Data Governance Network
Anchored by IDFC Institute

September 2020

Working Paper 10

Privacy in Healthcare: The Role of National Digital Health Blueprint

Alexander Fager and Prakhar Misra

IDFC Institute

Data Governance Network

The Data Governance Network is developing a multi-disciplinary community of researchers tackling India's next policy frontiers: data-enabled policymaking and the digital economy. At DGN, we work to cultivate and communicate research stemming from diverse viewpoints on market regulation, information privacy and digital rights. Our hope is to generate balanced and networked perspectives on data governance — thereby helping governments make smart policy choices which advance the empowerment and protection of individuals in today's data-rich environment.

About Us

IDFC Institute has been set up as a research-focused think/do tank to investigate the political, economic and spatial dimensions of India's ongoing transition from a low-income, state-led country to a prosperous market-based economy. We provide in-depth, actionable research and recommendations that are grounded in a contextual understanding of the political economy of execution. Our work rests on three pillars – 'State and the Citizen', 'Strengthening Institutions', and 'Urbanisation'. The State and the Citizen pillar covers the design and delivery of public goods, ranging from healthcare and infrastructure to a robust data protection regime. The Strengthening Institutions pillar focuses on improving the functioning and responsiveness of institutions. Finally, the Urbanisation pillar focuses on the historic transformation of India from a primarily rural to largely urban country. All our research, papers, databases, and recommendations are in the public domain and freely accessible through www.idfcinstitute.org.

Disclaimer and Terms of Use

The analysis in this paper is based on research by IDFC Institute (a division of IDFC Foundation). The views expressed in this paper are not that of IDFC Limited or any of its affiliates.

Design

Cactus Communications

Suggested Citation:

Fager, A., Misra, P. (2020). *Privacy in healthcare: the role of National Digital Health Blueprint.* Data Governance Network Working Paper 10.

Supported by a grant from Omidyar Network

Abstract

It is widely accepted that sharing of health data will reap benefits to relevant parties involved in the ecosystem. Yet, sharing of the data poses a threat to privacy of individuals. Given that health data is sensitive personal data, it holds a greater risk of harm to privacy should the necessary protocols to protect it not be in place. This paper evaluates the discreete components of the National Digital Health Blueprint, proposed by the Ministry of Health and Family Welfare under Government of India, in protecting privacy of individuals while allowing for sharing of health data. We look at Daniel Solove's taxonomy of privacy and evaluate the various features of the blueprint to see how it fares on protecting the specific types of privacies. We end with a few recommendations on how the mechanisms can be improved by changing/adding a few technical and institutional components proposed in the blueprint.

Table of Contents

1. Introduction **04**

2. Nature of Health Data and the Challenges it Poses for Policy **05**

2.1 Ethical Obligation to Share Data for Public Benefit 05

2.2 Obligation to Protect Patient Confidentiality 06

2.3 Risk and Uncertainty Inherent to Medical Care 06

3. Legal & Legislative Precedence for Health Data Sharing in India **07**

3.1 Technical Standardisation 07

3.2 Legislative Propositions 08

3.3 Judicial Proceedings 09

4. Finding the Balance between Privacy and Sharing **12**

4.1 Information Collection Privacy Harms 12

4.2 Information Processing Privacy Harms 13

4.2.1 Aggregation 13

4.2.2 Identification 14

4.2.3 Insecurity 14

4.2.4 Secondary Uses 15

4.2.5 Exclusion 16

4.3 Information Dissemination Privacy Harms 16

4.3.1 Breach of Confidentiality 16

4.3.2 Disclosure 17

4.3.3 Exposure 17

4.3.4 Increased Accessibility 18

4.3.5 Distortion 18

4.4 Invasion Privacy Harms 19

5. Conclusion: What next for NDHB? **20**

5.1 Willingness to Delete Data 20

5.2 Mandating Audit Trails and Digital Footprints 21

5.3 Collecting Meaningful Consent 22

5.4 Clarity on Process of Access to Health Data 22

References **25**

Acknowledgements 30

About the authors 30

Introduction

Over the past decade, governments have vigorously attempted to regulate the increasingly pervasive datafication of society and the economy. Regulatory attempts across jurisdictions are guided in large part by the purpose for which data is used and the risk of harm from its misuse (personal data vs. sensitive personal data).[1] Health/Healthcare data[2] falls in the category of sensitive personal data. The digital revolution has changed such data in multiple ways — primarily through an increase in sources of health data due to digitisation (Topol, 2012).

Before the digital revolution, health data was created almost exclusively in hospitals and clinics, through the interaction of a patient with the doctor. Other types of data were created in specific clinical trials or broad surveys, such as information on diet and exercise. All of this data was in the hands of the healthcare providers. Now, it is distributed across different stakeholders: companies, platforms, individuals, corporations, etc. The use of individual's data controlled by third-party entities can cause direct or indirect harms bringing into question erstwhile frameworks of governance for such data. The control over data is now much more distributed than before owing to the different types of data being generated. The increase in their concurrent points of creation with fitbits and smartphones collecting data of consumers adds to this phenomenon.

There is, today, a need to keep this data private and prevent harm to consumers. Yet, there are plenty of ways in which such data is helpful in its digital form — improving clinical care, diagnosis and patient health management. Digital governance must therefore walk the fine line between protecting privacy of individuals and sharing data to assist innovation and policymaking for public good. This qualitative difference between digital and analogue data makes this challenge far greater in former context. The volume, velocity, variety, value, and veracity of digital healthcare data changes both its potential and risk significantly. Much of the literature that we surveyed and many of the experts we spoke to emphasise supporting sharing of health data for public benefit over a regime that focuses on strict privacy protections.

India, to its credit, has recognised this problem and is gearing up to face it. In the past few years, a number of legislative and regulatory developments have altered the landscape: the recognition of a fundamental right to privacy, extensive deliberation around the nuances of the Personal Data Protection Bill, 2019 and the evolution of healthcare legislation. The National Digital Health Blueprint (NDHB) was also introduced to set up infrastructure for secure sharing of data. In this paper, we outline the steps that India has taken to protect health data privacy while balancing its public benefits and go deeper into the NDHB's role.

[1] Article 9 of the GDPR and chapters III and IV of the draft Personal Data Protection Bill in India echo these differences.

[2] Health data is any data “related to health conditions, reproductive outcomes, causes of death, and quality of life" - McGraw-Hill Concise Dictionary of Modern Medicine. McGraw-Hill. 2002. Moving forward, this paper uses the broadest form of defining health data, meaning that it is data that reveals information on the health condition of an individual or group, no matter its provenance. Although far from exhaustive, this includes data that might be garnered from sources as varied as wearable devices, social media, or internet searches. We recognise that this involves including both personal and non-personal data, which have different concerns, however think that that distinction is not very relevant to the analysis of the paper.

Section 2 explains the characteristics of health data that make it different compared to other types of data. We then look at technical, judicial and legislative developments in Section 3 to make the point that while sharing of health data is being promoted, measures to protect privacy may be falling short. This is followed by an evaluation in Section 4 of the safeguards in the National Digital Health Blueprint that aim to protect the privacy of patients. We end with Section 5 outlining suggestions for guidelines and frameworks that the government can use to better maintain this balance.

2. Nature of Health Data and the Challenges it Poses for Policy

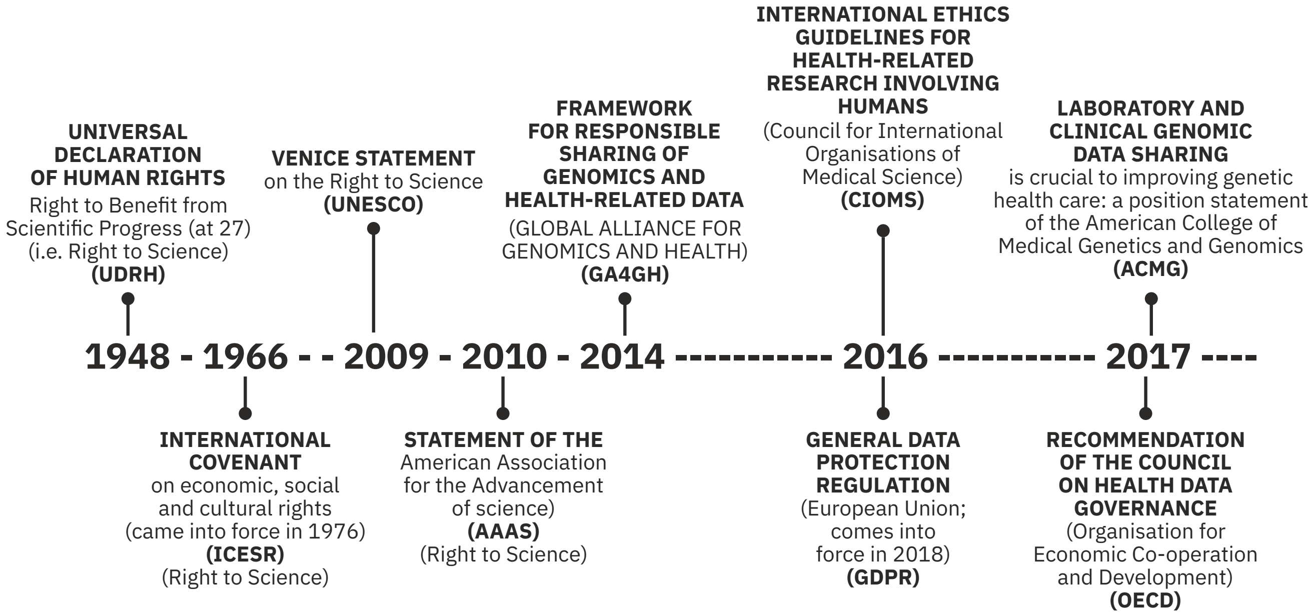
Health data is different from other data points as as it has two key characteristics — the limits on sharing affect the possibilities of saving lives and the cost of privacy is often externalised. Other data sets may have one or the other of these characteristics, but the combination is rare. Balancing patient care with patient privacy is thus incredibly important. We look at three characteristics of health data that make this trade-off difficult:

