Economic&PoliticalWEEKLY

Alarmingly Simplified Data Gathering

Vickram Crishna (vvcrishna@gmail.com) is an engineer and a petitioner in the Supreme Court against Aadhaar, the identification project under the aegis of the Unique Identification Authority of India.

The Aadhaar Effect: Why the World's Largest Identity Project Matters *by N S Ramnath and Charles Assisi, Oxford University Press, 2018; pp 328, ₹350.*

Dissent on Aadhaar: Big Data Meets Big Brother *edited by Reetika Khera, Orient Blackswan, 2019; pp 288, ₹406.*

We Are Data: Algorithms and the Making of Our Digital Selves *by John Cheney-Lippold, NYU Press, 2018; pp 268, $14.19.*

In 2009, without any public discussion, the Government of India issued an executive order setting up the Unique Identification Authority of India (UIDAI), creating it as an office under the Planning Commission. The first chairperson, Nandan Nilekani, until then chairperson of a private company, Infosys Ltd, was conferred the rank of a cabinet minister.

Given this status, he was able to quickly launch what was publicised as a portable identification system, claimed to provide India's first national digital identity for persons who were said to lack access to any documentation at all. This would, it was said, provide anytime access to guaranteed welfare benefits under a range of such schemes. Without this, it was said, the functioning of such schemes led to widespread "leakages," ensuring that vast amounts of moneys paid out by the state were not reaching the intended beneficiaries at all, and this would put an end to such problems.

In 2018, after a lengthy series of struggles in various courtrooms across the nation, the Supreme Court issued an omnibus order on a large number of individual public interest litigations, and against a series of appeals against repeated contempt of court by the government at various levels (both state and centre). The original executive order was, after seven years, replaced with legislation, but in the form of a money bill. This attracted further litigation against the manner of its passage into law, apart from numerous contentious issues with its framing.

Following the judgment, two books have been published directly covering the scheme and its impacts. The first is *The Aadhaar Effect* and the second, *Dissent on Aadhaar*. A third book, *We Are Data*, examines the regime of personal data created in digital form by such schemes, in tandem with other generators, many of which are commercial exploitations of online social services, offered free for use in return for almost unrestricted data access.

Usage of Data

This review article places the project and its impacts within the ambit of massive data collection and how it is being, can be, and, ominously enough, will probably be used, by coalescing these three studies.

"We believe this is a project in perpetual beta," concludes *The Aadhaar Effect*. To underline it, the cover title is casually misspelled "The Aadaar Effect," on the book spine. For many, this kind of shallow carelessness typifies a project that was conceived in haste, implemented post-haste, and will be forever in catch-up mode.

To the authors, however, the rapidity with which the use of the number was enforced across the country, coercing over a billion people to sign up for a new documented identity, in spite of the fact that over 99.97% were found to already possess at least one, is laudable.

Ten years down the line, the default use of the number remains, largely, presentation of "the card," a printed acknowledgment of enrolment issued as is by UIDAI. The mockery this makes, of a transformational shift towards nationwide digitisation of functional systems, does not seem to faze its promoters.

Portability also remains a question mark, as is exhaustively documented in *Dissent on Aadhaar*. When systems - demand authentication, as has been tried in both pilot and policy in some areas, notably in Jharkhand, it resulted in starvation deaths amongst those denied issue of welfare, as the direct outcome of inability to comply (Tewary 2018).

Reasons for non-compliance range from the failure to register biometrics, to failure of the system itself (lack of connectivity, lack of electricity, etc). Overwhelmingly, the "card," produced as a "proof" of identity, serves to clear more trivial barriers, such as airport entry, where identification documents are incessantly demanded. Efforts have been made in recent months to "enhance" the value proposition of the card by printing it with a QR code

that authenticates it, when scanned with the camera of a mobile phone, if the phone is equipped with the appropriate software.

