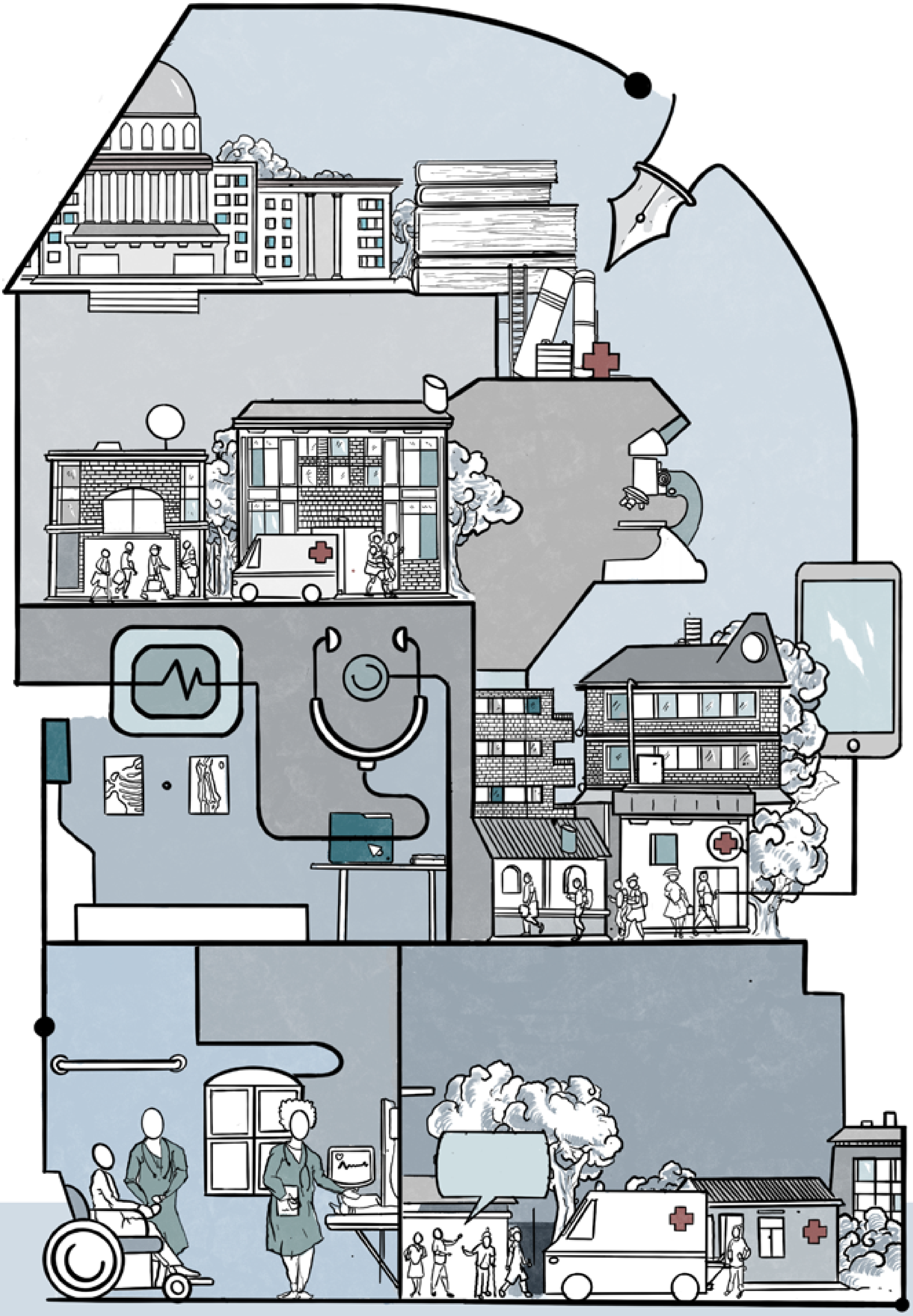
aapti institute



Health Data Stewardship: Learning From Use Cases

Health Data Stewardship:
Learning From Use Cases

aapti institute

Soujanya Sridharan, Siddharth Manohar
and Astha Kapoor

July 2021

Aapti is a public research institute that works on the intersection of technology and society. It examines the ways in which people interact and negotiate with technology both offline and online.

This is a work of independent research, funded by
The Rockefeller Foundation.

Foreword

As I read this report from the Aapti Institute, I recall a meeting during the early days of the Covid-19 pandemic. An advertising firm, inspired to fight the pandemic, had turned to the tools they had: Smartphone location data to tracking the movements of hundreds of thousands of people. They presented maps following people from a "superspreader" event back to their town and neighborhoods. They also rated, at a block level, levels of compliance with stay-at-home orders, suggesting that the data could be used by authorities to target outreach.

The conversation made clear the promise of data in a pandemic: The world desperately needed this kind of detailed data as it sought to understand and respond to a new threat. Contact tracing, learning who was most vulnerable, identifying effective treatments and interventions could all be accelerated by knowing details about millions of people.

However, I was also alarmed. There were real perils embedded in this detailed data. Fear of the virus was rampant, and there was—and still is—a real risk of inciting repression, discrimination, and even violence against the people being traced and their neighbours. Suspicion—both justified and unjustified—about how authorities and companies are using the data about use impedes the response to this day. And of course, data is only useful if it's relevant, if we can have confidence in its accuracy and representativeness, if it is widely accessible and usable for analysis, and if the limitations and risks of the resulting insights are well-understood.

That conversation also demonstrated the weakness of today's standard tools for maximizing the promise and minimizing the peril of data. The data I was seeing were from people who had "consented" and "been de-identified." Yet they were likely unaware of what they had consented to, and we could trace them to their block, if not their house. As advertisers were tracing the movement of people at national scale with only a vague sense of how the data could be responsibly applied to public health, public health authorities around the world were struggling to aggregate basic case and patient data needed they needed.

This report from the Aapti Institute calls for a new approach of "data stewardship" as a response to both the promise and peril of data. A stewardship approach recognizes that data is a collective resource for individuals and for communities, whose use creates real, practical benefits and risks. Stewards have many roles to play, from ensuring that data is in a usable format and accessible; to ensuring that individuals and communities feel confident that the way data is being with their participation and for their benefit.

The health sector carries special peril and promise, calling out for data stewardship. The report tackles these special problems, and highlights inspiring public and private stewardship efforts in the health sector: Data cooperatives for health research (MiData, global.health), collaboratives for resource planning (Health Data Collaborative), and personal data stores that allow individuals to direct their data to specific uses (Clue), and many more. All of these provide exciting alternatives to existing data sharing structures.

The report convincingly makes the argument that data stewardship is a journey that all of us who collect, share, and use data need to intentionally set out on, not a set of boxes to be checked. Nonetheless, the authors lay out a practical guide for navigating the complexity of stewardship. Better stewardship will require policy change, experimentation, new technical and institutional tools, education and awareness, and engagement with the context and communities in which data is being collected and used.

The benefits of stewardship are real and tangible: the data stewards they profile are speeding the discovery of treatments, adding new capabilities to our public health response, and enabling new applications in precision medicine, among other uses.

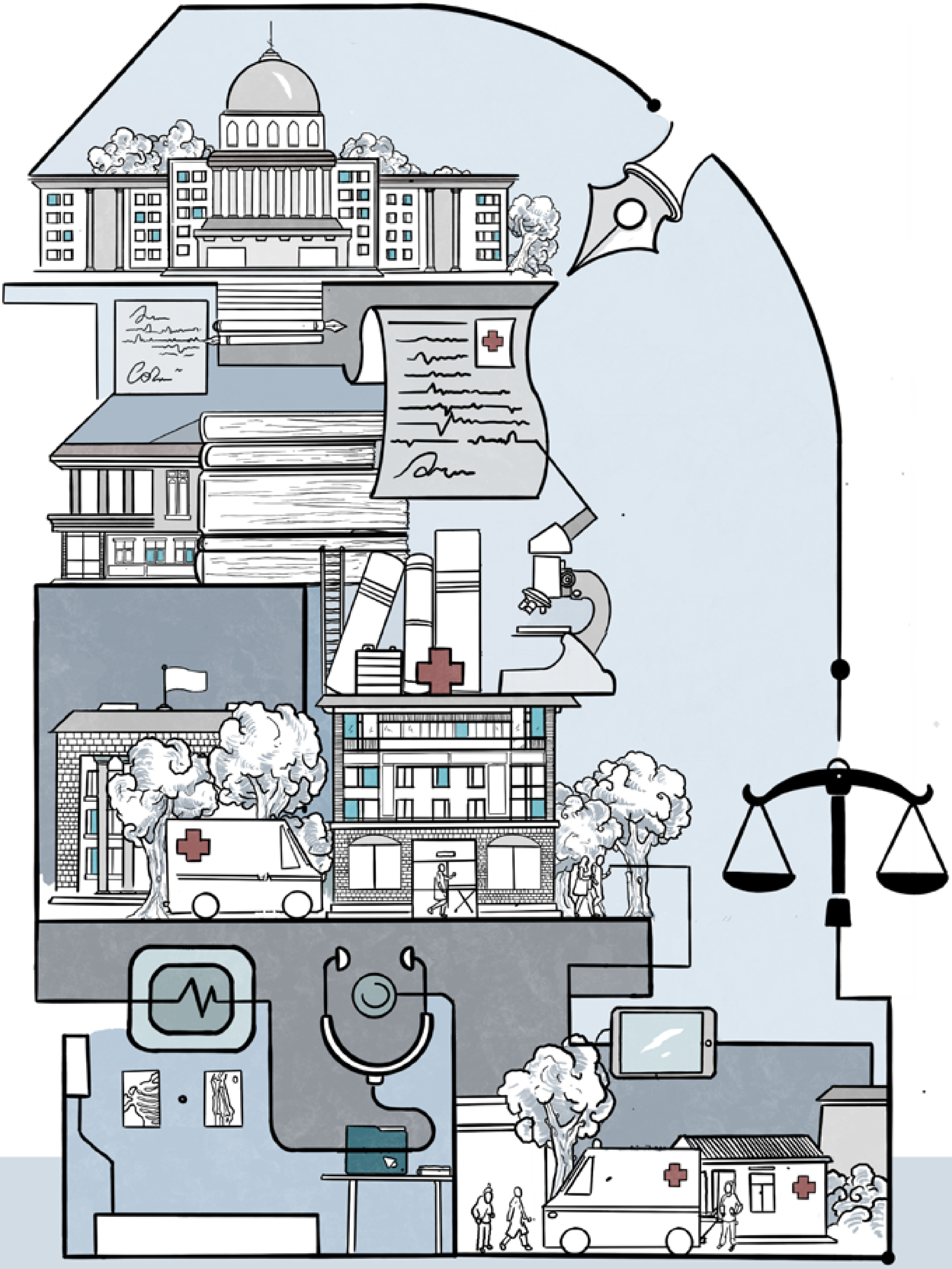
Our work at The Rockefeller Foundation is often inspired by the refrain that “data saves lives.” This research on data stewardship is an important step in creating the usability, accessibility, accuracy, and trust required to that make that statement true.

Kevin O’Neil

Director, New Frontiers
Rockefeller Foundation

Contents

aapti institute



Health Data Stewardship: Top-Down State Action for Public Benefit Data Sharing

Health Data Stewardship: Top-Down State Action for Public Benefit Data Sharing

aapti institute

Soujanya Sridharan, Siddharth Manohar
and Astha Kapoor

July 2021

Acknowledgements

This report draws from the expertise of numerous academic and industry experts and practitioners. We are grateful for their input during interviews and rigorous feedback sessions. We also thank Kevin O'Neil and Priyank Hirani of The Rockefeller Foundation for supporting this work.

Research assistant: Bilal Mohamed | **Editor:** Yana Bey
Report design: Meher Rajpal | **Illustration:** Siddhartha Iyer

Contents

Glossary

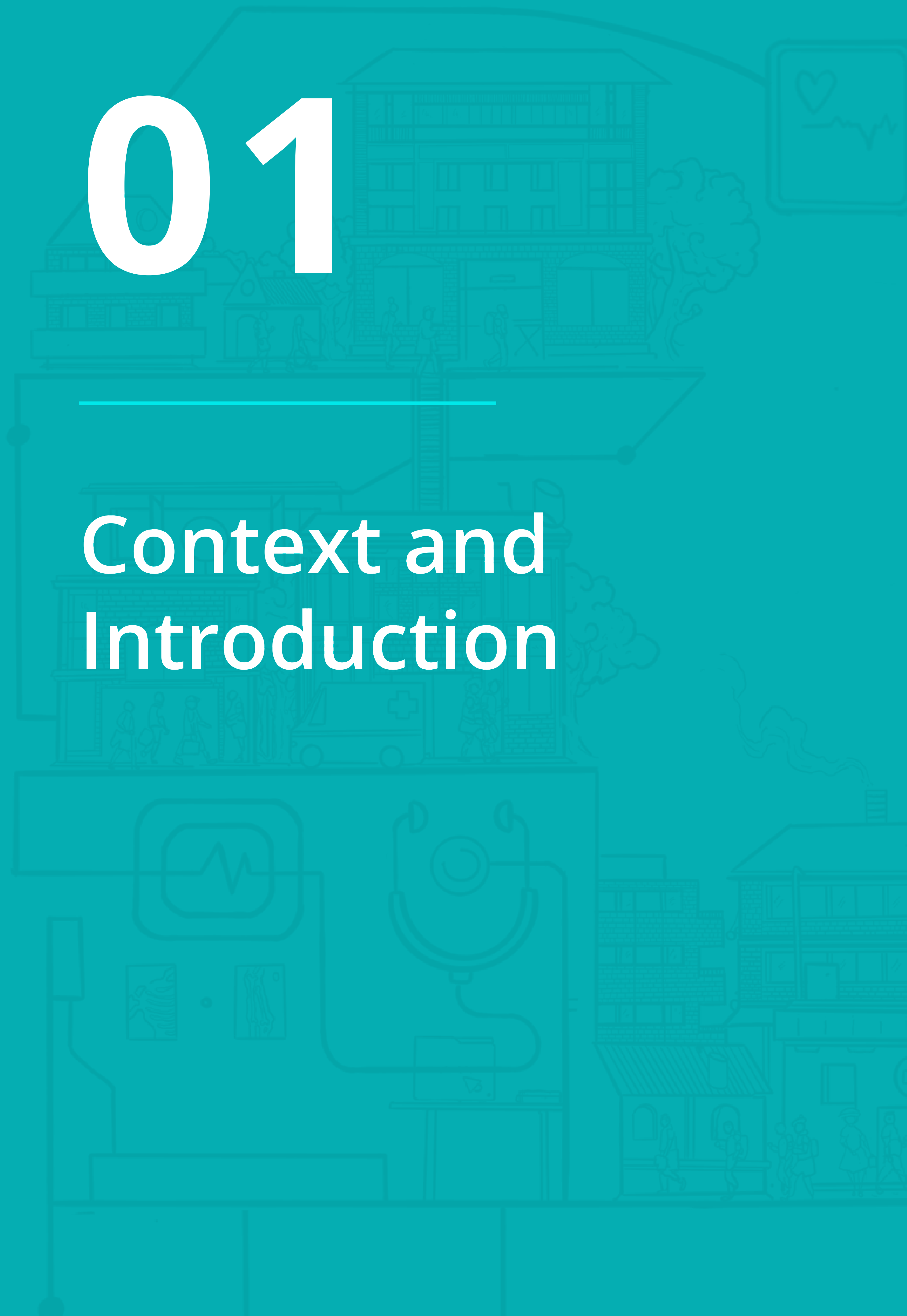
(CONCEPTUAL AND ABBREVIATIONS)

Data	Any representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means (as in the Information Technology Act, The Personal Data Protection Bill).
Consent	Given by the individual to an entity for processing of their data; must be free, informed, specific, clear and capable of being withdrawn (as in the PDP Bill).
Data principal / user	Persons, both natural and legal, to whom any data relates (as in the PDP Bill).
Personal data	Data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such natural person, whether online or offline, or any combination of such features with any other information, including any inference drawn from such data for the purpose of profiling (as in the PDP bill).
Non-personal data	Data that either never related to an identified or identifiable natural person, or data which may have initially been personal data, but was later anonymised through transformation techniques to the extent that individual-specific events are no longer identifiable (as in the Non-personal data Report).
Ecosystem	A composite term referring to the various stakeholders within the data economy, enabling infrastructures for data sharing, enabling legislation to data sharing, and the network of relationships between each of these - working together to make data sharing a possibility.

Bottom-up action for stewardship	This relates to the Activities undertaken by providers, civil society organisations and other relevant stakeholders such as health tech operators to institute human-centric mechanisms for data governance within their organisations.
Top-down action for stewardship	Collection of legislative, regulatory and policy frameworks implemented by state authorities and public institutions to operationalise data stewardship within each jurisdiction.

01

Context and Introduction



> Health data traditionally refers to medical records, public health registry data, and data produced in the context of biomedical or clinical research. Recent advancements in technology, however, have enabled individual medical records to be bundled in the form of comprehensive electronic health records.[1]

Patterns emerging from healthcare data can be turned into actionable knowledge which generates value for patients and healthcare organisations. It can, inter alia, i) increase effectiveness and quality of treatment by discovery of early signals and disease intervention, reduced probability of adverse reactions, etc.; (ii) widen possibilities for prevention of diseases by identifying risk factors; (iii) improve pharmaco vigilance and patient safety through the ability to make more informed medical decisions based on directly delivered information to patients; (iv) predict outcomes for infectious diseases; and (v) reduce inefficiencies by optimising operations.[2]

In the healthcare sector, providers process personal health information through blanket or implicit consent mechanisms which govern users' data flow.[3] Naturally, in this increasingly digitally mediated data-intensive society, current consent mechanisms are ineffectual in giving users autonomy on aspects impacting their privacy. In such a setting, data breaches in the healthcare sector do not just prove financially expensive for providers but also exacerbate the trust deficit between stakeholders. For instance, a 2019 IBM study revealed that healthcare companies, on average, incur $7 million in breach costs – a 10% increase from the previous year.[4] Recently, the Federal Trade Commission (FTC) penalised Flo, a company that runs a period and fertility tracking app, for sharing personal user information with Facebook, Google, and other organisations without authorisation.[5]

However, at the core, it is important to recognise the imbalance of knowledge and control over sensitive personal data in the digital ecosystem. Health data stewardship, in its role as a trusted intermediary, recognises this imbalance, and facilitates or holds consent and decision-making on behalf of users.

---

[1] Biomedical Big Data: New Models of Control Over Access, Use and Governance, Vayena and Blassime, December 2017, accessible at https://www.researchgate.net/publication/320236003_Biomedical_Big_Data_New_Models_of_Control_Over_Access_Use_and_Governance/.

[2] Study on Big Data in public health, telemedicine and healthcare, EU Report, December 2016, accessible at https://ec.europa.eu/health/sites/default/files/ehealth/docs/bigdata_report_en.pdf/.

[3] "Biomedical Big Data: New Models of Control Over Access, Use and Governance", Effy Vayena and Alessandro Blassime, Journal of Bioethical Inquiry, December 2017, accessible at https://www.researchgate.net/publication/320236003_Biomedical_Big_Data_New_Models_of_Control_Over_Access_Use_and_Governance/.

[4] "IBM Report: Compromised Employee Accounts Led to Most Expensive Data Breaches Over Past Year", IBM Newsroom, July 2020, https://newsroom.ibm.com/2020-07-29-IBM-Report-Compromised-Employee-Accounts-Led-to-Most-Expensive-Data-Breaches-Over-Past-Year/.

[5] "FTC Finalizes Order with Flo Health, a Fertility-Tracking App that Shared Sensitive Health Data with Facebook, Google, and Others", June 2021, US Federal Trade Commission, accessible at https://www.ftc.gov/news-events/press-releases/2021/06/ftc-finalizes-order-flo-health-fertility-tracking-app-shared/

Stewards enable sharing mechanisms which unlock and generate value whilst balancing individual data rights.[6] And in an area where huge amounts of sensitive personal data flow, health data stewards play a crucial role in fostering trust in processes of data use and analysis which otherwise exclude the key stakeholders i.e., individuals and communities.

And, as the bottom-up study delineates, engagement with stakeholders – through civil society organisations and ecosystem enablers – facilitates a participatory approach that enhances collaboration amongst different stakeholders in the healthcare ecosystem.[7] However, actualising this requires strengthening at an infrastructural and design level. This top-down approach necessitates legislative and policy interventions that define, codify, and incentivise data sharing. Internationally, states have adopted various measures to institutionalise and develop a data sharing ecosystem. For instance, the European Union (EU) has taken steps to build a European Health Data Space as part of the European Commission's 2019-24 priority vision – "a Europe fit for the digital age". By strengthening data quality and improving the interoperability of health data across various electronic records and health appliances, the EU seeks to derive value from health data by enhancing sharing capabilities.[8] Similarly, the Act on the Secondary Use of Health and Social Data, passed by the Finnish government in 2019, facilitates the secondary use of personal health and social data for "steering, supervision, research, statistics and development in the health and social sector".[9]

This study's comparative approach, therefore, seeks to analyse top-down policy interventions of four jurisdictions – at varying levels – to derive effective pathways for instantiating health data stewardship. The study also identifies potential gaps – legislative and regulatory – that need addressing for robust accountable systems to be created. For instance, data protection legislation that does not identify the right to portability and data re-use could create hurdles for allowing sharing of data for public good. Therefore, having legislative and policy interventions that draw the contours of what is permissible is not enough. It also places countries such as India, Sudan, and Egypt, which are yet to pass data protection legislation, in a unique position to imagine such legislation and policies as interventions that can enhance equitable redistribution of the value of data while also championing the individual's privacy and trust.[10]

---

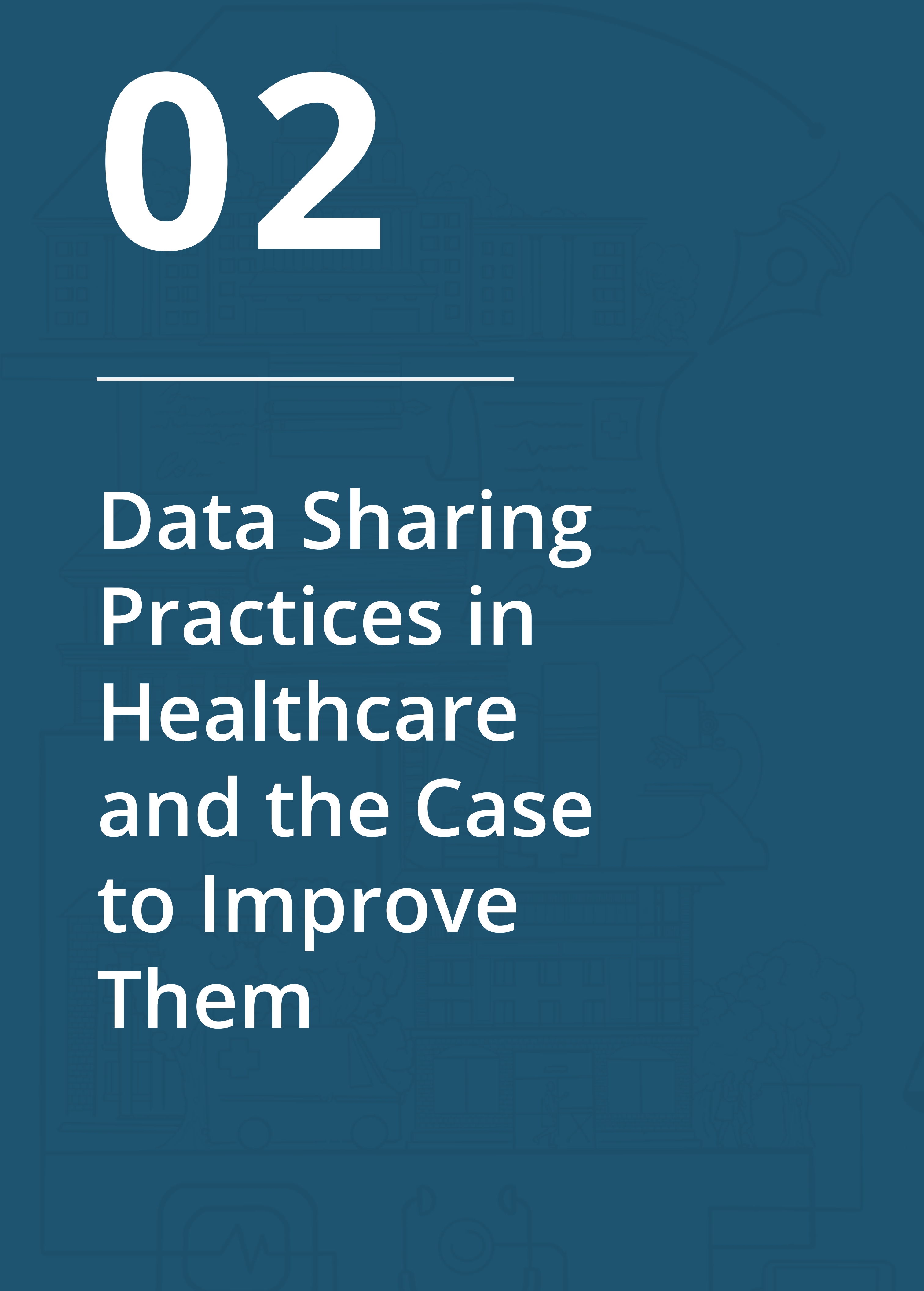
[6] For more information on the different models of stewardship, please refer to 'Data Stewardship: A Taxonomy', Aapti Institute, available at https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/

[7] For more information on how bottom-up stakeholder engagement can instantiate health data stewardship, see the allied report in this series, "Health data stewardship: Bottom-up stakeholder engagement".

