BMJ Global Health

Rebooting consent in the digital age: a governance framework for health data exchange

Nivedita Saksena ,[1] Rahul Matthan,[2] Anant Bhan,[3] Satchit Balsari[1,4]

To cite: Saksena N, Matthan R, Bhan A, *et al*. Rebooting consent in the digital age: a governance framework for health data exchange. *BMJ Global Health* 2021;**6**:e005057. doi:10.1136/bmjgh-2021-005057

Handling editor Soumitra S Bhuyan

Received 19 January 2021
Revised 5 April 2021
Accepted 4 May 2021

[1]Harvard TH Chan School of Public Health, FXB Center for Health and Human Rights, Boston, Massachusetts, USA
[2]Takshashila Institution, Bengaluru, India
[3]Centre for Ethics, Yenepoya (Deemed to be University), Mangalore, Karnataka, India
[4]Department of Emergency Medicine, Harvard Medical School / Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

Correspondence to
Dr Satchit Balsari;
balsari@hsph.harvard.edu

ABSTRACT

In August 2020, India announced its vision for the National Digital Health Mission (NDHM), a federated national digital health exchange where digitised data generated by healthcare providers will be exported via application programme interfaces to the patient's electronic personal health record. The NDHM architecture is initially expected to be a claims platform for the national health insurance programme 'Ayushman Bharat' that serves 500 million people. Such large-scale digitisation and mobility of health data will have significant ramifications on care delivery, population health planning, as well as on the rights and privacy of individuals. Traditional mechanisms that seek to protect individual autonomy through patient consent will be inadequate in a digitised ecosystem where processed data can travel near instantaneously across various nodes in the system and be combined, aggregated, or even re-identified.

In this paper we explore the limitations of 'informed' consent that is sought either when data are collected or when they are ported across the system. We examine the merits and limitations of proposed alternatives like the fiduciary framework that imposes accountability on those that use the data; privacy by design principles that rely on technological safeguards against abuse; or regulations. Our recommendations combine complementary approaches in light of the evolving jurisprudence in India and provide a generalisable framework for health data exchange that balances individual rights with advances in data science.

Summary box

- India's National Digital Health Mission envisions a systems of electronic health records where data are collated with the patient's consent.
- However, traditional mechanisms that seek to protect individual autonomy through patient consent are inadequate in a digitised ecosystem.
- It is impossible to truly foresee how data may be combined and recombined and eventually used, making rational choices about future use of data uninformed, if not ineffective.
- We examine the merits and limitations of proposed alternatives like the fiduciary obligations that hold data processors accountable; privacy by design (PbD) principles that rely on technological safeguards against abuse; and regulatory frameworks.
- We favour the creation of an enabling regulatory environment where PbD principles can be leveraged not only to allow safe data exchange, but also to embed enforceability at scale.
- Our recommendations combine complementary approaches in light of the evolving jurisprudence in India and provide a generalisable framework for health data exchange that balances individual rights with advances in data science.

INTRODUCTION

The COVID-19 pandemic has underscored the need for a robust digital health ecosystem to deliver telemedicine,[1 2] remote care[3 4] and supervised task-shifting.[5 6] In India, the urgent need to compensate providers for COVID-19 care through the national health insurance scheme, Ayushman Bharat, accelerated the institution of a federated digital health ecosystem, the National Digital Health Mission (NDHM).[7] Borrowing heavily from technological innovations in the financial sector, the NDHM seeks to use a 'consent manager' to regulate data exchange between patients, provider, payers and others. The volume of daily data transactions expected across this ecosystem servicing 1.3 billion people raises significant privacy concerns.

Health data have historically been protected through consent, de-identification and ring fencing of storage and access.[8] Advances in data science however render these traditional approaches ineffective, making it possible to re-identify individuals or groups with relative ease.[9 10] This paper begins by describing the use and limits of consent in contemporary clinical practice and research, followed by an examination of three proposed alternatives: (1) the placing of fiduciary obligations on data processors; (2) privacy by design; and (3) and expanding regulation. We call for a complementary, contextually intelligent approach that balances the need for privacy with the opportunity to responsibly use data to advance medicine and public health.

