HIGH-VALUE DATASETS: BOOSTING THE ECONOMY OR JEOPARDISING PRIVACY

*—Ishan Ashish & Harsh Anmol**

Abstract— With rapid digitisation in India, the volume of data is growing substantially. The Indian government has realised the potential of such digitisation in facilitating opportunities for better governance and growing innovation in sectors critical for societal transformation. Against this backdrop, the Government released the "India Data Accessibility and Use Policy" to "radically transforms India's ability to harness public sector data". The policy, if passed, would govern all non-personal data and information created/generated/collected/archived by the Government.

It aims to open Government data by default and to create High-Value Datasets (hereinafter, 'HVDs'), which will be shared freely within government departments and licensed to the private sector. Making data open is vital for leveraging its potential for Indian society and economy. This may involve enriching the research and developing new products and services. The impact of open data is essentially realised by application and depends upon various factors such as costs, data quality and documentation, and access modality.

These concerns must be addressed to maximise the impact of open data and lower market entry barriers for start-ups and Small and Medium Enterprises (hereinafter, 'SMEs'), resulting in increased business opportunities. To increase its impact, efforts should be focused on datasets with the greatest economic and societal potential. HVDs refer to documents, the re-use of which is associated with economic and societal benefits because they create value-added services and high-quality jobs. There exist several lacunae in the policy because the government fails to ensure that sharing of data respects citizens' right to privacy. There is a concern that open governmental data could create 360-degree profiles and enable mass monitoring as India lacks

* Students, National Law University and Research in Law, Ranchi.

legislation providing accountability and remedy for privacy violations. Further, the Government fails to recognise the existence of automated tools for the re-identification of anonymous data. In light of such issues, the paper aims to highlight the benefits of the HVDs for the economy and society and exhibit how the current approach of the government can lead to infringement of the right to privacy.

Keywords: High-Value Datasets, Small and Medium Enterprises, Open Data, Non-personal Data, 'Draft India Data Accessibility & Use Policy.

INTRODUCTION

The world is witnessing a digital era and unprecedented technological advancement. With such technological advances, our lives have become more entangled with data. In recent times, the need to make open data available for everyone is an imperative step to leverage its potential for society and the economy. For instance, the opening of data is essential to enrich research; it helps to encourage informed decision-making, and research and development, which eventually leads to the development of new products and services.[1] Open data generally refers to non-personal data that is accessible to all and can be freely used, re-used, and distributed by anyone.[2] The impact of open data is essentially comprehended through applications and depends on several factors, such as costs, quality of data and its documentation, or its modality of access.[3] To maximise the impact of these data, it is critical to identify and target datasets with the most significant perspective for society and the economy. In this regard, the Indian government released a policy proposal titled "*Draft India Data Accessibility & Use Policy, 2022"*[4] (hereinafter, 'draft policy'). This policy aims to radically transform India's ability to harness public sector data and create societal as well as economic transformation. High-value datasets can be defined as "*the documents the re-use of which is associated with important benefits for society, the environment and the economy, in particular, because of*

---

1 Esther Huyer and Marit Blank, *Analytical Report 15, High-Value Datasets: Understanding the Perspective of Data Providers* (*Publications Office of the European Union*, 25 February 2020) <https://data.europa.eu/sites/default/files/analytical_report_15_high_value_datasets.pdf> accessed 20 July 2022.

2 Antti Halonen, *Being Open About Data: Analysis of the UK Open Data Policies and Applicability of Open Data* (*Finnish Institute in London*, 2012) <http://fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf.> accessed 20 July 2022.

3 Esther Huyer and Marit Blank, *Analytical Report 15, High-Value Datasets: Understanding the Perspective of Data Providers* (*Publications Office of the European Union*, 25 February 2020).

4 Ministry of Information and Technology, 'Draft India Data Accessibility & Use Policy' (2022) <https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf> accessed 20 July 2022.

their suitability for the creation of value-added services, applications, and new high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets".[5]

Although opening of data has several benefits, it potentially threatens our right to privacy, especially the right to be forgotten, which is an essential part of Article 21. In the absence of a proper data protection law in the country, the sharing of non-personal data as well as all the information created/generated/collected/archived by the Government of India, will be unregulated. This paper aims to highlight the intricacies of open data, the economic and societal benefits of the opening of data, and identifying HVDs. Further, the paper highlights the conundrum about how these HVDs, without proper regulatory measures in place, can be a major threat to our right to privacy.

