The Third Way

The Third Way

India's Revolutionary Approach to Data Governance

Rahul Matthan

juggernaut

JUGGERNAUT BOOKS
C-I-128, First Floor, Sangam Vihar, Near Holi Chowk,
New Delhi 110080, India

First published by Juggernaut Books 2023

10 9 8 7 6 5 4 3 2 1

P-ISBN: 9789353452636
E-ISBN: 9789353451851

Illustrations by Priya Patil

Typeset in Adobe Caslon Pro by R. Ajith Kumar, Noida

Printed at Thomson Press India Ltd

For Appa

Contents

Terminology and Notes

Data protection regulations in almost every country follow a standard naming convention for the different parties concerned. The individual whose personal data has been collected is usually called the data subject, the entity that has either collected it or has effective control over it post collection is called the data controller, and any entity that performs any actions on the data (collection or any other form of processing) on behalf of the data controller is referred to as the data processor.

However, when Justice B.N. Srikrishna was tasked with drawing up a new regulatory framework for data protection in India, one of the first things he did was change this naming convention. The last time Indians were anyone's subjects, he said, was when the queen of England ruled India. And so, to underscore the central importance of the individual in data protection, the data subject is referred to, in India, as the data principal. Extending this principle even further, the data controller is referred to as the data fiduciary since it holds the data principal's personal data in trust.

In this book, I have used Justice Srikrishna's terminology, retaining the terms data principal and data fiduciary when

referring to entities in the context of their data protection rights and obligations. If you are more accustomed to European terminology, kindly adjust your frame of reference appropriately.

I've always been irritated by books where footnotes take up half the page. Or where they are piled up at the end – leaving the reader feeling even though they have reached the end, they have not read all the author has to say.

Notes to text, in my mind, should be entirely optional – available to readers who want to go deeper – and not unnecessarily add to the bulk of the book. It is for that reason that this book has neither footnotes nor endnotes. For readers who want access to sources or additional material, that material is available at www.thethirdway.in.

.

Introduction

In 2008, less than one in twenty-five Indians had a verifiable identity. Just about 17 per cent had a bank account and only 15 per cent used digital payments. However, by 2018, almost every one of the 1.4 billion citizens of the country had a digital identity; over 80 per cent had a bank account and 68 per cent of all payments were digital. According to the World Bank, between 2014 and 2017, one of every two new bank accounts opened in the world was in India. In December 2022 alone, 8 billion digital payment transactions were processed in India.

This dramatic transformation can be directly attributed to the digital public infrastructure (DPI) that India has built, one layer at a time, for the last fifteen years. The resulting ecosystem represents a unique approach to solving societal problems by using powerful digital ecosystems built using open, interoperable architecture for public benefit. These ecosystems have evolved across a range of different sectors – from financial services to healthcare – and taken together, offer services ranging from identity and credentials to payments and credit. Thanks to the widespread penetration of mobile data, these are accessible to everyone, anywhere and at any time.

The transformation this has wrought across Indian society has had powerful ripple effects throughout the economy. While affluent sections of society could always access services like these (they already had verifiable identities, bank accounts and credit cards for digital transactions), the deployment of these systems at scale has significantly benefitted the poor and the marginalized, allowing them to significantly better their lives.

All the literature on the subject has, so far, positioned India's DPI as a digital transformation story, one that describes how India, through an ingenious use of digital technology systems, has been able to solve intractable societal problems with an efficiency and economy that have helped it leapfrog several stages in its economic development journey. These stories focus on the technology India has built, its design, its features and the outcomes it has achieved.

However, there is more to India's DPI story than simply the digital transformation it has wrought. Contained within these digital systems is a robust data governance framework that operates by embedding legal and regulatory objectives directly into code. This means that instead of looking to impose obligations through the enforcement of laws, compliance can be achieved by getting participants to use these ecosystems. This gives the regulators who manage the protocols and specifications on which these systems function direct control and, often, the ability to influence policy outcomes in real time.