CONSENT AND ITS LIMITATIONS

Informed consent is a key tenet in medicine, and is often understood as the explicit documented approval given by a patient to receive medical interventions after having reflected on related benefits and harms.[11] The seeking of consent to collect and use patients' data—including from their medical records, radiological images and tissue samples—has historically been less explicit.[12]

Coercion and obfuscation

In most primary care settings in India, general practitioners seldom maintain any records, and consent is not sought when they do. Community health workers routinely collect large volumes of data without explicit consent or explanation about how the data will be used.[13] The NDHM's strategy document, however, envisions a systems of electronic health records where data are collated with the patient's consent.[12] In modern hospitals, if consent is sought for the collection or use of data, it is documented during patient registration or at the bedside just prior to interventions.[14] According to the more recent telemedicine guidelines, if a patient initiates a telemedicine consultation, her consent is implied and not required to be explicitly sought.[15]

Such rule-based consent for data collection, the cornerstone of the NDHM architecture, satisfies formal legal requirements but risks being coercive[15 16] and does not constitute what Faden *et al*[17] described as true 'autonomous authorisation'. The power hierarchy operating in such interactions likely impedes true autonomous decision-making and is particularly exacerbated when services are sought by individuals already discriminated against due to gender, caste or class.

Routinisation of consent

Health data are also increasingly exchanged across services such as wearables, applications and some point-of-care devices that are governed by weak data protection regulations. To counter coercion, the NDHM makes participation in the health exchange voluntary.[18] The language, length and complexity of consent documents[19 20] accessed through small screens on mobile devices or wearables with little or no true choices have rendered them irrelevant, opaque, non-comparable and inflexible.[21] While the NDHM requires that privacy notices be clear and concise, and consent be informed, it proceeds to list at least nine distinct categories of information that each notice must provide.[14] Long privacy policy notices result in 'consent fatigue' and exacerbate the 'routinisation' of consent, where its provision merely signals compliance to gain access, and it is no longer informed or meaningful.[22]

Attempts to address the challenges with consent for collecting data have included dynamic consent, shorter consent forms and multimedia aids, all with limited success.[23] Consent for exchanging or transmitting collected data has been addressed either by (a) de-identifying or anonymising data sets rendering the data 'non-personal', and outside of the purview of privacy protection regulations, or (b) seeking 'blanket' consent for any use, or 'broad' consent for any reasonably foreseeable secondary use of data.[8] We discuss the limits of both approaches below.

Side-stepping consent

Large anonymised aggregated human mobility data sets collated from social medial platforms and AdTech companies were used during the COVID-19 pandemic to estimate the impact of social distancing directives.[24 25] Social media users who consented to the secondary use of their data could not have foreseen its use for pandemic response planning. Ethics committees have allowed such use of data because these are no longer 'identifiable,' and because the research is in the public interest in the midst of an emergency. However, such data when combined with other data sets may violate individual or group privacy through inadvertent or intentional re-identification or inferences.[9 10 26]

It is impossible to truly foresee how geolocation data collected from cell phone towers or AdTech companies, or digital phenotypes (unique characteristics) deduced from health and lifestyle applications may be combined and recombined, making rational choices about future use of data uninformed, if not ineffective.[27 28] Nissenbaum, while noting the ever changing nature of data flow and the cognitive challenges it poses, concludes: 'Even if, for a given moment, a snapshot of the information flows could be grasped, the realm is in constant flux, with new firms entering the picture, new analytics and new backend contracts forged: in other words, we are dealing with a recursive capacity that is indefinitely extensible.'[29]

Downstream commercialisation of data also raises important questions about claims to profits. Consider, for example, a successful machine learning algorithm that has trained on a trove of archived roentgenograms from rural public hospitals and can now accurately detect cancers long before the expert radiologist's eye. The invention is sold by a start-up for a large amount of money. The original set of patients may be uncontactable or deceased, while their data are being monetised by a third-party in ways that may have been inconceivable at the time the X-rays were administered. What does the company profiting from their data owe them, if anything at all?

Broad and blanket consent

For health data exchanged among laboratories, pharmacies and physicians in the routine care or billing of patients, consent for secondary use is 'broad', and in India almost always implicit. India's Personal Data Protection Bill 2019 (PDP Bill)[30] currently tabled in the Indian parliament, however, prohibits organisations from asking for blanket consent. Organisations will not be able to make the provision of services dependent on consent to unrelated processing (ie, cannot ask data principals to 'pay' with their data) and cannot treat users' failure to

opt out of preset settings as implying consent. This makes obtaining consent for processing that does not provide direct, tangible benefits to the data principal difficult even for companies that have direct relationships with end users.

Industry advocates have argued that consent-heavy systems thwart innovation, preventing society from benefiting from the application of artificial intelligence and machine learning in the fields of medicine and public health. Users are less likely to consent to processing that offers them little in return or bother to opt into settings they are opted-out of by default, running the risk of inadvertently blocking the development of products and technologies that may generate public good.[31] Subsequent sections examine alternative approaches that seek to address these limitations.