[8] Press Release - European Health Union: Commission publishes open public consultation on the European Health Data Space, May 2021, accessible at https://ec.europa.eu/commission/presscorner/detail/en/ip_21_2083/

[9] "Secondary use of health and social data", Government of Finland, 2019, accessible at https://stm.fi/en/secondary-use-of-health-and-social-data/

[10] Data Protection and Privacy Legislation Worldwide", United Nations, accessible at https://unctad.org/page/data-protection-and-privacy-legislation-worldwide/



02

Data Sharing Practices in Healthcare and the Case to Improve Them

When speaking of health data, it is important to consider that one refers to not just the narrow category of electronic health records collected by clinics and hospitals about individual patients. Health data may include, at the individual level, other details about patients, their activities, and data collected from their personal devices. At a collective level, it may consist of derived data and inferred data processed by organisations with access to large quantities of health information.[11]

Data sharing remains a challenge in many contexts due to differing levels of accessibility in digital health infrastructure.[12] This is exacerbated by the lack of capacity or of a common ontology to make shared health data easily usable. The limit on usability of shared data discourages investment in creating sharing platforms or indeed even to share data at all.[13] There is also lack of awareness and capacity that limit the possibilities of leveraging data with consent. The absence of any collective engagement mechanism renders the process the atomised one we see today, based on ill-negotiated terms of consent that are often restrictive and as ambiguous as the entitlements given to end users.

These conditions will persist given the lack of a negotiation platform for users.[14] This is a gap in the regulatory tools available to enhance data sharing, and requires institutional support in order to materialise and remain effective. Part of meaningful consent is providing the opportunity to meaningfully choose between alternatives, and users are empowered to do this when operating as a collective. This is also exhibited in the activities of Ciitizen, which liaises with patient collectives and advocacy groups in order to focus on making data available for research on specific medical conditions.[15]

Setting up a responsible sharing framework can help increase transparency and accountability regarding data use and access, and increase trust in the sharing process.

---

[11] "Electronic Health Records: Then, Now, and in the Future", R.S. Evans, Yearbook of Med Informatics, 2016, accessible at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5171496/

[12] "Bridging the digital divide in health care", Anita Makri, The Lancet, 2019, accessible at https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30111-6/fulltext/

[13] "Why digital medicine depends on interoperability", Moritz Lehne & Ors., NPJ Digital Medicine, 2019.

[14] "Collective bargaining on digital platforms and data stewardship", Astha Kapoor, Friedrich Ebert Foundation research paper series, 2021, accessible at http://library.fes.de/pdf-files/bueros/singapur/17381.pdf

[15] See Ciitizen, accessible at https://www.ciitizen.com/

2.1

Need for new frameworks

There is a need for new data governance frameworks, to explore the economic structure of data markets and inherent power flows to redistribute data-derived powers. Non-economic harms of the current paradigm of data practices and alternative data governance structures such as stewardship must be examined.

The value proposition of data sharing for the state is significant. Collection of data in crucial sectors such as health and finance can help as a regulatory tool in evaluating what measures are needed. Geo-specific data and data on specific communities can help frame policy on improving availability of basic infrastructure and entitlements to underserved groups. It is relevant to point out there are converse risks of community-specific data which may be prone to misuse as tools of discrimination and end up exacerbating the problems sought to be solved.[16]

The availability of data has well-demonstrated value in humanitarian crises, including those of public health, where data collection can be helpful in informing authorities about the nature and spread of disease, and deciding necessary action.[17] The COVID-19 pandemic provided an example of this where data sharing was instrumental in keeping healthcare workers and authorities updated with crucial information. In the aftermath of this ad hoc progress made in understanding the value of data sharing for health, there have been greater calls for a more standardised set of principles in order to make useful data available for public health emergencies.[18]

New standards and structures around data sharing can help data subjects exercise more control over data. They can help increase transparency and accountability with regard to data use and access, and hike trust in the sharing process. Access to datasets in non-shareable domains and formats can be increased, at reduced cost and heightened efficiency; these measures in turn will improve trust in the data sharing process.

---

[16] "Big Data and discrimination: perils, promises and solutions. A systematic review", Maddalena Favretto and ors., Journal of Big Data, 2019, accessible at https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0177-4/ ; see also "Dissecting racial bias in an algorithm used to manage the health of populations", Ziad Obermeyer and ors., Science, October 2019, accessible at https://science.sciencemag.org/content/366/6464/447/.

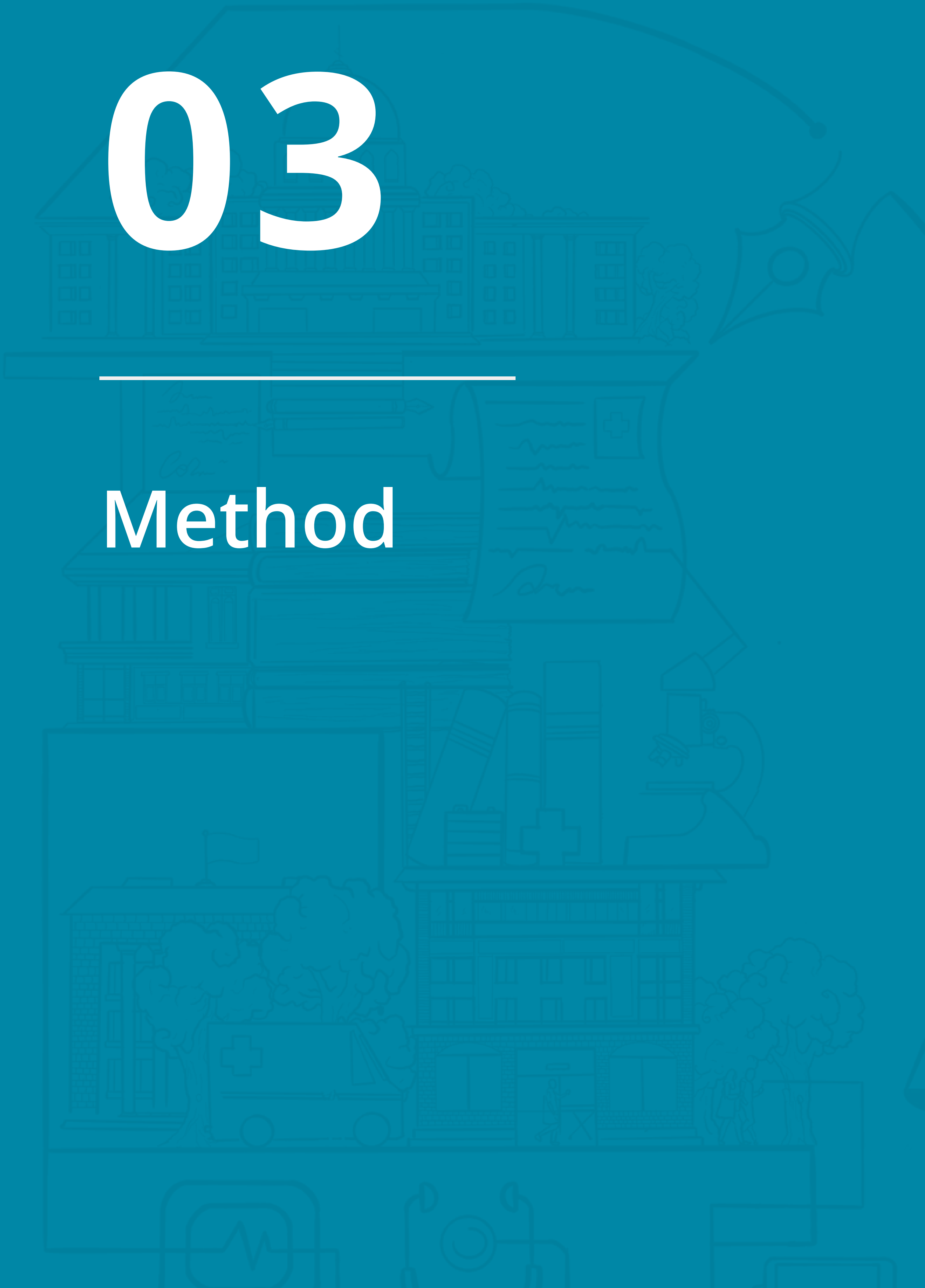
[17] "Make Data Sharing Routine to Prepare for Public Health Emergencies", Jean Paul Chretien and ors., PLoS Medicine, August 2016; "Data Sharing in a Time of Pandemic", Sarah Callaghan, Patterns, August 2020, accessible at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7427527/.

[18] "Transmission dynamics: Data sharing in the COVID-19 era", Randi E. Foraker and ors., Learning Health Systems, 2021, accessible at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7323052/.

The advantages that maximise the public value of data primarily include data rights of principals. Previously inaccessible data is safely made available in shareable formats, increasing collaboration on public questions and the ability to harness value from data.

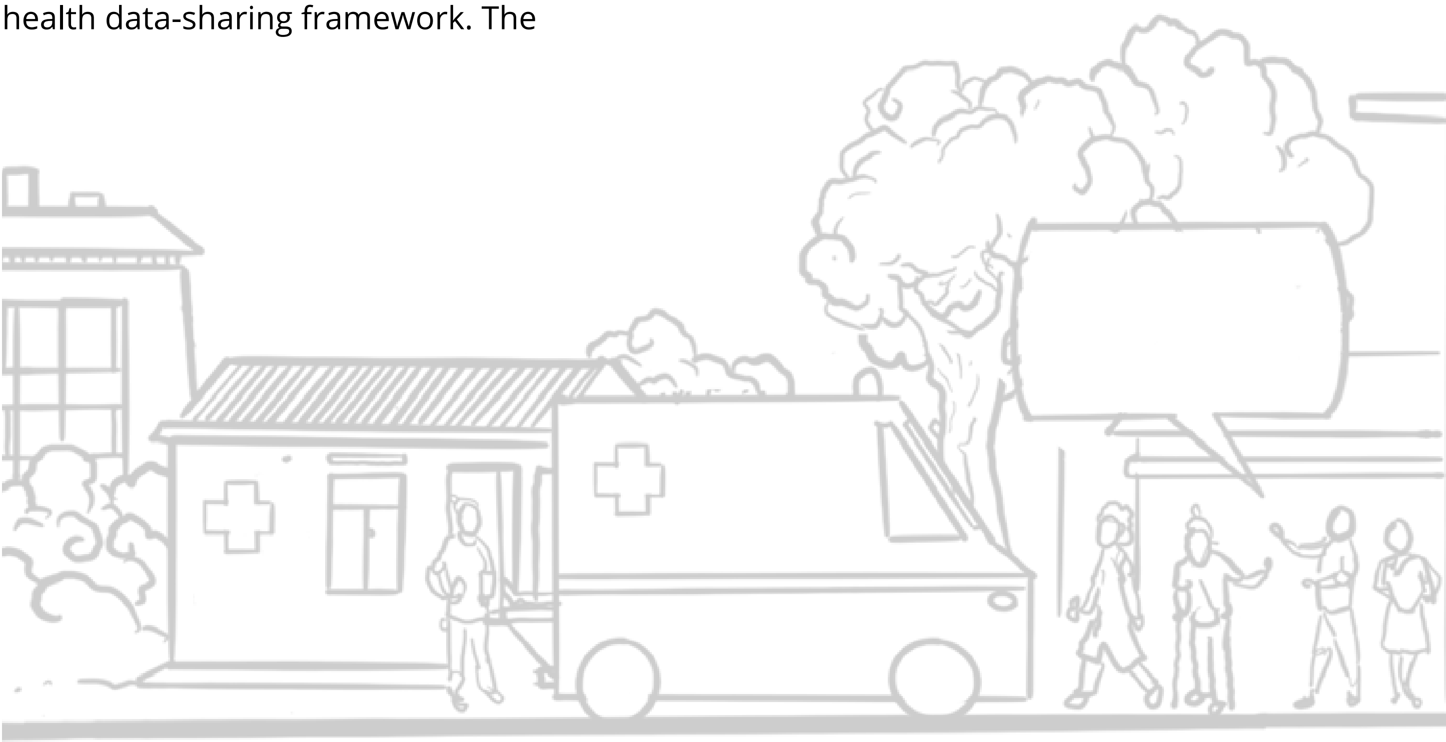
03

Method



Any meaningful realisation of health stewardship models requires an exploration of two contrasting yet complementary approaches[19]; a bottom-up approach[20], centred around data generators with a network of civil society organisations that facilitate the instantiation of these systems, and a top-down approach focused on the state's legislative and regulatory functions that define the contours of such systems. This paper seeks to explicate the latter by studying the legislative, policy, and regulatory frameworks for health data sharing in Estonia, Finland, India, and Ontario (Canada) – jurisdictions in varying stages of stewarding health data.

The legislative and policy exploration builds on the primary research focused on understanding the current health data-sharing framework. The primary research entailed in-depth interviews with various data stewards and ecosystem enablers across the world. Notes from these interviews were categorised against predefined codes that identified models, decision-making processes, and data storage/sharing mechanisms.[21] Insights from desk research and interviews with civil society organisations and experts in various jurisdictions showed health data sharing trends in different regions. Although the study began by exploring policy frameworks on health data sharing in more than 10 jurisdictions, it narrowed to Estonia, Finland, India, and Ontario for the final comparative analysis. Figure 1 maps the initial stage of research, highlighting legislation, policies, and strategy documents relevant to effectuate top-down data sharing.



---

[19] "The Nature of Policy Change and Implementation: A Review of Different Theoretical Approaches", Organisation for Economic Co-operation and Development, 2013, https://www.oecd.org/education/ceri/The%20Nature%20of%20Policy%20Change%20and%20Implementation.pdf.

[20] For more detailed insights on this approach, please refer to the accompanying report on "Health Data Stewardship: Bottom-up stakeholder engagement".

[21] To better understand data governance at an organisational level please refer to "Health Data Stewardship: Bottom-up stakeholder engagement".

REGION	REGULATION / DOCUMENT	STATUS
European Union	• General Data Protection Regulation • Digital Services Act • Data Governance Act • Regulation on Free Flow Data • European Health Data Space • European Strategy on Data	• *In force* • *Legislative Proposal* • *Legislative Proposal* • *In Force* • *Strategy Document / Framework* • *Strategy Document / Framework*
Finland	• Law on Secondary Use of Health and Social Data • Agreement on National Archives for Health Information (Kanta) • Electronic Prescriptions Act	• *In force* • *In force* • *In force*
Estonia	• Public Information Act (Information systems data exchange layer) • National Health Information System Regulation • Estonian Information System Authority	• *In force* • *In force* • *In force*
Ontario, Canada	• Personal Information Protection and Electronic Documents Act • Personal Health Information Protection Act • Digital Privacy Act (amendment to PIPEDA) • Ontario's Patient First Digital Health Strategy • Digital Health Playbook • Digital Health Information Exchange Policy • Data Trusts Consultation Paper	• *In force* • *In force* • *In Force* • *Strategy Document / Framework* • *Strategy Document / Framework* • *Policy Proposal* • *Strategy Document / Framework*
India	• The Personal Data Protection Bill • Report by the Expert Committee on Non-Personal Data Governance Framework • National Health Policy • National Digital Health Mission • Data Empowerment and Protection Architecture	• *Legislative Proposal* • *Expert Group Report / Framework* • *Strategy Document / Framework* • *In force* • *Policy Proposal*

Figure 1: *Snapshot of the legislative, regulatory, and policy developments on data rights and data sharing in the four jurisdictions. The chart also includes the EU to contextualise strategic decisions on data sharing at a broader continental level.*

The methodology relies on first principles, legislation, enabling tools, and regulatory functions as key markers for analyses of policy pathways (see Section 5). These are identified across six attributes, namely, community centredness, conditions for access, data protection, decision-making bodies, standards for interoperability, and regulatory sandbox. The attributes help contextualise the approaches taken and function as a basis for comparing the comprehensiveness of policies in each of these jurisdictions. Similarly, to assess how these policies effectuate, it was deemed necessary to compare and evaluate how these frameworks involved community organisations. The findings from this analysis have been distilled into a high-level checklist (see Section 6) which provides a step-by-step process for countries to instantiate stewardship through policy mechanisms.

04

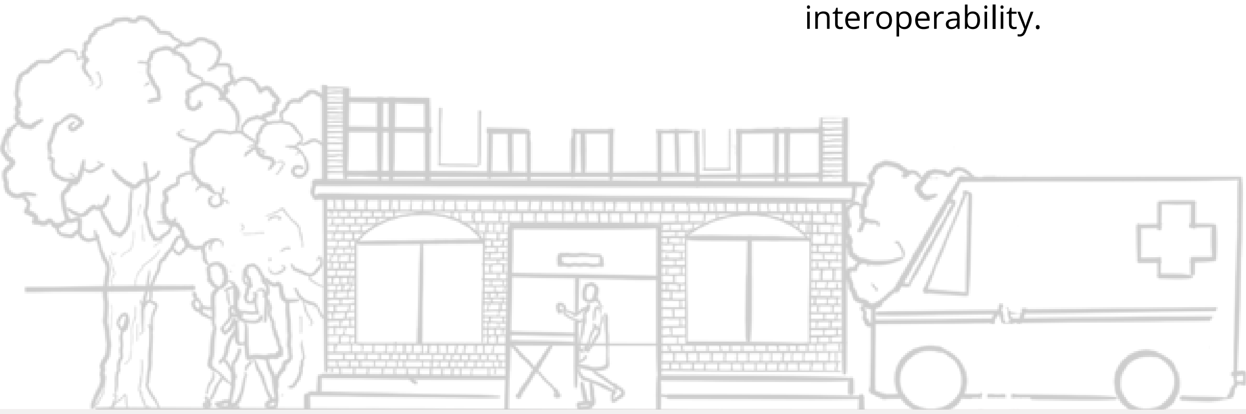
Data Stewardship in Healthcare

4.1

Stakeholders in healthcare

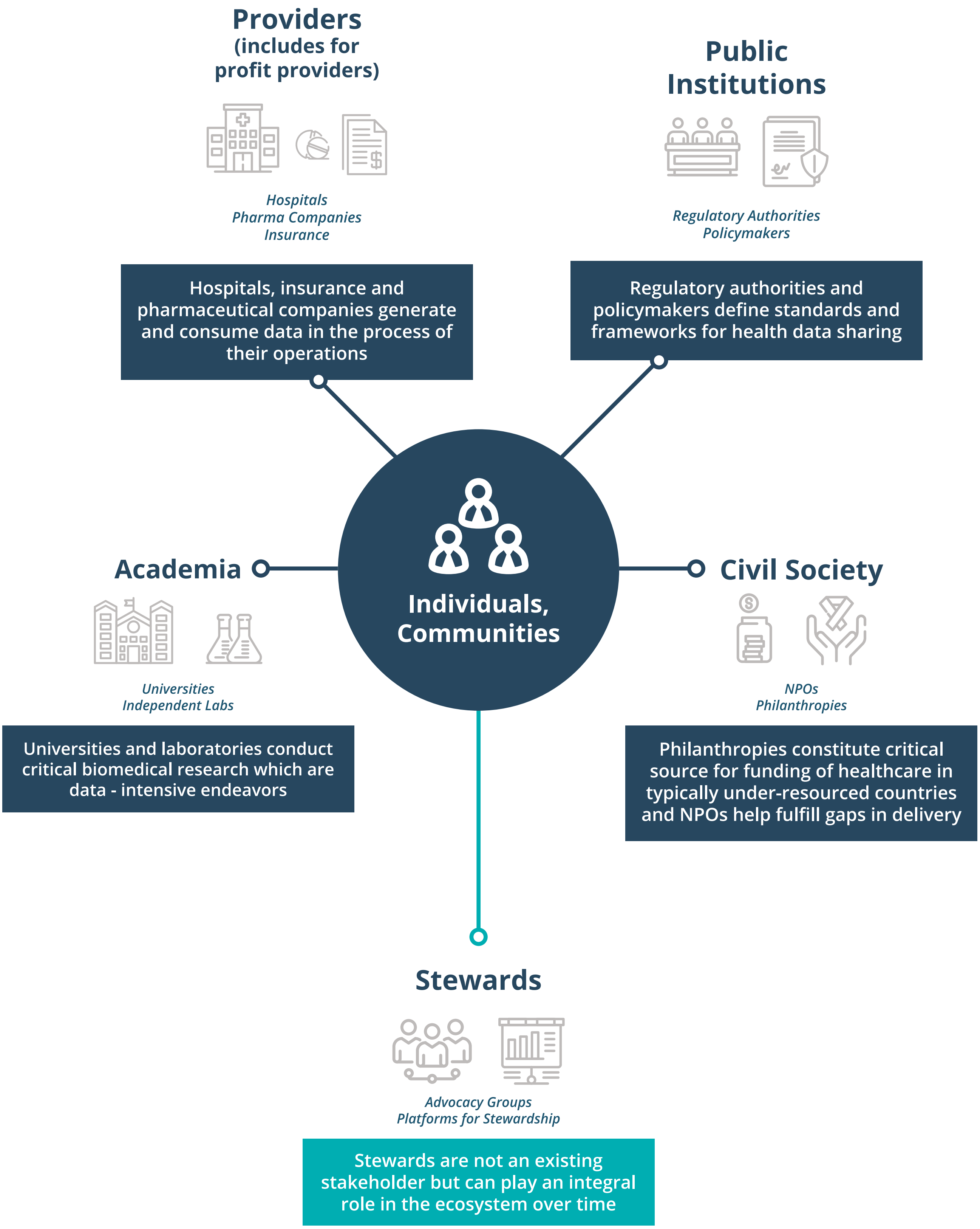
The study identified five sets of stakeholders in the healthcare ecosystem, namely, individuals and communities, providers, civil society, academia, and public institutions. Central to this ecosystem are individuals and communities who, as patients and caregivers, perform the role of primary data generators. They are catered to by service providers, both for-profit and not-for-profit organisations, such as hospitals, insurance companies, and pharmaceutical companies that collect and process the data generated during the course of their operations. Similarly, philanthropic and civil society organisations provide much-needed funding and redress gaps in healthcare delivery, thereby assuming proximity to the data generators. The secondary use of the data generated from both these functions/sources is immensely valuable for academia (universities and laboratories) in advancing health research.

