in colonial India. He has worked in government agencies, both in the US and India, on data science projects in education, health, and skill development at the city, state, and federal levels. He studied Information and Communication Technology in India before transitioning to the social sciences at the University of Chicago. He has previously published in the journal *South Asia* and is a contributing editor to the journal *Cultural Anthropology*.

Anirudh Raghavan has an MPhil in sociology and social anthropology from the Department of Sociology, Delhi School of Economics. His work concerns the history of the virus as a biological, social, and philosophical entity, epidemiology, medical surveillance, and the global biopolitics of disease. He is also interested in Big Data analytics in healthcare and new visual and experiential technologies such as Augmented Reality systems. He has previously written on Pokémon Go and the challenges it poses to the anthropological conception of the 'every day'. Currently, he is an independent researcher, learning Sanskrit and exploring the burning coal fields of Jharia.

Anumeha Yadav is a journalist and researcher who reports on the effects of social and economic policies in rural India. She has reported for two of India's leading national dailies: *The Hindu* (2012–2015), and *The Indian Express* (2005–2007). She has worked as a multimedia journalist for the news portal *Scroll.in*, where she reported on labor and social policy, and the impact of policy decisions on working class and rural communities. Yadav has published in news outlets including *The Caravan*, *HuffPost*, *The Guardian*, and *Al-Jazeera*.

Gaurav Godhwani is the director at CivicDataLab (https://www.civicdatalab.in/), a research lab harnessing data, tech, design, and social science to strengthen the course of civic engagements in India. He works to harness the potential of open-source movement to enable citizens to engage better with public reforms. He works closely with several governments, non-profits, think-tanks, media houses, and universities to enable data-driven decision-making at scale. He has been instrumental in starting initiatives like DataKind Bangalore (https://www.datakind.org/chapters/datakind-blr) and Open Budgets India (https://openbudgetsindia.org/). At CivicDataLab, he is focused on growing data innovations in sectors like Public Finance, Justice, Indian Languages, Urban Planning, and more.

Guneet Narula is an information technology consultant working in the development sector, with a focus on data and web technologies. He has worked in different subsectors like water, sanitation, education, agriculture, and nutrition in South Asia. Among others, he is associated

with Akvo, Fluoride Knowledge and Action Network, Oxfam India, Centre for Policy Research, International Campaign for Justice in Bhopal, and the Centre for the Study of Developing Societies. He is also an active contributor in open data, open maps, and open source communities in India, such as Datameet and Hillhacks. He is interested in the intersection of the open movement and socio-economic development in India. He is based in New Delhi and runs an agency called Sputznik (sputznik.com).

Karl Mendonca is a filmmaker and mixed media artist whose research and work explores experimental forms of collective and personal storytelling at the intersection of media studies and postcolonial studies. His films and installations have been shown at a number of galleries and film festivals, including the *Lower Manhattan Cultural Council*, *The Queens Museum of Art*, *the Oxford Film Festival*, *Stuttgart Filmwinter*, *Jersey City Museum*, and *Experimenta* (India). After completing an MA in Media Studies at The New School, Karl was as an adjunct faculty member at Eugene Lange, where he taught hybrid courses on media theory and production. He is presently a PhD candidate in Film & Digital Media at the University of California, Santa Cruz.

Lilly Irani is an Associate Professor of Communication & Science Studies at the University of California, San Diego. She also serves as faculty in the Design Lab, Institute for Practical Ethics, the program in Critical Gender Studies, and the Halıcıoğlu Data Science Institute. She is the author of 'Chasing Innovation: Making Entrepreneurial Citizens in Modern India' (Princeton University Press, 2019). Her research examines the cultural politics of high-tech work and the counter-practices they generate, as an ethnographer, a designer, and a former technology worker. She is a co-founder and maintainer of the digital labor activism tool Turkopticon. Her work has appeared in *ACM SIGCHI*, *New Media & Society*, *Science, Technology & Human Values*, *South Atlantic Quarterly*, and other venues. She has a PhD in Informatics from the University of California, Irvine.

Mahima Taneja is a recipient of the UGC Research Fellowship and is currently pursuing her PhD from Centre for Political Studies, Jawaharlal Nehru University, New Delhi. Her doctoral research focuses on the gender politics of urban leisure spaces in two Indian cities and engages with conceptual debates on publics, right to the city, affect, and citizenship. Her previous work has focused on feminist campaigns and movements in urban spaces, such as the Nirbhaya movement, Slutwalks, Kiss of Love, and Why Loiter, which she has presented at various national and international conferences and published in peer-reviewed journals like *Women's Studies International Forum*. In the past four years, while working with organizations such as Outline India, Institute of Social Studies Trust, and Institute of Economic Growth, she has consulted on multiple evaluation and research projects across different sectors, including water and sanitation (WASH), gender equality, smart cities, and urban infrastructure. Her forthcoming publication, funded through a Ford Foundation grant, focuses on understanding WASH programming and its monitoring framework through a Feminist Evaluation Approach.

Noopur Raval is a postdoctoral research fellow at the AI Now Institute at New York University. She received her PhD in Informatics from the University of California, Irvine. Her research looks at forms of platform labor in global contexts. She is especially interested in how globally distributed tech labor creates challenges for regulatory frameworks and daily ethical living.