OPEN DATA AND ITS INTRICACIES

India is not the sole country that has identified the advantages of opening data and identifying the HVDs. In 2020, the EU members found it vital to leverage its potential for European society and economy.[6] The Open data encompasses within it two specific concepts of openness, firstly, data is "*legally open*" which implies that the data is published under an open license and that the conditions for re-use are limited to attribution, and secondly, the data is *"technically open"* this means that the data is free to access for everybody, and the file format and its content are not restricted to a given non-open source software tool.[7] In 2009, Hans Rosling posited a niche argument on the potential of data. He coined the term "raw data, now", which was based on the principle that taxpayers have already paid for government-collected data, and thus, the general public should be allowed to re-use it.[8] In 2013, the G8 summit (now G7) defined the importance of Open Government data by creating the Open data charter. It emphasises the role that Open data can play in both governance and growth simulation.[9] Open data, when made freely accessible, can be re-used to create value, which in turn can lead to enriching research, informing decision-making, or developing new applications. It can reduce market entry barriers for start-ups and SMEs.

---

5 Council Directive (EU) 2019/1024 of 20 June 2019 on Open Data and the Re-use of Public Sector Information (recast) [2019] OJ L 172/56.

6 Esther Huyer and Marit Blank, *Analytical Report 15, High-Value Datasets: Understanding the Perspective of Data Providers* (*Publications Office of the European Union,* 25 February 2020) <https://data.europa.eu/sites/default/files/analytical_report_15_high_value_datasets.pdf> accessed 20 July 2022.

7 Dinand Tinholt and Wendy Carrara, 'Open Data Goldbook' (*European Data Portal*, 2018) <www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf> accessed 4 September 2022.

8 Hans Rosling, 'The Best Stats You Have Ever Seen' (*TED*, 2006) <www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen> accessed 4 September 2022.

9 Leigh Dodds, 'How to Write a Good Open Data Policy' (*Open Data Institute*, 26 December 2016) <https://theodi.org/article/how-to-write-a-good-open-data-policy/> accessed 4 September 2022.

Now as per the draft policy, the opening of data aims to harness public data for large-scale societal and economic benefit.[10] The idea of reaping the benefits of public data for better governance was discussed way back in 2010 in Finland and other European countries.[11] It was observed that opening public data resources contributes to at least three different goals: the transparency of democracy and administration, the creation of innovations and markets, and the increase of efficiency within the government.[12]

In the guidebook[13] published by the Finland's Ministry of Transport and Communication, the imperative need to create an ecosystem wherein the citizens, research institutes, corporate organisations, and government entities should be considered important when processing open data was discussed. It was observed that the creation of an ecosystem creates a new system that could aid in the utilisation of large-scale data, which, in turn, can create new services, research, and information that could potentially have some commercial value.

INDIA'S TRYST WITH OPEN DATA

In India, the Government in the year of 2012 issued the National Data Sharing and Accessibility Policy (hereinafter, 'NDSAP') to ensure proactive data sharing. The NDSAP policy emphasised high-quality public sector data and to achieve this goal the NDSAP policy identified quality as a key principle governing data access, recognising both quality control and quality upgradation as regulatory objectives.[14] However, the issue with India's tryst with open data public policy is that despite these policy manoeuvres, access to high-quality open public data remains difficult. This is evident from its ranking in the Global Open Data Index, which is 32 among 94 countries.[15]

The benefits of open data in different areas are conspicuous. However, the opening of data in itself is not sufficient to leverage its full benefits. Thus, it becomes imperative to identify quality HVDs that could aid in leveraging the

---

[10] Draft India Data Accessibility & Use Policy' (*Ministry of Electronics and Information Technology*, 2022) <https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf> accessed 20 July 2022.

[11] Antti Poikola, Petri Kola and Kari A Hintikka, 'Public Data: An Introduction to Opening of Information Resources' (*Ministry of Transport and Communication Finland*, 2010) <https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78201/Public_data_-_an_introduction_to_opening_information_resources.pdf?sequence=1> accessed 5 September 2022.

[12] ibid [13].

[13] ibid.

[14] Press Information Bureau, 'National Data Sharing and Accessibility Policy – 2012 Approved' (9 February 2012) <https://pib.gov.in/newsite/PrintRelease.aspx?relid=80196> accessed 5 September 2022.

[15] Open Knowledge Foundation, 'Global Open Data Index' <http://index.okfn.org/place/> accessed 5 September 2022.

potential of opening data.[16] Against this backdrop, the government decided to introduce the 'Draft India Data Accessibility & Use Policy, 2022'. While the policy occasionally refers to the need for data quality, it focuses more on the mechanism to process and disseminate high-quality public sector data. The intent of the policy is right, but there persist drawbacks that can hinder the accessibility of High-Quality Datasets to relevant stakeholders. The policy, for instance, does not define relevant data quality principles that can aid governmental agencies to make datasets more meaningful.[17] Furthermore, the policy has not provided a model definition of data quality. Without a proper definition of the same, maintaining and regulating the efforts to improve the quality of data cannot be sustained.[18] These drawbacks need to be resolved before the implementation of such a policy.

ACHIEVING SOCIO-ECONOMIC EQUALITY THROUGH HIGH-VALUE DATASETS

The G8 Governments, during the Conservative and Liberal Democrat Coalition Governments from 2010-15, published an Open Data Charter realising the enormous potential of Open Data. It said that the restrictions in access to data and data sharing from the end of governments and businesses which otherwise could improve the lives of people overseas make it a missed opportunity. For them, open data is "*an untapped resource with huge potential*." It can also increase transparency about the working mechanism of the government and businesses as it promotes accountability and good governance.