In this complex ecosystem, the introduction of stewards to perform the role of intermediaries in negotiating and navigating health data flows can act as a catalyst to unlock the social value of health data. The bottom-up report[22] delineates the benefits for the stakeholders in the ecosystem from instantiating of stewardship. Maximising gains of bottom-up stakeholder engagement, however, requires technical and regulatory interventions by public institutions. Policies that understand the harms of the current data flow practices and recognise the ability of an intermediary to address them will go a long way in instantiating sustainable health data stewards who safeguard individuals and communities from threats whilst also enhancing data sharing capabilities. Responsive regulators can meaningfully intervene to make data sharing conducive by setting common standards or frameworks which enable interoperability.



---

[22] For a better understanding of bottom-up approaches to stewardship in healthcare, please refer to the corollary of this report, “Health data stewardship: Bottom-up stakeholder engagement”



Providers
(includes for profit providers)

Hospitals
Pharma Companies
Insurance

Hospitals, insurance and pharmaceutical companies generate and consume data in the process of their operations

Public Institutions

Regulatory Authorities
Policymakers

Regulatory authorities and policymakers define standards and frameworks for health data sharing

Individuals, Communities

Academia

Universities
Independent Labs

Universities and laboratories conduct critical biomedical research which are data - intensive endeavors

Civil Society

NPOs
Philanthropies

Philanthropies constitute critical source for funding of healthcare in typically under-resourced countries and NPOs help fulfill gaps in delivery

Stewards

Advocacy Groups
Platforms for Stewardship

Stewards are not an existing stakeholder but can play an integral role in the ecosystem over time

Figure 2: *The healthcare ecosystem consists of five sets of stakeholders. The sixth category, stewards, though aspirational, performs the crucial role of intermediaries who structure and negotiate data flow among the other stakeholders.*

4.2

Value created by health data stewardship

The value created by stewardship stems from solving the problems that have been outlined in this paper. This brings value propositions to stakeholders across the spectrum in the health sector. Interviews with organisations working on beneficial sharing of health data highlighted explicit acknowledgement of the value proffered by a common and collaborative environment for exchanging usable data. The challenge of usability requires coordination – this is enabled by the platform of stewardship.

The value of data sharing in health is also of note for users and government bodies. Users are empowered in the market of health services in terms of how their data is managed and handled, when they are able to collectively decide on making their data available for beneficial purposes such as research on relevant diseases. Similarly, governments are able to harness the availability of data to better deal with public health emergencies and target availability of resources for emergency situations.[23]

Responsible stewardship of data is a framing that is congruent with the objectives of a framework for beneficial data sharing. Stewardship functions enable representation of community interests and collectivised negotiation with third parties. Stewardship also offers possibility for a decentralised system of data governance, where decision-making is delegated and decentralised, thereby reducing the regulatory burden on central bodies.

Value created by stewardship in turn can be categorised according to the ends served by it. These forms of value are harnessed by stakeholders in different capacities, discussed below. Broadly, stewardship creates three categories of value:

---

[23] "Make Data Sharing Routine to Prepare for Public Health Emergencies", Jean Paul Chretien and ors., PLoS Medicine, August 2016

I. Core Values

The core value of stewardship is to counteract the imbalances in relationships in the current model of health data sharing. This value operates at the level of the ecosystem, and benefits stakeholders across the value chain. It also helps fulfil state objectives of making innovation easier and empowering citizens regarding how their data is shared.

- **User agency:** Defined as the ability to exercise greater agency on data use, and protection. One of the primary objectives of stewardship is to empower the central stakeholders in personal data and community data. This enables the basic tenet of giving users decision-making power over their data. Initiatives such as Digi.me have a direct impact on user awareness and agency by providing the necessary tools for users to be informed on how their data is used and to change their preferences.[24]

- **Transparency and accountability:** Refers to clarity on how data is being collected and used, and ability to hold data users accountable regarding purpose of data use. Enabling transparency can assume a number of forms and is primarily about practice and procedure. A steward for health data may be set up by a government body or a not-for-profit, or legal obligations may be placed directly on healthcare providers and regulatory authorities. The merit of the system boils down to the specific instance: the presence of an expensively-assembled ethical review board, for example, would still be useless if its membership was compromised through external influence or conflicts of interest. While this is shaped by the rules set by authorities, the internal practices of organisations become relevant because of the nature of innovation in data practices – rules will inevitably not account for all harms. Thus, open access to the process for impacted stakeholders is crucial for accountability.

Similarly, simple measures such as availability of a process to revoke or modify consent through email may be a basic but effective measure in respecting and enforcing user preferences. Accessibility and ease of procedure count for a great deal when the objective is one of giving users the tools to engage with the process of data governance. Organisations such as LunaDNA[25] and Sage Bionetworks have set up processes that help researchers by making data available, but also have effective systems of informed consent, where users are made aware of the ramifications of

---

[24] "Securing Personal Data: Explaining modern techniques in commercial data privacy platforms", Gavin Ray, digi.me, accessible at https://digi.me/downloads/Executive_Briefing_-_Securing_Personal_Data_v2_18052021.pdf

[25] "Your Health Data, Our Principles: How To Feel Confident Sharing Your Health Data For Research", LunaDNA, accessible at "https://www.lunadna.com/lunadna-values-scorecard/ ; "LunaDNA review – Can you get rewarded for your DNA data?", Nebula Genomics, January 2021, accessible at https://nebula.org/blog/lunadna-review/

their choices – and their decisions are specific, not blanket.[26] Thus, users need not worry about their consent being assumed. This engenders a higher degree of trust in the organisation.

- **Instrumental value:** Relates to monetary and other value (ease of access, enhanced quality of goods/services) received by generators for use of their data. The broader outcomes of sharing health data include better quality in drug research and patient care measures. This is especially true when data is made available to projects focused on collating it for research on specific diseases. Initiatives such as Ciitizen[27] and Variant Bio[28] work on such initiatives of curated research and supporting community engagement, providing examples that illustrate the varied scale of such projects currently being implemented in the sector.

II. Sector-Specific Value

- **Expansion of health services:** Refers to collection and usage of datasets for informing healthcare operations and policy, to improve quality, affordability and delivery. Data collection and its output can fuel potential expansion of services if used well – the insights and analytics functions that would otherwise be out of reach for private and public healthcare providers can be made accessible through responsible stewardship in health, thereby empowering these service providers.[29]

- **Academic/research:** Represents research and development to fuel innovation for drug and treatment availability, quality and efficiency. The benefits of such value are long-term and relate to the primary value for patient groups and pharmaceutical companies. While regulation of this value is important, there is no doubt that increased and secure sharing of health data enables greater research, as there is exceeding demand from researchers for relevant and usable data.[30]

- **Access to healthcare:** Conceived as better access to existing services and aiding innovation for developing fresh ones. Relevant datasets can inform regulators and municipal authorities on the gaps in healthcare coverage, and areas and communities requiring more resources to be targeted for their well-being.[31]

---

[26] "The Elements of Informed Consent: A toolkit", Sage Bionetworks, July 2019, accessible at https://sagebionetworks.org/wp-content/uploads/2019/07/SageBio_EIC-Toolkit_V2_17July19_final.pdf

[27] See Ciitizen, accessible at https://www.ciitizen.com/

[28] "Ethics and Community Engagement at Variant Bio", Variant Bio, accessible at https://www.variantbio.com/pdfs/vb_ethics_community_engagement.pdf

[29] "Sharing Is Caring—Data Sharing Initiatives in Healthcare", Tim Hulsen, International journal of environmental research and public health, Volume 17(9), April 2020, accessible at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246891/

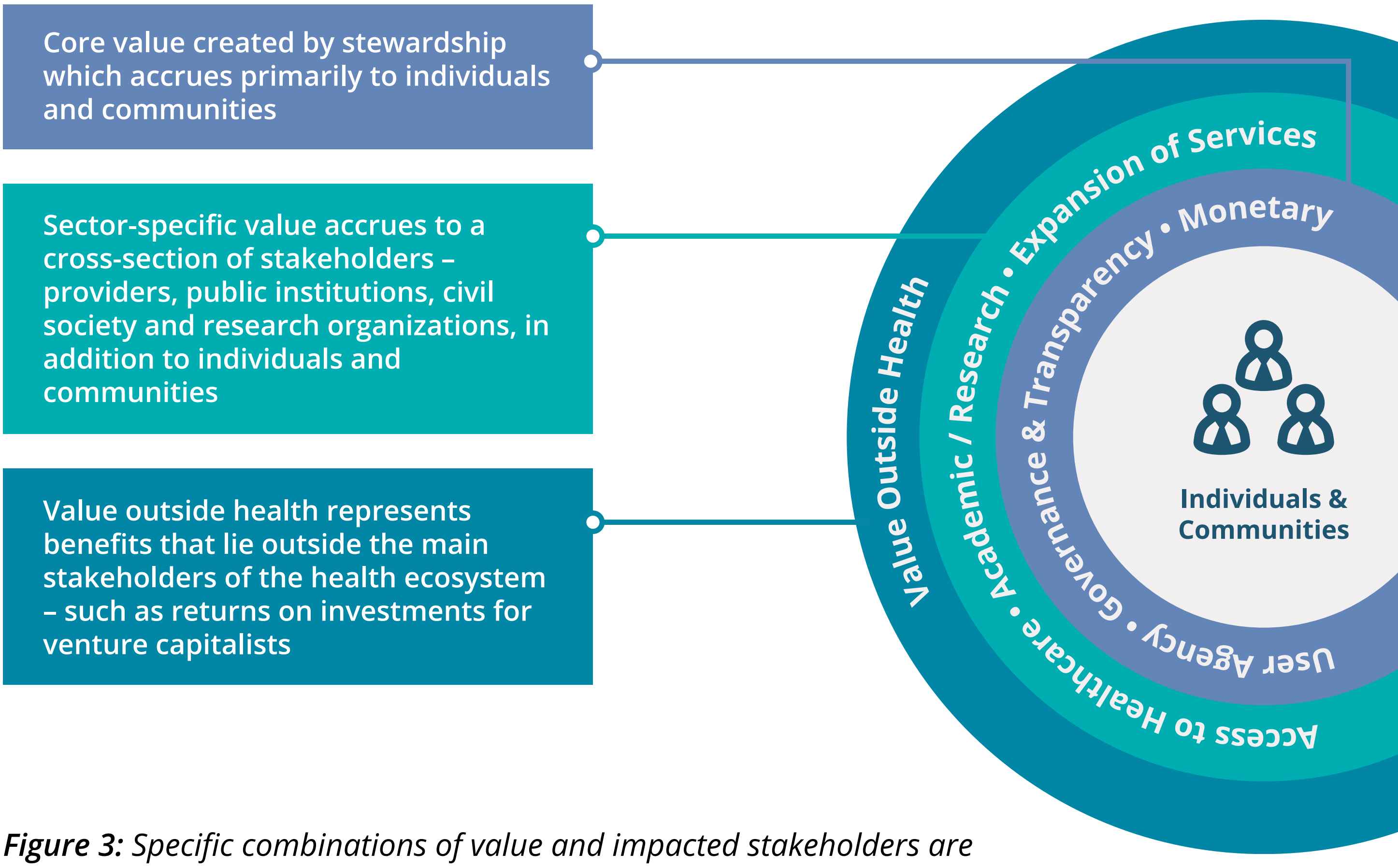
[30] "Big data analytics in healthcare: promise and potential", Wullianallur Raghupathi and Viju Raghupathi, Health information science and systems, 2014, accessible at: https://doi.org/10.1186/2047-2501-2-3/.

[31] "The big-data revolution in US health care: Accelerating value and innovation", McKinsey, 2013, https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care/

III. External Value, Outside Health Sector[32]

Value generated outside the health sector represents benefits such as revenue generated for investors in healthcare, with value accruing outside the ecosystem of stakeholders in healthcare.

Value created by health data that accrues outside the health sector has not been covered as part of this report. While such value doubtless exists, the incentives to create such value does not benefit the healthcare ecosystem, and incentivises actions that harm users groups. This framing finds itself in conflict with the interests of healthcare outcomes.[33]



Core value created by stewardship which accrues primarily to individuals and communities

Sector-specific value accrues to a cross-section of stakeholders – providers, public institutions, civil society and research organizations, in addition to individuals and communities

Value outside health represents benefits that lie outside the main stakeholders of the health ecosystem – such as returns on investments for venture capitalists

Figure 3: *Specific combinations of value and impacted stakeholders are better served by appropriate models of stewardship.*

---

[32] Note: External value is excluded from the analysis contained in this study as its benefits are realised outside the health sector, creating gains for investors and other actors not directly involved in the health data value chain

[33] "GPs urged to refuse to hand over patient details to NHS Digital", The Guardian, June 2021, accessible at https://www.theguardian.com/society/2021/jun/01/gps-urged-to-refuse-to-hand-over-patient-details-to-nhs-digital/ ; "GPs warn over plans to share patient data with third parties in England", The Guardian, May 2021, accessible at https://www.theguardian.com/society/2021/may/30/gps-warn-plans-share-patient-data-third-parties-england/

05

Policy Pathways to Stewardship

> The global regulatory environment is growing increasingly alive to the perils of an asymmetrical data economy in which power to control data is vested in a handful of corporations, rendering individuals and communities mere cogs in data machinery driven by monopolistic impulses and manipulation of users.[34]

This highlights the urgency for governance mechanisms that can check the power of corporations and redistribute value more broadly among stakeholders in the health data economy. The first and most comprehensive expression of this sentiment is the EU's General Data Protection Regulation (GDPR), 2016[35] which sought to impose controls on data processing, rooted in principles of individual harm, rights and privacy. This has spurred a flurry of personal data protection regulations that have been introduced or implemented in various jurisdictions outside Europe, such as India[36], Kenya[37] and Brazil[38], among others.

However, the GDPR and other legislation it has spawned suffer from a blinkered approach, the "privacy model" of data governance, which considers consent provisioning the only requirement for further processing of information.[39] It fails to compensate data generators – individuals and communities – for the value ensuing from downstream use of their data. Moreover, this model is primarily concerned with harms to individual privacy from data sharing, ignoring perspectives of communal harm.[40]

Alternatively, this study's proposed conception of stewardship is founded on the premise of the "accountability model" of data governance. Data is held in trust and stewarded by an intermediary who is bound by duties of care and loyalty towards the individuals and communities it represents. Consequently, under stewardship, the intermediary must act in the best interests of its member community and lends itself as well as controllers to being held accountable by the community. Similarly, member communities can stake claim to value derived from their data through an intermediary such as the data steward. As a result, stewardship enables individuals and communities

---

[34] "Step Up Or Break Up: The Challenge For Big Tech", Wal van Lierop, Forbes, October 2020, accessible at https://www.forbes.com/sites/walvanlierop/2020/10/09/step-up-or-break-up-the-challenge-for-big-tech/.
[35] Regulation (EU) 2016/679 - General Data Protection Regulation
[36] The Personal Data Protection Bill, 2019
[37] The Data Protection Act, 2019
[38] Lei Geral de Proteção de Dados, 2019
[39] "Rethinking personal data regulation in India", Amar Patnaik, The Indian Express, February 2021, accessible at https://www.newindianexpress.com/opinions/2021/feb/15/rethinking-personal-data-regulation-in-india-2264123.html/.
[40] "Six ways (and counting) how big data systems are harming society", Joanna Redden, The Conversation December 2017, accessible at https://theconversation.com/six-ways-and-counting-that-big-data-systems-are-harming-society-88660/.

to retain agency over their data, but also distribute value for public benefit purposes as determined by the communities.

To examine how the framework of health data stewardship is embedded within various legal regimes, the study investigates top-down approaches to data governance in a select cohort of countries – Estonia, Finland, India and the province of Ontario in Canada. The top-down approaches include enacted or proposed legislation, soft policy directives, and strategic and discussion papers released by state authorities within each country. The following is a jurisdiction-wise timeline of the policy pathways to stewardship and the broad timeline of legislating on data regulation:

5.1

Estonia

The country's X-Road data exchange layer is considered a technological revolution that not only ushered Estonia into the 20th century, but also laid the foundation of a new 'digital republic'.[41] This framework provides a secure technical and organisational infrastructure for internet-based data exchange such that 99% of all Estonian government services are made accessible online.[42] It follows a decentralised architecture embedded within an interoperable ecosystem in which citizens can access copies of their health and other information authorised through national e-IDs and digital signatures.[43] In such a context, the state functions as a steward of citizen data that is permitted for further use by dynamic consent controls accorded to citizens themselves. Figure 4 provides a brief timeline of the policy pathway to X-Road's implementation in Estonia.

---

[41] "Estonia is a 'digital republic' – what that means and why it may be everyone's future", Imtiaz Khan & Ali Shahaab, The Conversation, October 2020 accessible at https://theconversation.com/estonia-is-a-digital-republic-what-that-means-and-why-it-may-be-everyones-future-145485/.
[42] See https://e-estonia.com/.
[43] "Estonia – the Digital Republic Secured by Blockchain", PricewaterhouseCoopers, 2019, accessible at https://www.pwc.com/gx/en/services/legal/tech/assets/estonia-the-digital-republic-secured-by-blockchain.pdf/.

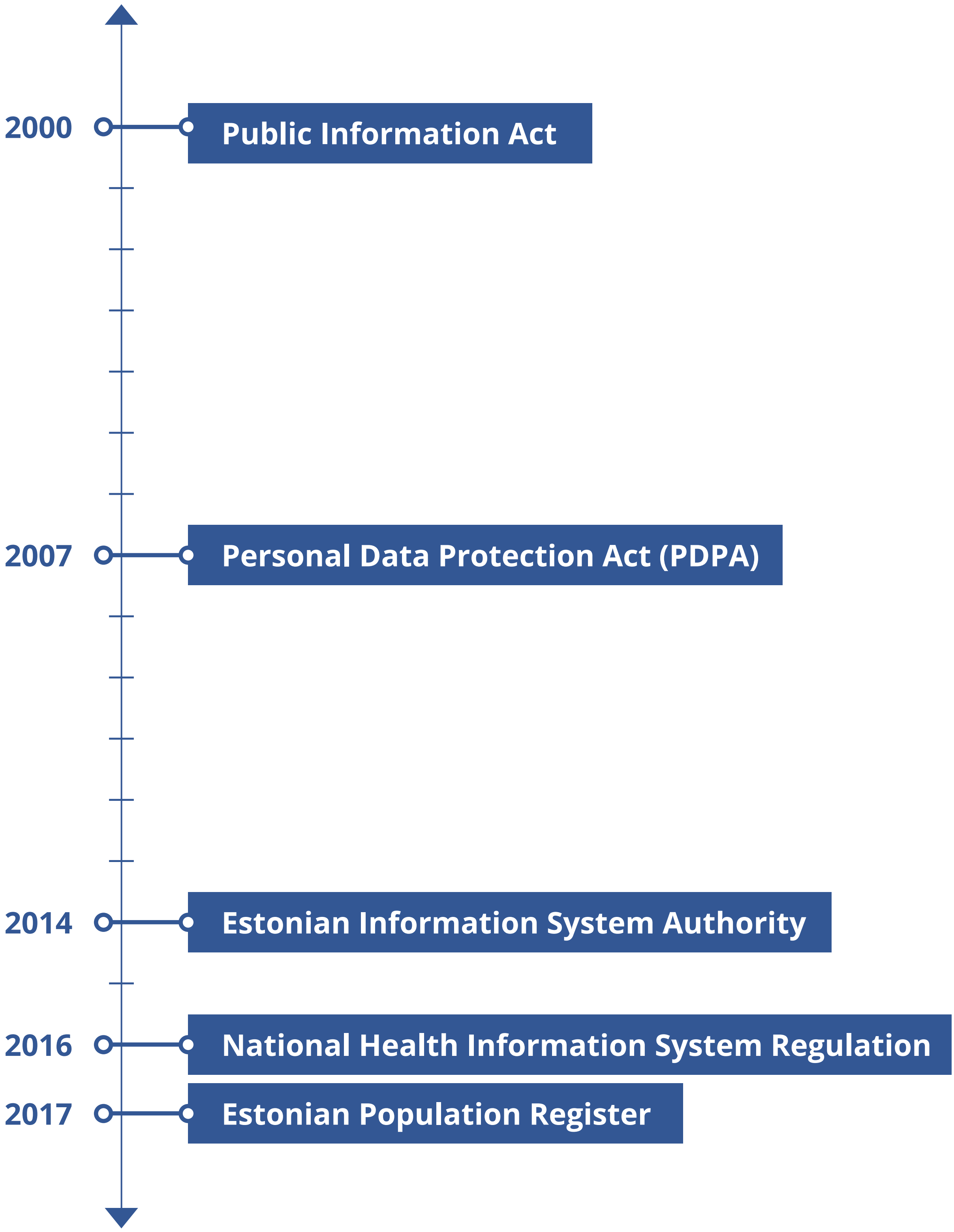


Figure 4: *Estonia has developed a public data exchange layer – the X-Road framework – which forms the foundation of a new digital republic in which the state acts as the steward of citizen data. It supplies the technical infrastructure to integrate ecosystem-wide data flow.*

5.2

Finland

Finland's long-standing history of data sharing for public benefit has its roots in digitisation pilots that have been underway for over two decades.[44] The Kanta platform provides a set of digital services for use in healthcare and social welfare, affording greater continuity of care to citizens as well as ease of access to timely information for providers, both public and private.[45] Consequently, citizens benefit from having full access to their data, which is linked to a variety of services – from e-prescriptions registries to pharmaceutical databases and patient data repositories, among others.

The abundance of data combined with political will to create evidence-based policies culminated in the enactment of the Act on the Secondary Use of Health and Social Data in 2019.[46] The Act establishes a new data permit authority – Findata[47] – bound by purpose limitations for downstream use of anonymised well-being data for research, policymaking and development interventions, in accordance with GDPR stipulations. Health and social data are collated from siloed sources across different agencies using the Kanta platform and authorised for sharing through data permits (to gain access to granular datasets) or data requests (to obtain statistical-level information and other insights from data).

---

[44] "The ePrescriptions System in Finland: A case study", Silviu Dovancescu and ors., RWTH Aachen University, accessible at https://www.wi.rwth-aachen.de/wi/theses/ePrescription.pdf/.
[45] For more information, see https://www.kanta.fi/en/what-are-kanta-services/.
[46] For more information, see https://findata.fi/en/uutiset/act-on-secondary-use-of-health-and-social-data-will-not-be-applied-to-clinical-trials/.
[47] For more information, see https://findata.fi/en/what-is-findata/.