Preeti Mudliar is an Assistant Professor at IIIT-Bangalore. Her research interests broadly center around gender, infrastructure, and digital media using ethnographic methods and analyses. Currently, her work is focused on people's acts of 'repair' and coping following biometric authentication failures in the public distribution system (PDS) in India. She holds a PhD in Communication Studies from the University of Texas, Austin.

Prerna Mukharya is the founder at Outline India (http://www.outlineindia.com), a research and development consultancy firm providing strategic advisory services and policy support since 2012 to international aid and donor bodies, NGOs, and national and state governments working to create social impact through data. She is now launching Track Your Metrics (TYM, https://www.trackyourmetrics.com/)—a tech tool for quick impact assessment by small NGOs and big funders alike, equipping the former with means to secure more funding and the latter to hold their grantees accountable. She featured on *Fortune* magazine's under 40 list in 2017 and 2018, among the top 25 businesswomen in India by *Forbes* magazine, and in the top 40 women changemakers in the 2018 annual list released by India Today. She has prior research experience working at the Centre for Policy Research and holds an MA in Economics from Boston University. She was a 2018 Chevening Fellow at Oxford University and a Raisina Fellow in 2017, awarded by the Ministry of External Affairs.

Puthiya Purayil Sneha works with the Centre for Internet and Society (CIS), Bangalore. Her work primarily engages with shifts in modes and practices of knowledge production in the humanities and arts with the digital turn. Her areas of interest include methodological concerns in arts and humanities, digital media and cultures, higher education and pedagogy, and access to knowledge. Her recent work has been on mapping initiatives in the field of digital humanities in India, with a focus on the digital transition in GLAM (Galleries, Libraries, Archives, and Museums) spaces, and writing on gendered labor and women's work in the digital economy.

Ranjit Singh (http://ranjitsingh.me/) received his PhD from the Department of Science and Technology Studies at Cornell University in August 2020. His research lies at the intersection of data infrastructures, global development, and public policy. His dissertation is an ethnography of India's biometrics-based national identification project, Aadhaar. It advances public understanding of the affordances and limits of biometrics-based data infrastructures in practically achieving inclusive development and reshaping the nature of Indian citizenship. His research draws on and contributes to science and technology studies; critical data studies; data ethics; studies of information and communication technologies for development (ICT4D); socio-legal and policy studies of data, algorithms, and computing; and studies of governance and bureaucracy. He is currently working as a researcher at Data & Society, an independent nonprofit research organization based out of New York City.

Sandeep Mertia is a PhD candidate at the Department of Media, Culture, and Communication and Urban Doctoral Fellow at New York University. He is an ICT engineer by training, and former Research Associate at The Sarai Programme, Centre for the Study of Developing Societies, Delhi. He is broadly interested in historical and anthropological research on computing, media-technologies, urbanism, and futures. His dissertation ethnographically

follows the emergent computational cultures at the intersections of the state and technology start-ups in India. His work has appeared in the *Economic & Political Weekly*, *The Fibreculture Journal*, *Computational Culture*, *Sarai*, *LSE Impact*, *Konrad-Adenauer-Stiftung*, *DataKind*, and other venues.

Sivakumar Arumugam has a PhD in Anthropology from Columbia University. His interests include the development of statistics and biomechanics in sports or, more broadly, the application of computation and quantitative methods in new areas of life that, retrospectively, find an unsurprisingly good fit. He is currently working on cryptocurrencies as a form of non-state money and projects that build on those monies to provide unregulatable banking services, including nascent claims to 'bank the unbanked.'

Theory on Demand #39
Lives of Data: Essays on Compu-tational Cultures from India
edited by Sandeep Mertia

"This remarkable collection is the first major portrait and assessment of the social and technical relationalities that constitute the ecology of big data in India today. Equally remarkably, the authors represent the first generation of scholars of digital media who speak through an Indian lens while being totally conversant with the cutting edge of global scholarship on big data." — Arjun Appadurai, Goddard Professor of Media, Culture, and Communication, New York University

"Wide-ranging and incisive, *Lives of Data* is essential reading for those who wish to understand the seductions and contingencies of being or becoming data-driven." — Lisa Gitelman, author, *Paper Knowledge* and editor, 'Raw Data' Is an Oxymoron

Lives of Data maps the historical and emergent dynamics of big data, computing, and society in India. Data infrastructures are now more global than ever before. In much of the world, new sociotechnical possibilities of big data and artificial intelligence are unfolding under the long shadows cast by infra/structural inequalities, colonialism, modernization, and national sovereignty. This book offers critical vantage points for looking at big data and its shadows, as they play out in uneven encounters of machinic and cultural relationalities of data in India's socio-politically disparate and diverse contexts.

Lives of Data emerged from research projects and workshops at the Sarai programme, Centre for the Study of Developing Societies. It brings together fifteen interdisciplinary scholars and practitioners to set up a collaborative research agenda on computational cultures. The essays offer wide-ranging analyses of media and techno-scientific trajectories of data analytics, disruptive formations of digital economy, and the grounded practices of data-driven governance in India. Encompassing history, anthropology, science and technology studies (STS), media studies, civic technology, data science, digital humanities, and journalism, the essays open up possibilities for a truly situated global and sociotechnically specific understanding of the many lives of data.

Printed on demand
ISBN: 978-94-92302-52-6

Institute of
network cultures