

Q: How would you describe the visualisation scene in India?

A: It would grow because of the same reasons worldwide, the sheer amount of data is growing rapidly.... I was walking in one of the schools and saw [a] huge pile of students' annual report card dump[ed], being a data junkie my heart sunk at seeing all valuable student data rot away silently.

Interview with Report Bee CEO Ananth Mani (Kirk, 2011)

IN SEPTEMBER 2015, INDIAN PRIME MINISTER NARENDRA Modi—once banned from the United States for his apparent role in orchestrating anti-Muslim pogroms in 2002—returned stateside, traveling to Silicon Valley to promote a vast flotilla of e-governance initiatives called Digital India. At its core is what has been widely termed the “JAM Trinity”: J for Jan Dhan Yojana, promising bank accounts to the poor; A for Aadhaar, the national biometric program promising to “de-duplicate” all duplicitous claims on state services in cash or kind; and M for mobile phones, the vehicle enabling the new “cashless society” JAM promises.

Digital India was rolled out just after Modi's first year in office. It appeared to centralize digital government, which for the past decade had been split at the national level between at least two ambitious programs, the National Population Register (NPR), tied to border security, and the Unique Identification Authority (UIDAI), with its “Aadhaar” ID form (*aadhaar* means “basis” or “foundation”). Each program promised to collect the biometrics of all Indian subjects, a process known as *capture*.

Digital India is under the purview of the Department of Electronics and Information Technology, or DeitY. The godly acronym existed before the 2014 election and is not an invention of Hindu right-wing ideologues within Modi's Bharatiya Janata Party (BJP). For both secular and religious identified blocs across parties, Digital India illustrated the emerging promise and debatable hubris of a new technocracy claiming self-consciously superhuman, panoptic powers. The infotech pantheon was *henotheistic*, in the sense claimed for Hinduism by the nineteenth-century Indologist F. Max Müller of a single deity uniting multiple divinities (1878). One may worship the goddess Siva, Vishnu, or myriad other valued divinities, Müller suggested, but one worships each as the One. Louis Dumont would later define such a relation between values as *encompassment* (1981).

In the digital pantheon over the past decade, the main divinities were UIDAI and NPR. Each promised a national archive of biometric governance that would identify all Indians: voluntarily for UIDAI, and by law for NPR. Each was building an archive to digitize traces of all persons in India, and each claimed the primacy of its archive against the other as the proper form and substance of a new kind of collective entity, what we might call *nation-as-archive*. Digital India and its JAM comprise an explicitly henotheistic mode of governance, encompassing both of these emergent, overlapping, and often competing biometric archives as a single political form, one closely identified with the PM and his charismatic authority.¹

If the population and its nation were mobilized as a visceral collective in the consolidation of European urban, colonial, and settler modernities through the emergence of statistical devices and the conception of a model, the nation-as-archive emerges as something else. We might turn to current historicizations of machine-learning approaches to big data by their architects—of big data constituting an emergent condition of plenitude organized less around statistical modeling than around data storage, curating, and algorithms enabling “visualization”—to convey a sense that the collective form at stake is an unrelentingly expanding mass of data in itself, a *different figure of mass* than that of the mass body and one that demands new conditions of governance. The point is not that such whiggish historicizations of big data (e.g., Gray 2009) are adequate to a history of reason or the archive, but that they offer a feel for the contemporary, for a widely available sense of collectives and their government as not only *dependent on* an immensity of information (the familiar ground of a biopolitics), but ontologically *constituted as* information.

The opening epigraph, from a boutique collection of “data visualisation stories from around the world,” gestures toward a collective form, one in which data—like organic matter—“rots,” in which the relationship between organic matter and data undergoes some kind of material-semiotic shift (Kirk 2011). The care of the child is here organized less around the rotting of sequestered or poorly distributed food stockpiles than around the rotting of piles of information. It is not only that such data is “dark,” in the sense of not yet monetized, but that its life

1 Digital India “cuts across multiple Ministries and Departments” and “weaves together a large number of ideas and thoughts into a single, comprehensive vision so that each of them can be implemented as part of a larger goal. Each individual element stands on its own, but is also part of the larger picture. Digital India is to be implemented by the entire Government with overall coordination being done by the Department of Electronics and Information Technology (DeitY). Digital India aims to provide the much needed thrust to the nine pillars of growth areas...” (DeitY 2015). It might be taken as pandering to left critique to note the requisite phallic language (thrusting pillars). But my provisional reading would be that such language mobilizes the foundational figure of the pillar and in effect links the imaginary of one program (UIDAI)—organized around an airborne and motile vision of platforms flexibly bearing the weight of the state and of the nation's biological need—to that of another program (NPR), organized around more conventional, grounded metaphors of the sovereign control of territory.