However, that postdates the existing usage of the "card," to the extent that, in submissions before the Supreme Court over the years, the government pleader continually referred to it as a card. Finally, in detailing the judgment through a judicial order, the Court itself, in its majority opinion, terms it a card. Note, the judgment itself is not a subject of either of the books, both of which were substantially written before the judgment was issued (*Dissent on Aadhaar* adds a few pages mentioning the judgment, but the main essays were written and edited earlier). This report, therefore, will avoid discussing the aftermath of the judgment, which was delivered in September 2018.

The third book, *We Are Data*, explicates why any judicial ruling that permits the identification scheme to continue makes the discussion of partial impacts irrelevant, in any case. While it is not particularly India-centric, it provides a huge insight to understanding the overt impacts of data discovery, which is inbuilt to the manner in which UIDAI has promoted the use of the UID (Unique Identity Number). This goes well beyond the layer made visible to most people, an identity card that opens various artificial barriers.

For most people, this aspect, some document that is recognised by others as a form of identification, is the most relevant and important. It is also the basis on which the project was initiated.

Unfortunately, it turns out it may not be the most critical reason for actually promoting this project. Across the world, multiple commercial services have been created that trade on the free access to data, in return for free furnishing of some service perceived as being useful. The penetration of these services in any geographically distinct population, however, hinges on the availability of other services, central to the policies by which those populations are governed. Without the services, people do not generate enough digital data.

The most important is connectivity, but connectivity is irrelevant without the availability of electricity. India, nearly two decades into the 21st century, lacks pervasive ("seamless," a term that implies 24×7 availability in all directions a person can move) facilities for both factors. This means that digital personal data is not driven bottom-up in this geography. The UIDAI project, therefore, is necessary to paper over this gap.

As it turns out, it is not justified on grounds of non-availability of an identification document in any form, for a majority of people in India. Submissions to the Supreme Court by the government over the years acknowledged that a vanishingly small percentage (0.03%) of people actually turned up at UIDAI enrolment centres to register, without at least one such document in hand. Yet, in 2018, *The Aadhaar Effect* blithely attributes (p 2) the lack of such documents as the overwhelming justification for spending public funds on the project.

The book continually bypasses such inconvenient realities (the "document gap" is only one of many). Despite this oversight, which appears to be deliberate, the book still has some interest, principal being the perhaps unconscious description of how a private-sector driven project, whose private sector gains are obvious (and laid out in painstaking detail in *We Are Data*), was packaged as a public sector scheme. Even more interestingly, it also describes how the tensions between private and public sector interests have eventually found the project apparently running in the public sector, while its gainful spin-offs are completely positioned within the private sector.

This factor has been commented on elsewhere, describing how centralised systems are routinely compromised by bureaucracies across the world. This has resulted in the generation of documented "successes" that are figments of imagination at best, flights of fancy sometimes, that lead to exceedingly dire consequences—large number of deaths, for instance, and destruction of an entire national economy—when used for planning and policy decisions in large countries.[1]

As related in *The Aadhaar Effect*, one of the objectives of creating a digitised identification system was to bypass the generation of such invented record-keeping. However, there was little or nothing in the design of the implementation that actually dealt with this issue. This is minutely covered in "Dissent on Aadhaar," from essays dealing with the effects of systems linked mandatorily to the number, demanding instant authentication of the individual in order to authorise the issue of welfare (such as foodgrains), to a description of how a much-ballyhooed digitisation of land records ultimately resulted in a service that, far from improving the system it replaced,[2] demanded and was dependent upon a host of new logistic hurdles.

In the case of welfare deliveries, no provision was made for tracking the actual amount of foodgrains delivered to the beneficiary, long reported (by the promoters) as one of the principal sources of malfeasance in the system. Instead, the onus has shifted on to the beneficiary, who is mostly illiterate, has poor quality biometrics, is unfamiliar with computers, and does not even know or understand that this shift in balance has taken place. In the meantime, while this exercise was being carried out, several states actually implemented alternative methods of welfare delivery tracking, focusing on earlier stages of distribution, resulting in close to total elimination of losses. Naturally, where such improvements have taken place, there is no significant purpose to digitising end-user identifications.