The use of accessible government data in creative ways can help in the development of innovative products that have the potential to make people's lives easier. In the private sector, such open data acts as a catalyst for innovation which can help generate new jobs and business opportunities. The G8 (Now G-7) realised the economic potential of the open data and, developed and followed a set of five principles that act as a foundation for the release, access, and re-use of data made available by them. One of the fundamental principles is *"Release of Data for Improved Governance."* They believe that the release of Open Data would strengthen our democratic institutions as it encourages better policymaking, which is true for countries across the world.[19] They also agreed to work on *"Releasing Data for Innovation"* according to which as

---

[16] Antti Halonen, *Being Open about Data: Analysis of the UK Open Data Policies and Applicability of Open Data* (*Finnish Institute in London*, 2012) <http://fininst.uk/wp-content/uploads/2017/09/being-open-about-data.pdf> accessed 20 July 2022.

[17] K S Roshan Menon, 'Building better datasets for India' (*The Hindu Business Line*, 18 April 2022) <www.thehindubusinessline.com/opinion/building-better-datasets-for-india/article65331546.ece> accessed 10 September 2022.

[18] ibid.

[19] Government of United Kingdom, 'G8 Open Data Charter and Technical Annex' (18 June 2013) <www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex> accessed 10 September 2022.

more people and organisations get adaptive to the use of data, the social and economic benefits will also become greater. For this, they agreed to work on increasing open data literacy to unlock its true potential.

Open Data-centric policymaking has been at Indian government's core for over a decade. The National Data Sharing and Accessibility Policy, which the Government of India issued in 2012, has vehemently stressed access to high-quality public-sector data, to which quality was identified as a key principle.[20]

For evidence-based policymaking, research is synthesised with policy formulation in a two-staged manner. At first, from the design stage, the policy decisions should be backed by sound data. Next, through data-collection exercises, an evaluation of implementation should be carried out.[21] The Central and State governments, through various initiatives and schemes, are making extensive use of data at all stages including formulation, implementation monitoring, and evaluation of the scheme. Data is at the core of various flagship programmes of the government, and it has been extensively used in the management of the COVID-19 pandemic as well.[22]

Data Analysis is also being done by the government of data collected through various surveys and censuses such as the National Crime Record Bureau's Crime in India Report, National Family Health Survey (hereinafter,'NFHS'), Census Data, etc. NCRB's data allows data-based policing, which can help focus on a specific area and allow police resources to be used more effectively.[23] NFHS provides key data on the social and economic determinants of health[24] and is a major household survey used to guide policymaking and programme planning for maternal and child health.[25] The United Nations has underlined the importance of census[26] in a handbook where it

[20] K S Roshan Menon, 'Building Better Datasets for India' (*The Hindu Business Line*, 18 April 2022) <www.thehindubusinessline.com/opinion/building-better-datasets-for-india/article65331546.ece> accessed 10 September 2022.

[21] Sohini Chatterjee, 'Hearsay or Fact: A Case for Evidence-Based Policymaking in India' *(The Wire*, 13 June 2022) <https://thewire.in/government/hearsay-or-fact-a-case-for-evidence-based-policymaking-in-india> accessed 10 September 2022.

[22] Dr Neeta Verma, 'Data-Driven Government' (*National Informatics Centre*, 22 May 2021) <www.nic.in/blogs/data-driven-government/> accessed 10 September 2022.

[23] Ram Pahal Pawar, 'Foreword: Crime in India-2020' (*National Crime Records Bureau*,1 September 2021) <https://ncrb.gov.in/sites/default/files/Foreword_ncrb-2020.pdf> accessed 10 September 2022.

[24] T K Rajalakshmi, 'NFHS-5 Findings: The Good, the Bad & the Ugly' (*The Frontline*, 17 June 2022) <https://frontline.thehindu.com/the-nation/public-health/the-good-the-bad-the-ugly-nfhs-5-report-documents-changes-in-india/article65463437.ece> accessed 10 September 2022.

[25] Rakhi Dandona and G Anil Kumar, 'Enhancing the National Family Health Survey-5 for Policy Making' (*Press Information Bureau*, 5 May 2022) < www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)31284-X/fulltext> accessed 10 September 2022.