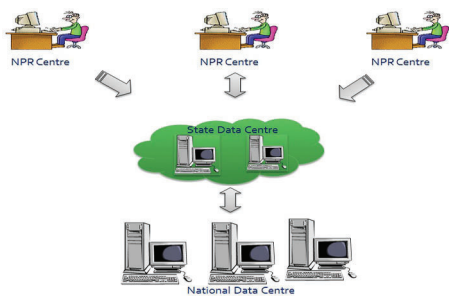


FIGURE 1. NPR Pyramid from Census of India. “The NPR Process”.



FIGURE 2. National Database.

extended and intensified the idea of a strong central government, here an inverted pyramid in which information appears to fall into a single repository. Identified with the passionate attachment to the singular nation and with a sedentary, hyperterritorial conception of those comprising it, this gravitational archive was, in Benedict Anderson’s, conceptualization, a bound seriality (1998: 29–45). It has been repeatedly represented as a central place or thing, collecting a wide range of territorial, demographic, and biographical information. National strength correlates with the quantity of information: multiple data fields for biography and territory fall together into one.

UIDAI’s public presentations of privacy protection similarly address its “highly secure data vault,” variably identified as its “Central ID Data Repository,” or CIDR.

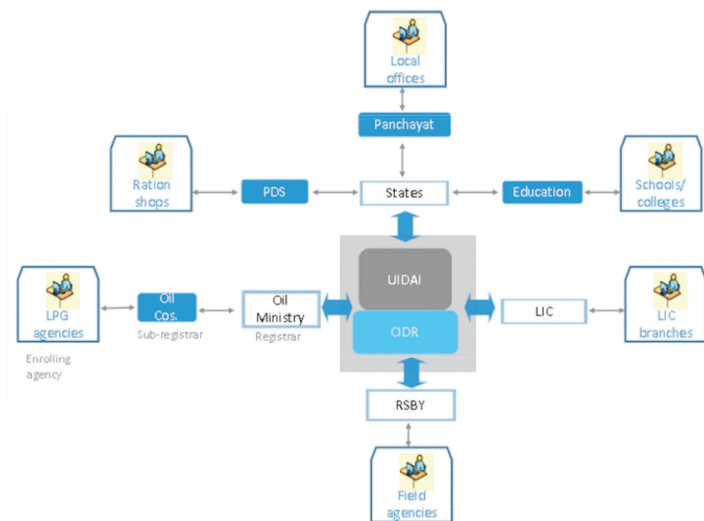


FIGURE 3. “The Unique ID Agencies” from UIDAI Strategy Overview, April 2010.

Eco System for Authentication Services from Aadhaar

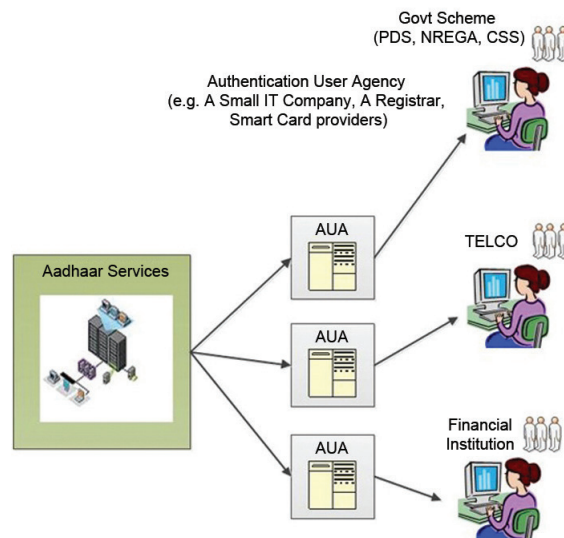


FIGURE 4. EcoSystem for Authentication.

festers or degrades.² Nation-as-archive similarly gestures toward an emergent terrain in which the nation is a database and governance depends on the care of its archive as a kind of living thing. This terrain involves a host of newly mobilized things: the silo and its loneliness; the loss and recovery of the social; security and its proliferating rationalities; and the transfer of “service” or “benefits” and the governmental problem of distribution.