Core Purpose

More importantly, it is clear from rosy-hued descriptions of the development of the project, that welfare efficiencies are not the core purpose of the project. They do, however, provide a seemingly credible reason for

the government to fund, and more importantly, insist upon the usage of such a system. To this end, both governmental and extra-governmental authorities (notably, the World Bank) have deliberately and fraudulently misrepresented welfare spends as savings,[3] leading to the impression that the project is working, despite its obvious failings (*Wire* 2017).

This creates a different set of problems with the manner in which the project has been implemented, which is not the subject of either of the two books mentioned above, though. To understand this problem area, it is necessary to have read the very first book brought out on the project, prior to the completion of the hearings. This is *The Sham ID, Called Aadhaar: Hoax of the Century* by (retired colonel) Mathew Thomas (2018).

Tireless investigations of the contracts signed by UIDAI, with a range of consulting firms and data processors, led to the discovery that not only was permission given to make copies of the data collected, but that the main contractors had serious links to foreign intelligence agencies, principal among them the United States (US)-based National Security Agency (NSA). The NSA has a mandate from the US government to surveil all electronic communications worldwide, with the exception of that of US citizens (Makkar 2017).

It is, however, not technically possible to definitively exclude the communications of US citizens from that of foreigners, when exhaustively funnelling in everything. The discovery that this was an ongoing illegality, following large-scale publication of communications declared secret by the NSA (by the actions of one of its former contractors, Edward Snowden), resulted in US courts ordering the immediate cessation of all activities that might result in US citizens being surveilled.

When questioned, the UIDAI, however, declared that it was unaware that the contractors were not Indian companies. This thin excuse is based on the fact that all bidding contractors were required to set up offices in India, prior to bidding. The contracts permit the data processing businesses to maintain copies of all data, and do not mandate supervision of such duplicate databases. For a long time, the UIDAI attempted to block release of such information, on the specious claim that the contracts also included clauses that might be confidential in terms of competitive business. However, the requests had not asked for release of such clauses in the first place, and the revelation of commercial secrets could have been justifiably redacted.

It appears likely that, for reasons that have not yet been forensically established, nor judicially enquired into, even by the Supreme Court, before which petitions were filed and included in the set adjudicated together in 2018, the entire database of Indian residents who enrolled has been shared with foreign interests.[4]

As it turns out, not all residents may be of interest to foreign intelligence agencies as such. But, on reading *We Are Data*, it becomes obvious that, in order to adequately identify persons of interest, it is necessary to have a much larger database, the universal set, from which relevant persons can be identified for flagging as worthy of tracking. That is the unsettling side of mass surveillance in the digital age, while an allegedly more benign purpose, identification of persons who might be commercially valuable, justifies the effort.

For most persons who are digitally active, it is a welcome innovation, begun in the very last year of the previous century, that free services of perceived usefulness are available. But, in return, wide access to personal data drives much of new wave global surveillance, the surveillance that is not directly the work of mandated “secret service” organs of various governments of nation states. Rather, it is the work of commercial organisations that have, in under two decades, become the largest purveyors of advertising in the world. Just two of these, Google and Facebook, have revenues and reach that put them head and shoulders above all known advertising businesses that dominated the 20th century.

Even advertising, although it may benefit commercially from access to targeting of a nature that puts the famous words of Lord Leverhulme in the shade (“Half the money I spend on advertising is wasted; the trouble is I don’t know which half”), is not the problem; at least, it is not the whole problem (Ogilvy 1963).[5] There is a strong possibility, discussed exhaustively in John Cheney-Lippold, that persuaders within advertising-driven services, such as social media and searches (the characteristics of the primary businesses of Facebook and Google, respectively), actually drive consumer behaviour.