[26] Dilip Mandal, 'Decade without Data– Why India is Delaying Census when US, UK, China went Ahead During Covid' (*The Print*,13 May 2022) <https://theprint.in/opinion/decade-without-data-why-india-is-delaying-census-when-us-uk-china-went-ahead-during-covid/954383/> accessed 10 September 2022.

states that evidence-based decision-making is a globally accepted process of efficient management of economic and social affairs and overall effective governance of societies today.[27]

The Ministry of Electronics and Information Technology, in July 2020, constituted an expert committee to study various issues relating to non-personal data,[28] headed by Mr. Kris Gopalakrishnan (Co-founder, Infosys).[29] He made a presentation on '*Data Regulation in India*' at the WTO Forum wherein he defined HVD as *"a dataset that is public-good and benefits the society at large. It helps in achieving a wide range of social and economic objectives including poverty alleviation, financial inclusion, agriculture development, skill development, etc"*.[30] Thus, the benefits of HVDs have been recognised by international organisations and the Indian Government as well. For this paper, the authors attempt to prove the significance of HVDs aid in the social and economic growth of the most downtrodden section of society by taking the recent policy of the Aspirational Districts Programme (hereinafter, 'ADP').

A primary example of data-driven governance and evidence-based policy at the district level in India is ADP, formulated by the NITI Aayog. Prime Minister, Mr. Narendra Modi, envisioned making India a $5 trillion economy by 2025.[31] However, GDP is not a good measure for economic activities.[32] To guarantee the availability of resources necessary for the citizens' well-being, the economy must grow, but, social well-being is something that needs to be taken into consideration, as social and economic development complement each other to promote a holistic development of the nation. Essentially, the goal of the country is not achieved by economic development alone.[33]

---

[27] United Nations Department of Economic and Social Affairs Statistics Division 'Principles and Recommendations for Population and Housing Censuses (*UNStats*, 2017) <https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf> accessed 12 September 2022.

[28] PRS Legislative Research, 'Revised Draft Non-Personal Data Governance Framework' (4 January 2021) <https://prsindia.org/policy/report-summaries/revised-draft-non-personal-data-governance-framework> accessed 12 September 2022.

[29] Ministry of Electronics and Information Technology, 'Report by the Committee of Experts on Non-Personal Data Governance Framework' (16 December 2020) <https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf> accessed 12 September 2022.

[30] Kris Gopalkrishnan, 'Data Regulation in India' (*WTO Forum,* 8 December 2020) <www.wto.org/english/res_e/reser_e/gopalakrishnan_dec20_e.pdf > accessed 12 September 2022.

[31] Rashesh Seth, 'One Vision and a Billion People's Dream for a $5 Trillion Economy' (*Firstpost*, 5 May 2022) <www.firstpost.com/opinion/india-5-trillion-economy-narendra-modi-10631691.html> accessed 5 October 2022.

[32] India Would Become a $5-Trillion Economy by 2026-27: CEA V Anantha Nageswaran' (*The Economic Times*, 14 June 2022) <https://economictimes.indiatimes.com/news/economy/indicators/india-would-become-5-trillion-economy-by-2026-27-cea-v-anantha-nageswaran/articleshow/92205688.cms?from=mdr> accessed 7 October 2022.

[33] Agyeya Tripathi, 'Social Progress Index- Moving beyond Economic Indicators' (*The Times of India*, 6 October 2021) <https://timesofindia.indiatimes.com/blogs/agyeya/social-progress-index-moving-beyond-economic-indicators/> accessed 27 December 2022.

Although India is on a high growth trajectory which is expected to lift millions of its citizens out of poverty, yet, the quality of life remains pathetic for many, which is not consistent with this growth story.[34] The country's persistent urban-rural divide and unequal progress have also been clearly shown in India's deplorable Human Development Index (hereinafter, 'HDI') ranking in recent years.[35]

The HDI for 2021/22 places India at position 132 among the 191 nations and territories. The HDI rankings of the majority of nations decreased in 2020 or 2021, reversing much of the progress made toward Sustainable Development Goals.[36]

RELEVANCY OF ASPIRATIONAL DISTRICTS PROGRAMME

The advantages of HVDs in policymaking and their ability to ensure that the benefits of these policies reach the most downtrodden and backward strata of society are undisputed. To further illustrate the benefits of HVDs, the authors in this paper have chosen the ADP and have attempted to exhibit how the HVDs have been phenomenal in aiding the programme to achieve its objectives. India can advance in the human development index by revamping the districts that have relatively less progressed towards achieving important social objectives.[37] This is where the ADP becomes relevant for India.

Under this programme, the Government of India every month identifies 115 districts and ranks them based on the progress that they have made on 49 key indicators under 5 broad socio-economic themes[38] which include Health & Nutrition, Education, Agriculture & Water Resources, Financial Inclusion & Skill Development and infrastructure.[39] These key parameters are the areas

[34] Press Information Bureau, 'Transformation of Aspirational Districts Programme' (5 January 2022) <https://static.pib.gov.in/WriteReadData/specificdocs/documents/2022/jan/doc2022153401.pdf > accessed 27 December 2022.

[35] Riuhin Deb, 'A Ray of Hope for Rural India: Aspirational District Programme (ADP)' (*Observer Research Foundation*, 22 September 2022) </www.orfonline.org/expert-speak/a-ray-of-hope-for-rural-india-aspirational-district-programme-adp/> accessed 21 December 2022.