DATA FIDUCIARIES

The consent-model places the onus of privacy on the data principal—the person whose data is being processed, absolving data processors from using the data responsibly. In the absence of laws, companies such as Amazon,[32 33] Microsoft[34] and Google[35 36] have published voluntary standards on fairness and ethics, largely focused on purging bias from artificial intelligence algorithms, among growing alarm that governments and private entities are expanding surveillance efforts and automating decision-making in ways that may be discriminatory.[37 38] Critics have argued that this approach avoids thornier questions about who should be allowed to use these algorithms, to what avail and why they were built in the first place.[39 40]

To compel data processors to use data in a privacy preserving manner, the proposed PDP Bill places fiduciary obligations on the processors, expecting them to serve as trustees and act in the best interest of the data principals. NDHM allows data exchange between fiduciaries via a consent manager that allows the acquisition of asynchronous consent, or without, when mandated by law, as in case of emergencies. In theory, such preauthentication should allow data principals to 'reflect on their choices' and make informed decisions about when and whom to share data with.[41]

The concept of the information fiduciary was proposed by Balkin and Zittrain in 2016 to place obligations on processors of data to adhere to purpose limitation (limit the scope of use), data minimisation (limit the collection to only what is necessary), storage limitation (limit the duration of use) and transparency.[42 43] By placing concrete obligations on data fiduciaries, the PDP Bill seeks to mitigate the vulnerability created by power and information asymmetries between individuals and health professionals, large corporations or the state.

Limitations

The fiduciary approach is not without its critiques. Drawing attention to the legally mandated obligation of corporations to their shareholders, Khan and Pozen[44] argue that data fiduciaries cannot always act in the best interests of data principals. Bailey and Goyal submit that a fiduciary duty to act in a person's best interests does not necessarily preclude the sale of their data for profit.[45] Except for data breaches, data fiduciaries are not mandated to report any legal transgressions under the PDP Bill.[46] A data principal can approach the 'Data Protection Authority' if she believes that a data fiduciary has violated obligations. Thus, the onus of detection and enforcement remains with the data principal and may pose a challenge for individuals.

Importantly, time and purpose limitations could replicate the kind of undue unintended burden the Health Insurance Portability and Accountability Act, the law governing healthcare data in the USA, has placed on data flow for routine academic research.[47] They are also in direct tension with the intent of machine learning algorithms to mine data to reveal novel biological or behavioural relationships.[48] The imposition on fiduciaries face some of the very challenges faced by consent-heavy frameworks: it is impossible to tell what the future holds for the data, and what future the data will reveal. Both benefits and harms may be impossible to predict, and the kind of consent-driven purpose limitation required by the NDHM while preventing harm, may also hinder scientific gain.

PRIVACY BY DESIGN

Privacy by design (PbD), a systems engineering approach first developed by Cavoukian in 1995, calls for proactive privacy preserving design choices embedded throughout the process life cycle.[49] Since the advent of electronic medical records (EMRs), experts have recognised the need for embedding technological safeguards to protect privacy and prevent data breaches.[50 51] Advances in data science help address several of the aforementioned limitations, by either manipulating the data through strategies like minimisation, separation or abstraction or regulating the process by defining conditions for control and notification.[51 52]

In many settings in India, personal data can often be easily accessed by people who do not need such access; for example, clinic-based facilitators that liaise with state or private insurance companies, insurance agents themselves and in the public sector, administrative officials. There is little recognition that such access, however unintentional or inadvertent, is unethical, and will very soon be illegal.[53] The NDHM strategy calls for PbD tools without providing greater detail.[12] We have described below the dominant tools in current use that apply PbD principles to address gaps in health data protection.[54] These examples are meant to be illustrative and are not exhaustive.

Data minimisation

When health data are collected, either through clinical operations or during research, there is temptation to

collect more and not less, given the opportunity costs associated with collecting these data. This results in exhaustive data sets archived in the public and private health sector that pose significant privacy risks.[53] Restricting data collection to the essentials has in fact been demonstrated to declutter and improve the user-interface, and consequently, user-experience and compliance, while reducing privacy risks.[55] While the NDHM espouses data minimisation, existing legacy digital public health systems continue to collect vast amounts of redundant data on millions of beneficiaries, without demonstrable justification.[14 53]

Role-based access

Role-based access is a standard feature in most advanced EMRs.[56] Open source tools like Dataverse provide scientists differential access to research databases as well.[57] Multi-authority attribute-based encryption schemes allow role-based models to scale by allowing access to users based on a set of attributes, rather than on individual identities.[58 59] For example, by virtue of being a verified clinician (regardless of who), physicians are generally able to look up most medical records at their institution easily; by virtue of being a public health administrator (regardless of who), officers should have no access to personal health information; and by virtue of being a research laboratory, the team would have access to authorised de-identified data, provided third-party regulators can affirm the veracity of each of their attributes (clinician, administrator, researcher).[49 60] The Account Aggregator, a similar consent management framework already in play in India's fintech ecosystem, lends itself to such selective, verifiable, pre-authenticated access as has been proposed at the backbone for the NDHM.[61] Since user-consent can be sought asynchronously (prior to actual data processing), this model somewhat mitigates inadvertent coercion associated with point-of-care consent seeking. The NDHM seeks to verify attributes by developing and maintaining 'registries' of providers.[62]