This is achieved by restricting choices perceived as being universally available (to the extent that they can be accessed via the World Wide Web). Such restrictions are enormous enough to actually mould and shape public opinion itself, in a remarkably grainy sense. Grainy, in this context, refers to the degree to which influenced audiences might be narrowed, and the purpose of this whittling down is to arrive at the absolute individual level.

Now, it turns out that this individual does not actually need to be the person that is “I.” It is, instead, the person that is the sum of data points that approximate, asymptotically, but in remarkable degree, to “I.” The purpose of digital identification then, is not so much the identification demanded by the nation state, the historical origin of identification, as distinct from the social origin of identity, but the shaping of a consumer of commercially viable goods and services. The approximation is asymptotic, only because we presently lack the technology to perfectly map every moment of every day of life in every detailed biological sense. The data-gathering points, however, are increasing rapidly. Even so, “we” are data, but data is not “we.”

Manipulation

The size of these sets is huge, by any standard excepting that of the genomically digitised real being. They enable daily manipulation of the data accessed by any person. In turn, this shapes the data “I,” to the extent that

the person "I" is blissfully unconscious of such manipulation. In modern research terminology, the "individual" person has been effectively replaced by the "dividual" set of data, one that emulates the person. It is this dividual that is useful to any corporate entity—government or private—that is both accessing the data from (and thereby, defining) the dividual, and is interested (motivated) to manipulating the data that will be accessed in future (meaning, the future that begins immediately after now) by the individual.

And, from the simple social media and search engines that kicked off this acceleration in both data-gathering and targeted data deliveries, the attractors have become much more persuasive. Health and quality of life, for instance, as promised by health tracking devices that offer familiar entry point services, such as wearable timekeeping. For example, this article[6] extols the ability of Apple Watch, a timekeeping wristwatch device made by the information technology consumer products and services giant, to prevent cardiac incidents, by issuing timely warnings to users. The device encourages users to share online biological indicators emanating from one's own body. Ownership of the data is far from assured: indeed, it is "necessary" to share it, on the grounds that accumulated user data will help make future health predictors even more accurate.[7]

Such devices, relatively expensive, may seem hardly relevant to countries like India, where potential users may be in low percentages. But this is an obvious fallacy: consider that estimates of India's middle class, perhaps 50% of all smartphone users in the country, exceed the entire populations of well over 90% of all other nations severally.[8] Targeting this scattered audience might have been a Herculean task in the mid-20th century, but, thanks to projects like this one of the UIDAI, which has linked the number to the phone numbers and banking accounts of a majority of the "worthy" residents, it is trivial. Furthermore, the practice of using wearables is itself a social phenomenon, one that is highly susceptible to influencers that are already obviously making use of the kind of data gathering—*sousveillance*—taking place constantly. With constant ongoing innovation in wearables, it is not inconceivable that some low-cost variant may soon find popularity (that is, millions of users) in India.

For present-day Indians, wearables and such instantaneous data-gathering devices of "deep" data may not, of course, be as relevant as the ability to manipulate open access to information, in a manner that leaves the individual unaware that the dividual is being constrained. For a nascent democracy like ours, plagued by poverty, illiteracy and lack of accessible media, the ability to manipulate those bereft of easy sources of information has long been demonstrated, as the repeated election of charismatic politicians shows. It is the ability to influence those who believe they are better informed that might lead to results that confound.

The public dissemination of data collected by UIDAI has been repeatedly established, very often by websites maintained by the government itself, at both the central and at state levels. This has usually been due to incompetence or poor technical design of web pages hosted by the government, but, very recently, criminal charges have been filed against a private company that evidently had access to the data of nearly 8 crore residents, collected (the case is under investigation, but an officer of the UIDAI has publicly commented that the data formatting appears remarkably similar to that of the UIDAI database maintained by Central Identities Data Repository (*Times of India* 2010).[9] The case complainant is, in fact, UIDAI itself, because one of the provisions of the law governing the authority reserves the right to file charges solely to itself.