[36] United Nations Development Programme, 'India Ranks 132 on the Human Development Index as Global Development Stalls' (*UNDP*, 8 September 2022) <www.undp.org/india/press-releases/india-ranks-132-human-development-index-global-developmnt-stalls?utm_source=EN&utm_medium=GSR&utm_content=US_UNDP_PaidSearch_Brand_English&utm_campaign=CENTRAL&c_src=CENTRAL&c_src2=GSR&gclid=EAIaIQobChMIu-3byqS8-gIVRSQrCh0ARQ37EAAYASAAEgI9e_D_BwE> accessed 21 December 2022.

[37] Ministry of Electronics and Information Technology, 'Report by the Committee of Experts on Non-Personal Data Governance Framework' (16 December 2020) <https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf> accessed 12 September 2022.

[38] Amitabh Kant, 'An Aspirational Journey for India's Underdeveloped Pockets' (*Financial Express*, 16 September 2022) <https://www.financialexpress.com/opinion/an-aspirational-journey-for-indias-underdeveloped-pockets/2204443/> accessed 27 December 2022.

[39] 'NITI Aayog, *Aspirational Districts Programme* <https://www.niti.gov.in/aspirational-districts-programme> accessed 22 September 2022.

where a HVD can benefit the public at large. This opens up a realm of possibilities for policy-makers, as it helps in understanding and identifying areas of intervention in a better manner, but also allows for an accurate assessment of the effectiveness and impact of interventions undertaken. By engaging in regular monitoring and using district ranking, the program hopes to engender a data-driven approach towards governance. The Champions of Change (hereinafter, 'COC') dashboard[40] is government's focused effort to emphasise outcome-based monitoring of public service provision. Based on COC data, three-year action plans for each sector were formulated by ministries and shared with all states and districts. NITI Aayog also conducts periodic review meetings to monitor the progress of implementing these plans. Data is at the center of these discussions and informs our decisions around policy and program priorities.[41] As a result of its success, the ADP Model launched in 2018, is being extended to block and city levels from this year.[42]

THE CONUNDRUM OF DATA LITERACY

The foregoing paragraphs exhibited the benefits of open data; however, the idea of open data has also faced several critical issues and open data advocacy has had to adapt in response to it. One of the most prominent issues faced by several countries is data illiteracy. India has taken a digital leap; however, digital literacy remains an issue in our country and has been proving to be detrimental to Indian companies.[43] Open data proponents have often made assumptions about the interests and potential of the citizens to act as potential end-users; however, initiatives to enhance the amount of open data available rarely take into account the resources needed to encourage the use of that data after it is released.[44]

Contemporary academicians have proffered many literacies needed in the modern world, the need for data literacy has also been recognised as one of the literacies which enhances the skills of the population. It has been argued that the rise of the web, in particular open data, has helped to put data literacy

---

[40] NITI Ayog, *Champions of Change Platform* (2018) <http://championsofchange.gov.in/site/coc-home/> accessed 16 September 2022.

[41] Amitabh Kant, 'Unlocking the Power of Data to Improve Lives: The Aspirational Districts Model', (*NITI Aayog*, 16 September 2022) <www.niti.gov.in/data-aspirational-districts-model> accessed 21 September 2022.

[42] Anubhuti Vishnoi, '213 Districts, 500 Blocks Identified to Scale up Aspirational District Model' (*The Economic Times*, 1 July 2022) <https://economictimes.indiatimes.com/news/economy/policy/213-districts-500-blocks-identified-to-scale-up-aspirational-district-model/articleshow/92582245.cms> accessed 21 September 2022.

[43] Hemai Sheth, 'Data-Literacy Gap Could Cost Indian Companies up to ₹33,216 Crore in Productivity' (*The Hindu Business Line*, 26 February 2022) <www.thehindubusinessline.com/companies/data-literacy-gap-could-cost-indian-companies-up-to-33216-crore-in-productivity-says-report/article30922038.ece> accessed 22 September 2022.

[44] Tim Davies and others (eds), *The State of Open Data: Histories and Horizons* (African Minds Publication 2019).

on the agenda of several organisations and agencies.[45] Although several definitions of digital literacy have been put forward by academicians, the definition proposed by D'Ignazio is pertinent as he defines it as "*the ability to read, work with, analyse, and argue with data*".[46] Further, the Data-Pop Alliance has propounded that data literacy as a term does not specifically deal with the necessity to adapt and update our understanding of the role of (open) data in the modern world, particularly as machine learning and other technologies change the use of data. They proffer that we should not refer to a sub-type of literacy (i.e., data literacy) but rather to the broader idea of "literacy in the age of data".[47]

While there exists a data literacy gap in our country,[48] however through collaborative efforts by the Government and privately owned organisations, these obstacles can be overcome. In order to do so, the method suggested by the organisation School of Data[49] is imperative. According to then, the organisations need to promote several data literacy capacity-building models, ranging from short-term efforts, such as workshops, community events, and datathons, to medium-term efforts, such as multi-week "lab" models and training programmes, and further, to long-term initiatives, such as fellowship programmes, research initiatives, and other longer-term projects.[50]

Further, efforts made by Internews' fellowships, which have been embedded in newsrooms for several weeks in places like Palestine[51] and Kenya[52] offer a longer-term model for building capacity by changing the processes of organisations to promote the use of data in everyday work. Thus, exhibiting how promoting data literacy at an organisational level can help in utilising data more efficiently.