This is a brand-new approach to data governance, one that represents a powerful new alternative to how the world has traditionally achieved regulatory objectives.

There are, broadly speaking, three actors in a digital ecosystem: market participants who use it to provide and receive services; regulators who establish the rules of the road, stipulating what can and cannot be done; and technology providers who build the applications, protocols and digital infrastructure that define the contours of the ecosystem. A well-functioning digital ecosystem allows each of these actors to operate at full potential. This is the role of governance.

This book looks at India's DPI from a governance lens. It describes how, in the hands of a thoughtful regulator, technology can be used to further policy goals. It describes how legal principles could be embedded directly into the fabric of the infrastructure so that the mere act of participating in the ecosystem guarantees compliance. It also describes how, by carefully arraying competing interests against one another, it might be possible to use the incentives that drive these different actors to achieve the stated objectives.

I will make my argument in three stages.

I will first discuss the challenges of data governance in the age of the internet, examining the problems that have arisen because of how data has insinuated itself into every aspect of modern society and the many legal and regulatory issues that regulators today have to deal with as a result. To do this justice, we will need to understand the history of computer and data technology, why it was created and what it has metamorphosed into.

We will reflect on how these new technologies have benefitted us and the harms that we now need to contend with. This will allow us to appreciate the need for data governance and the fine line that regulations need to walk between limiting what these technologies can be used for and ensuring that we can take full

advantage of all they offer. In the process, we will examine the two distinct approaches to regulation prevalent today and why we need an alternative.

We will then discuss the range of solutions that comprise the entirety of India's DPI. Since it is beyond the scope of this book to provide a comprehensive listing of all the different elements of India's DPI stack, we will take a whirlwind tour of a representative sampling. I have found it helpful to organize DPI in India according to the different stages of maturity that they describe.

When countries start on their DPI journey, they design the DPI to provide citizens with *access* to the DPI ecosystem. At this stage of the maturity journey, DPI typically consists of digital identity and verifiable credentials. Once these have been established, nations need to enable *engagement* among participants in the ecosystem. A wide variety of DPI makes this possible depending on the sectors to which they are applied. The most common examples of this are the fast payments systems that countries establish to enable digital payments, but this category could also include solutions that enable commerce, logistics, education and healthcare outcomes.

It is only after the population has been engaged in the digital ecosystem for some time that it can graduate to the next stage of maturity – *empowerment*. At this stage, the DPI is used to enable citizens to take advantage of the digital trails that they have laid down by establishing frameworks for consented data transfers that put them in control of how their data can be used within an ecosystem of service providers to offer opportunities over and above what is currently available to them.

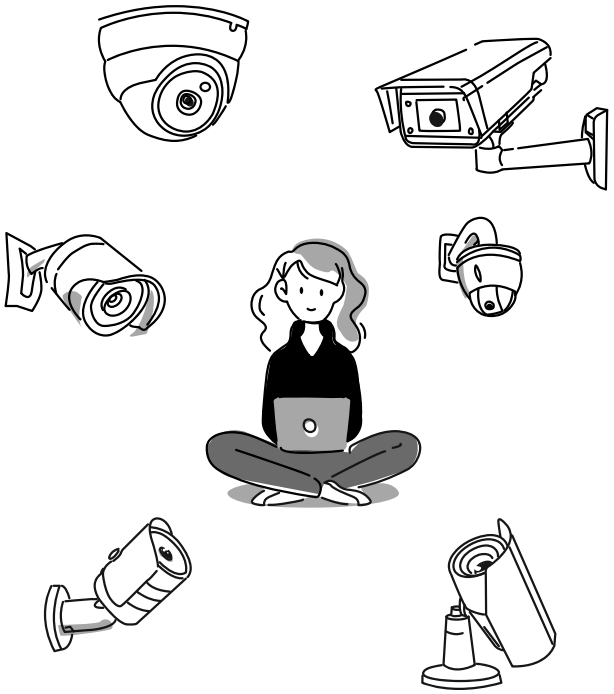
In the final section, I will discuss how India's DPI approach