Far from residents being protected, as envisaged under the Constitution, the only extant law in place currently places the onus for investigation and filing charges on the government itself, and within the government, solely upon UIDAI, for offences that involve data collected by the UIDAI. As has been made clear from records released by UIDAI itself, a significant number of cases filed charged journalists and publishers who revealed that such data thefts were taking place, while the total number of cases filed number far less than one hundred. However, a massive number of agencies contracted to UIDAI for data gathering have been dismissed for activities that include data fraud.[10]

No explanation for the lack of charges against such data fraudsters has been furnished till date.

In conclusion, a very large exercise has been publicly documented, through several books and voluminous literature, and its purpose seems to be somewhat different from what was publicly announced, and for which public funds were allocated. A country where personal data was both difficult and expensive to collect and collate has been rapidly converted, to the point where high quality data gathering has become alarmingly simplified.

Other nations had years and decades to come to terms with such massive data-gathering exercises, and built up, in fits and starts, some amount of political resistance and lawmaking[11] to attempt to contain it. For India, it has taken less than a decade, and there is little evidence of legal systems matching up to the challenge. Individuals need not be lured here, to turn over their data in return for superficially attractive commercial services, such as free email and ease of social communication, involving the sharing and exchange of rich media in small communities. The exercise here has been enforced and coerced, through direct and threatened denial of essential services, upon both needy (woefully poverty-stricken) and comparatively well-off sections of the population, by direct governmental intervention. In doing so, 1.3 billion people have been dragged into a net that only a few, mostly technologically aware but politically immature, can see, much less acknowledge. The net is ostensibly digital, but it has the potential to be not unlike stainless steel razor wire, when abused by interests inimical to a free democratic republic.

Notes

\1 Harari, Yuval Noah, *Homo Deus: A Brief History of Tomorrow*, HarperCollins, pp 402. In Part II: The Storytellers, from footnote 3 on page 166: citations from Jean C. Oi, *State and Peasant in Contemporary China: The Political Economy of Village Government,* and others, describe the generation of almost fictional records of agricultural output in China and Tanzania, two nations in which centralised reporting systems had been rigidly created by the 1960s. The farm outputs thus documented formed the basis for national policy that, in the case of China, led to a massive famine that reportedly killed tens of millions, as real grains were exported wholesale, leaving little for local consumption, and in Tanzania to the destruction of a viable and thriving local agricultural ecosystem, the mainstay of the economy.

2 The author of the impugned essay, *Inside the Plumbing of Technology Projects*, Jonnalagada K, provides an insight that technology projects in public service domains basically reroute responsibilities, from the local authority (the ration officer, for instance, in a district) to a technology designer. However, this person has no actual accountability for endpoint service deliveries. Worse, such public service projects rarely build in a self-correcting mechanism for public comment and correction, either during the design stage or following the rollout of the programme.

3 The *Wire*, an online publication, reported on 3 October 2017 that the World Bank had misrepresented Indian government welfare spending as savings accrued from the UID project. This was done by misquoting an actual study carried out where the welfare spending figure was estimated. This was then presented as savings, even as the study pointed to potential savings that might accrue if a number of other factors were actualised (but were not actually in place). Following the exposé, and after a series of sharp letters were sent to the institution, the report was reluctantly "corrected." However, until a year later, senior World Bank representatives continued to attribute Indian government spending estimates as savings, primarily to provide support for its own global programme, branded ID4D (identity for development). The motivation for such misrepresentation is very sharp: given the need to miscontrue identification systems, which are basically expensive technological interventions, as substitutes for identity, a social construct.

4 As reported by Ogilvy (1963). However, the attribution has been questioned, and may be apocryphal. Even so, it refers to an age well before digital communications became commonplace, in fact, even before the first telephone conversation took place.