2.1. Ethical Obligation to Share Data for Public Benefit

Private health data comes with an ethical obligation that counterbalances the imperative to respect the individual's privacy: sharing it for public welfare, wherever it may be possible. “Ethics is about goods that we have reason – and sometimes even an obligation – to pursue, such as the good of knowledge that can be used to bring about significant improvements in health.” (Vayena & Tasioulas, 2016). The public welfare effect of health data falls in two categories - one, the usage of health data to prevent immediate, direct negative effects as in the case of infectious disease, and two, the use of data over the long term to improve medical science and the delivery of care.

The first category is well established -- not just as an ethical concern but also in international protocols for public health authorities. As the World Health Organisation (WHO) states in its policy for data sharing in public health emergencies, “epidemiologic data belong to the countries where they are generated, but there was consensus that data should be shared by default, but with an opt.out policy, to ensure that the knowledge generated becomes a global public good” (World Health Organization, 2016). In the case of emergencies, while there are concerns about data integrity, distribution of benefits, and clear line of ownership, there is still a generally recognised obligation for data to be shared (Langat et al., 2011). Many jurisdictions, including India, have data protection exemptions for public health use, particularly in emergencies.

In the second category of using health data for medical advances, Vayena & Tasioulas (2016) argue that the health data rights framework should extend beyond privacy rights to encompass a more dynamic and holistic view of other rights that include the “right to science”. The right to science is seen as a positive right and is ratified by 165 countries as a part of the 1948 Universal Declaration of Human Rights. Therefore, using health data to further the cause of improving human health is a legitimate purpose in this understanding. This puts an additional burden on governments to secure the right kind of frameworks to govern health data. In fact, many organisations have already laid out rules and guidelines on this front; as an example, the OECD has come up with 12 principles around governance and use of health data. The Council of International Organisations for Medical Sciences and the Global Alliance for Genomics and Health have also furthered the ways and means by which health data can be shared for research purposes.


UNIVERSAL DECLARATION OF HUMAN RIGHTS Right to Benefit from Scientific Progress (at 27) (i.e. Right to Science) (UDRH)
VENICE STATEMENT on the Right to Science (UNESCO)
FRAMEWORK FOR RESPONSIBLE SHARING OF GENOMICS AND HEALTH-RELATED DATA (GLOBAL ALLIANCE FOR GENOMICS AND HEALTH) (GA4GH)
INTERNATIONAL ETHICS GUIDELINES FOR HEALTH-RELATED RESEARCH INVOLVING HUMANS (Council for International Organisations of Medical Science) (CIOMS)
LABORATORY AND CLINICAL GENOMIC DATA SHARING is crucial to improving genetic health care: a position statement of the American College of Medical Genetics and Genomics (ACMG)
1948 - 1966 - - 2009 - 2010 - 2014 ----------- 2016 ----------- 2017 ----
INTERNATIONAL COVENANT on economic, social and cultural rights (came into force in 1976) (ICESR) (Right to Science)
STATEMENT OF THE American Association for the Advancement of science) (AAAS) (Right to Science)
GENERAL DATA PROTECTION REGULATION (European Union; comes into force in 2018) (GDPR)
RECOMMENDATION OF THE COUNCIL ON HEALTH DATA GOVERNANCE (Organisation for Economic Co-operation and Development) (OECD)

Source: https://www.sciencedirect.com/science/article/pii/S2452310017300264

2.2. Obligation to Protect Patient Confidentiality

Prior to the digital revolution, health data and patient confidentiality were sacrosanct. This was based on the established principles enjoining doctors to maintain the privacy of patients' data (Ferguson, 2015). As Kaplan (2014) points out, "the World Medical Association's International Code of Ethics makes respecting the right to confidentiality a duty integral to a physicians' responsibility to patients," indicating the widespread acceptance of this burden on care providers.

Confidentiality is crucial to maintaining doctor-patient trust and enabling accurate diagnosis. Patients' decisions around reproductive procedures, life support and organ donation may reflect religious values which they want to keep private. There could be an underlying stigma to seeking out psychiatric help or cosmetic surgeries. Substance abuse treatments and chronic illnesses may create fear of discrimination. There are, plainly, sufficient reasons for a patient to demand confidentiality from a doctor (Mann, Savulescu & Sahakin, 2016). Balancing this need for confidentiality -- when a lack of trust between a patient and doctor could lead the former to withhold medically crucial information -- with legitimate breach of confidentiality imperatives is a complex regulatory issue.

2.3. Risk and Uncertainty Inherent to Medical Care

Medical care is fraught with risk and uncertainty which is reflected in the medical industry's practices, as noted by Arrow (1963). The nature of such care, the behaviour of the physician, uncertainty around outcomes and pricing practices — all point toward the difficulty in allowing markets in healthcare to work efficiently.

Rather than focusing on the rights of patients or the right to science for data that can be used for protecting and promoting public health, economic analysis is more germane to the system that facilitates the interaction of health data.

Information problems are a key impediment to effective decision making in healthcare services.[3] To simplify, consider three actors: patients, care providers, and insurers.

The patients receive care from doctors but insurers bear the cost by determining the likelihood of paying for the patients. The patient has information like pre-existing medical conditions, diet choices, exercise routine, substance use, perceivable symptoms etc. which is crucial for the insurers. The care providers have information on what care options are (important for the patient) and they tend to have decision-making power.

Sharing health data therefore becomes crucial in making markets work efficiently. Doctors can use data on lifestyle choices via wearables and monitors to improve the diagnosis. Such data sharing with insurers could also alleviate some of the adverse selection problem and even encourage risk-based pricing benefiting the consumer. This will make for a more efficient market. It is also possible that insurers use the data to their advantage, reducing consumer surplus. The potential for such discriminatory pricing schemes and the pricing out of some individuals may not be socially desirable. However, this would need to be addressed by a policy intervention.

The above three points - ethical obligations to use health data, the protection of patient confidentiality, and, risk and uncertainty of the healthcare market - highlight the challenges policy regimes face while dealing with health data. We turn to ways in which the Indian governance setup has dealt with these problems in the next section.

3. Legal & Legislative Precedence for Health Data Sharing in India

The regulation of the collection and use of health data in India has developed out of the institutions that conducted public health surveillance during British rule (Mushtaq, 2009). In the first decades of Indian independence, doctors codified this obligation into the Code of Medical Ethics, which asserted the need for doctors to respect the privacy of their patients but also to ensure that diseases were reported to public health authorities. Using contemporary notification systems, public health authorities in India were empowered to ensure that doctors and hospitals in the health system notified them of infectious diseases with the earliest laws and regulations of independent India (Tariq, 2015). It was not until the 1990s - post economic liberalisation and the adoption of information technology - that data collection in the health system began to develop rapidly. There have been several attempts to standardise India's health system and promote sharing data for public health benefit while balancing privacy safeguards. These efforts include technical, legislative and judicial actions, which are covered in turn below.

3.1. Technical Standardisation

Previous attempts to standardise the data use of India's healthcare system in the past two decades have met with varying levels of success. The Health Management Information System

[3] In "Uncertainty and the Welfare Economics of Medical Care," 1963, Kenneth Arrow outlined the information asymmetries that abound in the medical care sector. Particularly, he noted the uncertainty in the incidence of disease and in the efficacy of care, and argued that these drove the economic problems in the sector.

(HMIS) used by government hospitals has evolved from legacy systems implemented since the 1980s, but took off with the National Health Mission in the mid-2000s. Today, while it helps to provide vital statistics on population level issues such as maternal health and infectious disease, the data is broadly diagnostic rather than granular, and suffers from issues of implementation, participation, and comprehensiveness. Several independent evaluations of the system have shown that the capacity of this system is severely lacking in many districts, noting the lack of personnel, fiscal capacity and technological know-how (Saikia, Nandita & Husain, 2018; Krishnan et al., 2010).