can be used to address the challenges of data governance. I will argue that the design principles on which India's DPI has been built on can, in the hands of a thoughtful regulator, be used to effect policy outcomes by implicitly shaping how interactions take place within the ecosystem. Based on India's experience of operationalizing these networks, I will attempt to describe how it is possible to carefully align the incentives of various ecosystem participants to achieve desired outcomes, even though some of these outcomes might be inimical to the stated objective of these participants.

Finally, I will argue that it is possible to embed regulatory principles directly into the code of the technical infrastructure. I will do this by analysing India's Data Empowerment and Protection Architecture, which has implemented a consented data sharing framework by directly encoding data protection principles into its workflows.

Section 1

The Challenges of Data Governance



The Data Revolution

'Any sufficiently advanced technology is indistinguishable from magic.'

– Arthur C. Clarke

By the year 2000, we were already a technologically advanced species. We had tamed much of the planet, dedicating vast tracts of land to generate food, effectively eliminating, or at least significantly mitigating, the scourge of famine for much of the rapidly increasing global population. We had conquered the great distances that separated us with powerful machines that cross the seas and skies in relative comfort and safety. A massive global communication infrastructure had made it a trivial matter to speak to people anywhere on the planet. As a result, the world, which used to be a vast and largely inaccessible place, had become much, much smaller.

And yet, all our achievements up to that point, remarkable as they were, were about to pale in significance compared to what humanity was going to achieve over the next two decades.

While we've had the internet for decades, there were at the turn of the twenty-first century only 400 million people using it, nearly half of whom lived in North America and Europe. Today,

that number is fast approaching 5 billion, with the vast majority of internet users accessing it from Asia.

In 2018, the total amount of data consumed in the world was 33 zettabytes (33x10^21 bytes). By 2020, that rose to 59 zettabytes, and it is anticipated to reach 175 zettabytes in 2025. But apart from the sheer size of the number, what it is being used for and how, has changed so radically that our past selves would have considered much of what we take for granted today to be fantastical.

A large part of this transformation has been driven by the growth of mobile data. At the turn of the century, there were just 740 million mobile phone subscriptions worldwide. Today, there are more mobile phones than there are people on the planet. By the end of 2021, 4.3 billion people were using the internet on their mobiles, even though 3.2 billion people living within the footprint of a mobile broadband network were still not using it. In India, where the overall teledensity is 84 per cent of the population (over 1.1 billion people), 82 per cent of that is due to wireless access. Of India's 840 million broadband subscribers, 805 million access it over wireless broadband.

Most of these subscribers use the internet for instant messaging, voice and video calling and watching free videos. The single largest online activity is instant messaging, with nearly 62 per cent of global mobile internet users using some form of it daily. There has been an uptick in other uses too, thanks in no small part to the coronavirus (COVID-19) pandemic. Today, 38 per cent of mobile internet users report that they use the mobile internet for education at least once a week, an increase of over 10 per cent compared to 2019. Around 21 per cent use it to manage their health and 14 per cent to order goods or services.

After communication, the second-most popular activity is watching videos. YouTube, on its own, accounts for 25 per cent of global mobile traffic, with creators uploading 500 hours of videos to the site every minute. Over 2.6 billion people worldwide use YouTube every month, and 467 million of them are in India. This is nearly double the number of users from the second-largest market, which is the United States at 246 million. T-Series, India's largest music label and movie studio, has the largest number of subscribers globally, with 245 million subscribers to its YouTube channel at the time of writing. Every day, people watch over a billion hours of video, with more than 70 per cent of them doing so on a mobile device (there are over 10 billion YouTube installs on Android devices).