User preference

The General Data Protection Regulation in the European Union facilitates data access by requiring companies to provide a consent management platform to give users more control over their data, by selecting from a menu of data-use options.[14] In India, the Data Empowerment and Protection Architecture and the NDHM seek to empower users by allowing them to place revocable time and purpose limitations on the use of their data—the sorts of choices that would be extremely beneficial to patients.[63] In theory, patients would control who accesses their data at all times, would receive notification of third-party access (whether authorised or not), or be able to revoke access at will, when permitted by law.

Others have elaborated on the idea by allowing data principals to opt into certain 'data trusts' or stewards with pre-negotiated access controls, where general attributes can be used to guide future data sharing: for example, a patient may elect to always allow healthcare providers to access her data but always deny access to pharmaceutical companies regardless of the identifiability of the data.[64–66] This approach would entail data principals communicating their preferences to the consent manager to accordingly direct data toward select categories of data processers; for example, to clinical health information users, and say, public research agencies like the Indian Council of Medical Research, but not to pharmaceutical companies.[12] The asynchronous and one-time (but revocable and changeable) nature of the process—made possible by the consent manager framework—may allow users to make a more informed and coercion-free choice, if citizens are encouraged to actively enrol in the system prior to clinical care.

Differential privacy

The current NDHM guidelines require that all health information processors make aggregated data available. Not only are aggregation and anonymisation inadequate for protecting privacy for the reasons described above, but many aspects of clinical and population health will require non-anonymised, high resolution data to actually be useable and useful.[12] The NDHM's Health Data Management Policy prohibits inadvertent unforeseen re-identification while processing data.[14]

Differential privacy (DP) seeks to balance such access to rich data while preserving privacy. It achieves this balance by differentially introducing 'statistical noise' in the data set, depending on what is being queried and by whom, thus combining the aforementioned approaches. The 'noise' masks the contribution of each individual data point without significantly impacting the accuracy of the analysis. Moreover, the amount of information revealed from each query is calculated and deducted from an overall privacy budget to halt additional queries when personal privacy may be compromised. If effective, this approach will help alleviate some of the concerns about combining large data sets; its utility in the clinical setting is yet to be determined. There is precedent for DP as a model for collaborative research.[67] Open source platforms like OpenDP are likely to accelerate use of the application of DP across disciplines.[68] DP may however lead to noisy aggregates with poor utility for analytical tasks in public health.[69 70] Given the nascency of DP applications, it is premature to assess utility based on field-impact.

REGULATION

The jurisprudence on privacy is rapidly evolving in India (see table 1), and includes a landmark judgement of the Supreme Court affirming the right to privacy.[71] The PDP Bill seeks to regulate the collection and transfer of all personal data, including health information. The law requires consent from the data principal before processing their personal data, and because health data are considered 'sensitive' by the law, the data principal

Table 1 Existing framework for data protection in India

Document	Details	Type	Nature
Puttaswamy versus Union of India	Judgement of the Supreme Court of India affirming the right to privacy of all individuals under the Indian Constitution.	Law	Binding
Information Technology Act, 2000	Prescribes security practices for the protection of personal data. Requires that consent must be sought before the collection of any sensitive personal data.	Law	Binding and enforceable
HIV/AIDS Act 2017, Mental Healthcare Act, 2017, Transplantation of Human Organs and Tissues Act, 1994	Sector-specific laws that govern data related to the disease area. The requirements may be different from those under the Information Technology (IT) Act.	Law	Binding and enforceable
Personal Data Protection Bill, 2019	Proposed law that updates the IT Act and protects all personal data, establishes a data protection regulator and prescribes penalties for violations of these rules.	Bill; pending in parliament	Unenforceable till passed as law
Data Empowerment and Protection Architecture	Framework for data management and security issued by NITI Aayog, a government think-tank.	Draft report	Voluntary
National Digital Health Blueprint, NDHM Health Data Management Policy, NDHM strategy overview	Lays out the architectural framework for the digital health infrastructure under the NDHM.	Government reports	Voluntary
Report by the committee of experts on Non-Personal Data Governance Framework	This committee of experts was constituted by the Ministry of Electronics and IT to propose a governance framework for non-personal data. It has released a draft report for public comments (July 2020).	Draft government report	Recommendations to the government

NDHM, National Digital Health Mission.