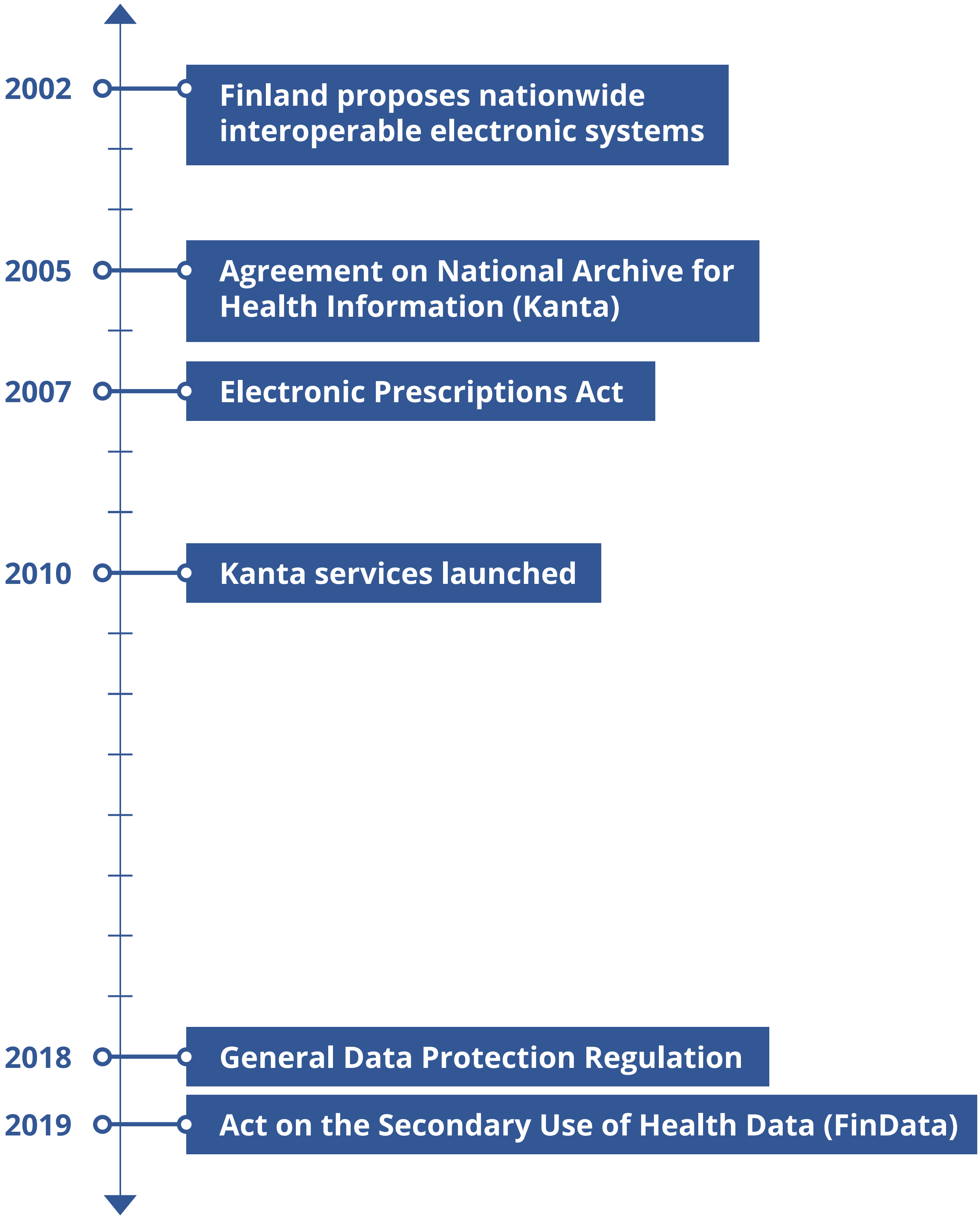


Figure 5: *Finland has unlocked the societal value of data through its integrated health systems developed over 20 years.*

5.3

India[48]

The Indian state's approach to data governance, although encouraging, is fragmented and scattered, in comparison with the pathways adopted by Estonia and Finland. The primary data protection legislation – the Personal Data Protection Bill, 2019[49] (PDP Bill) – is yet to be tabled in Parliament and as a result, citizens are afforded limited security for personally identifiable information.

Further, the health sector-specific guidelines under the National Digital Health Mission[50] (NDHM) provide the basis for India's digital health infrastructure. However, the policy is conspicuously silent on the value of health data and fails to outline procedures for its safe re-use. Without the provisions of the NDHM being harmonised with the PDP Bill, the subsequent rights to protection and portability offered to individual health data remain unrealised.

Lastly, the draft consultation report on governance of non-personal data[51] (NPD Report) recommends the recognition of community rights, but fails to outline comprehensive processes and structures for meaningful realisation. The NPD Report is considered controversial for its mandatory data sharing stipulations required of data-driven entities[52], trumping incentivisation and ecosystem approaches to sharing that are characteristic of global regulations such as the EU's proposed Data Governance Act, 2020.[53] Figure 6 outlines the timeline of legislation on stewardship adopted by Indian regulators.

---

[48] To better understand India's data regulatory ecosystem, please refer to a corollary of this report - "India: Health data stewardship landscape and recommendations"

[49] The Personal Data Protection Bill, 2019, Government of India, accessible at http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf/.

[50] National Digital Health Mission: Health Data Management Policy, Government of India, 2020, accessible at https://ndhm.gov.in/health_management_policy/.

[51] Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020, accessible at https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf

[52] "Tracking India's Approach to Data Governance: From Localization to Stewardship of Data", Jyoti Panday, Internet Governance Project February , 2021, accessible at https://www.internetgovernance.org/2021/02/09/tracking-indias-approach-to-data-governance-from-localization-to-stewardship-of-data/.

[53] Data Governance Act, 2020 (Draft proposal), European Commission, accessible at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0767/.

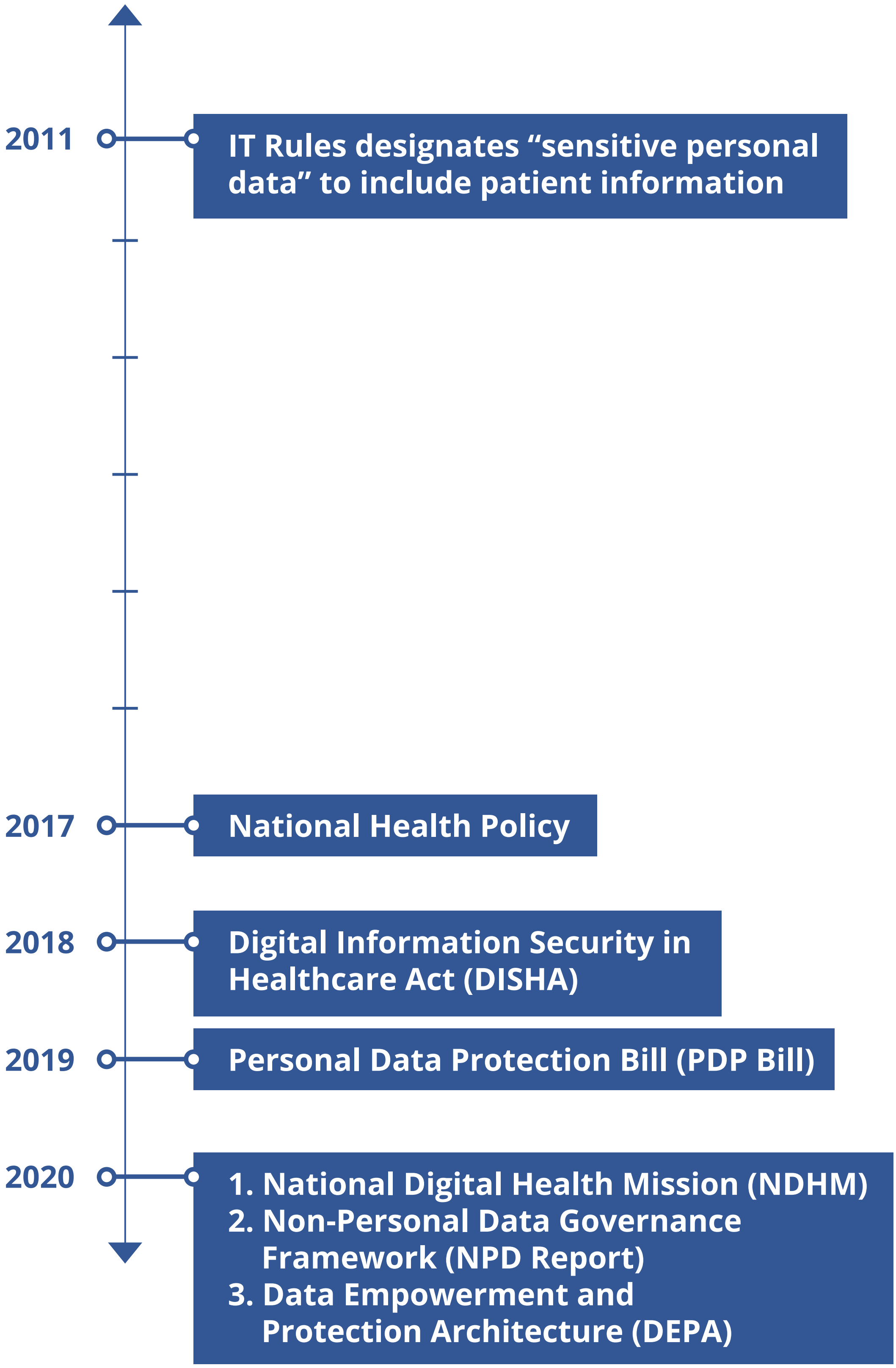


Figure 6: *India’s policy pathway is unique as it is the first jurisdiction to acknowledge community rights over data, as articulated in the NPD Report. This forms the basis for collective governance of data that is crucial for stewardship.*

5.4

Ontario, Canada

Ontario's health data ecosystem is chiefly guided by Canada Health Infoway – an independent, federally-funded not-for-profit organisation tasked with accelerating the adoption of digital health services.[54] Moreover, Ontario's experiments with Sidewalk Labs as a part of an effort to create a futuristic, data-driven smart city in Toronto[55] and limited experiments on health data interoperability (see Digital Health Playbook[56]) have spurred political and public interest in data stewardship.

Additionally, the Government of Ontario has released a draft discussion on the Digital Health Information Exchange Policy[57] that aims to furnish the technical infrastructure necessary to integrate its health systems – a crucial function supported by health data stewardship. Lastly, the thrust for instantiating stewardship is bolstered by the Government of Ontario's public consultation on revamping its data protection legislation – the Personal Information Protection & Electronic Documents Act (PIPEDA), 2000 – to make provisions for data trusts that would enable privacy-preserving and transparent data sharing for public benefit.[58] See Figure 7 for a timeline of Ontario's policy trajectory towards stewardship.

---

[54] For more information, see https://infoway-inforoute.ca/en/.

[55] Sidewalk Labs Toronto was abandoned in May, 2020 due to the economic uncertainty onset by the pandemic. However, it was a first-of-its-kind effort to use data for mobility and brought together public municipal authorities in Toronto alongside citizens and private entities (Alphabet Inc.). For more information about its plans and data use, see https://www.sidewalklabs.com/toronto/.

[56] The Digital Health Collection, Government of Ontario, 2019, accessible at https://health.gov.on.ca/en/pro/programs/connectedcare/oht/docs/dig_health_playbook_en.pdf.

[57] Digital Health Information Exchange Policy, Government of Ontario, 2020, accessible at https://www.ontariocanada.com/registry/showAttachment.do?postingId=32590&attachmentId=43697#:~:text=The%20purpose%20of%20this%20policy,and%20provincial%20digital%20health%20tools/.

[58] "Ontario Launches Consultations to Strengthen Privacy Protections of Personal Data", Ontario Newsroom, Government of Ontario, August 2020, accessible at https://news.ontario.ca/en/release/57985/ontario-launches-consultations-to-strengthen-privacy-protections-of-personal-data/.

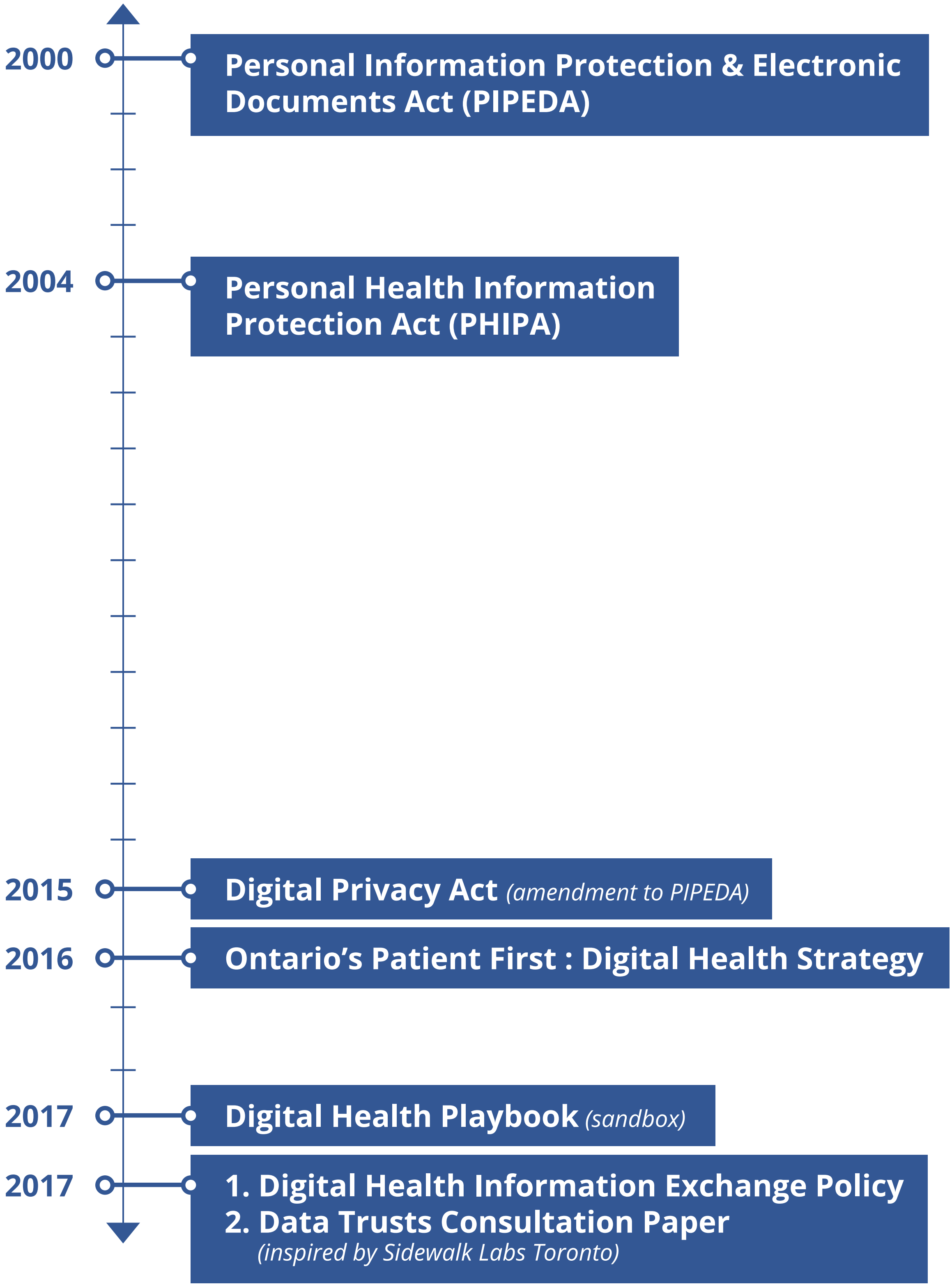


Figure 7: *Ontario’s experience with Sidewalk Labs Toronto and experiments in health data interoperability (Digital Health Playbook) have spurred political and public impetus for stewardship.*

While each of the pathways is at varying levels of implementing avenues for data stewardship, they stand out from other jurisdictions due to their focus on health data sharing. The analysis of policy pathways to stewardship within the aforementioned legal regimes yields indispensable insights on the value of health data as well as the terms, purposes and processes for sharing this information. Further, each pathway is examined along six definitive attributes that constitute essential elements for instituting stewardship in the health sector (see Figure 8).

The analysis of attributes is underpinned by four metrics – first principles, legislation, regulatory functions and regulatory sandbox – which comprise crucial characteristics of policy pathways to stewardship. A first principles approach to data sharing refers to essential, non-negotiable ideals such as community-centredness and delineation of lucid conditions for access to data that should be incorporated into the data regulation landscape. This is imperative to create top-down actions for stewardship that are grounded in recognition of data rights of individuals and communities.

Similarly, legislating on data protection constitutes a significant corollary to the first principles. Any policymaking on data governance must be directed towards protecting interests of communities and preventing adverse harms arising from data sharing. In a similar vein, legislation should recognise the role of individuals and communities as generators of data and confer appropriate rights regarding value derived from downstream use of such data.

Regulatory functions refer to those actions of the state that define and rebalance power relationships between stakeholders in the broader health data ecosystem. This is born of the realisation that data generators are left out of critical data decisions and are seldom compensated for their contribution to the data value chain. Consequently, regulation must lay down processes and structures for the participation of individuals and communities in decision-making, simultaneously recognising the right to portability and re-use of data. Stipulating technical standards within regulations is a vital component of promoting health data interoperability.

The last of the attributes examined – enabling tools – goes a long way in creating an environment that fosters responsible health data stewardship. As an agency-enhancing and empowering paradigm for data exchange, stewardship requires policy tools such as sandboxes to facilitate safe pilot testing and provide empirical inputs into the process of legislation. Such holistic feedback loops between practitioners and lawmakers is critical to build trust in stewardship.

Figure 8 illustrates the six attributes considered in this research, alongside an explanation of what they mean for stewardship with appropriate keys to understand their role within the data regulation ecosystem.

ATTRIBUTES	DESCRIPTION
1. COMMUNITY - CENTEREDNESS	• Acknowledge individuals and communities are primary generators of data • Uphold the rights of communities in data decisions; data sharing is consent-bound, through legislation
2. CONDITIONS FOR ACCESS	• Data access mediated by pre-defined terms about purpose of sharing, role of data requestor • Backed by individual or community consent which is dynamic and revocable
3. DATA PROTECTION	• Formalize mechanisms to ensure protection of individual's and Communities' data; include personal and non-personal data • Stipulate harms arising from data sharing as well as impose penalties and enforce recourse mechanisms
4. DECISION - MAKING BODIES	• Clear regulatory hierarchy; establishment of appropriate adjudicatory authorities – sector-agnostic and sector-specific • Empowered to prescribe remedies for harms arising from data sharing
5. STANDARDS FOR INTER - OPERABILITY	• Core digital infrastructure for sharing across systems and sectors within a jurisdiction's data ecosystem • Standards should be cheap, comprehensive and easily adaptable
6. REGULATORY SANDBOX	• Safe and trustworthy pilot environment for live testing of new products or services • Authorities may or may not grant regulatory relaxations for testing

■ First principles ■ Legislations ■ Regulatory functions ■ Enabling tools

Figure 8: *Policy pathways to stewardship include six core attributes that are implemented by state agencies as part of top-down actions to operationalise data stewardship.*

The four jurisdictions considered – Estonia, Finland, India and Ontario (Canada) – were examined for fulfilment of the six core attributes to determine the comprehensiveness of their policies on health data governance. Based on a comparative analysis of the different jurisdictions, Finland's journey towards health data stewardship emerged as the most advanced due to its long history of data sharing for public good.[59] Moreover, Finland's efforts have been spurred by successive pilots on digital health solutions[60] that helped build a state-sponsored enabling ecosystem for stewardship. On the other hand, India's policy pathways emerged as the least coherent owing to the lack of unified provisions across legislation and a glaring absence of common data ontology. Sector-specific guidelines on health stipulate technical standards for data interoperability that contribute to compliance burden on providers because the Indian government does not cover for the cost of transition to new technologies.[61] Consequently, the scope for operationalising health data stewardship is hampered by a ragged data regulatory ecosystem.

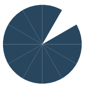
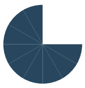
Figure 9 illustrates a broad overview of the four pathways to stewardship and their fulfilment of the six attributes. Since this is an evolving landscape, the analysis is not exhaustive. It serves to surface the commonalities and divergences within different policy pathways to operationalise health data stewardship.

---

[59] "Good practices on B2G data sharing: Finnish forest data ecosystem | Shaping Europe's digital future", News Article, European Commission, March 2021, accessible at https://digital-strategy.ec.europa.eu/en/news/good-practices-b2g-data-sharing-finnish-forest-data-ecosystem/

[60] "The ePrescriptions System in Finland: A case study", Silviu Dovancescu and ors., RWTH Aachen University, accessible at https://www.wi.rwth-aachen.de/wi/theses/ePrescription.pdf

[61] To better understand India's data regulatory ecosystem, please refer to another section of this set of reports - "India: Health data stewardship landscape and recommendations".

	INDIA	ONTARIO (CANADA)	FINLAND	ESTONIA
COMMUNITY-CENTEREDNESS	Communities are primary beneficiaries of value from data	Fails to recognize community rights over data	Community participation lacking; sharing is top-down action	Lacking in community governance; limited benefit from reuse of data
CONDITIONS FOR ACCESS	Defined by PDP Bill, NPD report; broad mandatory sharing requirements	Consent-driven sharing of health data, no guidelines for access to non-personal data	Access determined by purpose; limited access - anonymised data; backed by right to portability	Visibility to data subjects, providers and persons fulfilling legal duties; backed by right to portability
DATA PROTECTION	Draft legislation underway; offer comprehensive protection for users	PHIPA contains clear identification of user rights and harms	GDPR comprises clear identification of user rights and harms	GDPR and PDPA contain clear identification of user rights and harms
DECISION-MAKING BODIES	National Health Authority is sector-specific regulator	No sector-specific authority; decisions taken by privacy ombudsman	Data Permit Authority authorizes licenses for secondary use of health data	Decision-making by umbrella agency – Estonian Information System Authority
STANDARDS FOR INTEROPERABILITY	NDHM stipulates FHIR, SNOWMED and ICD	FHIR is prescribed under Ontario HIE framework	Follow internal ontology based on Kanta guidelines	Follow internal ontology based on X-road framework
REGULATORY SANDBOX	Permitted under NDHM Sandbox Overview	Facilitated through Digital Health Playbook	No sandbox; FinData based on previous pilots	No sandbox
COMPREHENSIVENESS OF POLICIES				

🟦 First principles 🟦 Legislations 🟦 Regulatory functions 🟦 Enabling tools

***Figure 9:** Based on comparative analysis, Finland's journey towards health data stewardship emerges as the most advanced while India's policy landscape is the least coherent of all pathways.*

06

Role of Public Institutions in Operationalising Stewardship

Top-down state actions for stewardship require fulfilment of a set of roles and responsibilities associated with the state. Most importantly, the framework of health data stewardship elucidated within this study places patients, caregivers, advocacy groups and related communities at the heart of the health data ecosystem. Therefore, all state action must acknowledge the primacy of their position as the originators of data and redirect data governance to ultimately benefit communities. Public authorities should recognise the powerful position they enjoy within this ecosystem, defining the principles, standards and frameworks to mediate relationships between individuals and communities on the one hand, and providers, civil society, academia and related health tech entities on the other.

Figure 10 outlines the roles and responsibilities assigned to public institutions for instantiating top-down stewardship in the healthcare sector. Roles refer to the functions to be discharged by the state while the accompanying responsibilities supply minimum standards for implementing a role. The roles are indicated through circles at the perimeter of Figure 10 and the corresponding responsibilities are illustrated within the boxes mentioned alongside each circle.