We can, today, find videos on just about any topic online, from music videos and movies from every corner of the planet to historical footage, educational content, social commentary and memes. There exists such an abundance of choice that it is impossible for anyone to consume all that is on offer. As a result, access to knowledge has truly become unlimited. Anyone can benefit from the expertise of the greatest minds that ever existed by simply surfing on their mobile phones at next to no cost.

And yet, the internet is so much more than just a source of information. Thanks to the ever-increasing abundance of mobile bandwidth, cloud storage and real-time computing, the internet has enabled the development of powerful technologies that power much of our daily lives in magical ways, so much so that the world we inhabit today would have seemed fantastical to someone even just twenty-five years ago.

Getting Around

Take navigation, for example. There was once a time when we had to plan carefully before going somewhere, getting clear directions as to how to get there, knowing that if we made a mistake, we were sure to get lost. There were times that I have gone back without finding what I was looking for because there was no way to ask for directions and course correct when I was obviously in the wrong place. Today, we just whip out our mobile phones, punch in our destination and wait scant seconds until an app plots the two or three routes to our destination, no matter where on the planet it might be. We get into our vehicles and head to where we are going without even checking to see if the route we are headed down is the most efficient, confident in the knowledge that even while we are en route, the applications on our phones will be monitoring the traffic ahead of us and will inform us of alternate routes, should there be another way to get to our destination sooner.

Many of us don't even worry about active navigation anymore. If we want to go somewhere, we simply summon transport that can take us there, knowing that there is an application that will identify our current location, deliver a car and driver to our doorstep, ready to take us wherever we want to go. Chores like shopping have become a thing of the past. We no longer need to go to stores to pick up what we need. We can simply order things from the comfort of our homes, no matter whether we are picking it up from our local grocery store or a bespoke boutique halfway across the planet.

All this is possible because the mobile devices in our pockets are not just communication devices – they are powerful

multifunctional computers. They contain sensors that can locate us precisely, no matter where on the planet we might be, allowing us to navigate and to be the targets of navigation. They have gyroscopes and accelerometers built in that can, from how they jiggle in our pocket, sense when we are going for a stroll and when we are working out. This is why we no longer fear new places, assured that anytime we head out to a new destination, our phones will tell us how to get to wherever we want to go.

We've come to take all this for granted, but none of this would have been possible without our ability to gather vast amounts of data and use it to generate valuable insights. The maps we use on our phones today were generated, for the most part, by harnessing inputs from an army of volunteers who used early versions of these maps and helped improve their accuracy by correcting mistakes and identifying points of interest. Even today, improvements in most map applications are user-generated. The route information, along with travel times and traffic predictions, are the result of real-time calculations carried out using live inputs from the hundreds and thousands of commuters using those roads at that very moment, comparing the speed at which they are currently travelling with the speed that the algorithm had indicated you would have to drive at to get to your destination in time.

Everything Is Smart

But many of us, not satisfied with this level of data-fication, have gone one step further and embraced wearables, a wide variety of different devices that can be worn on our person, tracking our physical parameters in real time, 24/7. These include

smartwatches that, in addition to telling the time, track our heart rates, oxygen saturation and atrial fibrillation to know if we are cycling or swimming, whether we have had a fall, or are about to have a heart attack. We also have rings and chest straps that offer increasing levels of accuracy to help monitor our sleep or our athletic performances. And for those who want to hack their vitals, there are constant glucose monitors that collect a minute-by-minute record of blood glucose from interstitial fluid.

In much the same way, our homes are filling up with smart devices. Our televisions understand us when we speak to them and can curate content for us across a range of different channels, organizing them all into a video feed of what we most like to watch. Some are smart enough to understand which family member is watching and can reorganize the feed to that person's preferences. Various other smart devices participate in our daily lives: smart speakers play music for us, remind us of what we need to do and shop for us when the cupboard is bare; doorbells can identify who is at the door, allowing us to screen visitors; and a range of smart plugs, light bulbs and electronic utilities can switch themselves on and off autonomously to improve the energy efficiency of our homes and offices or to simply improve our mood. All these devices are internet-enabled and optimally intelligent, capable of carrying out simple workflows either by themselves or as part of a cloud-based solution. They typically use information they gather from us and our environs to offer a level of convenience and comfort that would have been unheard of even a decade ago.