Contemporary to the HMIS system, private healthcare providers have implemented their own data systems with varying levels of interoperability, protection of privacy and capacity. Some private healthcare providers have looked to HIPAA standards from the United States to guide the creation of their data systems while protecting patient privacy -- but there is no comprehensive evaluation of these private systems or the level of implementation. The technical aspects of these systems are not entirely ungoverned. The Ministry of Health and Family Welfare introduced the Electronic Health Record (EHR) Standards[4] in 2013 and 2016, which aim to bring about standardisation and interoperability. Again, there is a paucity of evaluation and evidence of low implementation.

These technical standards have moved the needle on promoting the sharing of health data for public benefit and privacy protection, but in a disjointed fashion. HMIS and the digitisation of data flows for evaluating public health are a step in the direction of facilitating sharing -- a crucial obligation echoed in Section 1. On the other hand, EHR standards and the use of HIPAA standards create technical environments that are meant to protect patient privacy -- also an important characteristic of health data talked about earlier. Given these tensions in the current state of this health data system, proposals have been made to create an integrated ecosystem. We will look at the legislative suggestions first.

3.2. Legislative Propositions

The Information Technology Act, 2000 and the Information Technology (reasonable security practices and procedures and sensitive personal data or information) Rules, 2011[5] govern the data protection and sharing rules in India of all data, including health. In 2019, the central government proposed the National Digital Health Blueprint (NDHB)[6] as a means of creating a nationwide system for digital health data. The NDHB draws significantly from the Indian government's previous experiences (like Aadhaar) with digital infrastructure. It proposes a comprehensive digital health ecosystem that is undergirded by a government operated system of exchange, with standards of interoperability and consent being implemented as core to the system. The NDHB system builds largely on the proposals of NITI Aayog's National Health Stack proposals (2018)[7], itself built off goals found in the National Health Policy (2017).[8]

[4] https://www.nhp.gov.in/ehr_standards_mtl_mtl

[5] https://meity.gov.in/sites/upload_files/dit/files/GSR313E_10511(1).pdf

[6] https://www.nhp.gov.in/NHPfiles/National_Digital_Health_Blueprint_Report_comments_invited.pdf

[7] https://niti.gov.in/writereaddata/files/document_publication/NHS-Strategy-and-Approach-Document-for consultation.pdf

[8] The National Health Policy (2017) sets a goal of "the deployment of digital tools for improving the efficiency and outcome of the healthcare system," with an aim of "establishing federated national health information architecture,to roll-out and link systems across public and private health providers at State and national levels." Following this, the National Health Stack (2018) by NITI Aayog proposed a national digital infrastructure for health that included registries, federated personal health records, and a unique health ID, principles that were adopted in the NDHB.

Stemming from the National Health Policy in 2017, the DISHA Act[9] was drafted by the Ministry of Health and Family Welfare in March, 2018. The Act itself aimed to standardise and control the collection, sharing and processing of health data to enable health information exchange between hospitals and clinics. It offered strong protection to individuals and paid attention to specific purposes for processing healthcare data. However, in 2019, the Ministry announced[10] that the DISHA Act will be subsumed under the 'Data Protection Framework on Digital Information Privacy, Security & Confidentiality' Act. This was mostly done to avoid duplication of effort.

The government's recently launched Ayushman Bharat scheme suggests that a digital backbone is critical to governance. It aims to improve healthcare provision by expanding insurance coverage to nearly half the population through Pradhan Mantri Jan Arogya Yojana (PM-JAY)[11]. It also aims to improve primary care availability through increasing the number of Health and Wellness centers. Given that the government will supply such a large proportion of healthcare, there is a potential for a rise in cost. NDHB's system is seen as a key part of implementing Ayushman Bharat and controlling those costs: digital transactions will facilitate the payments and care of patients in the system, help control fraud, root out inappropriate use of resources and improve the productivity of doctors (Niti Aayog, 2019). Information on healthcare interventions and the long-term health of patients can help to ensure that public insurance programs like PM-JAY only pay for effective intervention, preventing long-term cost escalation (Raghupati & Raghupati, 2014).

From the NDHB and its aims, it is clear that the government supports health data sharing and looks to make India a digitally ready country in the coming years. This, of course, has implications for privacy, which we argue (in section 3) are met to some degree by the blueprint but there is more that the government can do (which we talk about in Section 4).

The Personal Data Protection Bill, 2019 (PDP Bill)[12] is important legislation in this regard. Recognising the unique aspects of health data, the bill categorises health data as sensitive data, which necessitates that the data be treated with additional parameters in regards to consent and storage. The bill also recognises the need to use personal health data without consent in the case of emergencies - both for personal treatment and public health emergencies. In the 2018 draft,[13] it included proportionality (discussed in the next section) as a condition in delimiting exemptions for government processing of personal data without consent. However, this was changed effectively in the 2019 version, giving the government complete legitimacy to use data for any such purposes it deems fit. At the time of writing, the PDP Bill did not include proportionality as a condition for exemption; instead, exemption only had to meet the condition of being “in the interest of sovereignty and integrity of India, the security of the State, friendly relations with foreign States, public order. ” There have been other rules like the e-pharmacy rules which the government has introduced to regulate the e-portals that sell medicines. However, these have often been in direct clash with the PDP Bill or even the DISHA Act.

[9] https://www.nhp.gov.in/NHPfiles/R_4179_1521627488625_0.pdf

[10] https://pib.gov.in/Pressreleaseshare.aspx?PRID=1578929

[11] https://www.pmjay.gov.in

[12] http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

[13] https://meity.gov.in/writereaddata/files/Personal_Data_Protection_Bill,2018.pdf

[14] https://www.medianama.com/wp-content/uploads/1890431.pdf

Through these legislations it again becomes clear that the balance between privacy and sharing of data is tilted towards the sharing of data. The government seems to have committed itself to data sharing for public benefit given the broad mandates in legislation and jurisprudence that encourage this.

It must be noted that the NDHB is still a blueprint and does not have legal standing. It is, in fact, the PDP Bill that is likely to govern the use and misuse of health data. However, the NDHB lays the foundation and provides a framework for using data for improving patient care, stimulating medical research and governing big data in healthcare. It is likely that the Data Protection Authority, once formed under the PDP Bill, will formulate the regulations for sharing and processing of sensitive personal data, including health data. In that case, the prototype and framework laid out in the NDHB will be of even more relevance prior to issuing such rules and norms. Thus, analysing which aspects of NDHB allow for privacy protections is important so that those elements can be retained and improved whenever the legislation has passed. The NDHB is quite vast and offers a skeletal framework to introduce digitisation of healthcare in India and its relevance to data sharing, and by implication data privacy, is only going to gain more traction.

3.3. Judicial Proceedings

In India, the harmonisation of the two competing principles of public good and the protection of privacy have been most explicitly laid out through Supreme Court judgements. In Justice K. S. Puttaswamy (Retd.) and Anr. vs Union Of India And Ors. (2017), the Supreme Court affirmed that the right to privacy was a fundamental right. Furthermore, it also recognised the limits to this right and noted that privacy should only be abrogated in the case of pursuing legitimate state goals with legal backing and in proportion to the aim. Four aspects highlighted in Puttaswamy (2017)[15] are that the proportionality should have a legitimate goal, a rational connection, the necessity of the abridgment of privacy, and the balance of the impact.[16] These lay out explicit guidelines for evaluating how privacy should be balanced against “legitimate” goals.

For health, legitimate goals of the state would be the promotion of general welfare and protection of citizens against direct harms. An example of a direct connection is that data on, say, one's travel history would be rationally connected to a new outbreak of flu while information on sexual identity would not be. The necessity of the abridgment of privacy refers to the need for the data instead of some other means of achieving the legitimate goal; a health data example would be the lack of a need for accessing a patient's entire health history in the case of an infectious disease. The balance of impact refers to the impact of the privacy violation on one set of persons against others: for example, accessing data on people in a single area where a disease outbreak took place should not have an undue impact on those people.

[15] https://indiankanoon.org/doc/91938676/

[16] To clarify, there is the proportionality test which includes these four aspects; however, the overall test is threefold, being that the state action has legal backing, is a legitimate goal of the state, and is proportional.

The legitimate goal evaluation is a particularly salient test in the judgement: does the data collection, which poses a risk to privacy rights, have a legitimate goal in terms of the state's aims? As long as the goal falls in this realm, this condition can be satisfied. However, this aim must be articulated: one of the provisions of the Aadhaar Act that the court read down was Section 57, as it allowed government and private entities to use Aadhaar for identification for any purpose, which the court found to be broad and not meeting a legitimate goal (indeed, any goal at all).