---

45 Mark Frank and others, 'Data Literacy: What is it and How Can we Make it Happen?' (2016) 12(3) The Journal of Community Informatics 4-8.

46 Rahul Bhargava and Catherine D'Ignazio, 'Designing Tools and Activities for Data Literacy Learners' (*Oxford, UK: Web Science'–Data Literacy Workshop*, 2015) <https://www.media.mit.edu/publications/designing-tools-and-activities-for-data-literacy-learners/> accessed 25 September 2022.

47 Rahul Bhargava and others, 'Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data' (*Media MIT*, 2015) <www.media.mit.edu/publications/designing-tools-and-activities-for-data-literacy-learners/> accessed 25 September 2022.

48 Ram Sagar, 'Data Literacy fora Digital India: Why Should You be Serious About your Data' (*Analytics India Magazine*, 2021) <https://analyticsindiamag.com/data-literacy-digital-india-serious-about-data/> accessed 25 September 2022.

49 Dirl Slater, 'Research Results Part 2: The Methodologies of Data Literacy' (School of Data, 14 January 2016) <https://schoolofdata.org/2016/01/14/research-results-part-2-the-methodologies-of-data-literacy/> accessed 25 September 2022.

50 ibid.

51 Eva Constantaras, 'Using Data Journalism to Probe Economics in the West Bank' (School of Data, 2014) <https://schoolofdata.org/2014/08/27/using-data-journalism-to-probe-economics-in-the-west-bank> accessed 25 September 2022.

52 Ida Jooste, 'Data Journalism: The New Normal in Kenya' (Internews, 2016) <www.internews.org/story/data-journalism-new-normal-kenya> accessed 25 September 2022.

Therefore, to maximise the utility and efficiency of open data, we need to create an ecosystem where data literacy is emphasised and improved.

THE QUANDARY OF PRIVACY

Privacy concerns are at the forefront of the debates about data. While the economic and social benefits of HVDs are undeniable, the benefits should not come at the cost of the privacy of individuals. The Right to Privacy is a multi-dimensional concept, in India, it has been recognised as an integral part of Article 21.[53] Open data programmes urge the release of government datasets in re-usable formats under open licences. They also seek to make data findable and datasets interoperable to maximise their re-use both alone and in combination with other datasets. Many government datasets do, however, contain information about identifiable people. Furthermore, information about individuals' use of government services is among the most important types of government data.[54]

DE-IDENTIFIED DATA AND ITS INADEQUATENESS

The draft policy released, though, intends to share non-personal datasets.[55] Non-personal datasets which were initially personal have been anonymised (through the use of specific techniques to ensure that individuals to whom the data relates cannot be identified).[56] Despite the security of anonymisation, the possibility of inflicting harm to the original data principal prevail, as no anonymisation technique is flawless.[57] It has been suggested that using the term '*anonymised data*'can create a fall sense of security as it is almost impossible to ensure that after the anonymisation of personal data, it truly is and will remain anonymous.[58] Rather, a better term to use when trying to anonymise personal data is "*de-identified data*", that conveys the idea that the personal identifying information has been removed. However, caution must be exercised as there may still exist some unrecognised identifying information, or data, or any kind of inferences that could be re-identified when added or combined with other data.[59]

---

[53] *KS Puttaswamy v Union of India* (2017) 10 SCC 1 *SC*.

[54] Elena Simperl, Richard Gomer and Kieron O' Hara , *Analytical Report 3: Open Data and Privacy* (*Luxembourg: European Data Portal*, 2020) <https://data.europa.eu/sites/default/files/open_data_and_privacy_v1_final_clean.pdf> accessed on 25 September 2022.

[55] Ministry of Information and Technology, 'Draft India Data Accessibility and Use Policy' (2022) <https://www.meity.gov.in/writereaddata/files/Draft%20India%20Data%20Accessibility%20and%20Use%20Policy_0.pdf> accessed 20 July 2022.

[56] PRS Legislative Research, 'Revised Draft Non-Personal Data Governance Framework' (4 January 2021) <https://prsindia.org/policy/report-summaries/revised-draft-non-personal-data-governance-framework> accessed 12 September 2022.

[57] ibid.

[58] Aarathi Ganesan, 'Without a Data Privacy Law, India Must Consider Hazards of 'Deanonymisation' of Non-Personal Data' (*The Wire*, 8 October 2022) <https://thewire.in/rights/deanonymisation-in-non-personal-data> accessed 19 November 2022.