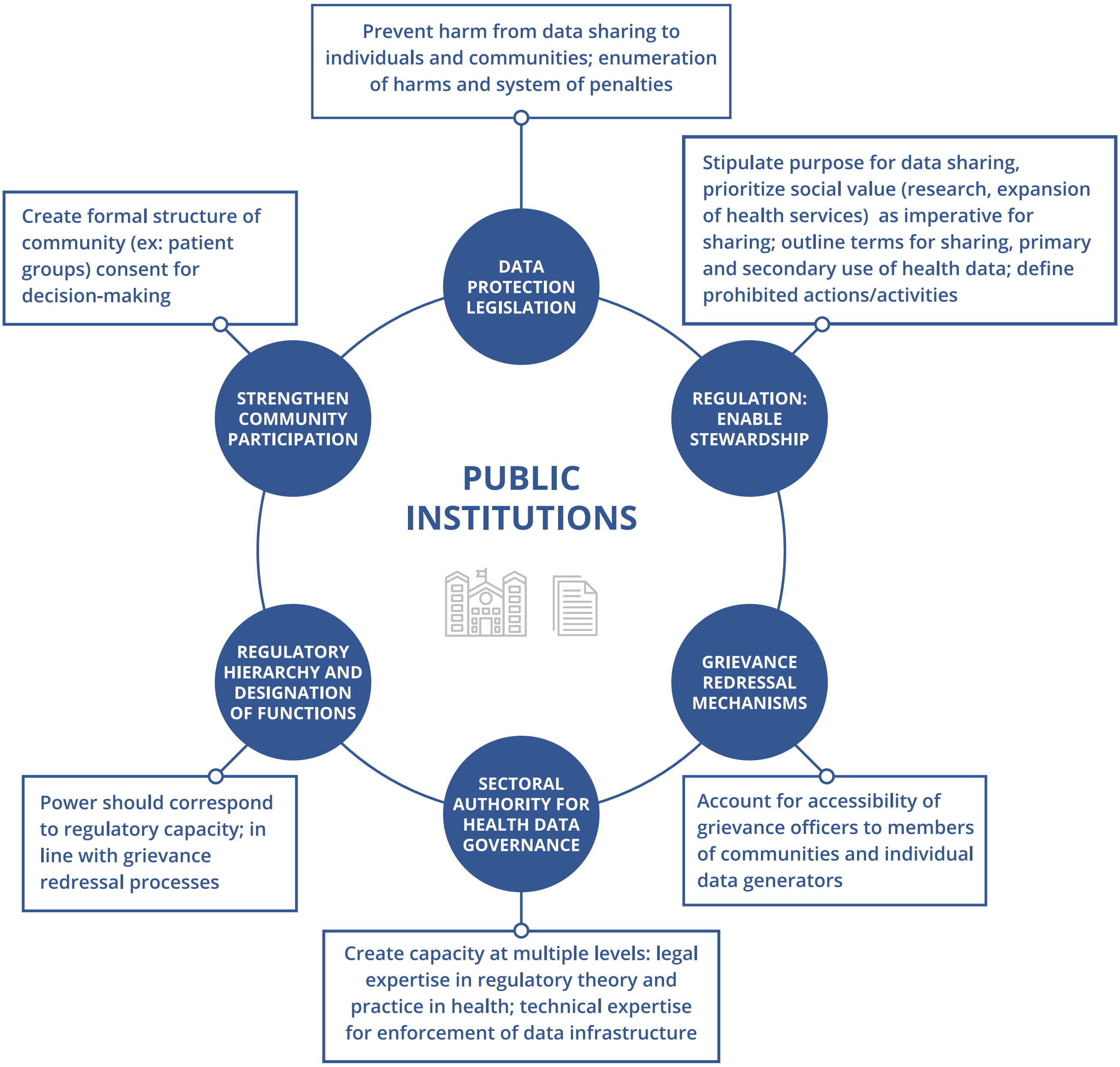


Figure 10: *Public institutions play a critical role in mediating the interaction between communities and other stakeholders such as providers, civil society and academia in the health data ecosystem. Such institutions define frameworks, principles and standards for operationalising top-down data stewardship.*

According to the roles and responsibilities of public institutions identified in this report, state actions can be divided into three broad categories – legislative frameworks, regulatory structures and enabling environment. Each of these categories has to work in tandem with the others to produce a resilient sharing environment that is necessary for institutionalising stewardship at the policymaking level. The three categories are correlated inasmuch as their alignment can contribute to seamless health data flow within and across jurisdictions. Figure 11 provides an explanation of the three categories of state actions that should be taken into account while implementing data stewardship in healthcare.

1. Legislative Frameworks

This includes policies for data governance set forth by public institutions within each jurisdiction. Legislative actions include the establishment of individual and community rights over data. This is supported by data protection frameworks that seek to protect communities from harm due to breach of consent. Similarly, granting individuals the right to portability and re-use of data is a precursor to stewardship that hinges on making data available for public benefit use. In turn, the use of data for public benefit requires a clear articulation of first principles, values and purposes for unlocking data which should be enshrined within legislative frameworks on health data stewardship.

2. Regulatory Structures

This refers to composite spaces set up by the state for interaction among communities to deliberate and exercise preferences regarding data. Building appropriate regulatory structures is immanent to empowering communities to participate in decision-making and reduce the consequent burden on state agencies to represent community interests. Decentralising data decisions through clear articulation of sharing norms guided by consent-driven and purpose-specific clauses is foundational to institutionalising stewardship through policy action.

3. Enabling Environment

Relates to those actions of the state that open up data for technical and regulatory innovation. Setting up a safe pilot environment for low-risk testing using anonymised and synthetic health datasets is salient to surfacing the practical considerations involved in instituting stewardship. Similarly, periodic consultation with communities, civil society and private entities – stakeholders affected by health data stewardship – to examine legislative frameworks and regulatory structures provides feedback for evidence-based policymaking on data exchange. State subsidies that may incentivise uptake of technical sharing formats such as FHIR, SNOMED, etc., would fall under this category of actions directed at implementing top-down stewardship.

Figure 11: *To institute stewardship through policy, various categories of state action should align – legislative frameworks, regulatory structures and an enabling environment.*

The findings from the analysis of policy pathways to stewardship identified across jurisdictions exhibited a diversity of mechanisms that state authorities have employed to create human-centric approaches to data governance. These actions have been distilled into a high-level checklist comprising a combination of legislative frameworks, regulatory structures and enabling tools to institutionalise stewardship through state policy. Figure 12 illustrates eight consecutive steps within the checklist that state institutions can adopt to facilitate participatory data governance.

		STEPS		EXPLANATION
DATA SHARING	01	Articulate clear position on data as relating to the individual and as a commodity	▶	GDPR views data subjects as “owners” of personal information; NPD (India) views ”community” as owner of value derived from data
	02	Create data protection legislation – define rights of data subjects	▶	Individual patient rights over data, i.e. right to access, delete and repurpose data collected by providers and state
DATA GOVERNANCE	03	Define harms arising from data sharing and prescribe remedies/penalties	▶	Reidentification, unauthorized use, enabling private surveillance and diminished access to health resources or services
	04	Delineate role of community; establish procedure for their participation in data governance	▶	Terms of engagement between communities and data requestors – through representation or direct participation
	05	Create decision-making structures for data sharing, including sectoral regulator for health	▶	Appointment of independent bodies with specific guidelines; performs quasi-judicial role. Ex: NHA (India), Info and Privacy Commissioner - Ontario, Estonian Information Systems Authority
DATA STEWARDSHIP	06	Establish or prescribe standards (ex: FHIR, SnowMED, ICD) for health data interoperability	▷	Data portability across sectors and stakeholders. Cost of formatting data should not fall on users and requestors. Cost to be borne by state/large-scale data processors
	07	Clearly define limits and conditions for mandatory state access to and use of data	▶	Predictability in use of state power; clear and proportionate conditions for state access; sectoral rules by regulators
	08	Define terms and purposes for secondary use of health data by stakeholders other than the state	▶	Set clear condition and purpose to authorize third party access to data. Prohibit specific access/purposes which carry risk for users and community. Ex: FinData legislation, NPD Report
	09	Establish regulatory sandbox to minimize risk of harms	▷	Safe environments for data innovators with limited datasets to minimize any possible harm while encouraging experiments and research into models of health data stewardship

▶ Implemented by public institutions

▷ Implemented in collaboration with private/non-profit sector

Figure 12: *The checklist surfaces policy opportunities that public authorities can leverage to institutionalise data stewardship in the health sector.*

Digital enclosures[62] stifle the innovative potential of technology to deliver better health outcomes. Disrupting the consensus anchored in narrow conceptions of value to unlock data for public benefit relies on top-down interventions. The checklist illustrates the contours of policy prerogatives to reconstitute data regulation in ways that make empowering individuals and communities the lynchpin of state action.

---

[62] Surveillance in the digital enclosure", Mark Andrejevic, Surveillance in the Digital Enclosure, The Communication Review, December 2007, accessible at http://dcac.du.ac.in/documents/E-Resource/2020/Metrial/31SagorikaSingha1.pdf

07

Way Forward

> This paper has presented potential solutions to the problems that plague data sharing in health.

The issues of the ecosystem have been outlined, identifying the main hurdles of lack of usability and availability of data. Implementing principles of data stewardship has been suggested. The value proposition of stewardship lies in creating a framework for stakeholders in the health data ecosystem to work with one another to make their work easier whilst benefiting the sector as a whole. Patients, doctors and clinics find value in making data shareable and gaining insights that further their objectives. The governance value of stewardship makes this possible, and affords stakeholders the process necessary for working with others and for end users' rights and interests to remain central to the regulatory framework. Stewardship takes a number of forms, and its use and features will depend on the specific context of its deployment. In order to make stewardship feasible, there needs to be in place a minimum set of protections and regulatory guardrails in order to prevent abuse of the system by any of the stakeholders. These regulatory measures serve to enforce the data rights of end users, enable community participation in decision-making, and allow data to be shared for community health benefits.

To enable this shift towards responsible data sharing in health, public institutions must play a pivotal role in building the necessary frameworks. This includes laying down strong and enforceable rules on protecting users against harms arising from usage of their data. With risks of harm still at play, users will be unwilling to share data. To engender trust in the ecosystem, the framework must enable healthcare companies to work with community and patient groups, to decide on practices and implement preferences of user groups. To make data shareable and usable, once shared, common standards are required, which is achievable through collaboration and consultation and the availability of opportunities for this.

Alongside, communities need tools to engage with decision-making structures in a manner that is accessible in their particular context and network of relationships.[63] This is necessary for them to express their preferences in data governance issues. A good policy framework for responsible data sharing will provide for institutional structures and accessible procedures to enable this community engagement to be a crucial part of the feedback loop for regulators on policy. Community organisations and non-profits will have a significant role to play in data governance going forward, and stewardship is a framework that is able to take this into account.[64]

---

[63] "Democratic Data: A Relational Theory For Data Governance", Salome Viljoen, Yale Law Journal, 2020, accessible at https://ssrn.com/abstract=3727562/

[64] "Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance", Sylvie Delacroix and Neil D Lawrence, International Data Privacy Law, Volume 9, Issue 4, November 2019, accessible at https://doi.org/10.1093/idpl/ipz014

The framing of regulatory measures working alongside community engagement points to the need for a combination of top-down and bottom-up approaches.[65] Top-down measures consist of rule-making, appointing bodies in charge of the sector and specific functionaries in the framework. Bottom-up measures are designed to give impacted user groups the tools to engage with this regulatory framework to keep its operation in alignment with their interests and preferences.[66]

These measures would bring regulators and authorities increased clarity on how data is being collected and used. It would empower users and public health officials to have a say in third-party data use. Processes that increase participation of users and communities and employ informed consent through active communication channels would find it much simpler to keep users informed and abreast with developments in their activities. For example, in contexts where data is available regarding a spike in any given disease in an area following a natural disaster or a pandemic/ epidemic, information on what areas need emergency care will be available on a real-time basis. This avoids the need for ad hoc measures as were seen during the challenges of the COVID-19 pandemic.[67]

To create a smoothly functioning health data stewarding ecosystem where stakeholders are able to work with one another, initiative is needed at multiple levels, as this paper has outlined. Given necessary protections, regulatory tools to engage public feedback, and community participation, health data sharing can be made beneficial for its primary constituents – patients and their caregivers.

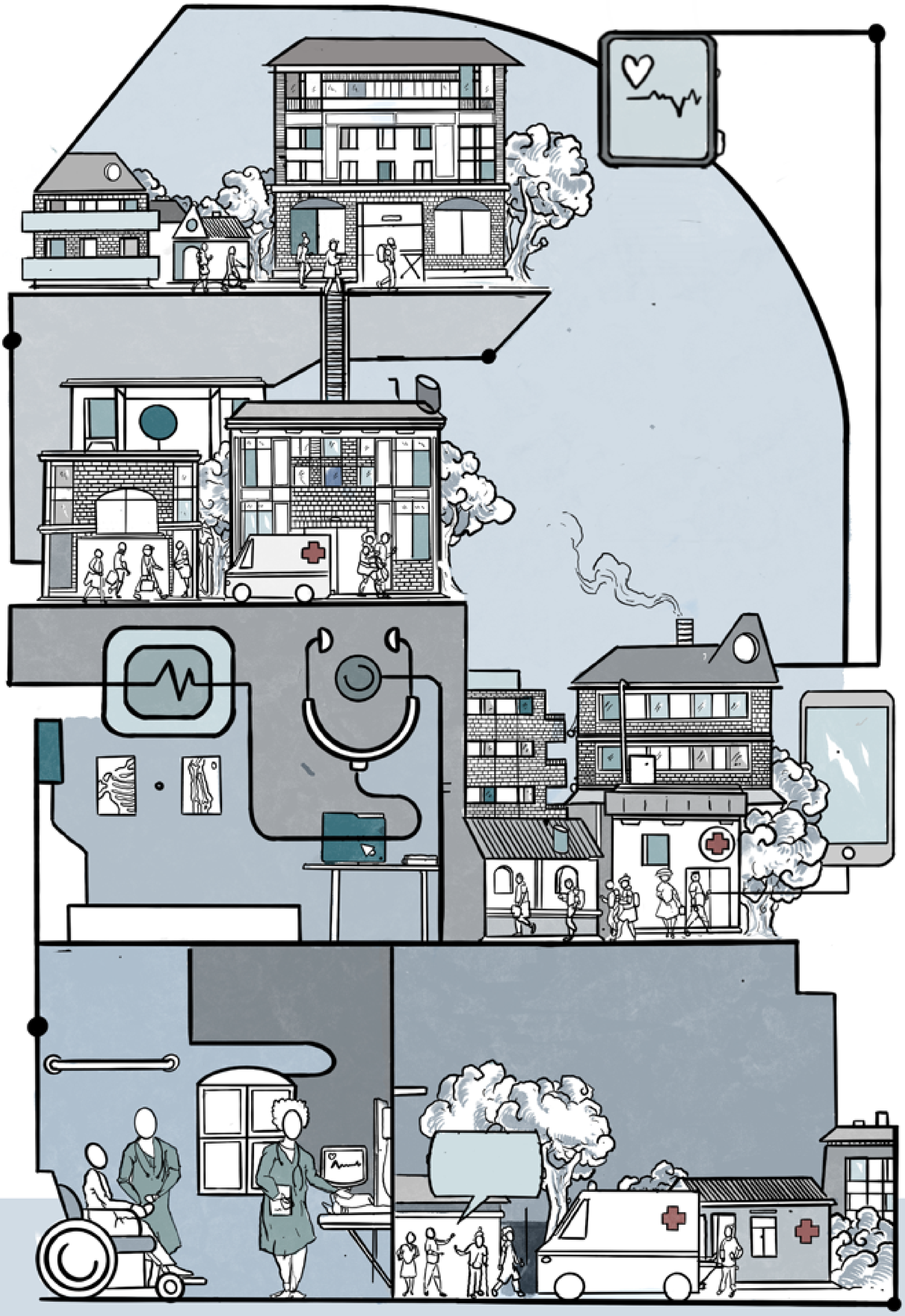
---

[65] For more detailed insights on this approach, please refer to the accompanying report on "Health Data Stewardship: Bottom-up stakeholder engagement"

[66] "Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance", Sylvie Delacroix and Neil D Lawrence, International Data Privacy Law, Volume 9, Issue 4, November 2019, accessible at https://doi.org/10.1093/idpl/ipz014

[67] "Transparency during global health emergencies", The Lancet Digital Health Editorial, Volume 2 Issue 9, September 2020, accessible at https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30198-9/

aapti institute



Health Data Stewardship: Bottom-Up Stakeholder Engagement

Health Data Stewardship:
Bottom-Up Stakeholder Engagement

aapti institute

Soujanya Sridharan, Siddharth Manohar
and Astha Kapoor

July 2021

Acknowledgements

This report draws from the expertise of numerous academic and industry experts and practitioners. We are grateful for their input during interviews and rigorous feedback sessions. We also thank Kevin O’Neil and Priyank Hirani of The Rockefeller Foundation for supporting this work.

Research assistant: Bilal Mohamed | **Editor:** Yana Bey
Report design: Meher Rajpal | **Illustration:** Siddhartha Iyer

Contents

Glossary

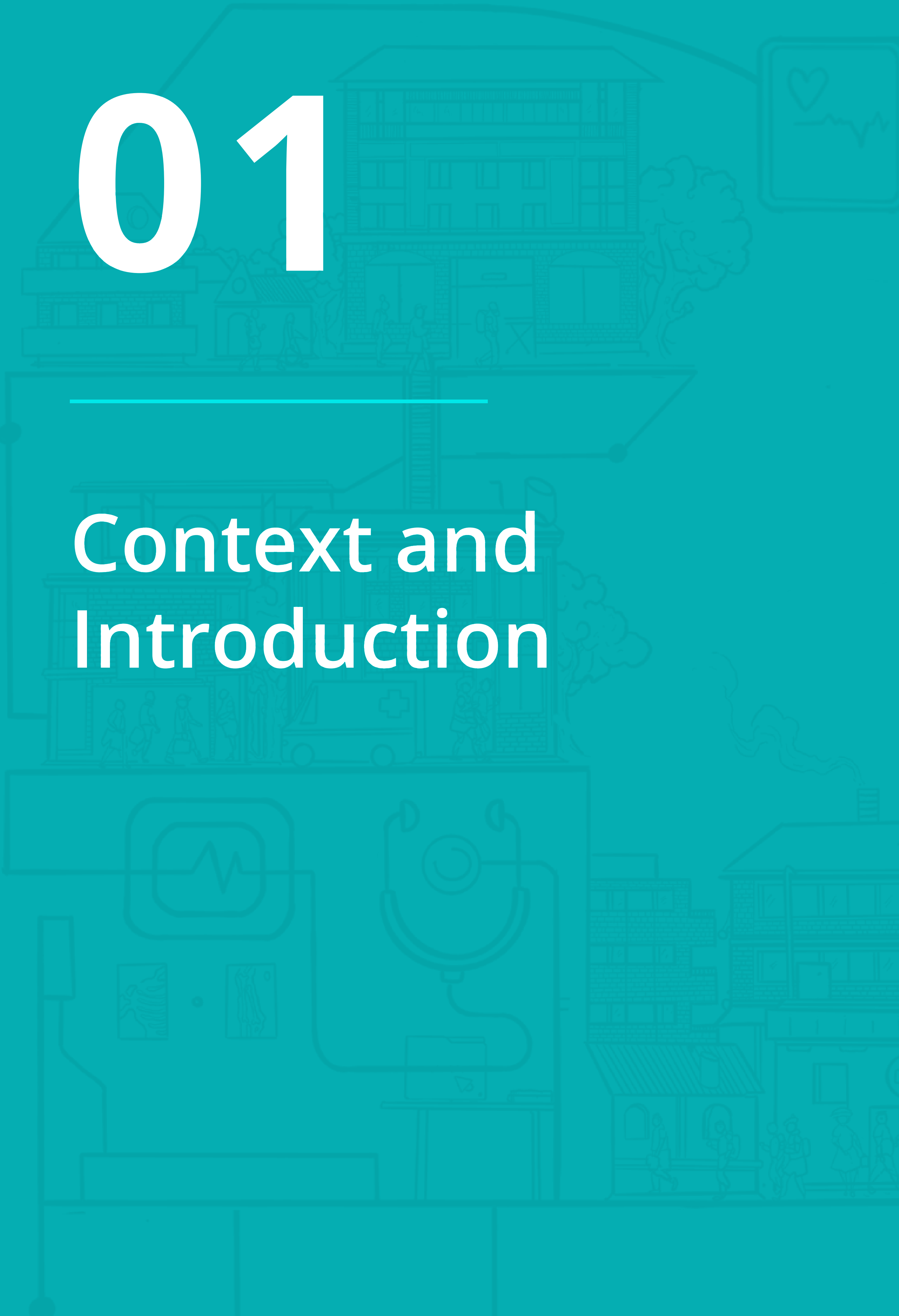
(CONCEPTUAL AND ABBREVIATIONS)

Data	Any representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means (as in the Information Technology Act, The Personal Data Protection Bill).
Consent	Given by the individual to an entity for processing of their data; must be free, informed, specific, clear and capable of being withdrawn (as in the PDP Bill).
Data principal / user	Persons, both natural and legal, to whom any data relates (as in the PDP Bill).
Personal data	Data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute or any other feature of the identity of such natural person, whether online or offline, or any combination of such features with any other information, including any inference drawn from such data for the purpose of profiling (as in the PDP bill).
Non-personal data	Data that either never related to an identified or identifiable natural person, or data which may have initially been personal data, but was later anonymised through transformation techniques to the extent that individual-specific events are no longer identifiable (as in the Non-personal data Report).
Ecosystem	A composite term referring to the various stakeholders within the data economy, enabling infrastructures for data sharing, enabling legislation to data sharing, and the network of relationships between each of these - working together to make data sharing a possibility.

Bottom-up action for stewardship	This relates to the Activities undertaken by providers, civil society organisations and other relevant stakeholders such as health tech operators to institute human-centric mechanisms for data governance within their organisations.
Top-down action for stewardship	Collection of legislative, regulatory and policy frameworks implemented by state authorities and public institutions to operationalise data stewardship within each jurisdiction.

01

Context and Introduction



> Health data stewardship[1] is a mechanism that structures data flow according to community-centric imperatives, creating models of information exchange grounded in promotion of societal value.

It assumes importance in the healthcare context due to the highly sensitive nature of the information, potentially including confidential data on medical history, racial or ethnic background, and financial status of patients and their communities.

