

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 mathematician¹

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 (UID-AI)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.