But the benefits of modern technology are far greater than just the gadgets they have enabled. Our lives today are greatly enriched by the power of data that we can now aggregate in vast

'data lakes' in the cloud, which can be easily accessed and analysed. All the devices and gadgets we have surrounded ourselves with suck data from us, piping it into these massive data warehouses where they are pooled with similar data from others from all over the planet. Because of the sheer volume of these massive data sets and the number of data fields they comprise, it is possible to discern patterns in the data that would have otherwise remained invisible. The trends that emerge allow us to predict events with a degree of certainty that would have otherwise been impossible. We have learnt to use powerful algorithms to harness such data sets, training them to corral insights out of that data – typically products or services we may like to purchase and activities we might want to perform.

We also take for granted the algorithmic recommendations we receive, that nudge us to a particular course of action that might not have occurred to us without prompting. When we listen to music, we are offered suggestions about the next track and, more often than not, I find myself trying the algorithm's recommendations. When I finish reading a book, I get a variety of suggestions about my next good read, each as eclectic as the one I just read. When we're done with watching a movie, almost before the credits finish rolling, we are presented with suggestions as to what we might want to watch next. Little wonder that Netflix users are happy to follow the algorithm's recommendations over 80 per cent of the time.

Health and Wealth

There are numerous other ways in which data is used for our benefit that have nothing to do with the commercial applications

most visible to us. Take health, for example. Thanks to the veritable explosion of wearable devices, many of us now have a longitudinal record of various health parameters collected at intervals of as short as a minute apart. If these are plotted on graphs and tied together, they provide us with trendlines offering insights about our health. A recent study found that data from wearables could be used to predict the onset of influenza-like illness up to a week before symptoms appeared by analysing changes in heart rate and sleep patterns – offering hope of a new kind of early warning system in the event of another pandemic.

Data has also helped scientists make significant breakthroughs in our understanding of the human genome. Thanks to dramatic improvements in computational power and processing capabilities, it has become much easier to sequence the human genome and decipher its meaning. Scientists can now identify specific genes that contribute to certain diseases more accurately and develop personalized treatments. CRISPR gene editing technology has made it possible for scientists to edit an individual's DNA, potentially curing genetic diseases and even preventing certain conditions from developing in the first place. For example, in an episode of my podcast *Ex Machina*, I tell the story of how advancements in these technologies offered a young boy in Bengaluru, India, the chance at life which would otherwise be cut short because of a genetic defect that had already put him into a wheelchair at a very young age.

Lenders can use data to get new insights into potential borrowers to help them better understand their ability to repay. Credit bureaus have been collecting information on borrowers for decades, using this data to build a credit history and a record of payment behaviour. But what is different is the range of

new and often unexpected data sources that lenders can now reliably lean on to determine creditworthiness. Many of these alternative credit-scoring models can consider data about the actual transactions the borrower has undertaken or the regularity with which they pay their utility bills. The more sophisticated of these algorithms also draw insights from unlikely parameters such as whether a potential borrower types in capital letters or the average battery life on their phone. For people who don't have a traditional credit history – and no direct way to get one – these algorithms are a viable path into the formal economy. Given that in India, only 14 per cent of small businesses have access to credit, leading to an astounding $530 billion credit gap, there is almost no option but to find ways to use these unusual technological workarounds.

Teachers can use data to gain insights about their students to improve their educational outcomes. By collecting information about the performance of their students across a range of different parameters, they can build a more granular picture of each student's strengths and weaknesses. This, in turn, can help them re-orient their teaching methodology to help each student improve. For instance, if a school can analyse data about reading speed, fluency and comprehension, it can identify patterns that can accurately predict which students are most at risk of falling behind and design interventions early enough to help struggling students catch up.