Beyond the broad judgements on privacy, the judiciary has also ruled on specific cases about the use of health data. In the case of *Patient X v Hospital Z*,[17] the courts have ruled that the right to privacy is curtailed when private health information is directly relevant to the transmission of disease to their spouse. This provides a clear precedent for care providers to pass along information to those that may be at risk of direct harm. However, this judgement did not resolve several questions. For example, while the care providers did not err in passing along information to those at risk of direct harm, the judgement did not clarify if there was privacy harm in regards to the further disclosure to others in the patient's community. For example, the the Human Immunodeficiency Virus and Acquired Immune Deficiency Syndrome (Prevention And Control) Act, 2017[18] does not clarify if disclosure beyond the interested parties is permitted or constitutes a privacy harm.

Finally, legislation that deals with emerging technologies like DNA sequencing has not aligned with the regulation of health data more generally, as in the case of the draft DNA Technology Regulation Bill[19] up for consideration in the Budget Session and being examined by the Parliamentary Standing Committee at the time of writing. The bill seeks to regulate the storage of genetic information and the creation of national databases for DNA use in criminal matters. However, concerns have been raised as to the relationship of such a database and genetic testing laboratories with health concerns for using genetic information to improve medical outcomes.

For this paper, the judgement in Puttaswamy (2017) will be used as a guide to evaluate the balance of health data use and privacy, but relevant differences will be noted. The first is that for private entities, the PDP Bill 2019 regulates the collection and use of data at the point of collection for personally identifiable data, but does not regulate data that has been anonymised. For the purposes of this paper, the frameworks for the appropriate use of health data are applied to both government collected or controlled data in anonymised and non-anonymised forms, as well as private data in anonymised forms.

It is much harder to make a rational connection between the purpose and the abridgement of privacy in many cases, especially in health care. Here, it is sufficient to outline that wherever the state collects data, that data should be connected to the purpose at hand. To illustrate, since the purpose of Aadhaar was to uniquely identify individuals in order that they may receive their benefits, there was a connection between the need for unique identification and the biometric data collected. A counter-example would be that there is no need for information on, say, a person's voting record, in order to meet the goal of uniquely identifying a person.

[17] https://indiankanoon.org/doc/382721/

[18] https://lms.naco.gov.in/frontend/content/4%20HIV%20AIDS%20Act.pdf

[19] https://www.prsindia.org/billtrack/dna-technology-use-and-application-regulation-bill-2018

Finally, the abridgement of privacy is evaluated with respect to the balance of impact on the rights holder themselves. This means that there should not be a disproportionate impact on the person whose privacy has been affected by the data collection. In the judgement, the court notes that if the abridgement of privacy were to have a disproportionate effect on an individual or group such as the marginalised or poor, then that abridgement would be disallowed. From this last point, an additional subsidiary principle might be added under the public welfare principle for evaluating data usage. Notwithstanding that the only test for checking data use is that fundamental rights are not harmed, data use should also be equitable. In practice, the use of data should be for public welfare and should not be against any group such as to cause them disproportionate harm. The difficulty arises in weighing out benefit against the harms caused, which we turn to in the next section.

4. Finding the Balance between Privacy and Sharing

In his seminal work on privacy, Solove (2002) makes the argument that it is best to understand privacy rights as a “family” of related concepts and outlines six categories that form a family of rights: the right to be left alone, the ability to limit access to the self, the right to secrecy, the right to control personal information, the right to protect one's personality and individuality, and the right to have control over one's intimate aspects of life. This largely maps to what Mittelstadt and Floridi (2016) found in their review of literature concerning ethical concerns in data use in biomedical contexts. They highlight the main concerns of informed consent, privacy (including anonymisation and data protection), ownership, the epistemology of the data, and big data divides between those that can harness data and those who cannot.

While the above literature talks about considerations in general, these are exacerbated with health data - given the obligations to share and protect patients at the same time. Below, we follow Solove with his work on a taxonomy of privacy in order to classify privacy harms, which he categorises under information collection, information processing, information dissemination, and invasion. We look at these four types of harms in healthcare and see how NDHB engages with these in order to prevent them.

4.1. Information Collection Privacy Harms

The most common interaction which results in the collection of health data is between the patient and doctor. Traditionally, this collection of data is understood to take place in a fiduciary context, which has been extended in modern health systems beyond the doctor-patient interaction to the interaction with a care provider in general (Higgins, 1989)[20]. New technologies, such as wearables, can monitor and collect data on a person's health over time, adding new dimensions to health data collection. Users can choose to use wearables, and agree to the terms that the wearable provider has laid out for the product's data collection. Google Flu Trends demonstrates that data can be collected on one's health status through other behaviours beyond those mentioned above (Lazer et al., 2014).

[20] Public health officials, however, mostly depend on the processing of the information collected by the patient interaction with the care provider and the reporting of notified diseases to public health authorities. Thus, public health surveillance is able to collect information via the methods described above of the patient-doctor interaction and care provider-patient interaction.

The NDHB incorporates consent into such collection at several different points to prevent privacy harms. Consent is one of the few mechanisms that help maintain the balance between the obligation to share and protect patient confidentiality — both of which are core to health data. Consent also recognises and establishes ownership at some level without needing to articulate it, giving individuals autonomy and building trust in the system.

In NDHB, the core design principle is "Managing the consents for collection... of personal/ health data, to ensure privacy and confidentiality." In the design of the system, the NDHB notes again that consent has to be given at the point of collection (i.e. the facility level) and proposes that as not all facilities may be able to effectively manage consent, the NDHM should provide "Consent Management-as-a-service" for facilities that may not be able to manage it themselves.[21]

The problem here is truly preserving privacy. What is "meaningful consent" in terms of the collection of health data? The first part of this troubling question is understanding that large segments of the population lack access to quality education and may find it difficult to understand what the implications of having their health data collected are (Kadam, 2017). Even well-informed individuals may be reluctant to question their doctors and care providers on the collection of their health data. In the health system, care providers have expertise which places them in the position of being knowledgeable about what information is needed and should be collected. This power dynamic is at the root of why the doctor-patient relationship is framed as a fiduciary relationship (Scambler, 2008). However, because of this, it brings into question how meaningful the patient's consent is in the first place (Annas, 2017). The PDP Bill does, however, flag that consent needs to be 'meaningful' in a given context for the processing of personal data. It also takes care of some of the above issues with data trust scores that can be assigned to a data fiduciary. Requirements of explicit consent and certain standards for sensitive personal data are in line with the NDHB and its stress on collection of consent.

4.2. Information Processing Privacy Harms

4.2.1. Aggregation

Aggregating information can reveal new facts about a person that they may not wish to be disclosed.[22] The aggregation of information in regards to a person is not a harm in and of itself — a care provider can only make sense of a chronic disease, for example, by aggregating data from different discrete points in time. Under the NDHB, the aggregation of an individual's health data is managed through the mechanism of a Personal Health Record (PHR), which is actually a series of record pointers to the individually recorded medical events for a given patient.[23] In this system, a single event such as a doctor's visit is only recorded in the PHR as a record pointer, while the data about that event is held at the doctor's facility. Aggregation is then controlled by the "consent manager" through which the patient can control information sharing and access. The PDP bill also discusses the importance of consent managers and views them as a subset of data fiduciaries. They are an "accessible, transparent and interoperable platform" which enable the data principal to "gain, withdraw, review and manage" their consent.

[21] It also suggests that National Digital Health Mission (NDHM) standards for collection should follow ISO Health Informatics standards for the management of consent and the Electronic Consent Framework laid out by MeitY.

[22] Solove, 2005

[23] Note that in other literature on digital health, longitudinal records may be referred to as Electronic Health Records (EHR)

However, the plans do not detail what consent management will look like for patients. Instead, the document refers to the Consent Framework found in MeitY's Electronic Consent Framework (Technology Specifications vl.1).[24] These standards imply that data use may be determined at the time of collection and consent can be obtained at that time, as mentioned previously in the privacy risks of information collection. Given this design, it is not clear if patients will have to consent for future uses of their data at the time of collection or if they will have to do so every time their data is used for new purposes.

4.2.2. Identification

"Identification" is the action by which information about an individual is connected with the person. As health data is pertinent to a specific physical body, that body needs to be identified as the one to which the particular information relates.

The NDHB proposes that a unique identifier be created for the health data ecosystem. This identifier is referred to as the Unique Health Identifier (UHID) and is meant to ensure that medical records are accurate and that consent can be appropriately obtained.[25] Two principles are laid out in the UHID's design:

(a) No denial of health service to anyone in any scenario
(b) No scope for medical errors arising out of wrong identification of the patient

Identification could potentially harm individuals by connecting a person with data about them against their will. The NDHB's system prevents this harm through UHID in that the individual's consent is needed to connect any data. The risk, then, comes in abuse of the system beyond the parameters of its design; that is, if it is used to identify and link people to data without their consent.