Consequently, stewardship is particularly relevant for the health sector where data sharing for public good and innovation is at odds with the data rights and privacy of patients.[2] Moreover, during the current pandemic when digital health solutions hold out the promise of compensating for failings of overwhelmed health systems[3], their lack of transparency has been a concern. A flagrant disregard for user agency is a dominant feature of such solutions which depend on instituting new data pipelines to support their services.[4] For instance, contact tracing applications employed by public health authorities rely on unauthorised access to geolocation data from mobile network operators, without obtaining the explicit consent of users, to monitor population movements, as was discovered in the case of the Government of Pakistan.[5]

While data sharing promises to significantly improve healthcare quality and health systems' performance[6], unchecked sharing has adverse ramifications for access and affordability of health services for individuals, particularly when such information exists in silos, dispersed across multiple sources. For instance, an investigation by the Financial Times revealed that individuals in the UK using health websites and applications are vulnerable as cookies on the websites track their symptoms, diagnoses, ovulation cycles and drug names. This data, treated with de-identification techniques, is shared with advertisers and other third parties in ways that compromise the agency and data rights of the individuals. Explicit consent for processing of sensitive health data, as mandated by the European Union's

---

[1] For more information on the different models of stewardship, please refer to 'Data Stewardship: A Taxonomy', Aapti Institute, available at https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/

[2] "Health data stewardship: What, Why, Who How", Primer by National Committee on Vital and Health Statistics (2014), retrieved June 30, 2021 from https://www.ncvhs.hhs.gov/wp-content/uploads/2014/05/090930lt.pdf

[3] "In WHO Global Pulse Survey, 90% of Countries Report Disruptions to Essential Health Services since COVID-19 Pandemic", World Health Organization, August 31, 2020, retrieved June 30, 2021 from https://www.who.int/news/item/31-08-2020-in-who-global-pulse-survey-90-of-countries-report-disruptions-to-essential-health-services-since-covid-19-pandemic

[4] "Digital technologies in the public health response to COVID-19", Budd et al, Nature Medicine (August 07, 2020), retrieved June 30, 2020 from https://www.nature.com/articles/s41591-020-1011-4

[5] "Pakistani government access to mobile records may violate constitution", Privacy International (March 30, 2020), retrieved June 30, 2021 from https://privacyinternational.org/examples/3632/pakistani-government-access-mobile-records-may-violate-constitution

[6] "Sharing Is Caring-Data Sharing Initiatives in Healthcare", Hulsen, International journal of environmental research and public health MDPI (April 27, 2020), retrived June 30, 2021 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246891/

General Data Protection Regulation, 2016, is seldom sought in cookie notices.[7]

As a result, individuals and communities – primary generators of this information – have little recourse to control how their data is used by entities concerned with collecting and processing it. Moreover, compliance-bound obligations only ensure that patient consent for data sharing is obtained at the point-of-care through single blanket authorisation forms. Without clarity on the purposes and parties involved in the process of sharing, patients and advocacy groups are blocked from participating in data decisions.[8]

This is compounded by opaque data sharing agreements[9] (DSAs) between contractual entities that extract data from users and share it for commercial purposes. Often, prioritisation of commercial incentives undermines the inherent value of data in promoting social welfare. Potentially valuable information that could guide health policy and action is locked within DSAs enjoying monopolistic intellectual property protection that prevents data from being harnessed for public good.[10]

The experience of developing countries during the COVID-19 pandemic is a telling reminder of the pitfalls of a data sharing system that privileges entrenched financial interests. Calls for a TRIPS waiver, put forth by the governments of India and South Africa[11], would make vaccine-related technology and data widely available to accelerate response to the emergency. Regrettably, this move was blocked by pharmaceutical companies who prized intellectual property gains over equitable access to critical preventive care.[12] Open science and public health have been undermined in a manufactured crisis which continues to block the use of data for societal benefit.

The growing realisation that access to data is highly lopsided, combined with patient apprehensions about trust, privacy and misuse of data, has been reinforced by reports of abuse. In 2017, the NHS shared details about 1.6 million patients with Google's DeepMind, a move that violated the UK's own privacy laws.[13] GlaxoSmithKline's investment

---

[7] "How Top Health Websites Are Sharing Sensitive Data with Advertisers", Harlow, Max, and Murgia, Financial Times (November 13, 2019), retrieved June 30, 2021 from https://www.ft.com/content/0fbf4d8e-022b-11ea-be59-e49b2a136b8d

[8] "Public and patient involvement in health data governance (DATAGov): protocol of a people-centred, mixed-methods study on data use and sharing for rare diseases care and research", Freitas et al, BMJ Open (March 15, 2021), retrieved June 30, 2021 from https://bmjopen.bmj.com/content/11/3/e044289

[9] "Big Data & Issues & Opportunities: Data Sharing Agreements", van Asbroeck, Bird & Bird (April, 2019), retrieved June 30, 2021 from https://www.twobirds.com/en/news/articles/2019/global/big-data-and-issues-and-data-sharing-agreements

[10] "We must have a #PeoplesVaccine, not profit vaccine", Byanyima, UNAIDS Opinion (December 09, 2020), retrieved June 30, 2021 from https://www.unaids.org/en/resources/presscentre/featurestories/2020/december/20201209_we-must-have-a-peoples-vaccine

[11] "India, others push for TRIPS waiver scope beyond Covid vaccine at WTO", Raghavan, The Indian Express (May 23, 2021), retrieved June 30, 2021 from https://indianexpress.com/article/business/economy/india-others-push-for-trips-waiver-scope-beyond-covid-vaccine-at-wto-7326370/

[12] "How Bill Gates Impeded Global Access to Covid Vaccines", Zaitchik, The New Republic (April 12, 2021) retrieved June 30, 2021 from https://newrepublic.com/article/162000/bill-gates-impeded-global-access-covid-vaccines

[13] "Google DeepMind NHS App Test Broke UK Privacy Law." BBC News (July 3, 2017), retrieved June 30, 2021 from https://www.bbc.com/news/technology-40483202

in 23andMe[14], a personal genomics enterprise selling direct-to-consumer DNA kits, grants access to genetic data of five million users, and raises concerns about the commodification of health information in a milieu where users cannot exercise meaningful control over downstream use of their data or be compensated for their information. More recently, India's COVID-19 vaccination programme ran into controversy when the country's nodal health agency – the National Health Authority – used vaccine registration as an opportunity to roll out Unique Health IDs[15], a move that not only violated the consent of citizens but also enabled mass surveillance by the state.

The above instances have contributed to renewed awareness about data safeguards and magnified the need for inclusive, democratic mechanisms for data governance. To this end, data stewardship – a community-centric mechanism of data governance – in healthcare presents an alternative paradigm that can inculcate trust in the process of sharing by expanding community engagement while incentivising stakeholders to participate in such arrangements. Moreover, stewards can reliably unlock the value of health data for social benefit while balancing imperatives for promoting innovation and upholding privacy of individuals and communities.

Accordingly, to bring data stewardship to fruition, coordinated action at two levels becomes necessary: top-down policy measures by the state and bottom-up engagement by communities and stakeholders within and across organisations. Significantly, bottom-up stewardship initiatives undertaken by civil society organisations and select private enterprises demonstrate the merits of a participatory framework that helps communities gain meaningful control over their information. This is essential to ensure that data generators – individuals and communities – are involved at every step of the data value chain, channelling their information for productive ends such as research, drug development and delivery of patient-centric care.

The primary objective of this report is to make a case for the bottom-up instantiation of data stewardship in healthcare. The health sector is primed for stewardship, given the long-standing history of sharing between various actors – from patients to providers who share anonymised health information with academia and public agencies for use in research and policy planning. Such relationships are navigated through complex data sharing agreements that treat individuals and communities as mere inputs in the process of information exchange, consequently invisibilising their role as

---

[14] "A Major Drug Company Now Has Access to 23andMe's Genetic Data. Should You Be Concerned?", Ducharme, TIME (July 26, 2018) retrieved June 30, 2021 from https://time.com/5349896/23andme-glaxo-smith-kline/

[15] "Took Covid vaccine using Aadhaar? Your National Health ID has been created without your permission", Dogra, India Today (May 24, 2021) retrieved June 23, 2021 from https://bit.ly/3yB2TMB

the primary generators and owners of the data.

The study explores mechanisms for data sharing that subvert an essentially extractivist model of the health data value chain, in an effort to restore the agency of individuals and uphold community interests as a guiding imperative for data sharing. Data stewardship is one such paradigm that goes beyond compliance-bound practices, such as mere consent provisioning for data sharing, to unlock the value of data for social benefit while protecting individuals' rights.

> **Data stewardship is one such paradigm that goes beyond compliance-bound practices, such as mere consent provisioning for data sharing, to unlock the value of data for social benefit while protecting individuals' rights.**

02

Method

This study was conducted along two prongs: primary research involving in-depth interviews with civil society and private organisations[16] as well as experts working on health data stewardship. Detailed manuscripts of these interviews were analysed using a specific, predefined set of codes relating to the structure, governance, technical features and participation of data generators (individuals and communities). This helped derive insights for actualising models of bottom-up stewardship in healthcare.[17]

The primary research was corroborated by desk research that entailed a legal and policy analysis of the health data sharing ecosystem across select jurisdictions – India, Ontario (Canada), Finland and Estonia. The choice of specific regimes was to demonstrate the policy pathways to health data stewardship that underpin top-down interventions[18] by the state to create an enabling ecosystem[19] for secure and responsible health data sharing.

However, the focus of this study encompasses a detailed analysis of bottom-up models of stewardship in healthcare. The research has been instrumental in revealing a diversity of arrangements among stakeholders to share health information. Differing contexts demand different levels of representation and control exercised by data generators, proving to be crucial determinants of steward model type. As a result, there are no "one size fits all" prescriptions that can be made for stewards; it is crucial to arrive at this decision through an analysis of its internal data practices and priorities (See Section 4).

---

[16] For a detailed list of organisations interviewed as a part of this study, please refer to Annexe 1

[17] This study on health data stewardship borrows from Aapti Institute's ongoing engagement with Omidyar Network on The Data Economy Lab

[18] For more information on top-down policy interventions required to instantiate stewardship in healthcare, please refer to a corollary of this study, "Top-down state action for public benefit health data sharing"

[19] To better understand what an ecosystem approach to health data stewardship entails, refer to our position paper, "Enabling environment for health data stewardship"

The stewardship models interviewed for this research include:

(a) Data trust

Refers to stewards which are governed by the legal framework of trusts and carry fiduciary responsibility towards the users represented. Data sharing is driven by predefined purposes. Example: The Open Data Institute's trusts pilots on data about food waste, wildlife poaching and mobility in London are first of their kind initiatives to test the common law framework of trusts in the context of data.[20]

(b) Data collaborative or exchange

A model of stewardship in which two or more entities with a common goal or agenda come together to aggregate and share data to serve mutually beneficial ends. Example: Ciitizen[21] is a private collaborative platform that enables individual patients and larger patient advocacy groups to collect, digitise, and share their health information for research with pharmaceutical companies as well as hospitals for treatment management of individuals.

(c) Data Cooperative

Relates to a model of data stewardship in which sharing decisions are made on the basis of votes by members of the organisation. Example: SalusCoop[22] is a health data cooperative whose members can review medical research proposals and consent to share their data for specific projects. It also enables democratic decision-making through a general assembly of members accorded one vote each.

(d) Personal data store/ data repository

Refers to digital platforms in which individuals or groups store their data and enjoy granular consent controls to dictate sharing. Example: Digi.me[23] is a personal data store through which individuals can store and share data in exchange for monetary compensation.

(e) Ecosystem enabler

Entails organisations providing decentralised infrastructure to facilitate data collection, standardisation and, ultimately, the creation of stewards. Example: Dimagi[24] is an ecosystem enabler that provides open source products such as health information systems to facilitate disease management and clinical decision support.

---

[20] For more information on ODI's data trusts pilots, please visit https://theodi.org/project/data-trusts/#1554903732788-679e5312-2203

[21] For more information on Ciitizen, please refer to https://www.ciitizen.com/

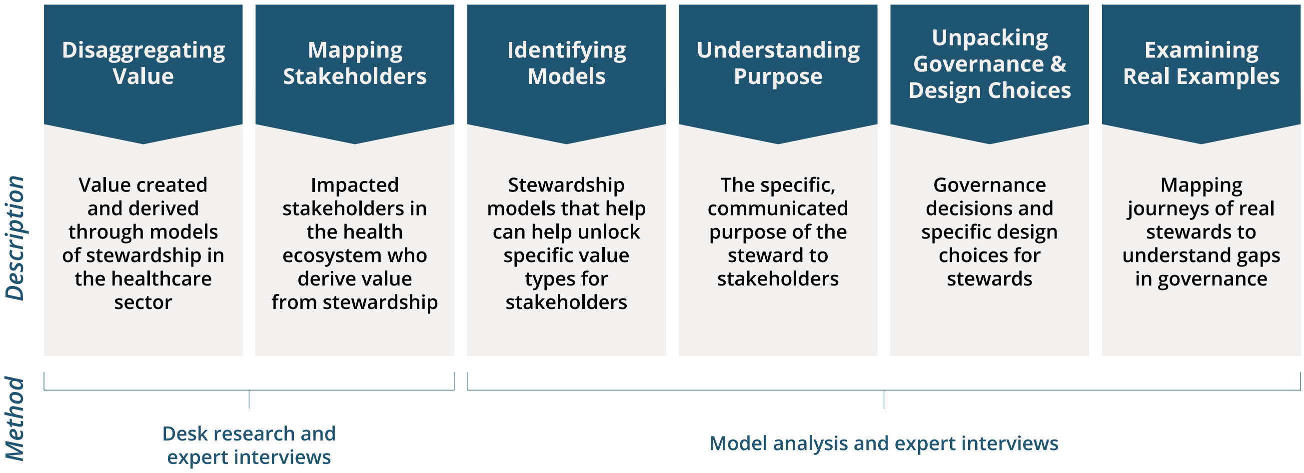
[22] To better understand Salus.coop's model, please visit https://www.saluscoop.org/

[23] More information on Digi.me's activities can be found at https://digi.me/

[24] To understand Dimagi's work in the context of health data ecosystems, please refer to https://www.dimagi.com/

Crucially, these user-centric models of stewardship have emerged as vehicles for meaningful participation by patients, advocacy groups and marginalised communities in data decisions. Further, the models have performed the twin tasks of unlocking data's societal value by empowering communities whilst concomitantly creating public good solutions – evolving a potentially expansive set of data sharing mechanisms engaging patients, providers, academics, and pharmaceutical companies.

Accordingly, we have devised a framework to translate value created by health data stewardship into governance considerations that organisations would encounter in their bid to implement bottom-up stewardship:



	Disaggregating Value	Mapping Stakeholders	Identifying Models	Understanding Purpose	Unpacking Governance & Design Choices	Examining Real Examples
Description	Value created and derived through models of stewardship in the healthcare sector	Impacted stakeholders in the health ecosystem who derive value from stewardship	Stewardship models that help can help unlock specific value types for stakeholders	The specific, communicated purpose of the steward to stakeholders	Governance decisions and specific design choices for stewards	Mapping journeys of real stewards to understand gaps in governance
Method	Desk research and expert interviews		Model analysis and expert interviews			

Figure 1: *The framework furnishes a rubric to translate this research into practical decisions and governance considerations that arise when organisations move from mere data sharing to focused models of health data stewardship.*

This framework is based on a review of 100 use cases of organisations which steward data and, more significantly, 24 in-depth case studies in the health sector that highlight structural considerations and design choices essential to build a steward.[25]

The subsequent sections of this report provide an analysis of the health data ecosystem: beginning with stakeholder mapping, going on to a disaggregation of value created by health data stewardship and culminating with a taxonomy of core stated purposes of stewards that guide internal data governance practices.

---

[25] The 100 use-cases examined as a part of this study belong to various sectors - health, mobility, sustainability and education. Interviews and case studies were conducted as a part of Aapti Institute's ongoing engagement with Omidyar Network on The Data Economy Lab.

03

Health Data Ecosystem

This study has identified five sets of stakeholders in the healthcare ecosystem, interacting on data issues:

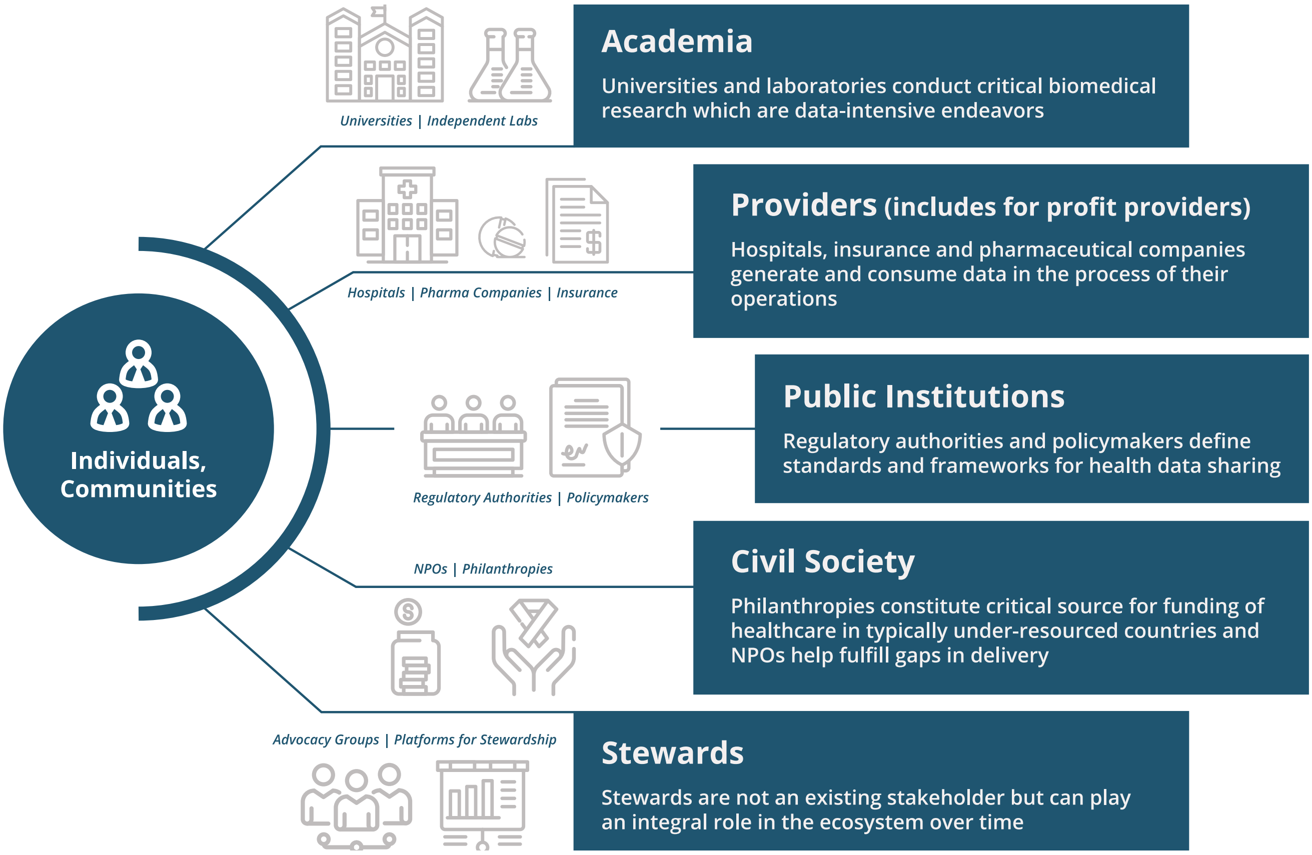
(a) **Patients and their caregivers** who comprise the category of **individuals and communities** that are the primary generators of data.

(b) **Providers**, including for-profit organisations, such as hospitals, clinics, pharmaceutical companies and insurance agencies, which generate and consume data in the process of their operations.

c) **Public institutions** which include policymakers and regulatory authorities who determine standards and legal frameworks for health data sharing.[26]

(d) **Civil society members** like philanthropic and not-for-profit organisations which provide critical sources of funding and delivery of healthcare in under-served contexts. Consequently, these organisations generate a wealth of health information.

(e) **Academia,** encompassing universities and laboratories that conduct critical biomedical research which are data-intensive endeavours.



Academia
Universities and laboratories conduct critical biomedical research which are data-intensive endeavors

Providers (includes for profit providers)
Hospitals, insurance and pharmaceutical companies generate and consume data in the process of their operations

Public Institutions
Regulatory authorities and policymakers define standards and frameworks for health data sharing

Civil Society
Philanthropies constitute critical source for funding of healthcare in typically under-resourced countries and NPOs help fulfill gaps in delivery

Stewards
Stewards are not an existing stakeholder but can play an integral role in the ecosystem over time

Figure 2: *The healthcare ecosystem consists of five sets of stakeholders. The sixth category of stewards, though aspirational, performs the crucial role of intermediaries who structure and negotiate data flow between other stakeholders.*

---

[26] Note: The role played by public institutions within the health data ecosystem is excluded from this study. Their corresponding legislative functions are explored independently in paper, "Health data stewardship: Top-down state action for public benefit data sharing".

Moving forward, the health data ecosystem could benefit from instantiating stewardship to deliver precision care that not only improves patient experience but also reduces cost of services. Further, leveraging technical and regulatory innovation like stewardship drives stakeholder engagement to unlock the social value inherent in health information,[27] feeding research and development of new drugs, treatment techniques and predictive analytics to determine public health priorities and investment. Broadly, the value created by health data stewardship[28] has been disaggregated into three categories: core values, sector-specific values and external values.

I. Core Values

- **User agency:** Defined as the ability to exercise greater agency on data use and protection.
- **Transparency and accountability:** Refers to clarity on how data is collected and used, ability to hold data users accountable regarding purpose of use.
- **Instrumental value:** Relates to monetary and other value (ease of access, enhanced quality of goods / services) received by generators for use of their data.

II. Sector-Specific Values

- **Expansion of health services:** Refers to collection and usage of existing datasets which may be deployed in health care operations and for informing policy to improve quality and affordability.
- **Academic / research:** Represents research and development to fuel innovation for drug and treatment availability, quality and efficacy.
- **Access to healthcare:** Conceived as better access to existing services and aiding innovation for development of new ones.

III. External Value[29]

- **Value outside health:** Represents benefits such as revenue generated for investors in healthcare, exists outside the ecosystem of stakeholders.

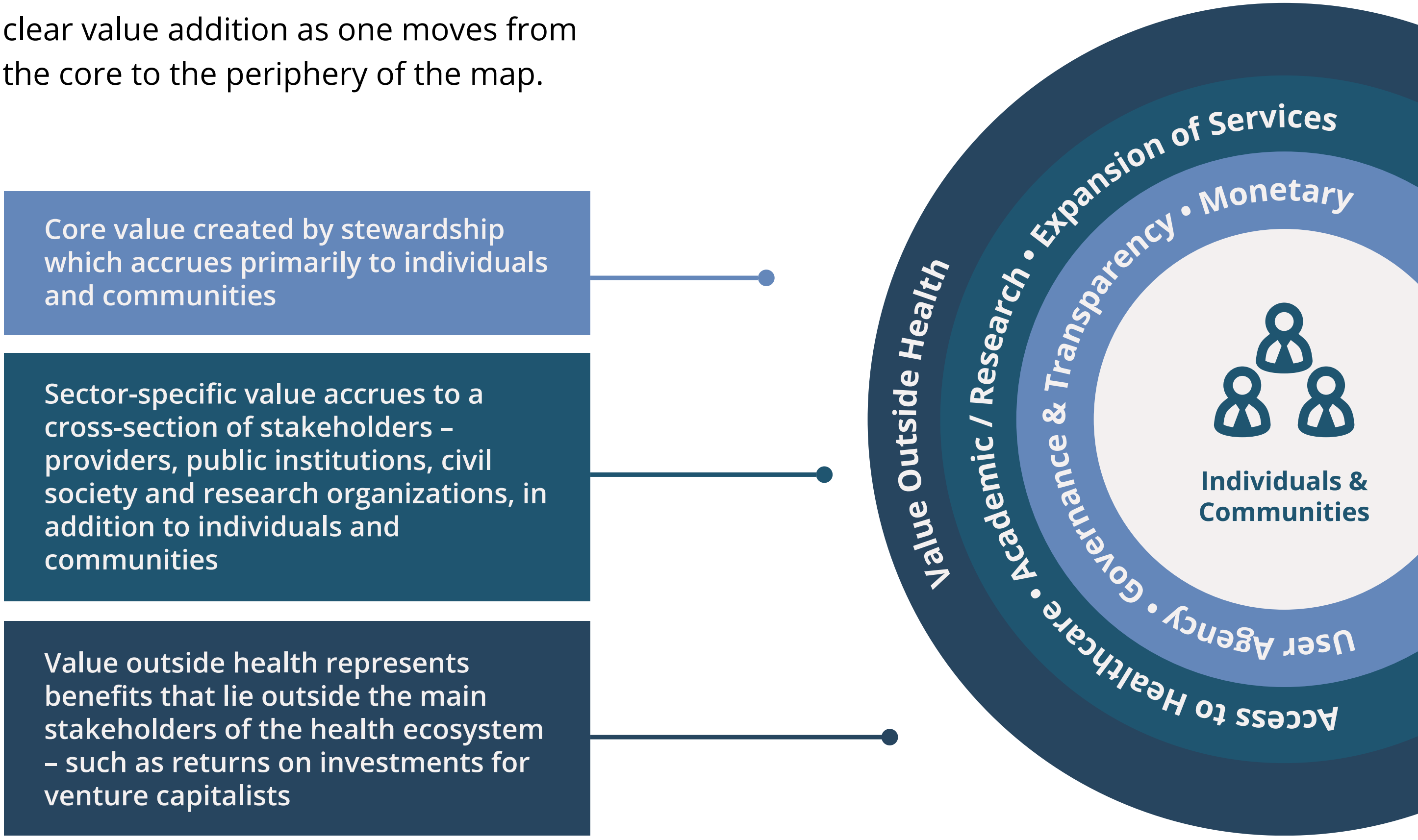
---

[27] "Public Value: How can it be measured, managed and grown?", Mulgan et al, Nesta (May 2019), retrieved June 30, 2021 from https://media.nesta.org.uk/documents/Public_Value_WEB.pdf

[28] "Understanding Value in Health Data Ecosystems", Marjanovic et al, RAND Health Quarterly (January 29, 2019), retrieved June 30, 2021 from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5798965/

[29] Note: External value is excluded from analysis in this study as its benefits are realised outside the health sector, creating gains for investors and other actors not directly involved in the health data value chain

The next level of analysis involves mapping stakeholders in the different categories of value created by stewardship. This is crucial to understand how each category affects different subsets of stakeholders, with clear value addition as one moves from the core to the periphery of the map.