[59] ibid.

In today's time the Government have several databases which consists of our personal data, such as e-SHRAM in health, Ayushman Bharat Digital Health and many more. The government may use anonymisation techniques to convert personal datasets into non-personal datasets. However, the fear of encroachment into privacy may still loom large. There also exist operational challenges regarding the anonymisation techniques, as identifying personal datasets and releasing them by using anonymisation techniques may be time and resource-intensive.[60] Thus, there is a possibility that in some cases government resources may not be sufficient for the task. The vexing issue of encroachment upon individual privacy by the governmental database can be understood better with the help of an illustration. In the recent communal riot of 2020 in Delhi,[61] reports[62] suggested that data from the vehicle registration database of the Ministry of Road Transport and Highways's, namely *Vahan*, could have been used to identify the owners of vehicles belonging to a particular community to cause damage to their vehicle during the riots.

Since the right to privacy is a multi-dimensional concept, one such concept is recognised in the case of *K.S. Puttaswamy v. Union of India*[63] is of the right to be forgotten. The concept of the right to be forgotten has also been upheld in *Subhranshu Rout v. State of Odisha*[64] wherein the High Court pointed out the need for legislating the "*right to be forgotten*" as it was not possible in every case for a victim to approach the courts. In this regard, the research of Luc Rocher[65] shed light as they noted results point to the likelihood that even thoroughly sampled anonymised datasets won't meet the GDPR's current standards for anonymisation, which pose a severe technological and legal challenge to the de-identification release-and-forget paradigm. Right to be forgotten is not only about remembering and forgetting, or about the harms of extensive digital memory. The right to be forgotten is a corollary of informational autonomy or informational self-determination and represents the control an individual should be able to exercise over their data.[66] Now, when such personal information is published and re-used for commercial purposes, the citizens have a right to be forgotten i.e., they can ask for the removal or "erasure" of such personal

---

60 Tim Davies and others (eds), *The State of Open Data: Histories and Horizons* (African Minds Publication 2019).

61 Soibam Rocky Singh, 'A Report on the North-East Delhi Riots' (*The Hindu*, 13 October 2022) <www.thehindu.com/news/national/a-report-on-the-north-east-delhi-riots/article66007158.ece> accessed 18 November 2022.

62 Aditya Chunduru, 'VAHAN Data Used to "Target" Muslims' (*Deccan Chronicles*, 29 February 2020) <www.deccanchronicle.com/nation/current-affairs/290220/vahan-data-used-to-target-muslims.html> accessed 18 November 2022.

63 *KS Puttaswamy v Union of India* (2017) 10 SCC 1 (SC).

64 *Subhranshu Rout v State of Odisha* 2020 SCC OnLine Ori 878 (Ori).

65 Luc Rocher, Julian M Hendrickx and Yves- Alexendare de Montjoye, 'Estimating the Success of Re-identifications in Incomplete Datasets using Generative Models' (*Nature Communications*, 2019) <www.nature.com/articles/s41467-019-10933-3> accessed 18 November 2022.

66 Vini Singh, 'Striking the Right (to be Forgotten) Balance: Reconciling Freedom of Speech and Privacy – Dignity – Autonomy' (2021) 8(1) NLUJ Law Review.

information. The right to be forgotten and the right to privacy are two important limbs of Article 21 that cannot be properly exercised unless a proper regulation or legislative mechanism substantially empowers the citizen to safeguard their rights and provides remedies in case of a violation of these rights.

NEED FOR EFFECTIVE LEGISLATIVE MECHANISM

In India, there exists no such legislative mechanism as the Data Protection Bill which was withdrawn by the government.[67] Nevertheless, the Indian Government has introduced a new draft on data protection and privacy.[68]

The new draft on data protection and privacy is more simplified than its previous version and is easy to comprehend. However, the new bill[69] does not differentiate between personal and non-personal data. The draft essentially focuses on personal data which is intrinsically linked to privacy and can lead to an individual's identification. However, as technology has evolved over the years, the immense benefits of processing data have been recognised by multiple players in the data ecosystem. That's where the whole conversation around Non-Personal Datasets and data governance rests upon—the processing of data that are not largely protected under the 'burdens' of data protection laws.[70] Thus, it becomes easier to de-anonymise non-personal data and carry out any activity, which essentially defeats the very purpose of having a data protection bill as one way or the other, the personal data of citizens is in jeopardy and their right to privacy can still be infringed.

The need of the hour is to recognise the threat emanating from non-personal datasets and further widen the applicability of the new draft on data protection and privacy. Lessons can be taken from the European Union's General Data Protection Regulation, 2018 (hereinafter, 'GDPR'). The GDPR under recital 26 states that the GDPR will not be applicable to anonymous data. While the GDPR does not define anonymised data, it does include prerequisites which need to be met before any data can be anonymised. The EU expert advisory group on data protection provides three techniques under which anonymisation can be done namely: randomisation, generalisation, and masking.[71] The GDPR

---

[67] Vrinda Bhandari, 'Withdrawal of the Personal Data Protection Bill was a Bad Move' (*The Hindu*, 10 August 2022) <www.thehindu.com/opinion/lead/withdrawal-of-the-pdp-bill-was-a-bad-move/article65750995.ece> accessed 18 November 2022.