4.2.3. Insecurity

"Insecurity" as a privacy harm comes from negligence in protecting private information from disclosure to unauthorised entities or persons.[26] Insecurity and subsequent unauthorised disclosure can lead to direct harms, such as social stigmatisation. For example, the disclosure resulting from insecurity of information about an individual's HIV/AIDS status can directly lead to social exclusion and other harms (Mahajan et al., 2008). Beyond direct disclosure harms, there is also the possibility of further use of data that leads to potential harm under other categories, such as aggregation. Protecting patient confidentiality is one of the key obligations of the medical fraternity.

[24] Along with its subsequent revision(s)" and to the ISO standards "ISO/TS $^{17975:2015}$ Health Informatics - Principles and data requirements for consent in the collection, Use or Disclosure of personal health information."

[25] The UHID is detailed as needing demographic information for the identification of the individual in order to create the record. The NDHB suggests that multiple ID forms will be leveraged in designing and implementing the UHID, giving examples such as existing identifiers like Aadhaar and PAN card.

[26] Solove, ^{2005}

lead to social exclusion and other harms (Mahajan et al., 2008). Beyond direct disclosure harms, there is also the possibility of further use of data that leads to potential harm under other categories, such as aggregation. Protecting patient confidentiality is one of the key obligations of the medical fraternity.

Several aspects of the NDHB's design follow principles for increasing the security of the system such as a federated architecture, anonymisation near the source, the auditing of security measures, and the creation of a Security Operations Center[27] (Balsari et al., 2018). The federated architecture in particular is designed to minimise insecurity. It also brings with it the possibility of community ownership. The NDHB plans for health data to be stored by the care provider of the facility at which the data is collected. This minimises the need for central repositories of data that may be breached or the leakage of data while in transit from facilities to central repositories. Thus, most health data stored at state and national levels will simply be links pointing to records that are held at facility levels.

While anonymisation is a useful tool for ensuring that data is not connected to individuals, it is not a foolproof mechanism (Culnane, Rubinstein & Teague, 2017). Recognising this, the NDHB calls for the anonymisation of data at the facility level, decreasing the risk of re-identification from data in centralised systems.[28] Finally, the NDHB plans include a security operations center and security auditing. Security auditing is a step in the right direction for providing independent oversight of the system.

4.2.4. Secondary Uses

The patient relationship with the care provider proceeds under the assumption that the information collected will be used for diagnosis and care planning. Using the data beyond the defined purposes would constitute secondary uses, which is a privacy harm as it threatens the understanding of purpose use at the time of collection.

How does the NDHB account for this potential privacy harm? Since the harm derives from fear and uncertainty about how one's data will be used in the future, the consent management system allows patients to feel secure in their knowledge of how their data will be used for secondary purposes, as they will have to grant consent for such uses. Similarly, the NDHB notes that anonymised data can be used for research "if it duly follows the principles so defined" and leaves the definition of such principles under a list of "recommended work on further standards." Under these rules, then, it seems like anonymised data can be used for secondary purposes as long as it fits under the principles of research. In the draft Personal Data Protection Bill, however, section 91 allows the government to get access to non-personal data (aggregated or anonymised) for secondary uses without any limitations.

As with data collection, the meaningful consent of patients to their data being used for secondary purposes is questionable. Additionally, while patients have control over secondary uses if the data identifies them, they do not have full control if the data has been anonymised. This does not allow them

[27] Given that the Security Operations Centre is largely conceived as a non statutory body, more work will be needed to think on the structures and protocols to serve its purpose.

[28] To imagine the operation of this principle, say that an actor with ill-intent gains access to a state-level database on a particular type of health data. Often, in order to re-identify individuals, another dataset would be needed in order to connect that health data with the individuals to which it applied. The actor would have to gain access to many different datasets at different facilities in order to re-identify people in the data that they obtained.

to limit the uses of such data. For example, a person who is a member of a minority (e.g. stigmatised groups like sexual minorities at high-risk of contracting HIV/AIDS) may not want even anonymised data on their health status used by researchers if they feel that this can result in conclusions and information being gathered on that minority. Similarly, the public may be concerned about the use of their data by private companies for research and subsequent creation of products or services for private profit. For what its worth, this is not a problem to NDHB itself - no legal regime around the world allows for individual consent to one's own data. That said, we think this should change.

4.2.5. Exclusion

Exclusion is "[...] a harm created by being shut out from participating in the use of one's personal data, by not being informed about how that data is used, and by not being able to do anything to affect how it is used." Thus, the exclusion harm can be said to be a primary one of agency - it occurs when patients do not have any say in how their data is used.

The NDHB design and the PDP, 2019, should provide the patient with sufficient means to participate in decisions on the use of their data. The consent manager system, in particular, is meant to ensure that patients participate in the decisions about the use of their data. Similar to the previously flagged issues with consent, the ability of patients to understand how their data is used is highly questionable. As might be expected, patients generally do not have an understanding of health as a discipline, so if they do not have an understanding of how their data is meaningful when it is being used, they are not participating meaningfully when they assent to its use.

4.3 Information Dissemination Privacy Harms

4.3.1. Breach of Confidentiality

In the case of breach of confidentiality and health data, this clearly occurs when care providers provide information on a patient's health to parties that were not part of the confidential agreement that is part of the fiduciary trust established between them. As the NDHB protects patients' rights to their data and requires their consent, the risk of breach of confidentiality comes in the misuse of datasets in the system. The federated design helps to ensure that malicious actors who may gain ingress will be stopped at different levels of the system, containing the breach. However, the NDHB can do little to address the possibility that doctors and care providers may breach confidentiality at the facility level, outside the NDHB ecosystem. Instead, it is current regulations and law that would punish care providers who unduly disclosed health data to unauthorised persons.[29]

[29] The relevant protections include the Code of Medical Ethics, Supreme Court jurisprudence on privacy, and the upcoming PDP 2019 draft.

4.3.2. Disclosure

In differentiating disclosure from breach of confidentiality, Solove sees disclosure as revealing information that will do damage to someone's reputation or others' understanding of them.[30] One factor to consider is the information that search engines and social media companies may have on an individual that can imply information about the individual's health status. These companies operate under a consent paradigm, and most consumers consent to their data being used and shared with third parties when they use such services. This exposes consumers to potential disclosure harms. There is some debate as to the extent of disclosure — if information was already available to some of the public, does disclosing it to more of the public constitute a privacy harm? In health, for example, a person might suffer from a medical incident in public. It is unclear if the further disclosure of information that incident would be a privacy harm.

The PDP 2019 draft also introduces a fiduciary relationship between a consumer and companies that have personal data on them. Under this proposed legislation, companies that have personal data will be obliged to treat that data under the fiduciary standard; the disclosure of this information, then, would constitute privacy harms under breach of confidentiality. Essentially, this ensures that all personal data will fall under this paradigm.

Certain cases of disclosure may constitute privacy harms but the line drawn between personal data and anonymised personal data is permeable. Individuals may not want data disclosed that connects or correlates certain attributes and outcomes to one another; this can represent an unwanted disclosure. For example, theoretically, health data could reveal the prevalence of sexually transmitted diseases among a sexual minority; the disclosure of this information even if it is anonymised may still cause a privacy harm in that it can affect the perception of that sexual minority.

It remains to be determined in the courts as to how far the protection of disclosure goes in non-fiduciary relationships. For example, in the case of Mr. 'X' vs Hospital 'Z' while the court did determine that the hospital did not err in revealing information on the patient's HIV/AIDS status to their fiance, it did not clarify if a privacy harm had occurred with regards to the further disclosure of that status to others in the community. Similarly, The Human Immunodeficiency Virus and Acquired Immune Deficiency Syndrome (Prevention And Control) Act, 2017 does not clarify if disclosure beyond the interested parties is permitted or constitutes a privacy harm. In order for clarification on the extent of privacy rights in regards to patients' health data and sharing, legislation and regulation should address this aspect of disclosure.

4.3.3. Exposure

Solove recognizes that a specific type of privacy harm comes about from the revelation of intimate activities. Exposure harm is "[...] the exposing to others of certain physical and emotional attributes

[30] Breach of confidentiality focuses on the relationship of trust that is broken while disclosure focuses on the information itself that is made available, affecting others' perceptions. This can take place regardless of the existence of a trust-based relationship between the party to whom the information pertains and the party disclosing the information. Most health data, however, is gathered under relationships that are operated under the fiduciary paradigm; there are few examples of data that might be disclosed that don't operate under the standards of harm that come under breach of confidentiality.

about a person. Grief, suffering, trauma, injury, nudity, sex, urination, and defecation all involve primal aspects of our lives—ones that are physical, instinctual, and necessary. We have been socialized into concealing these activities."[31]

In this sense, health data systems may contain information on expected activities that individuals may still wish to keep concealed and avoid exposure. Currently, the only paradigm for protecting patients against exposure exists as a derivative of the consent framework and the aforementioned protections against unwanted disclosure. The limitation here is that the system does not address the collection of data itself as a potential exposure privacy harm. Given the distribution of knowledge and power in the health system, patients may feel pressured to share and allow the collection of data on intimate matters that give a sense of exposure. As an example of socio-cultural senses of exposure, ASHA health workers are women as gender norms permitted these female workers to interact with female community members to provide care, whereas interaction with male workers was seen as violating a sense of privacy (Ved et al., 2019). Protocols around data collection should take into account such socio-cultural factors in order to facilitate data use without creating privacy harms.