In the first the decade of the 2000s, both NPR and UIDAI found different paths toward capturing the biometrics and variable amounts of biographical data of as many residents of India as possible, and each entity vied henotheistically to encompass the other’s archive. NPR’s conception of archive was centralized, as Figure 1 suggests.

As its architects imagined, the “National Data Centre”

2 This rendering of data as organic—or conversely, this digitization of rot—reprises the familiar racialized organicism of the (post)colony as garbage (Anderson 2010; Chakrabarty 1991; Kaviraj 1997).

These presentations intensify in response to civil society criticism of UIDAI and its Aadhaar that centers on privacy. But Aadhaar’s architects, in contrast, stress how *little* information UIDAI collects into the CIDR, and how this minimalist archive is more relevant as a platform (a more useful translation of *aadhaar*) that links together myriad “silos” of information, forming an “ecology” or “federation.”

When pyramids do appear in UIDAI’s self-representation, these are often turned on their side; it is less a repository through the sovereign force of gravity than a catalytic enabler of a range of goods. The box in Figure 4 labeled “Aadhaar services” places the secure central archive as part of an already distributed ecology charged with redistribution.

NPR and Aadhaar invert the relation of citizen and resident in different ways. What would become NPR began after the 1999 Kargil war with Pakistan as an effort to create a biometric identity database to distinguish Indian citizens in Kashmir from presumptive infiltrators. Biometrics carried future promise and were linked

to multiple biographical data to stress a proper relation to space: border security would be effected by linking the collection of a hyperterritorializing plenitude to the promise of indelible physical traces. This conception of archive was progressively scaled up over a decade, from the Indian Muslim to the Indian citizen, from Kashmir to the nation.

But how to achieve this larger scale? Bureaucrats and contracted experts associated with the Interior Ministry proposed piggybacking NPR on the Census of India. The Census was not an archive of citizens, but of *residents*: whoever was enumerable across the terrain of the nation. Using the Census's preexisting infrastructure to achieve the needed archival scale meant that *residence* and not *citizenship* became the condition of biometric subjectivity. The focus on internal security specified the unit of biometric data collection as the citizen, with biometrics offering the promise of distinguishing that citizen from its double or "duplicate": the fake citizen or terrorist.

NPR was never closely associated at the executive level with the emerging constellation of information technology capital and expertise. It drew not only upon the preexisting infrastructure of the Census, but also upon preexisting standards of administration, hierarchy, and contract in the creation of new governmental forms. Like other administrative units, it was subject to the familiar accusation of corruptibility, not only at the level of bureaucratic procedure but within the constitution of the digital archive.

By contrast, UIDAI organized itself around corruptibility as a problem. It moved away from standard governance—viewing corruptibility as requiring human solutions—to corruptibility as a machine engineering problem, one of databases, not bureaucracies. It is commonly narrated as being the brainchild of Nandan Nilekani, a founder and the CEO of the IT outsourcing giant Infosys, a company that pioneered a range of identity instruments to organize and credentialize IT service labor. Like many nouveau hyper-rich, Nilekani was troubled by the persistence of a massive and "leaky" state bureaucracy and its cozy relation to a small coterie of elite family capitalists, a situation preventing the efficient management of poverty and weakening entrepreneurialism. Nilekani offered a blueprint for completing the country's neoliberal transformation in his 2008 bestseller *Imagining India* and was brought into the previous Congress Party-led government to create Aadhaar.

Nilekani's concept in brief is that India's future increasingly depends upon the distribution of "service," principally forms of welfare in kind or, increasingly, cash, but corruption "leaks" out a significant proportion of this wealth, both through rent-seeking by petty bureaucrats and other office-holders charged with service distribution and through the production of "duplicates," fake or copied identities in the list of persons or households entitled to a service. The conception of service is organized around a biopolitical figure of bare life: of *residents* within or moving across a terrain who must be supplemented by services to survive and to thrive. Aadhaar's early critics from the political right worried that its basis only in residency (as opposed to citizenship) would enable

undocumented Bangladeshi migrants to gain official status and receive undeserved state services by getting Aadhaar numbers.