must be informed of any potential harm to them resulting from the processing of their data. It also requires fiduciaries to introduce technical safeguards through anonymisation and to adopt measures to prevent unauthorised access and misuse of the data, thus presenting a non-prescriptive opportunity to adopt privacy preserving design. Proposed policy and technology frameworks including the National Digital Health Blueprint, and its related strategy and policy documents, ascertain that the data principal 'owns' the data by authorising access to, from and across various 'health information processors', by providing consent for such transactions.[18]

The binary (personal vs non-personal) classification approach is likely to be rendered inadequate with novel applications of seemingly non-personal data. Data from accelerometers and gyroscopes of mobile phones, or from phone usage patterns, can be used to construct fairly accurate and unique 'digital phenotypes' of individuals.[72] Data that have been 'irreversibly' anonymised and do not fall within the scope of the PDP Bill are addressed by a proposed Non-Personal Data Governance Framework.[73] It recommends that fiduciary obligations remain in place even when personal data are anonymised, and that data principals should provide consent for both, anonymisation and the use of the anonymised data. The framework seeks to create differentially accessed data commons distinguished by source of origin of the data: whether from individuals, communities, public domains or private entities. While this may indeed be the holy grail of data access, the implementation path is uncertain in the absence of regulations.

Table 2 summarises the strengths and limitations of the four aforementioned approaches, none of which can alone provide satisfactory data access while preserving privacy. Socioeconomic realities, technological ubiquity and the scope and nature of the regulatory environment will help communities calibrate the approaches that will best suit them. We favour the creation of an enabling regulatory environment where PbD principles can be leveraged not only to allow safe data exchange, but also to embed enforceability at scale.

DISCUSSION

The accelerated growth of data science in recent years has resulted in large shifts in societal responses to new technologies. The growing excitement over the interoperability of mobile applications was replaced with collective concern about data-grabs and unforeseen use of personal data. Just as several early adopters of virtual assistants have unplugged their devices and turned off their cameras, and many WhatsApp enthusiasts are migrating to Signal, it is not unreasonable to expect an ebb and flow in society's embrace of health data exchange, as expectations and fears change with time. Low adoption of digital contact tracing applications during the coronavirus pandemic reflected the low levels of trust in technology platforms and in governments.[74] The technology and policy frameworks that eventually define health data ecosystems must therefore not only account for these tides but also acknowledge the local social contexts in which they are developed.[75]

Table 2 Strengths and limitations of proposed approaches to protect personal health data

Approach	Strength	Limitation
Consent-framework	1. Traditionally and widely used as a tool to ensure patient autonomy and (despite its limitations) prevent exploitative practices. 2. In common use by medical practitioners during the provision of routine healthcare, or researchers during research projects. 3. The ethical and legal framework for consent is well established. 4. No additional costs need to be incurred as it is already a part of patient care.	1. It currently takes the form of lengthy and complicated consent forms that the patient may not properly read or understand. With consent needed for many actions during a medical procedure, it may sometimes be given without due consideration or out of habit. 2. In the context of healthcare, a power differential exists between the patient and medical provider. It is therefore unclear how truly autonomous consent is. 3. It is impossible for the patient to consent to all the possible uses of the data which might not be known at the time that it is being collected. Re-consent may not be possible if data has been anonymised or the patients might not be contactable. This may hinder medical research and the development of novel technologies.
Fiduciary obligations	1. Instead of the onus for data protection being on patients, shifts this burden onto entities collecting, storing and using the data. 2. Particularly useful where the ability of the patient to provide informed consent is impaired such as in the context of de-identified or anonymised data where there is a potential for a privacy violation if the data is made identifiable or is de-anonymised.	1. It may be difficult for a data principal to detect that an entity processing their data has violated a fiduciary duty. 2. These obligations may conflict with legally enforceable duties that corporations owe to their shareholders. 3. Might be difficult to enforce since large quantities of data would have to be regulated. In India, it will require a strong and independent data protection authority.
Privacy by design	1. Reduces the chance of human-induced errors by baking privacy preserving practices and features into the technical architecture.	1. There is currently a lack of expert consensus or comprehensive guidelines from data protection authorities on the kinds of safeguards that must be incorporated in enterprise architecture for healthcare. 2. Might increase operational costs for healthcare organisations. This would disproportionately affect smaller organisations. 3. Has not yet been formally incorporated into the information systems of major health information technology companies or health systems of countries.
Regulation	1. Provides clear guidelines to protect the privacy rights of people and an environment of legal and operational certainty for entities processing data. 2. Rights can be enforced using legal mechanisms and penalties may be imposed for egregious violations of data protection obligations.	1. Regulations may differ in different countries, increasing costs of compliance for entities operating internationally. 2. If the regulations are too burdensome, it may limit innovation. 3. Large costs imposed by data protection regulators may affect smaller organisations but would be insignificant for big companies like Facebook and Google. 4. Since privacy is legally understood as an individual right, it may be difficult to protect group privacy under this framework.