4.3.4. Increased Accessibility

The recognition of increased accessibility as constituting a potential harm to privacy is not well-established.[32] Technically, the implementation of the NDHB system does not increase the accessibility of data to any actors in the system, besides the patient themself. This is because of the consent paradigm that constrains the increased accessibility created by the system's technological design. However, even though consent will empower patients, they may not be aware of the increased privacy harms to which they are being made vulnerable. The PDP Bill addresses some of this (as discussed in the earlier sections) by enforcing clarity in notices and allowing meaningful and informed consent to be collected from the principals.

There is a related issue that connects the harm of increased accessibility with security. The system outlined in the NDHB, as mentioned above, contains many measures to ensure its security. Nonetheless, the system still increases the accessibility to information for malicious actors. Therefore, the implementation of information systems must weigh the consequences of making information more available and the true utility such information has for health benefits.

4.3.5. Distortion

The dissemination of false information about a person is categorised as "distortion" under Solove's classification system. With health data, this, of course, can be seen with the distribution of false information on an individual' s health status or information about one's health. However, given the nature of health data systems, distortion harms have the potential to extend well beyond the primary

[31] In Solove's explanation, this differs from disclosure in that these are expected activities of any person and revealing them is not about the data so much as the exposure in of itself.

[32] Solove quotes the judgement in United States Department of Justice v. Reporters Committee for Freedom of the Press in order to show how increased accessibility threatens privacy even if such information was already made available : "[...]there is a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearinghouse of information."

victims. Instead, there is a great public risk that is derivative of distortion, which comes from the use of distorted information in data systems that are analysed to create medical knowledge used for care decisions. The reliability of that knowledge and thus the reliability of care decisions made from that knowledge are distorted.

The Code of Medical Ethics covers the rights of patients when it comes to distortion of their health data. The Code explicitly sets out the punishment of being deleted from the practitioners register if a practitioner misrepresents data on documents provided to the state or regulated by law. While these cases of distortion are explicitly disallowed, the fiduciary relationship implies that the practitioner cannot misrepresent data to other actors in the system generally.

4.4 Invasion Privacy Harms

Such a category of privacy harms deals with decisional intrusion — unwanted incursion by an authority such as the state into a person's personal life decisions." The invasion of privacy by decisional intrusion in healthcare consists of decision-making that takes place on behalf of patients without their permission. Defining the boundary of "unwanted" is the first important step for delimiting this privacy harm. Many health decisions made by public health authorities on behalf of protecting public welfare constitute an incursion of individual decisional autonomy. Technically, this means that there is an invasion of privacy if such public health motivated decisions interfere with individual autonomy — but as discussed above, such intervention has been justified in both normative terms as well as the specific parameters of Indian law and jurisprudence.

Nonetheless, while the normative principles allow for public health authorities to take decisions on behalf of individuals, further detail is needed to delineate between public welfare needs and the need to protect individual decisional autonomy. One parameter for determining that difference is the impact on others; for example, quarantines are direct invasions of decisional autonomy but are justified on the grounds of preventing direct harms to others by the spread of infectious diseases (World Health Organization, 2016a).

In the case of quarantine, the state is empowered by the Epidemic Diseases Act, 1897 (Ahmad, 2015). The act empowers public authorities to take such steps to prevent the spread of infectious disease. This, of course, constitutes a privacy harm, albeit a justifiable one. The disclosure harm was clearly felt during COVID-19 when lists of quarantined people were widely circulated either by the state or private actors like Resident Welfare Associations. This harm was also felt when quarantine stickers were pasted on doors, listing names of people and other private details.

However, one issue here is the lack of clarity in current legislation as to what authorities may do when outbreaks do not qualify as full-fledged "disasters." (World Health Organization, 2015). If there is a reason for taking public health actions that are not as extensive as a quarantine, how should that balance against the limitation of decisional privacy? There is a need for the Indian health system to come up with frameworks for such limitations.

[33] To the state such as courts or public health authorities and regulated by law such as mental diseases, notified infectious diseases, births, etc.

Indian jurisprudence allows for decisional intrusion into a patient's choice to disclose their status of having a sexually transmitted disease in order to inform those that may be directly affected. The Human Immunodeficiency Virus and Acquired Immune Deficiency Syndrome (Prevention And Control) Act, 2017, attempts to limit the decisional intrusion in this case by requiring the doctor to first consult the patient and advise them to inform their sexual partners; the decision is taken out of the patient's hands only if they do not inform those who face a risk of harm. Still, they do not have a choice of withholding the information, only how it will be delivered. Indian law and jurisprudence should go further in clarifying the situations beyond sexually transmitted diseases that such decisional interference takes place.

Beyond these situations, health data usage, in and of itself, will affect how decisions are made, and thus will affect the meaning of decisional autonomy for patients in a deeper manner. In the NDHB's envisioned system, the collected data will be used to create insights and help in care decisions.The paradigm of consent means that as long as patients consent to the use of their data in these systems, there is no privacy violation. However, this brings up again the question of meaningful consent — patients may have no understanding of how these decisions are being made and depend on the fiduciary relationship with the care provider.

In particular, new data-dependent technologies are capable of systematically posing a threat to the decisional autonomy of an individual's experience of healthcare service. This is due to the nature of algorithm-based systems. Essentially, new machine learning algorithms are only as good as the data on which they are trained, and if the real world biases that influence the collection of that training data are not addressed, the resulting algorithms can replicate those biases. Recently, researchers in the US found hospital decision-making systems were systematically biased against black Americans. The findings concluded that the algorithms used were less likely to recommend equally sick black patients for improved care programs as white patients (Obermeyer et al., 2019). This means that when a person's data is used to make a decision, it may be placed into a system that makes use of biased data in order to make that decision. Of course, this sort of decisional interference is not limited to algorithmic decisions — care providers often already make decisions from biased viewpoints — but the scale of data use in decision systems brings the problem to a new level. It is also easier for those biases to be exposed by careful analysis of the data and decision-making algorithms, creating the opportunity for more accountability.

5. Conclusion: What next for NDHB?

The NDHB, as it stands today, is still a blueprint and will have to be aligned with the PDP bill as and when that is passed. This paper has charted out a brief history of how legal and legislative moves seem to promote sharing of health data, while embedded in the architecture of the NDHB (and other medical laws) and some salient ways by which harms due to privacy can be curtailed. Apart from the effort made by NDHB in section 3, we think there are additional efforts that the government can take in order to garner support and build trust with people on the protection of their rights and liberties while allowing innovations in healthcare.

5.1 Willingness to Delete Data

As Solove's taxonomy affords a granular view of what specific harms may look like, it is important to take a step back and state in simple terms what patients' concern is: they worry that they have a lack of agency in preventing the misuse of their data. This concern is both an individual concern, in that patients' privacy is harmed if their individual data may be misused, and a societal one, in that data on the health of citizens as a whole can be misused.

The current NDHB plan provides for many useful mechanisms for individuals to assert their rights on the use of their data. The consent manager, "right to be forgotten" and other design principles discussed before address these concerns at the individual level. This aligns the Indian system with regulation in other countries to a right to data deletion.[34]

However, while these measures will help to ensure that individuals can assert their rights over their health data, it may not instill trust in the system's overall use of citizens' data. To instill trust in the system, the NDHB should explicitly commit to deleting centrally held data or record locators in the case of their misuse (keeping in line with the relevant clauses of the PDP Bill). Consider the backtracking of Google in a data analysis project with the US National Institutes of Health. When it seemed that the data use would expose personal data, Google cancelled the project and deleted its copies of the data. Similarly, a data processing partner of the NHS in the UK, Health IQ, was forced to delete health data it held outside the UK. This suggests that one of the keys to gaining legitimacy for the use of health data is satisfying the condition in the "social license" for using a public good, which is the reciprocity of public authorities being able to disempower the data processor by forcing them to delete data (Carter, Woods, & Dixon-Woods, 2015; Powles & Hodson, 2017).

5.2 Mandating Audit Trails and Digital Footprints

The NDHB provides that the use of a citizen's data should be recorded in such a manner that it is auditable. As the NDHB moves forward in implementation, the manner by which citizens can access this audit trail should be made explicit ab initio so that citizens understand how their data is being used. This will build trust in the system and provide citizens with an understanding of how their data is used. For example, the system would quickly show that information on notifiable diseases is shared with appropriate government authorities, building confidence in the public health system.