5 Hall (2017). There is a movement towards "wearables," computing devices that are normally attached to the body by various means, either on a strap or as clothing, even to the extent of being actually surgically embedded. This is different, although doubtless may be considered related, to the surgical implantation of prostheses and medicated devices, since such devices are empowered with either direct or networked computing power of their own, and could be further enhanced to allow some independent decision-making, based on the data gathered and analysed.

6 However, protection of that data from casual access is not so simple. In some countries, it may be necessary, given strong local laws. In India, where protection of such data is regarded by the Constitution as a fundamental right, there is in fact no effective civil or criminal code, or even jurisprudent precedent, for penalising individual or corporate entities who deliberately or accidentally disseminate such data, unless such dissemination also includes activities that might constitute other forms of crime, such as theft, for which routine statutes are in place. The fundamental right to privacy has no matching laws to penalise its breach. However, the wanton creation of rules and laws to breach privacy is possible to contain, by approaching high courts and the Supreme Court, *post facto*, in case the safeguards built into legislative practice prove insufficient. This has been the case following the executive order of 2019 that created the UIDAI, and the subsequent passage of the money bill that was converted into a law by signature of the President, in 2016. Parts of that law were struck down in 2018 on grounds of unconstitutionality, and review petitions that question the remaining law are pending, as of this writing.

7 WP(C) 9143 of 2014 before the Delhi High Court). Unlike the large set of cases "clubbed" together before the Supreme Court that were eventually adjudicated in 2018, this case has remained unheard, before a staggering number of successive justices who either retired or were removed from the case. This is a matter of record.

8 There are many estimates circulating, both for the fuzzily defined "middle class," and the actual number of smartphone users, in India, and have been for decades. The reasons for this are not hard to comprehend: neither is India's "middle class" equivalent to that of the middle class of countries that industrialised 50 to a 100 years earlier, nor do patterns of ownership and usage of phones in India match those found in other countries. Phone usage, for instance, varies widely from a single phone used by many members of a family, to two or more phones used by a single person. In both cases, users might belong to the economically defined "middle class," but for whom disposable income may not actually be available for opportunistic spending, in the same manner as might be found routinely for individuals in say, Europe or the United States. Phone usage patterns, both for the ostensibly simpler "featurephone," and for the more adaptive smartphone, are likely to be culture-specific, and India is in itself a grouping of highly individualised cultures.

9 The text of the letter addressed to the concerned police station, which is in the public domain, makes it clear that the data has been hosted with Amazon Web Services, outside India. It speculates that the data was acquired by the accused company illegally, from either UIDAI itself, or from the various state resident hubs. This acknowledges, incidentally, that data gathered by UIDAI and claimed (in the proceedings before the Supreme

Court) to be hosted under conditions of great security at Central Identities Data Repository was shared with multiple agencies outside of the control of the Central Government, apparently with poor safeguards and little concern for potential misuse.

10 https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. The European Commission revised, in 2018, its common data protection rules, applicable both for persons and entities residing within the nations who subscribe to the European Union (EU), and also to those outside, who want to do business with persons and entities in the EU. Such business may directly or indirectly involve data handling both within the EU and outside, and the new rules lay down the legalities and conditionalities of doing so. In particular, the rules lay down very stringent penalties for mishandling of such data, and breaches will be punitively punished. Some commonplace businesses, delivering services online, found themselves unable to comply, and shut down their offerings within the EU. Despite this well-publicised exercise, which took years in the making, there has been almost no commensurate protective legal framework development in India, which has claims to being an IT-savvy nation. We continue to lack a basic data protection framework, and an exercise conducted through 2018, with the purpose of starting such a legal framework, has been widely criticised for the emphasis it lays on protecting data-handling businesses, rather than the persons whose data is involved.

11 Speech of Ravi Shankar Prasad in Parliament on 10 April 2017 stating that 34,000 agencies have been blacklisted