We have traditionally hired for jobs based only on resumes and interviews, even though we know that these are, at best, poor predictors of how a candidate will perform in a given job. By studying data from various sources, recruitment agencies can get a more complete picture of a candidate's qualifications,

experience and preferences. Recruitment agencies are increasingly collecting data on personality traits, work style and company culture to assess a candidate's fit for both the role and the organization. Google famously discovered that the most effective teams are those made up of people with diverse backgrounds and perspectives, leading to a fundamental shift in how the company hired.

Examples, thus, abound of how data at scale, harnessed to technology capable of processing it, can offer actionable insights across a range of different sectors. A couple of decades ago, even if this was not accessible at scale, technology had progressed to the point where these possibilities were beginning to reveal themselves.

But while it might have been possible for the more prescient among us to have anticipated many of these advances, the one thing no one would have even dreamed of was how artificial intelligence would come to transform our lives by doing things that rival the capability of human minds.

If there ever was a technology so advanced as to be indistinguishable from magic, it would, at the turn of the century, have been artificial intelligence (AI).

Machine Intelligence

It all started with text. And search engines.

When the internet exploded, it rapidly grew so large that it soon became impossible to find anything useful unless someone showed you where it was. Search engines capable of doing this well became the gateways to our internet experience. Success began to be defined by how well the algorithm could understand

what humans meant when they typed a query. If they were going to give us the precise result we needed, they needed to first understand what we were looking for.

And so, the IT industry dedicated itself to refining search engine algorithms to solve this problem. Since all the world's search queries ran through their pipes, all they needed to do was process this information and apply probabilistic rankings to possible outcomes. Over time, they became preternaturally good at this, to the point where not only were search engines able to understand what we were thinking when we typed a query into the search field, they were also able to predict our question before we finished typing it. This autocomplete functionality in search soon spilled over to various other areas. Our SMS and instant messaging applications began suggesting short responses to the messages we received and our email programs, which were able to process the history of our correspondence, suggested more extended responses.

What started as a simple autocomplete functionality has, today, evolved into much, much more. Voice assistants that were long viewed as little more than gimmicks – novelty items that could set the alarm or tell you what the weather would be – are today an integral part of our lives. We use them to order groceries, control our smart homes and even carry out financial transactions. Where they previously struggled to understand anything other than English spoken in a Western accent, they now have no problem understanding not only English spoken in a wide range of accents but virtually every major language on the planet. The impact of this on accessibility alone is significant as people who only speak their mother tongue can now use technology that was previously beyond their reach.

AI can also now generate, from scratch, images and videos that are indistinguishable from those made by humans. One of my favourite pastimes is using AI algorithms like Midjourney to create high-quality images as illustrations for my writing, for no better reason than it is fun to see what the algorithm comes up with. I am very rarely disappointed. I know that similar capabilities exist in video creation and it is only a matter of time before I graduate to trying my hand at that. Now computer-generated imaging has become such an integral part of the movie industry today that it is impossible to tell real actors from those entirely generated by CGI. The ability to generate high-quality videos based solely on the text prompts of a creator will soon radically transform the industry. There is no better evidence of that possibility than *The Crow*, a short film generated entirely by text-to-video AI, which won the Jury Award at the Cannes Short Film Festival in 2022.

If all of this leads you to believe that the AI advances are of little use for anything more than entertainment, you need look no further than how it is transforming the field of medicine. Image recognition algorithms are being used with incredible success in radiology. When applied to analysing X-rays, CT scans and MRIs, these algorithms have proved capable of detecting abnormalities invisible to the human eye. They can detect and identify pulmonary nodules, colonic polyps and micro calcifications which indicate different forms of cancer. In the case of skin cancers, they can detect a wide variety of sizes, shades and textures of suspected lesions better than a trained dermatologist could, to the point where they are sometimes capable of characterizing a tumour as malignant.