Core value created by stewardship which accrues primarily to individuals and communities

Sector-specific value accrues to a cross-section of stakeholders – providers, public institutions, civil society and research organizations, in addition to individuals and communities

Value outside health represents benefits that lie outside the main stakeholders of the health ecosystem – such as returns on investments for venture capitalists

__Figure 3:__ Specific combinations of value and impacted stakeholders are better served by appropriate models of stewardship.

Organisations can use this exercise to identify distinct models of stewardship best suited to them, based on the stakeholder-value combination they choose to optimise for by revising their data practices. Accordingly, each combination of impacted stakeholder and value type lends itself to certain suggested models represented in Figure 4.

Impacted Stakeholders

Different types of value

	Individuals/ Communities	**Providers**	**Public Institutions**	**Non-Profits**	**Academia**
Academic or Research	*PDS / Data Cooperatives*	*Data Collaborative / Exchanges*	*Data Exchanges / Open Data*	*Data Exchanges / Data Trusts*	*Data Collaborative / Exchanges*
Access to Healthcare	*PDS / Data Coop / Data Exchanges*		*Data Collaborative / Exchanges*	*Data Collaborative / Exchanges*	
Expansion of Private / Public Health Services	*Data Cooperatives / Data Collaboratives*	*Data Exchanges / Repositories*	*Data Exchanges / Open Data*		
User Agency	*Data Cooperatives*	*Data Collaborative / Exchanges*	*Data Exchanges / Open Data*	*Data Cooperatives*	*PDS / Data Cooperatives*
Governance & Transparency	*PDS / Data Cooperatives*	*PDS / Data Cooperatives*	*PDS / Data Cooperatives*	*PDS / Data Cooperatives*	*PDS / Data Cooperatives*

Figure 4: *Each model presents a set of pros and cons. Stakeholders can optimise for specific values and purposes by choosing from the above set of recommendations.*[30]

---

[30] Note: Models left blank without any recommendations indicate gaps identified during analysis of interviews with data stewards in the health sector. Further research in these areas is required to propose recommendations for designated combinations of impacted stakeholder and value-type

04

Health Data Stewardship

As data value chains continue to evolve against the backdrop of the pandemic, there is a need to forge inventive solutions which place patients and their caregivers at the heart of health data governance frameworks. Balancing data openness and innovation to inform public health policy, research and disease management without compromising rights, agency and security of data generators (individuals and communities) is the primary aim of health data stewardship – a mechanism whose tenets are fairness, transparency and accountability.

Building inclusive governance frameworks depends, in turn, on the core stated purpose of the stewarding entity. An appraisal of the core stated purposes of different entities interviewed for this study revealed that this decision guided subsequent choices on data governance mechanisms employed by the steward. In many ways, it represents the first step in building a steward. Organisations were found to prioritise the ends they wished to serve as a touchstone for their interactions with individuals and communities.

The core stated purposes of different stewards have been bracketed in six divisions:

1

Individual Data Empowerment

Data generators are the primary decision-makers within the stewarding entity. They decide how to draw instrumental value through monetisation or exchange for goods and services. They also retain the right to access and share data securely across platforms and withdraw consent for sharing. For instance, Digi.me[31] is a personal data store which provides complete user control over data and places sharing decisions firmly in the hands of generators.

2

User-Driven Research

Collaboration for health research by the steward is driven by user group participation and is framed around their needs – research agenda and method are designed in consultation with communities of data generators. Variant Bio[32] is a data collaborative that works with historically marginalised populations and communities to facilitate people-driven therapeutics. Communities are engaged prior to the beginning of projects, and their data is collected and used within a framework that ethically focalises community concerns.

---

[31] For more information, please visit https://digi.me/

[32] For more information, please visit https://www.variantbio.com/

3

Collective Bargaining and Group Enfranchisement

The steward builds a shared data community and empowers it to bargain collectively for mutual benefit. This is crucial to manage asymmetries of power within the data economy where control over data is otherwise vested in big technology platforms. Data cooperatives such as MiData[33] allow communities to deliberate on how to share their health data and to what end, granting each member of the cooperative one vote to make democratic decisions.

4

Creation of Commercial Value

The steward facilitates collection, storage and standardisation of data to promote portability as well as increase its monetary value. Both stewards and data generators can draw instrumental value from sharing. BurstIQ[34] is a data connector / consent manager which employs blockchain to manage health data and share it through micro transactions for monetisation. Users retain granular ownership and revocable consent controls to permit verified, secure data services.

5

Public Good Solutions

Channelling data for social good is the primary goal of the steward. Users gain from extended reach, efficiency and effectiveness of healthcare delivery and experience. Ecosystem enablers such as Medic Mobile[35] are focused on capacity building of health workers. They provide cost-effective, open-source technical infrastructure to collect data and deliver services for communities residing in typically under-resourced contexts.

6

Epistemic Value

The steward forges partnerships with health research institutions who pledge to produce better health outcomes for communities. As a data repository, Clue[36] shares anonymised, aggregated user data (after obtaining consent) from its ovulation tracking application for trustworthy research initiatives. The platform has leveraged the social value of data by enabling its re-use for under-studied aspects of reproductive health.

---

[33] For more information, please visit https://www.midata.coop/en/home/

[34] For more information, please visit https://www.burstiq.com/

[35] For more information, please visit https://medic.org/

[36] For more information, please visit https://helloclue.com/

Further, organisations were found to dissect the governance features – termed "design choices"[37] – to formulate a set of decisions and functions ultimately defining their identity as stewards. Specifically, the design choices represent a series of interdependent decisions that not only expand the research on stewardship, but also bring clarity to the practical considerations necessary for its implementation. The consequent diversity of choices on governance available to individuals and communities will, over time, contribute to democratising the process of data management.

For ease of representation, we have divided data governance into six constituent buckets – steward structure, sharing controls, consent, incentives, technical features and business model. Each bucket has been further unpacked into design choices that a stewarding entity must consider while implementing community-centric mechanisms for data governance. The resultant mind map of governance features and related design choices is illustrated in Figure 5.[38]

---

[37] The research on design choices relied on interviews of stewards operating across multiple sectors, from mobility and smart cities to health, education, consumer experience and sustainability, among others. The subsequent analysis has been adapted to the health sector, substantiated by 24 in-depth case studies of health data stewards. This is part of Aapti Institute's broader engagement with Omidyar Network examining stewardship through The Data Economy Lab.

[38] The mind map is a product of interviews of stewards operating across multiple sectors, from mobility and smart cities to health, education, consumer experience and sustainability, among others. The subsequent analysis has been adapted to the health sector, substantiated by 24 in-depth case studies of health data stewards. This is part of Aapti Institute's broader engagement with Omidyar Network examining stewardship through The Data Economy Lab.

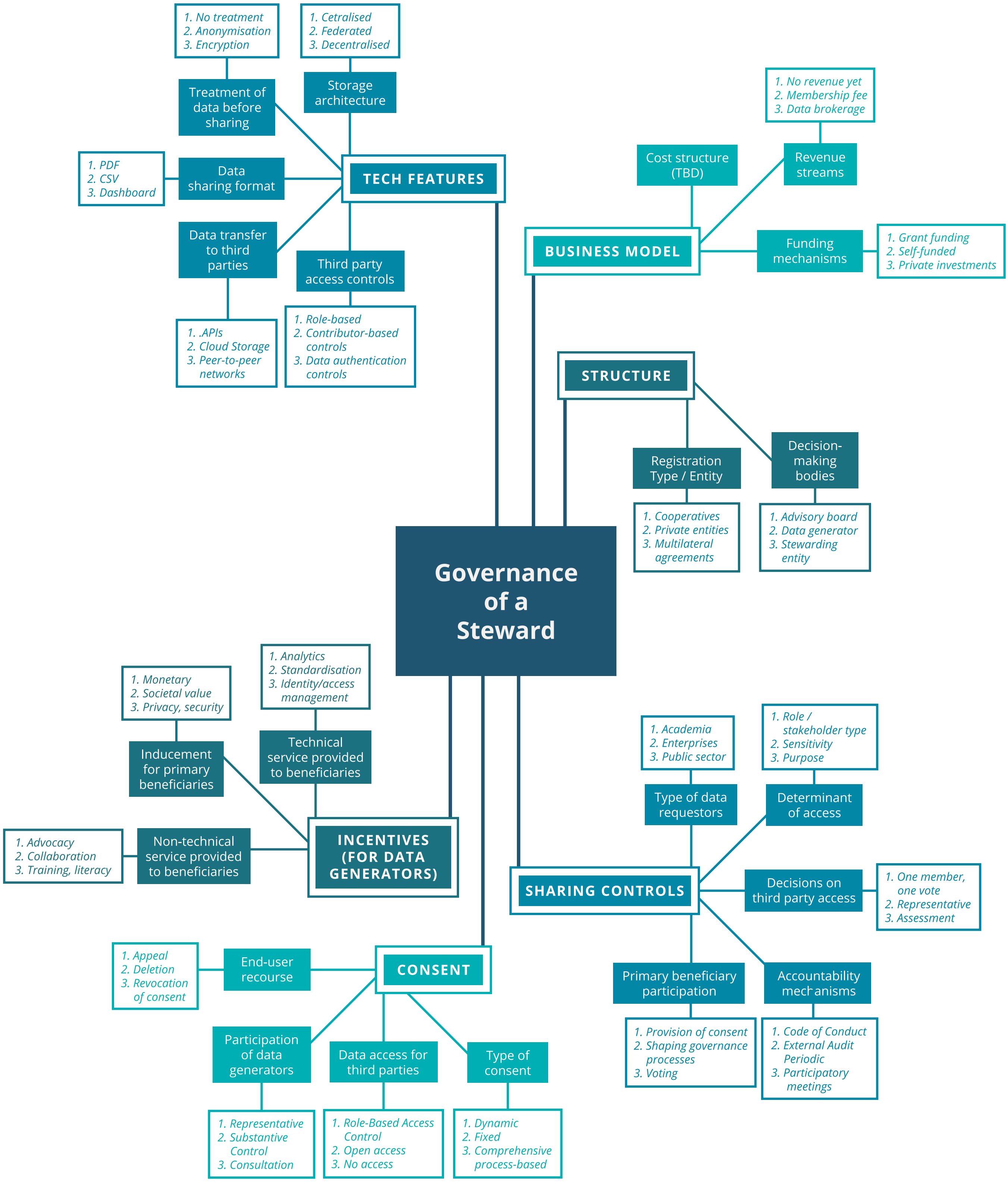


Figure 5: *Each bucket of governance produces a set of issues that the stewards must make choices about, flowing from purpose, value and the stakeholders it seeks to benefit.*

05

Ecosystem Review: Best Practices, Roles and Responsibilities for Stakeholders

5.1

Best practices and deep dives into features of a "good data steward"

A revealing aspect of this research revolves around the analysis of best practices in governance across the health data stewardship ecosystem. Practitioners were found to go beyond mere compliance to institute higher-order conditions that facilitated greater community engagement and participative governance. While it is true that compliance-bound governance has drawn attention to discourse around data rights, questions of agency remain obfuscated within policy. Therefore, it is incumbent upon organisations to prioritise user empowerment and build collaborative models of data governance.

In fact, extension of democratic controls over the process of data governance protects community interests in the same manner as afforded by various models of stewardship. By essentially adapting a regime of rights, ethical considerations around accountability, anti-discrimination and procedural fairness are focalised within the process of data exchange.[39] This has resulted in a variety of practices being implemented by different models stewarding health data, primarily directed towards preserving the community agency and public benefit aspects of sharing health information. The diversity of "best practices" that emerges consequently disrupts the impulse to "apply" these practices as uniform principles across organisations. Instead, this research has demonstrated that best practices in governance underpin certain features intrinsic to building a "good health data steward". In turn, chosen patterns of features constitute organisational data governance practices. An explanation of the features can be found in Figure 6.

---

[39] "Everyone should decide how their data are used – not just tech companies", Sadowski et al., Nature (July 1, 2021), retrieved July 2, 2021 from https://www.nature.com/articles/d41586-021-01812-3

FEATURES	DESCRIPTION
1. Availability of reliable, quality and usable datasets	• Stewards collaborate, maintain quality of data, make sharing easier. • A standardized and repeatable set of terms for access and usage benefits stakeholders. • Collaborative data access reduces cost of access
2. Data audit, ethics and assessment	• Ensures that stewards can be trusted with the information they handle. • This addresses any public interest questions that could potentially arise. • Inculcates transparency in the actions of the steward.
3. Clear incentive structure	• Either incentivise users to share data; or have a model where steward itself is incentivised to partner with third parties to make data available; this way, the steward is sustainable an not open to compromise by divergent interests.
4. Accountability and grievance redressal	• Open channels of accountability for users to access information on the steward's activities, express their preferences, protects stakeholder interests ensures meaningful consent. • Measures such as disclosures to beneficiaries and periodic review by independent review committees are demonstrations of accountability.
5. Financial sustainability	• Long-term financial health accounted for, a steward's collapse or closure can harm the populations that depend on the initiative in order to run.
6. Community - centered purpose	• A steward is, by definition, required to have generation of social value. This feeds is irrevocably linked to the purpose behind initially establishing the steward.

Figure 6: *Governance "best practices" underpin certain features intrinsic to building a "good data steward".*

However, it is important to acknowledge that at first glance the description of features might run the risk of seeming esoteric and create roadblocks for organisations exploring stewardship models. Transitioning to stewardship is a journey every organisation has to traverse, beginning with a problematisation of its existing data practices and culminating in a recalibration of priorities to reflect greater accountability and transparency in health data governance. In an effort to make this analysis tactile and actionable, we have disaggregated each governance feature into three sub-elements – starter, medium and advanced levels.

Further, conversations with existing stewards in the health sector exhibited deeply nuanced data practices in place within organisations to deliver on each of the features that go into building a "good steward". Organisations which have just begun their journey towards instantiating stewardship, based on their needs and preferences, have to make certain decisions on their internal data governance frameworks. But they have little understanding of what these decisions might appear like. The table in Figure 7 presents findings from deep dives into the features of a "good steward", enabling organisations to choose from a range of options that would determine their data governance priorities.

FEATURES	STARTER	MEDIUM	ADVANCED
Availability of reliable, quality and usable datasets	Stewarding entity acts as a single point of data access for requestors	Steward defines and implements formats and minimum conditions for storing and sharing data	Steward creates purpose-led terms for data sharing with a focus on protection of community interests
Data audit, ethics and assessment	Carry out compliance bound practices for self-assessment of data security at the organisational level	Engage an external data auditor for periodic evaluation and communicate results to key stakeholders	Conduct comprehensive audits by internal and external entities. Involve members of community in audit process
Clear incentive structure	There is broad but unspecific commitment to public value and user agency; activities of steward are opaque	Steward provides disclosure on funding and financial sustainability – highlights any possible conflicts of interests	Steward is committed to create alignment of funding and partnerships to user interests
Accountability and grievance redressal	Steward appoints an office for grievance redressal to act on complaints	There is an open channels of communication between steward and users; community may be consulted on sharing decisions	Steward establishes systematic grievance redressal process; follows consistent partnership with community
Financial sustainability	Brokerage earned through data sharing transactions only source of revenue for steward	Steward monetises data for public value e.g. selling data to municipalities for resource planning	Membership based remuneration for steward - relatively independent of prospects for monetisation through sharing
Community - centered purpose	Steward's commitment to community is perfunctory; retains sole control over decision - making	Communities enjoy consultative role; final decisions left to steward	Active engagement with community members to define and translate their preference on societal benefits

Figure 7: *Organisations, based on their needs and priorities, can decide which levels make sense for which feature.*

Additionally, instantiating stewardship at the organisational level was found to create demonstrable gains along multiple dimensions as entities moved from one level to another. At this point, it is important to note that trust in the process of data sharing underpins all activities of the steward. The pre-eminent and over-arching benefit of moving up each level is to maximise "trust" reposed in the steward, by all relevant stakeholders.

The broad contours of gains afforded by a responsible, fair and transparent stewardship framework can be delineated through five facets, providing compelling incentives for its instantiation within organisations, as explained in Figure 8.

GAINS	STARTER	MEDIUM	ADVANCED
Collaboration *Opportunities to work with new stakeholders*	**New partners** • Discover new partners and relationships – arrange in stakeholder specific relationships	**Broaden scope of relationships** • Enhance efficiency and transparency of relationships – increase focus to public value	**More investment & support** • Advance relationships to more investments and greater support
Network *Increase membership, use cases*	**Minimal usability** • Limited networks - don't create value to communities • Lack of scalable solutions	**Increase adoption** • Focus on interoperability to create tertiary connections in the community	**Network effects** • Through greater adoption of the steward • Community engagement for impact
Reputation *Enhance visibility, role in ecosystem*	**Low trust** • Poor quality and features limit trust and possible adoption of steward by community / stakeholders	**Increased trust through transparency** • Fulfill stakeholder interests and representation to increase value	**Policy positions** • Growth in reputation to play greater role in policy-making and standard-setting – represent communities
Instrumental *Increase investments, revenue and innovation*	**High costs of opportunity / innovation** • Curb opportunities for instrumental gains - partnership are small, need to be created repeatedly	**Value added functions** • Responsible for greater data sharing can develop value added functions, • Enhance efficiencies - better study of needs of market	**Cost effective, efficient operations** • Expertise in stewardship help make operations more efficient, cost-effective; • Innovate more opportunities
Compliance *Save prospective costs from regulations*	**Compliance** • Basic mandatory requirements of disclosures and transparency – compliance with current laws/regulations	**Compliance plus** • Ensure user interests in a fiduciary capacity – beyond compliance best practices	**Future ready compliance** • Regulatory roles being carved out for a responsible intermediary; stewards can potentially grow into a statutory role

Figure 8: *Analysis of the multifarious gains proves that individuals and communities benefit significantly from interaction with an advanced steward.*

5.2

Roles and responsibilities of stakeholders

Instantiation of bottom-up stewardship by organisations requires fulfilment of a set of roles and responsibilities assigned to each stakeholder within the health data sharing ecosystem. Ecosystem support was found critical for harnessing network gains afforded by stewardship.[40] As organisations expand and nuance their data governance practices to facilitate increased stakeholder engagement, the value created by health data stewardship is enhanced for patients and the broader communities they belong to. Data unlocked through ecosystem collaboration can effectively address gaps in healthcare delivery to tailor solutions to benefit individuals and communities.

Significantly, the framework of stewardship envisaged by this study places individuals and communities at the heart of the healthcare ecosystem because empowering them to effectively take control and manage their data is the goal of stewardship. They entrust providers, civil society and academia with their data in exchange for innovation in drug development and treatment management, enhanced quality of services and reduced cost of accessing healthcare, thus unlocking the wider societal value of data.

In reality, however, patients and caregivers have been denied visibility of sharing decisions[41], with limited avenues to engage with other stakeholders and register their preferences. Therefore, it becomes necessary to facilitate multi-stakeholder collaboration among healthcare service providers, and civil society organisations as well as individuals and communities to instantiate responsible stewardship at the organisational level. Figure 9 provides an illustration of what multi-stakeholder collaboration within the healthcare ecosystem looks like.