UIDAI's own concern with wastage was not the unsubstantiated specter of the migrant, but the general corruption or "leakage" of legitimate claims on distribution by most persons. Archives, and in particular databases, are rendered efficient and governable through consistent "de-duplication," ensuring that all items in the collection are "unique" and thus curtailing leakage.

De-duplication is a technical term that addresses problems of storage efficiency, of record variability and the need for correction, and of security from duplicate (e.g., stolen) identifying objects.

Efficiency: "de-duplication is a task of identifying record replicas in a data repository that refer to the same real world entity or object and systematically substitutes the reference pointers for the redundant blocks; also known as storage capacity optimization" (Faritha Banu and Chandrasekar 2012:364).

Correction: "data sources are independent... [adopting] potentially inconsistent conventions" (Maddodi et al. 2010:664), so to build an effective "data warehouse," data "has to be transformed and cleaned before it is loaded into the warehouse" (Chaudhuri et al. 2006). Data may differ across source archives because of different schemas by which they were formed, and thus cleaning involves "schema extraction and translation" (Thakare et al. 2015:10). Data difference may not only involve the cultural difference of distinct schemas, but also the problems introduced into any given source archive by human error, which constitute "dirty data" (Maddodi et al. 2010:664). The distinction between what makes data untranslatable, requiring schema extraction, and what specifically renders it dirty is not always clear in this literature. Archives, as products of assemblage, appear to present translation as both a semiotic and arguably a moral problem.

Security: the presence of duplicates in an archive when each of those duplicates refers to the same object (say a given resident of India) and provides a means for different users of the archive to make different claims as or for that object, as, for example, when the hero or villain in a movie gains access to the nuclear arsenal through a duplicated identity.

In creating UIDAI, Nandan Nilekani argued that for India to become more like China, a developing economy powerhouse, it needed to be de-duplicated *as a nation*. Neoliberal efficiency, the security of the commonweal in the face of mass corruption, and the translation problem of what we might term *history-as-assemblage*, were all gathered up into a single technocratic repertoire. Corruption was rendered as a matter of either duplication from above, the large-scale seeding of an archive with duplicates inserted by powerful interests exercising control over it, or duplication from below, the fake identities upon which persons—urban migrants, slum dwellers, landless laborers—unrecognizable within the formal ar-

chive may depend.

If duplication from above depends on control of archival infrastructure, UIDAI proposed a radically new and independent archive. The problem for its engineers was the social itself, the network of interest and biographical relations that limit fair and efficient distribution and produce leaks. They proposed collecting as little biographical and locational information about persons as possible—assurance would depend on biometrics and not biography—to produce a deterritorialized archive cut off from the duplicative nature of the biographical and social. The subject of this archive was a body offering ten fingers and two eyes, officially a “resident”: incorruptible and free from political tampering because the Aadhaar numbers issued to all residents of India would convey no information, no history. Each time this resident sought a service, the plan

Despite UIDAI’s insistence on an ecology of multiple silos federated through its Aadhaar network (as opposed to an NPR-like National Data Centre), and its claim that it archived almost no personal information besides a registrant’s biometrics, the drive by its engineers to make Aadhaar the platform for any “service” from food subsidy to credit suggested that any form of value in belonging to the nation would need to come through Aadhaar. This was a new kind of citizenship: UIDAI lacked any statutory right under Indian law to mandate its Aadhaar identity, to serve as the necessary platform of service delivery, or to be the vehicle of de-duplication.

Some UIDAI engineers that I interviewed formally responded to their progressive critics that the UIDAI archive, unlike NPR, eschewed biography, did not in itself collect information on Aadhaar registrants, and would

The subject of this archive was a body offering ten fingers and two eyes, officially a “resident”: incorruptible and free from political tampering because the Aadhaar numbers issued to all residents of India would convey no information, no history.

presumed, he or she would present a body part and the system would return a “yes” or “no”: you are you, or you are not you. As more and more persons were signed up, and as more and more services were linked, India would be de-duplicated.