They must also account and accommodate for the inequities that digitisation can exacerbate. Despite the perceived ubiquity of cell phones in India, only 502.2 million adults own smart phones, with the elderly, disabled and poor—those with likely the greatest health needs—having the least access.[76] Internet access in rural India is limited to one in three persons,[77] with significant gender disparity.[78] Data literacy and analytical capabilities are limited to a few institutions of higher learning, precluding the vast majority of local healthcare

institutions and public health agencies from leveraging the gains of readily available data, while posing not insignificant privacy risks. In the absence of demonstrated public health or clinical utility, process and algorithmic transparency will be key.[79]

The NDHM framework places consent at the centre of all exchange. The asynchronous authentication process permitted by the consent manager may in fact allow such consent to be non-coercive and meaningful if its scope and limits are transparently and effectively communicated to India's diverse range of users. Some argue that modern privacy laws place an undue burden on technology companies, inadvertently pushing out smaller players and giving larger data brokers a competitive advantage.[31] The PbD frameworks proposed by the NDHM must therefore be expanded beyond aggregation and anonymisation, to responsibly allow a broader community of scientists to access the vast data streams NDHM would generate, without harming individuals or groups.

The proposed regulatory changes seek to simultaneously protect data principals while liberating access to non-personal data. On the one hand, the strict consent-heavy purpose limitation may thwart innovation unless supplemented with notification to data principals during unplanned reuse. On the other, advances in data science applications may render the simple dichotomy between personal and non-personal data insufficient; risking all big data being classified as personal, since NDHM includes data that may inadvertently result in re-identification.

CONCLUSION

Consent will remain the bedrock of information exchange in medicine for the foreseeable future. In its current avatar, however, consent is flawed and must be improved by applying intelligent design to limit our ability to select harmful options. Legislation that mandates transparency and accountability will likely generate the trust needed to improve the adoption of digitisation. And trust will be the foundation of the kinds of data commons that must be built to advance the science of medicine and the health of populations.

Twitter Nivedita Saksena @NiveditaSaksena and Satchit Balsari @Satchit_Balsari

Contributors NS and SB conceptualised and wrote the first draft of the manuscript. RM and AB revised it critically for important intellectual content. All authors agreed with the conclusions of this article. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf; NS and SB report grants from Tata Trusts and Dell Giving outside the submitted work.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article.

ORCID iD

Nivedita Saksena http://orcid.org/0000-0001-6992-5368

REFERENCES

1 Agarwal N, Jain P, Pathak R, *et al*. Telemedicine in India: a tool for transforming health care in the era of COVID-19 pandemic. *J Educ Health Promot* 2020;9:190.
2 Mahajan V, Singh T, Azad C. Using telemedicine during the COVID-19 pandemic. *Indian Pediatr* 2020;57:658–61.
3 Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. *JMIR Public Health Surveill* 2020;6:e18810.
4 Dagan A, Mechanic OJ. Use of ultra-low cost fitness trackers as clinical monitors in low resource emergency departments. *Clin Exp Emerg Med* 2020;7:144–9.
5 Naslund JA, Shidhaye R, Patel V. Digital technology for building capacity of Nonspecialist health workers for task sharing and scaling up mental health care globally. *Harv Rev Psychiatry* 2019;27:181–92.
6 Robertson FC, Lippa L, Broekman MLD. Editorial. task shifting and task sharing for neurosurgeons amidst the COVID-19 pandemic. *J Neurosurg* 2020:1–3.
7 National digital health mission. Available: https://ndhm.gov.in/ [Accessed 17 Jan 2021].
8 El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. *BMJ* 2015;350:h1139.
9 de Montjoye Y-A, Hidalgo CA, Verleysen M, *et al*. Unique in the crowd: the privacy bounds of human mobility. *Sci Rep* 2013;3:1376.
10 Na L, Yang C, Lo C-C, *et al*. Feasibility of Reidentifying individuals in large national physical activity data sets from which protected health information has been removed with use of machine learning. *JAMA Netw Open* 2018;1:e186040.
11 Brach C. *Making informed consent an informed choice*. Health Affairs Blog, 2019.
12 National Health Authority. *National digital health mission strategy overview*, 2020.
13 Shah A. Using data for improvement. *BMJ* 2019;364:l189.
14 Health Data Management Policy. *National digital health mission*, 2020.
15 Medical Council of India, 2020. *Indian medical Council (professional conduct, etiquette and ethics) (Amendment) regulations*, 2020.
16 Ebeling MFE. *Coercive consent and digital health information. healthcare and big data: digital Specters and phantom objects*. New York: Palgrave Macmillan US, 2016: 67–94.
17 Faden RR, Beauchamp TL, King NMP. *A history and theory of informed consent*. Oxford University Press, 1986.
18 Ministry of Health and Family Welfare. *National digital health blueprint*, 2019.
19 Manta CJ, Ortiz J, Moulton BW. From the patient perspective, consent forms fall short of providing information to guide decision making. *J Patient Saf* 2016. doi:10.1097/PTS.0000000000000310
20 Pandiya A. Readability and comprehensibility of informed consent forms for clinical trials. *Perspect Clin Res* 2010;1:98–100.
21 Walker K. *The costs of privacy*. 25. Harv JL & Pub Pol'y, 2001.
22 Ploug T, Holm S. Informed consent and routinisation. *J Med Ethics* 2013;39:214–8.
23 Kaye J, Whitley EA, Lund D, *et al*. Dynamic consent: a patient interface for twenty-first century research networks. *Eur J Hum Genet* 2015;23:141–6.
24 Jeffrey B, Walters CE, Ainslie KEC, *et al*. Anonymised and aggregated crowd level mobility data from mobile phones suggests that initial compliance with COVID-19 social distancing interventions was high and geographically consistent across the UK. *Wellcome Open Res* 2020;5:170.
25 Kishore N, Kiang MV, Engø-Monsen K, *et al*. Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis. *Lancet Digit Health* 2020;2:e622–8.
26 Culnane C, Rubinstein BI, Teague V. Health data in an open world. *arXiv* 2017 https://arxiv.org/abs/1712.05627