The government should consider the technical aspects of this audit trail. The design of the system should be made available for public scrutiny so that citizens can be sure that the system accurately records any access to their data.[35] Additionally, in implementing the NDHB system, the NDHM should look to institute mechanisms and protocols for checking the accuracy of data to prevent and take action against the distortion of data.[36]

[34] Such as the PDP Bill and the GDPR regulations in the European Union.

[35] There is a possibility here of utilizing new technologies such as distributed ledger systems like blockchain to ensure that any data access is recorded.

[36] Additionally, distortion may not take place only with the misrepresentation of personally identifiable data - there is the possibility of the distortion of data that is aggregated and anonymised, and which may cause privacy harms in that the members of the misrepresented group will have their information represented wrongfully. For example, a malicious actor could distort the aggregated health data of a population affected by say, a mining operation nearby. This, of course, qualifies as a distortion and violates the privacy of the individuals as a group whose data has been affected.

The NDHB should also consider how citizens might address the misuse of their data. If, for example, a citizen audits their data and feels it has been misused, they would have to figure out if they wished to address their grievance to their care provider, the National Digital Health Mission's privacy operation center, the proposed system of redressal in the draft Personal Data Protection Bill, 2019, or to the judicial system. Citizens should be provided with a clear path for addressing the misuse of their health data. The government might even consider empowering a public authority with conducting audits of citizens data on their behalf, in a regulatory and proactive manner.

5.3 Collecting Meaningful Consent

There are two main avenues for addressing the meaning of consent beyond the framing of the fiduciary relationship. The NDHB in its implementation should provide guidelines for ensuring that at the stage of giving consent, easily understandable language is used, following global practices and suggestions (Falagas et al., 2009; Grady, 2015). It should also consider that in order to build public trust in the use of their data, public awareness campaigns and tools should be used to ensure citizens know how to access information on the use of their data and what they are consenting to.

Further, in order to balance between the public benefit of health data use and privacy, the NDHB should consider bundling consent at the stage of data collection. Under the current formulation, it appears that each time a citizen's data is used for non-emergencies, they will have to consent to its use, unless it is anonymised. The bundling could be done in a tiered manner based on the potential uses of that data for public benefit (outlined in the framework below). This will allow for smooth use of the data that has substantial public benefit without burdening the research system or the citizen unduly. On the other hand, the system should consider implementing consent mechanisms even for the use of anonymised data for public purposes. This might still be one of the tiered buckets, but it should not be excluded entirely from the citizen's agency.

5.4 Clarity on Process of Access to Health Data

Currently, authorities are able to access health data in the case of emergencies that pose a threat to public health. Each authority has its own regulations for determining public health threats and for accessing the data that they need to make decisions in these circumstances. The PDP Bill has sections 35 and 91 which form the basis for data sharing among public health authorities. However, it is not clear from the NDHB what the process would be for public authorities to access such records in this system; will public health authorities access data on, say, an emergent disease when care providers and other indicate there may be an issue, or will they passively collect data on any serious infectious diseases? In order for the public to trust the system, public health authorities should delineate how they will collect information, and where the lines are for this data use. More clarity is needed from a health data perspective and one that aligns with the clauses of the PDP Bill.

The NDHB should also provide a general framework for its future regulation of data that will be used for medical research. There will have to be an explicit balance struck with each type of data used between the potential for privacy harms or misuse and the public benefit. Such a balance should constitute a "privacy test" as Dyke, Dove & Knoppers (2016) outline: data which meet certain standards of "sensitivity" should be accorded extra protection and strict guidelines for their use. Even in the cases of anonymisation, there is a need for this type of framework to be applied given the risk of de-anonymisation. By having guidelines for use, research can be carried out with an appropriate regard for that risk.

Such a framework for the guidelines of health data use for research should also involve a public conversation on who should benefit from the data use. While public research institutions may use the data to innovate medical advances, it is also possible that private organisations will seek to conduct research with such data to create products or services for profit. The NDHB could leave it up to citizens to negotiate their consent to data use in exchange for profit on their own, but the experience of Internet companies in the last few decades suggest the citizens are not prepared to demand a share of the monetary benefits of their data. The NDHB should recognize that this profit motive exists, and create a clear mechanism for navigating it.

Below, we give a brief sketch of a framework such evaluation could utilise to balance the harms and benefits of a given type of data. This is by no means exhaustive but rather indicative of the work that is needed. On the “X” axis, we categorise the data type by its likelihood of causing certain harms, and on the “Y” axis, we categorise the potential public benefit of using a data type. A governance regime would need to use such categorical evaluation of each type of data that might be shared in order to determine who can access it. Additional axes might be added, such as an evaluation of profitability or ownership, as appropriate. Such a framework will be helpful for authorities in power in making judgements around sharing what kind of data with which person/organisation. It may also lend itself helpful in underscoring the kinds of safeguards required for different sets/categories of data. A harm-benefit characterisation will be helpful in navigating the trade-offs while acknowledging the essential characteristics of health data as explained in Section 1.

Example of a Framework:

Category of Potential Benefit	**Category of Risk of Harm**			
	Largest Likely Harms: Personally Identifiable Health Data on its win has potential for harm	Personally Identifiable Health Data in combination with other personal data has potential for harm	Health Data from multiple people in combination has potential for harm	Least Likely to be Harmful: Data that is publicly available
	I	**II**	**III**	**IV**
Most Beneficial: Data can actively be used to prevent, mitigate, or alleviate significant harms				
Data can be used in research to create solutions for preventing, mitigating, or alleviating harms				
Data has potential for benefit but no active use cases				
Least Beneficial: Data provides little to no health benefits for individuals or populations				

What is positive about the current legislative and legal regime in India is the recognition that privacy will have to be protected while allowing for data sharing to enable digital innovations. While there are steps being taken to operationalise this, the government will have to be careful and cognizant about the next steps inorder to prevent backlash and ensure citizen-state trust remains.

References

Ahmad, Tariq. (2015) India: Legal Responses to Health Emergencies. United States of America Library of Congress, Law Library of Congress. https://www.loc.gov/law/help/health-emergencies/india.php

Annas, G. J. (2017). Informed consent: charade or choice?. The Journal of Law, Medicine & Ethics, 45(1), 10-11.

Arrow, K. J. (1978). Uncertainty and the welfare economics of medical care. In Uncertainty in economics (pp. 345-375). Academic Press.

Balsari, S., Fortenko, A., Blaya, J. A., Gropper, A., Jayaram, M., Matthan, R., ... & Mandl, K. D. (2018). Reimagining Health Data Exchange: An application programming interface–enabled roadmap for India. Journal of medical Internet research, 20(7), e10725.

Bresnick, J. (2017, June 5). Understanding the Many V's of Healthcare Big Data Analytics. Retrieved from https://healthitanalytics.com/news/understanding-the-many-vs-of-healthcare-big-data-analytics

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care. data ran into trouble. Journal of medical ethics, 41(5), 404-409.

Chakravarthi, I. (2018). Regulation of Private Health Care Providers in India: Current Status, Future Directions. Indian Journal of Public Administration, 64(4), 587-598.

Culnane, C., Rubinstein, B. I., & Teague, V. (2017). Health data in an open world. arXiv preprint arXiv:1712.05627.

Desai, V. (2019, August 26). Wearable Technology: Redefining Real-time Healthcare Monitoring. Retrieved from https://www.healthcareexecutive.in/blog/wearable-technology

Donaldson, M. S., Lohr, K. N., & Bulger, R. J. (1994). Health Data in the Information Age: Use, Disclosure, and Privacy—Part II. JAMA, 271(18), 1392-1392.

Douglas MacMillan, G. B. (2019, November 15). Google almost made 100,000 chest X-rays public - until it realized personal data could be exposed. Retrieved from https://www.washingtonpost.com/technology/2019/11/15/google-almost-made-chest-x-rays-public-until-it-realized-personal-data-could-be-exposed/

Dyke, S. O., Dove, E. S., & Knoppers, B. M. (2016). Sharing health-related data: a privacy test?. NPJ genomic medicine, 1(1), 1-6.

Falagas, M. E., Korbila, I. P., Giannopoulou, K. P., Kondilis, B. K., & Peppas, G. (2009). Informed consent: how much and what do patients understand?. The American Journal of Surgery, 198(3), 420-435.

Ferguson, A. H. (2015). The Role of History in Debates Regarding the Boundaries of Medical Confidentiality and Privacy. Journal of medical law and ethics, 3(1-2), 65.

Fry, E. (2019, April 3). Who Should Own Your Health Data? Retrieved from
https://fortune.com/2019/04/03/personal-health-data-ownership/

Furukawa, M. F., Raghu, T. S., & Shao, B. B. (2010). Electronic medical records and cost efficiency in hospital medical-surgical units. INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 47(2), 110-123.

Grady, C. (2015). Enduring and emerging challenges of informed consent. New England Journal of Medicine, 372(9), 855-862.