De-duplication did not require UIDAI: multiple parallel and derivative data-cleaning projects emerged around the same time. The customer list of a small cooking gas distribution agency near Delhi whose owner and manager I know was considerably reduced when every gas customer had to present proof both of identity and of residence to the agency, which was then turned over to state auditors. Whether or not people presented their Aadhaar numbers or other forms of legitimate ID as proof, the exercise de-duplicated the list by more than half. The “corruption” of households availing themselves of multiple subsidized gas cylinders was curtailed, as was that of gas deliverymen siphoning off small amounts of gas or police pressuring gas deliverymen for a cut.

Nilekani’s dream to remake India demanded de-duplication of service distribution at a massive scale. The archive had to scale up to the nation. Unlike NPR, the Census was inadequate for UIDAI to produce an identity archive—to produce India—at such a scale. Rather, public-private partnerships (PPPs) were set up in most Indian states to be independent from current bureaucracy: subcontractors were paid per new biometric registrant, profit expansion and not national infrastructure drove scale, and the server network was designed to test and retest subcontractor skill and honesty.

As Aadhaar became both a promise of inclusion for those too marginal to have access to earlier modes of identity, and a threat of Big Brother as the universal platform making life through service possible, it began to appear the very condition of citizenship. Civil society activists on the left argued that UIDAI would not just link the “silos” of individual service distribution programs through its platform, but produce a condition of total *convergence*.

maintain a federation of silos, not the feared convergence. Privately, however, some UIDAI engineers told me that if politicians wanted to use Aadhaar to converge silos, they could. When I reported this internal concern to one of the most senior UIDAI engineers, he got upset: we have created a corruption-free identity, he said. But if politicians and social forces misuse it, there is a limit to what we can do.

Nilekani and his team fought to prevent the powerful senior officials aligned with the Interior Ministry and NPR from getting control of UIDAI. For UIDAI, the state security apparatuses—including NPR and other central repositories of identity—were each silos that could be more effectively governed if Aadhaar became their universal platform. For the NPR team, UIDAI was simply a different and parallel way to gather data, and if it promised efficiency, it was yet another contracted tool for national information to be encompassed by the demands of the National Data Centre.

Digital India’s publicity in 2015 offered an explicitly flexible account of information and its government. Existing bureaucratic structures across the range of state service were expected to open themselves to new norms of easy access, no longer dependent upon the power of the bureaucratic office and of its rent seeking. Existing archives of identity could be flexibly deployed to manage and audit this access. Concerns with both *physical* leakage—the wrong people on the wrong side of the border—and also with economic leakage—the proliferation of duplicates wasting the commonweal—were to be secured through the interrelation of what we might call *neo-Aadhaar* and the Modi persona itself.

Under the previous, Congress Party-dominated central government, Nilekani and his allies in the Indian Planning Commission—the dominant mandarinate of the development state—envisioned UIDAI’s success on the model of other PPPs free from the bureaucratic (“social”) entanglements of lesser arms of the state. UIDAI was set

up in relation to the Planning Commission and the Finance Ministry, but was largely autonomous from them. UIDAI evaded the party politics of the parliamentary system and was not constituted as a statutory body. But as Aadhaar increasingly came to be constitutive of a new form of citizenship, its critics launched a series of court cases challenging its legality. In 2014 and 2015, the Supreme Court of India affirmed that no service could require that people register for an Aadhaar card.

There were other challenges. Before Narendra Modi won the 2014 election, Aadhaar's fate seemed politically as well as constitutionally unclear: the program was closely identified with Congress President Sonia Gandhi, and Nilekani himself, despite his frequent disavowal of social and political corruption, had been pressed to run for office. He, like Congress, lost.

But Modi, victorious, would go on to embrace Aadhaar with a vengeance. News accounts and popular stories began to circulate about the new PM's panoptic ability to know what was going on in all senior political and bureaucratic offices, and that he was having Aadhaar scanning devices placed in every major government office to ensure that officials were present and that their output could be measured. Aadhaar, with its reputation under Nilekani for placing the nation-as-archive outside of and protected from the bureaucratic office—that is, the conventional institutions of the state—was being brought in to manage those very institutions. If Aadhaar had been designed to disentangle office from service, it was now synonymous with a new government of office. Beginning in late 2014, I heard an emergent class of panoptic Modi joke in which an official skipping office duty, breaking a rule, or not following the PM's instructions would suddenly get a phone call from Modi himself.