27 Sloan RH, Warner R. Beyond notice and choice: privacy, norms, and consent. *J High Tech L* 2014;14.
28 Solove DJ. *The digital person: technology and privacy in the information age*. NYU Press, 2004.
29 Nissenbaum H. *A contextual approach to privacy online*. Digit Enlight Yearb, 2012: 219–34.
30 The Personal Data Protection Bill. *The personal data protection bill 373 of 2019: LOK Sabha*, 2019.
31 Chivot E, Castro D. *The EU needs to reform the GDPR to remain competitive in the algorithmic economy: center for data innovation*, 2019.
32 Partnership on AI. Available: https://www.partnershiponai.org/about/ [Accessed 17 Jan 2021].
33 NSF. *Nsf program on fairness in artificial intelligence in collaboration with Amazon (FAI): national science Foundation*, 2020.
34 Microsoft. Microsoft AI principles. Available: https://tinyurl.com/y3g4zz54 [Accessed 17 Jan 2021].
35 Google. Ai at Google: our principles, 2018. Available: https://www.blog.google/technology/ai/ai-principles/ [Accessed 17 Jan 2021].
36 Olson P. Google Quietly Disbanded another AI review board following Disagreements: the wall Street Journal, 2019. Available: https://www.wsj.com/articles/google-quietly-disbanded-another-ai-review-board-following-disagreements-11555250401?st=8djs0dshthp1z05 [Accessed 17 Jan 2021].
37 Murray SG, Wachter RM, Cucina RJ. Discrimination by artificial intelligence in a commercial electronic health record—a case study. *Health Affairs Blog* 2020;10.
38 Obermeyer Z, Powers B, Vogeli C, *et al*. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019;366:447–53.
39 Whittaker M, Crawford K, Dobbe R. *Ai now report 2018*. New York, NY: AI Now Institute, 2018.
40 McLennan S, Lee MM, Fiske A, *et al*. Ai ethics is not a panacea. *Am J Bioeth* 2020;20:20–2.
41 Coiera E, Clarke R. e-Consent: the design and implementation of consumer consent mechanisms in an electronic environment. *J Am Med Inform Assoc* 2004;11:129–40.
42 Balkin J, Zittrain J. *A grand bargain to make tech companies trustworthy*. The Atlantic, 2016.
43 Balkin JM. *The fiduciary model of privacy*. 134. Harvard Law Review Forum, 2020.
44 Khan LM, Pozen DE. A skeptical view of information fiduciaries. *Harv L Rev* 2019;133:497.
45 Bailey R, Goyal T. *Fiduciary relationships as a means to protect privacy: examining the use of the fiduciary concept in the draft personal data protection bill, 2018*, 2019.
46 Reddy P. Personal data protection bill: notification of data breaches: Bloomberg Quint, 2020. Available: https://www.bloombergquint.com/opinion/how-the-personal-data-protection-bill-tackles-data-breaches [Accessed 17 Jan 2021].
47 Nosowsky R, Giordano TJ. The health insurance portability and accountability act of 1996 (HIPAA) privacy rule: implications for clinical research. *Annu Rev Med* 2006;57:575–90.
48 Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics* 2016;22:303–41.
49 Cavoukian A. *Privacy by design: the 7 foundational principles: the International association of privacy professionals*, 2009.
50 Privacy RTC. Information technology, and health care. *Commun ACM* 1997;40:92–100.
51 Semantha FH, Azam S, Yeo KC, *et al*. A systematic literature review on privacy by design in the healthcare sector. *Electronics* 2020;9:452.
52 Privacy design strategies. *IFIP international information security conference*. Springer, 2014.
53 Sahay S, Mukherjee A. *Where is all our health data going?* 55. Economic & Political Weekly, 2020.
54 van Rest J, Boonstra D, Everts M. *Designing Privacy-by-Design*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014: 55–72.
55 Balsari S. Will AI help universalize health care? *BMJ Opinion* 2019 https://blogs.bmj.com/bmj/2019/09/23/satchit-balsari-will-ai-help-universalize-health-care/