[68] Aashish Aryan, 'Government Releases Digital Personal Data Protection Bill Draft' (*The Economic Times*, 18 November 2022) <https://economictimes.indiatimes.com/tech/technology/government-releases-digital-personal-data-protection-bill-draft/articleshow/95599120.cms> accessed 22 November 2022.

[69] Digital Personal Data Protection Bill 2022.

[70] Aarathi Ganesan, 'Without a Data Privacy Law, India Must Consider Hazards of "Deanonymisation" of Non-Personal Data' (*The Wire*, 8 October 2022) <https://thewire.in/rights/deanonymisation-in-non-personal-data> accessed 19 November 2022.

[71] Sharp Cookie Advisor, 'Anonymization and GDPR Compliance: An Overview' (*GDPR Summary*, 21 July 2022) <www.gdprsummary.com/anonymization-and-gdpr/> accessed 29

is stringent regarding anonymised data, even when the datasets are eliminated of identifiers, if they do contain data that could lead to re-identification, then this 'anonymised' dataset would fall within the provisions of the GDPR.

Currently, the new bill provides that the new data protection act will be applicable to the processing of personal data collected from data principals online and secondly it will be relevant to such personal data collected offline, is digitised.[72] The old bill of 2019 provided different forms of data that come under the ambit of 'personal data' and lists 'inferences' under it.[73] However, the new bill completely negates such inferences. Inferences in this context refer to the unverifiable inference and predictions drawn by Big data analytics and artificial intelligence about the behaviours, preferences, and private lives of individuals. These inferences are based on extremely varied and feature-rich data with unpredictably high value and they open up new possibilities for biased, discriminating, and privacy-invading profiling and decision-making.[74] Clause 2(13) of the new bill of 2022 does not refer to the inferences specifically and restricts personal data's ambit to only those forms of data which make a person identifiable. The omission of classifying inferences as personal data is a regressive move, as explained (earlier) as even after the de-identification process there might be some inferences that could trace back the data principle and significantly jeopardise their right to privacy. The new bill needs to rectify this lacuna and extend the definition of personal data to include inferences within its ambit to ensure that individuals are provided with the full ambit of rights associated with personal data.

The new bill also needs to identify the risk associated with the anonymisation of personal data into non-personal data, a dataset can be considered 'anonymous' only if each individual is protected. The new data protection act also needs to include accepted techniques of anonymisation of data and specify under which circumstances data can be categorised as anonymised data, where there is a reasonable risk of identification, the data shall be deemed to be personal data and should be treated as such.

Furthermore, the draft policy defines data anonymisation as "*the process of converting personal data to a form in which a data principal cannot be identified, which meets the standard of irreversibility specified by the competent authority*". This definition is ambiguous as the government has not specified the standard of anonymisation required therein. It has been entirely left at the whims and fancies of the executive authority. Thus, neither the new bill on data protection nor the draft policy describes what anonymisation techniques

November 2022.

72 Digital Personal Data Protection Bill 2022, s 4.

73 Personal Data Protection Bill 2019, s 3(28).

74 Brent Daniel Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' [2016] Sage Journal <https://journals.sagepub.com/doi/full/10.1177/2053951716679679> accessed 30 November 2022.

can be used to convert personal data to a form where the data principal's identity cannot be identified.

The government needs to give clarity on what kind of anonymisation technique is permissible. Clarity on the issue is required as it will fundamentally increase and build the trust of the citizen and privacy preservation and security provisions pivot on trust (e.g., one will allow only those whom one trusts to enter one's zone of inaccessibility; one will not feel secure unless one trusts the security provider).[75]

CONCLUSION

High-Value Datasets have been a boon to the economy and society all over the globe. The world has moved to the era of Open Data which has enabled countries to identify such important datasets that benefit society and the economy. HVDs are a revolutionary change which if effectuated properly can substantially help in aiding and achieving the goal of a 5 trillion economy. However, to achieve the same, India needs to overcome several obstacles that the paper has highlighted. The need of the hour is to focus on increasing data literacy to utilise the full benefits of Open Data and HVDs; also the need to strike a balance between privacy and the economic benefits of these HVDs. The privacy of the citizens cannot be jettisoned to leverage the economic potential of these datasets.

---

[75] Wanbil W Lee, Wolfgang Zankl and Henry Chang, 'An Ethical Approach to Data Privacy Protection' (2016) 6 ISACA Journal <www.isaca.org/resources/isaca-journal/issues/2016/volume-6/an-ethical-approach-to-data-privacy-protection> accessed 10 December 2022.