Beginning in July 2015, Digital India loosened this close connection between the panoptic Modi and Aadhaar. Whether or not Aadhaar itself would be the primary identity archive for the new e-governance seemed less important, particularly given its questionable legal future, than did its formal apparatus: biometrics, ever more universal scanners, and some kind of henothetically constituted lattice of future identity archives serving as the platform layer for the state and for finance. In July, I heard stories of a “secret” pact between Modi and Nilekani to keep Aadhaar's powerful linkage of the nation's silos intact. Over the next months, UIDAI and its Aadhaar program were placed within JAM, a commitment to shift all service to direct cash transfer via the explicit trinity of universal bank accounts, Aadhaar biometric scanning to ensure deduplication, and mobile phones as the sites across which the anticipated regime of microcredit and microspending would be enacted.

It is clear is that the division between a centrist and state-based national archive (the National Data Centre of NPR) and an exceptional nation-as-archive located across a vast federation of silos managing welfare, health, education, credit, labor, and so forth (Aadhaar as universal platform beyond the reaches of state corruption) no longer seems to hold. Modi as panopticon may have diminished somewhat, but the controversial leader's image and persona girding a new ethic of state office has been linked to Nilekani's promise of a guarantee of identity and service based on the separation of service and office. ■

LAWRENCE COHEN is Sarah Kailath Professor of India Studies in the Department of Anthropology at the University of California, Berkeley.

BIBLIOGRAPHY

- Anderson, Benedict Richard O’Gorman. 1998. *The Specter of Comparisons: Nationalism, Southeast Asia, and the World*. London: Verso.
- Anderson, Warwick. 2010. “Crap on the Map, or Postcolonial Waste.” *Postcolonial Studies* 13(2):169–178.
- Chakrabarty, Dipesh. 1991. “Open Space/Public Space: Garbage, Modernity and India.” *South Asia* 14(1):15–31.
- Chaudhuri, Surajit, Ganti, Venkatesh, and Kaushik, Raghav. 2006. “Data Debugger: An Operator-Centric Approach for Data Quality Solutions.” *Data Engineering* 29(2):60–66.
- Department of Electronics and Information Technology (DeitY). 2015. “How Digital India will be realized: Pillars of Digital India,” <http://www.digitalindia.gov.in/content/programme-pillars> (Accessed Feb 22, 2016).
- Dumont, Louis. 1981. *Homo Hierarchicus: The Caste System and Its Implications*. Chicago: University of Chicago Press.
- Faritha Banu, A., and Chandrasekar, C. 2012. “A Survey on Deduplication Methods.” *International Journal of Computer Trends and Technology* 31(3):364–369.
- Gray, Jim. 2009. “Jim Gray on eScience: A Transformed Scientific Method.” *The Fourth Paradigm: Data-Intensive Scientific Discovery*, edited by Tony Hey, Stewart Tansley, and Kristin Tolle, pp. xvii–xxxi. Redmond, WA: Microsoft Research.
- Kaviraj, Sudipta. 1997. Filth and the Public Sphere: Concepts and Practices about Space in Calcutta. *Public Culture* 10(1):83–113.
- Kirk, Andy. 2011. “Data Visualization Stories from India, By Ananth Mani (Interview),” *Visualizing Data*, August 30th, 2011. <http://www.visualisingdata.com/2011/08/data-visualisation-stories-from-india-by-ananth-mani/> (Visited Feb 22, 2016).
- Maddodi, S., Attigeri, G. V., and Karunakar, A. K. 2010. “Data Deduplication Techniques and Analysis.” *Emerging Trends in Engineering and Technology, 3rd International Conference*, pp. 664–668. New York: Institute of Electrical and Electronics Engineers.
- Müller, F. Max. 1878. *Lectures on the Origin and Growth of Religion: As Illustrated by the Religions of India*. London: Longmans, Green and Co.
- Nilekani, Nandan. 2008. *Imagining India: Ideas for the New Century*. New Delhi: Penguin Books.
- Thakare, Manisha R., Mohod, S.W., and Thakare, A.N. 2015. “Various Data-Mining Techniques for Big Data.” *IJCA Proceedings on International Conference on Quality Up-gradation in Engineering, Science and Technology* 8:9–13.