Gupta, A. (2019, May 2). Max Bupa's big question: If staying fit gets you 10% off on health insurance, will you buy? Retrieved from https://prime.economictimes.indiatimes.com/news/69137477/pharma-and-healthcare/max-bupas-big-question-if-staying-fit-gets-you-10-off-on-health-insurance-will-you-buy

Hall, K. (2016, August 26). NHS slaps private firm Health IQ for moving Brits' data offshore. Retrieved from https://www.theregister.co.uk/2016/08/26/health_iq_rapped_by_nhs_digital_data_security/

Hickok, E. (2018, August 5). The DNA Bill: Another Invasive, Imperfect Database! Retrieved from https://www.bloombergquint.com/opinion/the-dna-bill-another-invasive-imperfect-database

Higgins, G. L. (1989). The history of confidentiality in medicine. Canadian Family Physician, 35, 921.

Hoffman, S., & Podgurski, A. (2009). E-Health hazards: provider liability and electronic health record systems. Berkeley Tech. LJ, 24, 1523.

Is Data Privacy Real? Don't Bet on It. (2019, August 23). Knowledge@Wharton, Wharton School of Business, University of Pennsylvania. Retrieved from
https://knowledge.wharton.upenn.edu/article/data-privacy-real-dont-bet/

Jain, A. (2016, September 17). The 5 V's of big data. Retrieved from https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/

Kadam, R. A. (2017). Informed consent process: A step further towards making it meaningful!. Perspectives in clinical research, 8(3), 107.

Kaplan, B. (2016). How Should Health Data Be Used?: Privacy, Secondary Use, and Big Data Sales. Cambridge Quarterly of Healthcare Ethics, 25(2), 312-329.

Knoppers, B. M., & Thorogood, A. M. (2017). Ethics and big data in health. Current Opinion in Systems Biology, 4, 53-57.

Krishnan, A., Nongkynrih, B., Yadav, K., Singh, S., & Gupta, V. (2010). Evaluation of computerized health management information system for primary health care in rural India. BMC health services research, 10(1), 310.

Langat, P., Pisartchik, D., Silva, D., Bernard, C., Olsen, K., Smith, M., ... & Upshur, R. (2011). Is there a duty to share? Ethics of sharing research data in the context of public health emergencies. Public Health Ethics, 4(1), 4-11.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. Science, 343(6176), 1203-1205.

Mabiyan, R. (2018, February 22). Hope to see Indian healthcare sector becoming HIPAA compliant soon: Winmagic's COO - The Economic Times, HealthWorld. Retrieved from https://health.economictimes.indiatimes.com/news/health-it/hope-to-see-indian-healthcare-sector-becoming-hipaa-complaint-soon-winmagics-coo/63026276

Mahajan, A. P., Sayles, J. N., Patel, V. A., Remien, R. H., Ortiz, D., Szekeres, G., & Coates, T. J. (2008). Stigma in the HIV/AIDS epidemic: a review of the literature and recommendations for the way forward. AIDS (London, England), 22(Suppl 2), S67.

Malin, B. A., Emam, K. E., & O'Keefe, C. M. (2013). Biomedical data privacy: problems, perspectives, and recent advances.

Marjanovic, S., Ghiga, I., Yang, M., & Knack, A. (2017). Understanding value in health data ecosystems.

Ministry of Health and Family Welfare, Data Ownership of EHR. (2015, June 3). Retrieved from https://www.nhp.gov.in/data-ownership-of-ehr_mtl

Minor, L. B. (2017). Harnessing the power of data in health. Stanford Med. Heal. Trends Rep.

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. Science and engineering ethics, 22(2), 303-341.

Mossialos, E., Wenzl, M., Osborn, R., & Sarnak, D. (2016). 2015 international profiles of health care systems. Canadian Agency for Drugs and Technologies in Health.

Mukul. (2018, December 7). EHRs in India: Challenges and Opportunities vis-a'-vis' Ayushman Bharat. Retrieved from: https://ehealth.eletsonline.com/2018/12/ehrs-in-india-challenges-and-opportunities-vis-a-vis-ayushman-bharat/

Mushtaq M. U. (2009). Public health in british India: a brief account of the history of medical services and disease prevention in colonial India. Indian journal of community medicine : official publication of Indian Association of Preventive & Social Medicine, 34(1), 6–14. https://doi.org/10.4103/0970-0218.45369

Niti Aayog (2019). Health System for a New India: Building Blocks. Chapter 5: Reimagining India's Digital Health Landscape: “Wiring” the Indian Health Sector, Streveler, D, & Gupta, P.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. Health and technology, 7(4), 351-367.

Porsdam Mann, S., Savulescu, J., & Sahakian, B. J. (2016). Facilitating the ethical use of health data for the benefit of society: Electronic health records, consent and the duty of easy rescue. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160130.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health information science and systems, 2(1), 3.

Saikia, Nandita & Husain, Zakir & Bora, Rimon. (2018). CHALLENGES OF THE HMIS IN INDIA A case study of Udham Singh Nagar, Uttarakhand. 10.13140/RG.2.2.28064.69128.

Scambler, G. (2008). Sociology as Applied to Medicine, Chapter 4: The Doctor-Patient Relationship E-Book. Elsevier Health Sciences.

Sharma, N. C. (2018, July 16). Adoption of e-medical records facing infra hurdles: Report. Retrieved from https://www.livemint.com/Politics/CucBmKaoWLZuSf1Y9VaafM/Adoption-of-emedical-records-facing-infra-hurdles-Report.html

Sharma, S. (2016). Problems of the Health Management Information System (HMIS): the experience of Haryana. Ajay Shah's blog. https://blog.theleapjournal.org/2016/06/problems-of-health-management.html

Singh,A.(2018,December,12). Mr X vs Hospital Z disclosure of dreadful diseases. Retrieved from https://blog.ipleaders.in/mr-x-vs-hospital-z-disclosure-dreadful-diseases/

Solove, D. J. (2002). Conceptualizing privacy. Calif. L. Rev., 90, 1087.

Solove, D. J. (2005). A taxonomy of privacy. U. Pa. L. Rev., 154, 477.

Supreme Court of India, (1998, September 21).“Mr. 'X' vs Hospital 'Z'”. Retrieved from https://indiankanoon.org/doc/382721/

Supreme Court of India, (1998, September 21).“Mr. 'X' vs Hospital 'Z'”. Retrieved from https://indiankanoon.org/doc/382721/

Topol, E. (2012). The creative destruction of medicine: How the digital revolution will create better health care. Basic Books.

Topol, E. (2019). Deep medicine: how artificial intelligence can make healthcare human again. Hachette UK.

Upshur, R. (2003). The ethics of quarantine. AMA Journal of Ethics, 5(11), 393-395.

Vayena, E., & Tasioulas, J. (2016). The dynamics of big data and human rights: The case of scientific research. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160129.

Ved, R., Scott, K., Gupta, G., Ummer, O., Singh, S., Srivastava, A., & George, A. S. (2019). How are gender inequalities facing India's one million ASHAs being addressed? Policy origins and adaptations for the world's largest all-female community health worker programme. Human resources for health, 17(1), 3.

Winter, J. S., & Davidson, E. (2017). Investigating values in personal health data governance models. In Americas Conference on Information Systems (Vol. 2017).

Wolford, Ben. (2019, February 13). Everything you need to know about the "Right to be forgotten". . Retrieved from https://gdpr.eu/right-to-be-forgotten/

World Health Organization. (2015). International public health hazards: Indian legislative provisions. World Health Organization.

World Health Organization. (2016a). Guidance for managing ethical issues in infectious disease outbreaks.

World Health Organization. (2016b). Policy Statement on Data Sharing by the World Health Organization in the Context of Public Health Emergencies

Acknowledgements

The authors would like to thank an anonymous reviewer for their comments. We'd also like to thank Anirudh Burman, Dhruv Joshi, Anja Kovacs, Rahul Matthan, Sarayu Natarajan and Niranjan Rajadhyaksha for providing useful comments. We'd also like to thank Vikram Sinha for editorial comments. In addition, we'd like to acknowledge the feedback received from the participants at the Third Quarterly Roundtable of the Data Governance Network held on 13 December, 2019 where we presented a proposal for this paper and at the Fourth Quarterly Roundtable of the Data Governance Network held on 24 March, 2020 where a draft of this paper for presented. All errors are the authors' own.

About the Authors

Alexander Fager is an Analyst and Princeton in Asia Fellow at IDFC Institute. He graduated with a Bachelor's Degree from the Woodrow Wilson School of Public and International Affairs at Princeton University in June 2019.

Prakhar Misra is Senior Associate at IDFC Institute. He also holds a Masters in Public Policy from the University of Oxford where he was a Chevening Scholar. He was Chanakya Scholar at Meghnad Desai Academy of Economics where he studied Economics and Finance. He holds a Bachelor's in Engineering from MS Ramaiah Institute of Technology where he graduated as the best outgoing student of his class.

datagovernance.org dgn@idfcinstitute.org

@datagovnetwork /datagovnetwork /datagovnetwork