56 de Carvalho Junior MA, Bandiera-Paiva P. Health information system Role-Based access control current security trends and challenges. *J Healthc Eng* 2018;2018:6510249
57 King G. An introduction to the Dataverse network as an infrastructure for data sharing. *Sociol Methods Res* 2007;36:173–99.
58 Wickramasuriya J, Venkatasubramanian N. *Dynamic access control for ubiquitous environments*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004: 1626–43.
59 Yu S, Ren K, Lou W. *Security and privacy in communication networks*. Berlin, Germany: Springer, 2009.
60 Jiang S, Zhu X, Wang L. EPPS: efficient and Privacy-Preserving personal health information sharing in mobile healthcare social networks. *Sensors* 2015;15:22419–38.
61 RBI. Non-Banking Financial Company - Account Aggregator (Reserve Bank) Directions, 2016. Available: https://www.rbi.org.in/Scripts/bs_viewcontent.aspx?Id=3142 [Accessed 17 Jan 2021].
62 NDHM. NDHM health facility registries FAQ. Available: https://ndhm.gov.in/home/health_facility_registry_faq [Accessed 17 Jan 2021].
63 Niti-Aayog. Data Empowerment and protection architecture: draft for discussion, 2020. Available: https://niti.gov.in/sites/default/files/2020-09/DEPA-Book_0.pdf [Accessed January 17, 2021].
64 Delacroix S, Lawrence ND. Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law* 2019;9:236–52.
65 Tentori M, Favela J, Rodriguez M. Privacy-Aware autonomous agents for pervasive healthcare. *IEEE Intell Syst* 2006;21:55–62.
66 Rosenbaum S, Governance D. Data governance and stewardship: designing data stewardship entities and advancing data access. *Health Serv Res* 2010;45:1442–55.
67 Dwork C. *Differential privacy: a survey of results. International Conference on theory and applications of models of computation*. Springer, 200: 1–19.
68 OpenDP. OpenDP: building an open-source suite of tools for deploying differential privacy, 2020. Available: https://projects.iq.harvard.edu/opendp [Accessed 17 Jan 2021].
69 Santos-Lozada AR, Howard JT, Verdery AM. How differential privacy will affect our understanding of health disparities in the United States. *Proc Natl Acad Sci U S A* 2020;117:13405–12.
70 Mervis J. Can a set of equations keep U.S. census data private? *Science* 2019.
71 Justice K. S. *Puttaswamy (Retd.) and Anr. vs Union of India. Supreme Court cases: Supreme Court of India*. 1, 2017.
72 Onnela J-P, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016;41:1691–6.
73 MEITY. *Report by the Committee of experts on Non-Personal data governance framework: Ministry of electronics and information technology*. Government of India, 2020.
74 Knowledge Wharton. Combating COVID-19: Lessons from Singapore, South Korea and Taiwan: Knowledge@Wharton, 2020. Available: https://knowledge.wharton.upenn.edu/article/singapore-south_korea-taiwan-used-technology-combat-covid-19/ [Accessed 17 Jan 2021].
75 May C, Mort M, Williams T, *et al*. Health technology assessment in its local contexts: studies of telehealthcare. *Soc Sci Med* 2003;57:697–710.
76 Silver L, Smith A, Johnson C. *Mobile connectivity in emerging economies: Pew research Cente*, 2019.
77 Kawoosa VM. *Connectivity gets better but parts of India still logged out*. Hindustan Times, 2020.
78 Bhuyan A. 4 in 5 Bihar women have never used the Internet December 16, 2020. Available: https://www.indiaspend.com/gendercheck/4-in-5-bihar-women-have-never-used-the-internet-702855 [Accessed 17 Jan 2021].
79 Sarbadhikari SN, Pradhan KB. The need for developing Technology-Enabled, safe, and ethical workforce for healthcare delivery. *Saf Health Work* 2020;11:533–6.