---

[40] "The next wave of healthcare innovation: The evolution of ecosystems", Singhal et al., McKinsey (June 23, 2020), retrieved June 25, 2021 from https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-next-wave-of-healthcare-innovation-the-evolution-of-ecosystems

[41] "What Is the NHS 'Data Grab' and What Does It Mean for Patients?", Kent, Medical Device Network (June 17, 2021) retrieved June 30, 2021 from https://www.medicaldevice-network.com/features/nhs-data-grab-gpdpr/

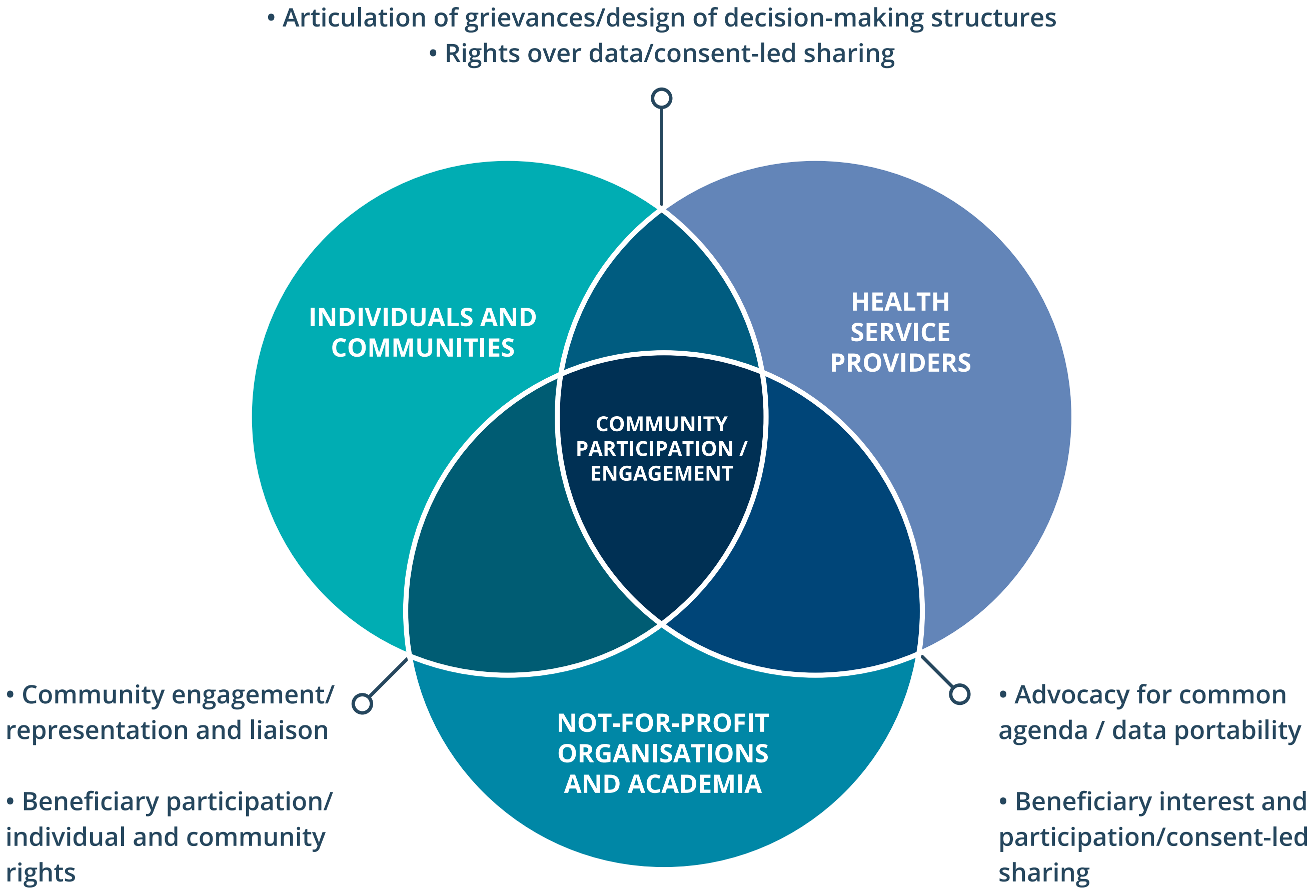


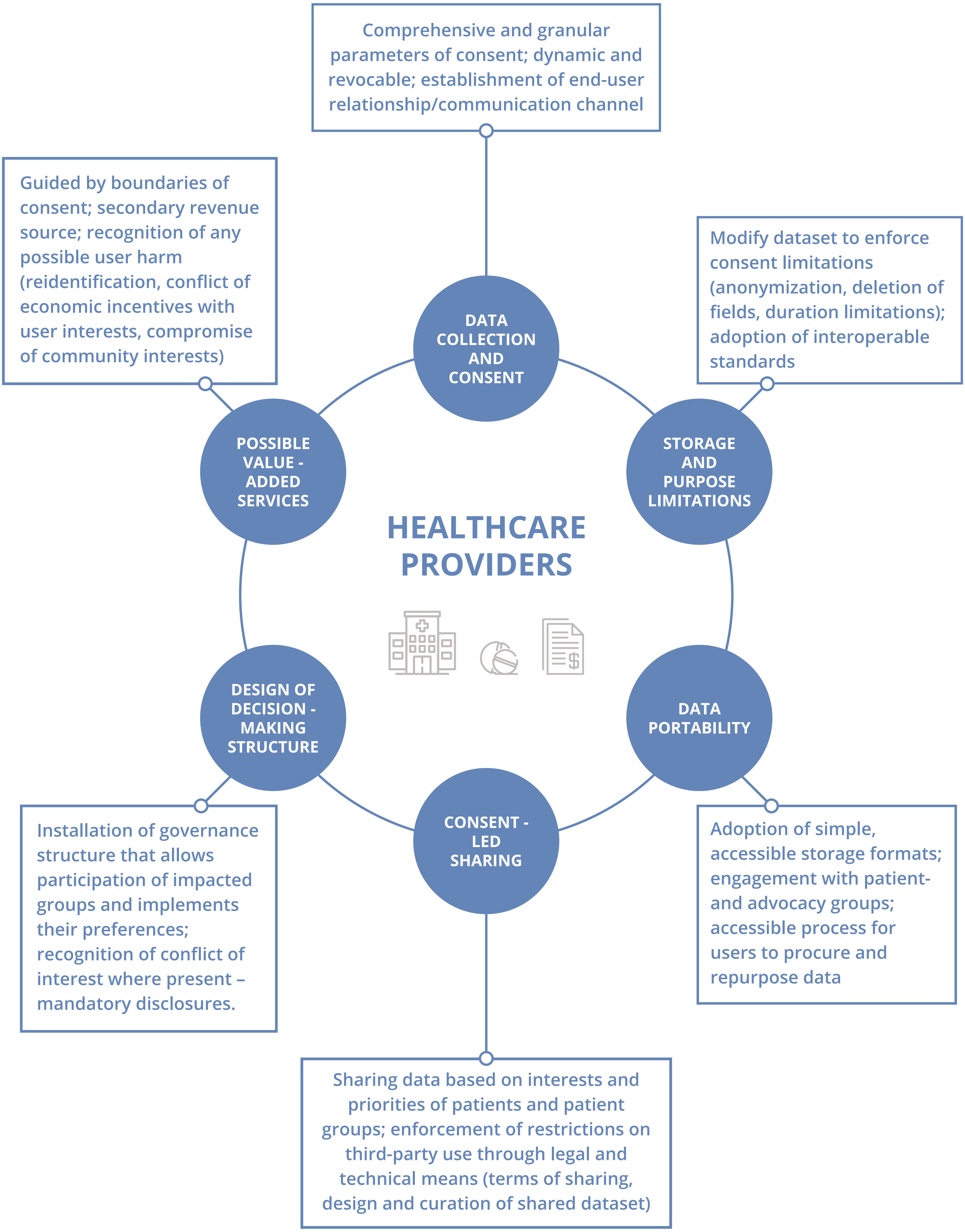
Figure 9: *Overlaps indicate complementary roles played by the three sets of stakeholders. Community participation / engagement is a central tenet of stewardship.*

Further, the study includes an analysis of the specific roles and responsibilities of each stakeholder. In this conceptual framework, role refers to the functions carried out by stakeholders while responsibility indicates a minimum standard for fulfilling a role.[42] Accordingly, each stakeholder plays a distinct role in the ecosystem, bringing different values and perspectives to the health data value chain. Outlined below are the roles of stakeholders, alongside the functions and responsibilities to be discharged as part of fulfilling the roles.

Providers constitute the first link in the data value chain, collecting and processing vast amounts of sensitive health information as part of care delivery. Therefore, introducing dynamic consent controls is imperative to ensure that patients and their caregivers can participate meaningfully in sharing decisions to register their preferences. For providers, instituting coherent internal data governance frameworks and end-use limitations on downstream uses of data should be the focus while instantiating stewardship. Their roles and accompanying responsibilities are represented in Figure 10.

---

[42] Note: Roles are denoted by coloured circles and the blurbs explain the accompanying responsibilities.



Comprehensive and granular parameters of consent; dynamic and revocable; establishment of end-user relationship/communication channel

Modify dataset to enforce consent limitations (anonymization, deletion of fields, duration limitations); adoption of interoperable standards

Adoption of simple, accessible storage formats; engagement with patient- and advocacy groups; accessible process for users to procure and repurpose data

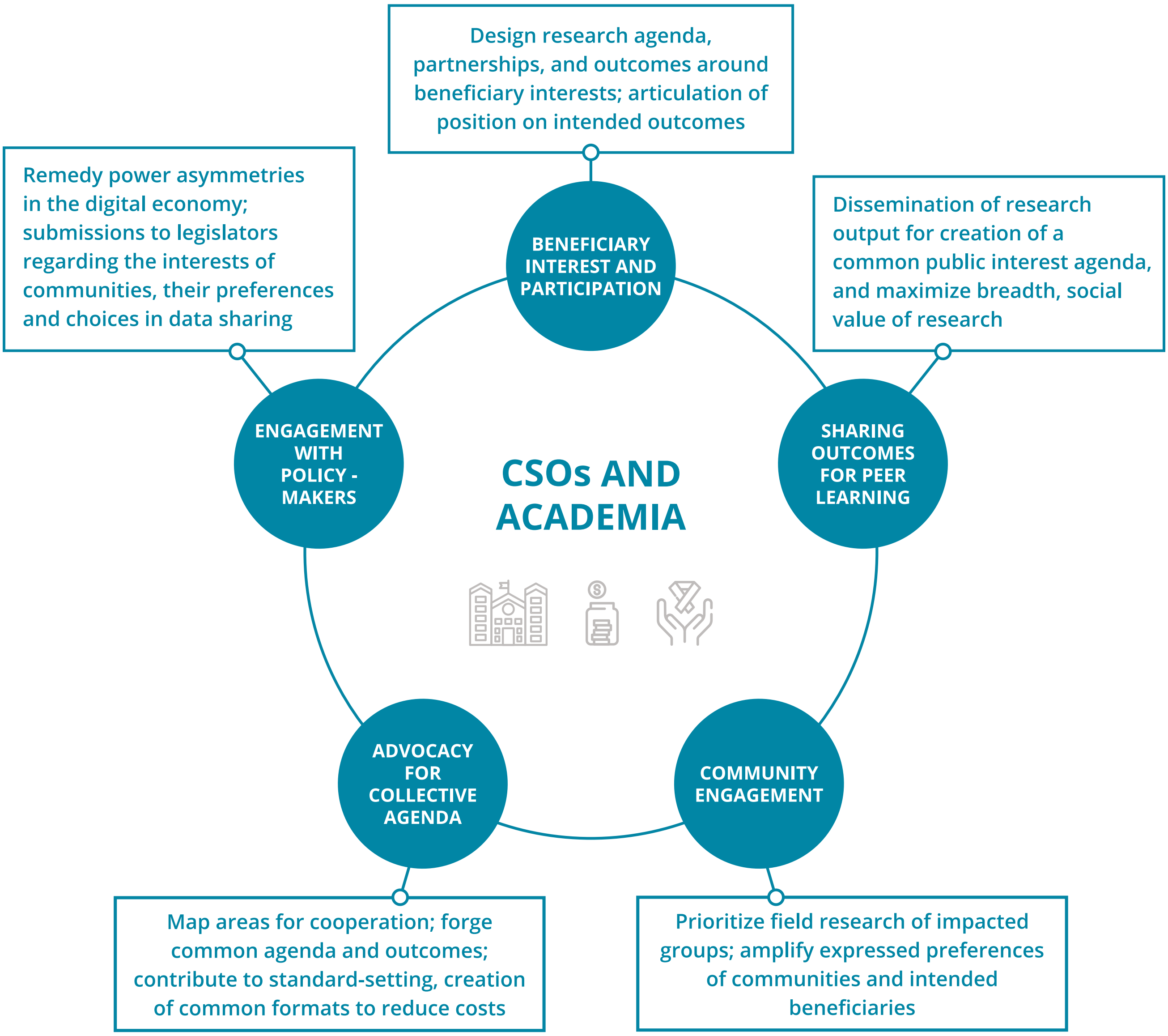
Sharing data based on interests and priorities of patients and patient groups; enforcement of restrictions on third-party use through legal and technical means (terms of sharing, design and curation of shared dataset)

Installation of governance structure that allows participation of impacted groups and implements their preferences; recognition of conflict of interest where present – mandatory disclosures.

Guided by boundaries of consent; secondary revenue source; recognition of any possible user harm (reidentification, conflict of economic incentives with user interests, compromise of community interests)

Figure 10: *As part of their functions, healthcare providers should focus on revising internal data governance practices to facilitate greater community participation in data sharing.*

On the other hand, civil society organisations and academia play similar roles within the health data ecosystem in that they help fill gaps in healthcare delivery, research and innovation, particularly in under-served contexts. Consequently, the two actors interact with communities actively, engaging in complex relationships of negotiation and representation that leave them flush with a wealth of health information. Stewarding this data in a manner representative of community interests is critical to unlock societal benefit in a secure and responsible manner. This research has demonstrated that civil society organisations and academia would benefit from introducing collaborative exercises to promote data literacy, strengthen community participation and liaise with policymakers in their attempts to instantiate bottom-up data governance. Figure 11 outlines the roles and responsibilities of civil society organisations and academia within the health data ecosystem.



Design research agenda, partnerships, and outcomes around beneficiary interests; articulation of position on intended outcomes

Dissemination of research output for creation of a common public interest agenda, and maximize breadth, social value of research

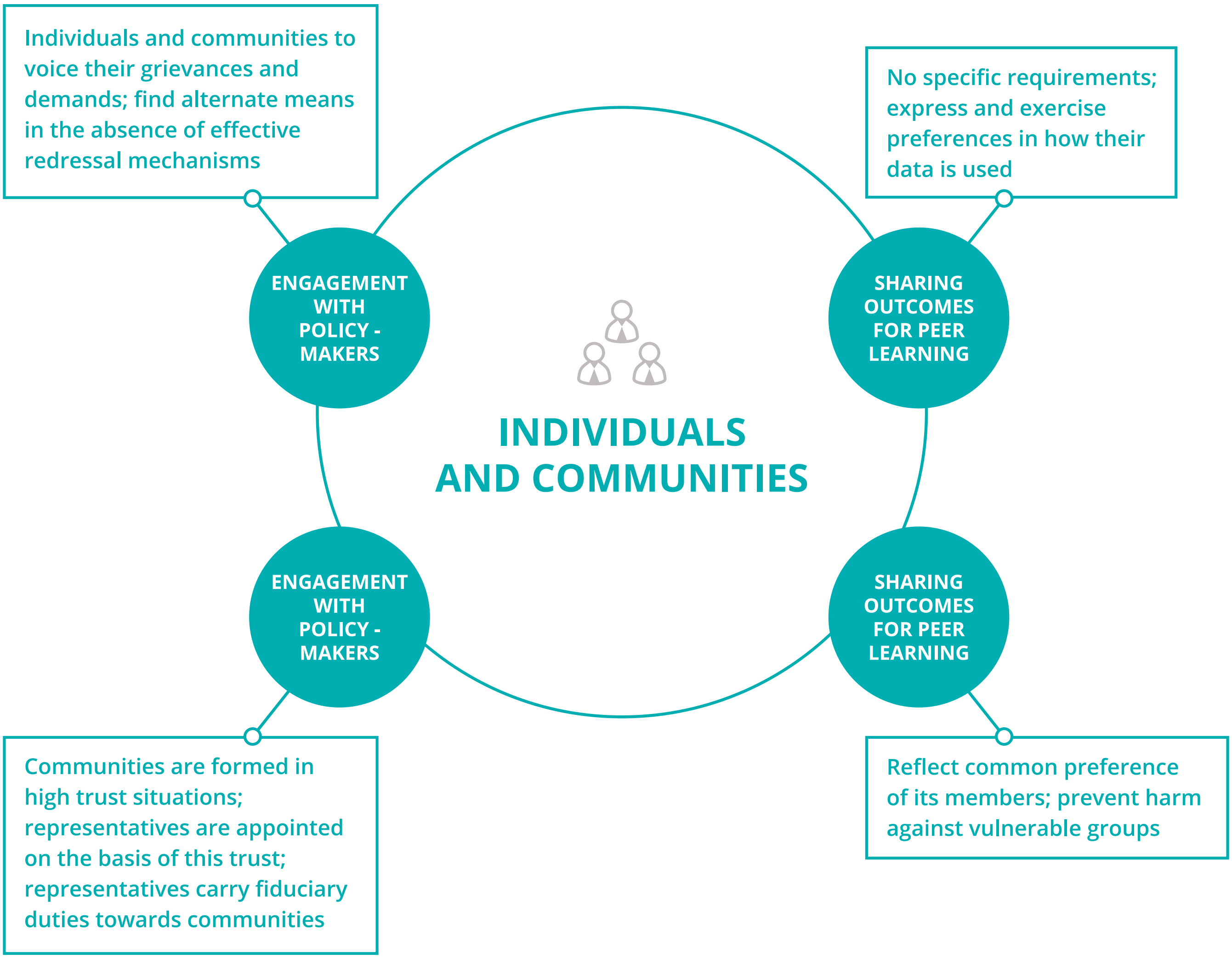
Prioritize field research of impacted groups; amplify expressed preferences of communities and intended beneficiaries

Map areas for cooperation; forge common agenda and outcomes; contribute to standard-setting, creation of common formats to reduce costs

Remedy power asymmetries in the digital economy; submissions to legislators regarding the interests of communities, their preferences and choices in data sharing

***Figure 11:** Civil society organisations and academia have emerged as crucial vehicles for representation of community voices. Their activities should be directed towards empowering communities to derive maximum societal benefit from data.*

Lastly, instantiating participatory models of stewardship at the organisational level gives rise to reciprocal responsibilities for individuals and communities. Leveraging their positions as generators of information and harnessing the power of collective bargaining are indispensable for meaningfully asserting their data rights. The conferring and recognition of data rights, in turn, helps communities hold other stakeholders – public institutions, providers, civil society organisations and academia – accountable. Significantly, this research has found that enforcement of communities' right to privacy as well as the right to the value derived from data constitutes the basis of an enabling environment for stewardship. Accordingly, a rights-based approach to community engagement within the larger health data ecosystem is illustrated in Figure 12.



Individuals and communities to voice their grievances and demands; find alternate means in the absence of effective redressal mechanisms

No specific requirements; express and exercise preferences in how their data is used

Communities are formed in high trust situations; representatives are appointed on the basis of this trust; representatives carry fiduciary duties towards communities

Reflect common preference of its members; prevent harm against vulnerable groups

Figure 12: *Individuals and communities are at the heart of the health data ecosystem. Stakeholders' actions must be directed towards preserving their agency and promoting their participation in decision-making.*

06

The Way Forward

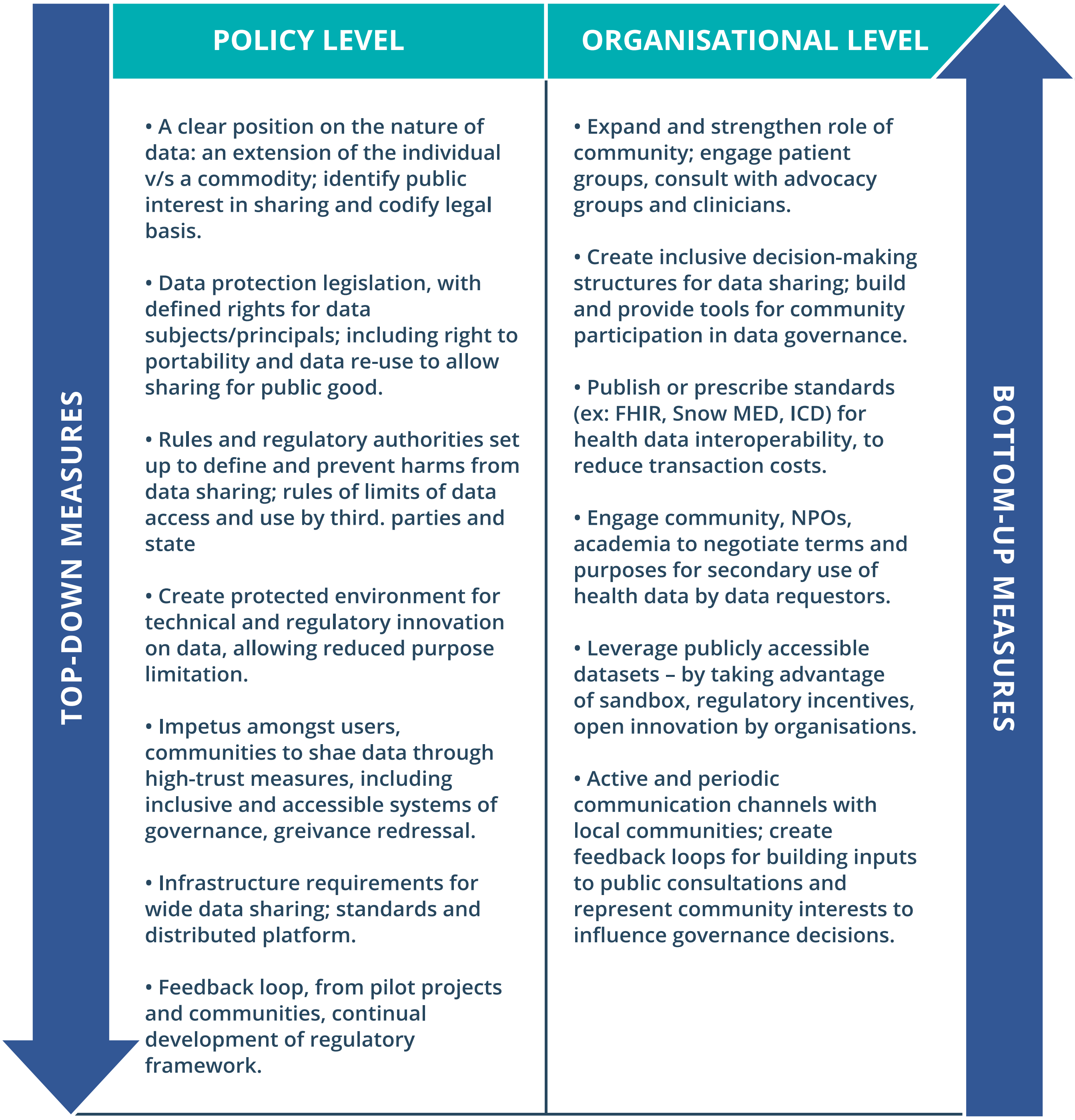
This study appraises the landscape of health data stewardship and aims to provide a comprehensive guide for instantiating stewardship within organisations. The value addition of stewardship combined with an analysis of what it could mean for stakeholders creates compelling incentives for its uptake. Translating theory into practice is another important facet of this analysis that is best demonstrated through an exploration of best practices and deep dives into features of a "good steward" alongside roles and responsibilities of stakeholders in the health data ecosystem.

As the findings suggest, bottom-up engagement allows for participatory forms of collaboration among stakeholders, with the aim of upholding the interests of communities during health data sharing. On their part, individuals and communities stand to gain through enhanced access to health services as well as greater transparency and accountability, facilitated by stewardship. Through the presence of dynamic consent controls and purpose-led sharing, stewards are able to create user-centric models of data governance that extend beyond compliance-bound obligations to enhance the agency of individuals and communities. Such an approach, that prioritises accountability, is cognisant of the limitations of existing models of consent and attempts to focalise transparency as an ethical imperative to be cultivated in the process of health data sharing.

Significantly, this study has highlighted the salient role of civil society organisations (CSOs) in bridging the trust deficit between policymakers, providers, and communities by addressing concerns around informational asymmetry, digital literacy, and community participation. Too often, health policy and care provisioning are decided unilaterally by policymakers and experts, uninformed by the needs of communities and wider populations who are ultimately affected by these actions.

In such contexts, CSOs play the role of critical intermediaries between communities and other stakeholders to ensure that the benefits of an expanded health system reach those it claims to serve. Retrospectively, communities constitute and rely on civil society to create momentum to participate in health system governance and build bottom-up momentum for stewardship. Thus, CSOs represent sites of resistance, forging bottom-up solidarities that can act as powerful counters to the influence of providers and policymakers within the health data ecosystem.

At the same time, insights from conversations with practitioners have been instrumental in manifesting the shortcomings of an approach to stewardship relying exclusively on bottom-up engagement for its realisation. On the contrary, this study reinforces the importance of top-down policy interventions by public institutions to strengthen calls for instantiation of equitable data governance frameworks within organisations. Creating functional feedback loops between accountable regulatory systems, health service providers and communities of patients and caregivers has emerged as vital to building a resilient steward.



POLICY LEVEL

- A clear position on the nature of data: an extension of the individual v/s a commodity; identify public interest in sharing and codify legal basis.
- Data protection legislation, with defined rights for data subjects/principals; including right to portability and data re-use to allow sharing for public good.
- Rules and regulatory authorities set up to define and prevent harms from data sharing; rules of limits of data access and use by third. parties and state
- Create protected environment for technical and regulatory innovation on data, allowing reduced purpose limitation.
- Impetus amongst users, communities to shae data through high-trust measures, including inclusive and accessible systems of governance, greivance redressal.
- Infrastructure requirements for wide data sharing; standards and distributed platform.
- Feedback loop, from pilot projects and communities, continual development of regulatory framework.

ORGANISATIONAL LEVEL

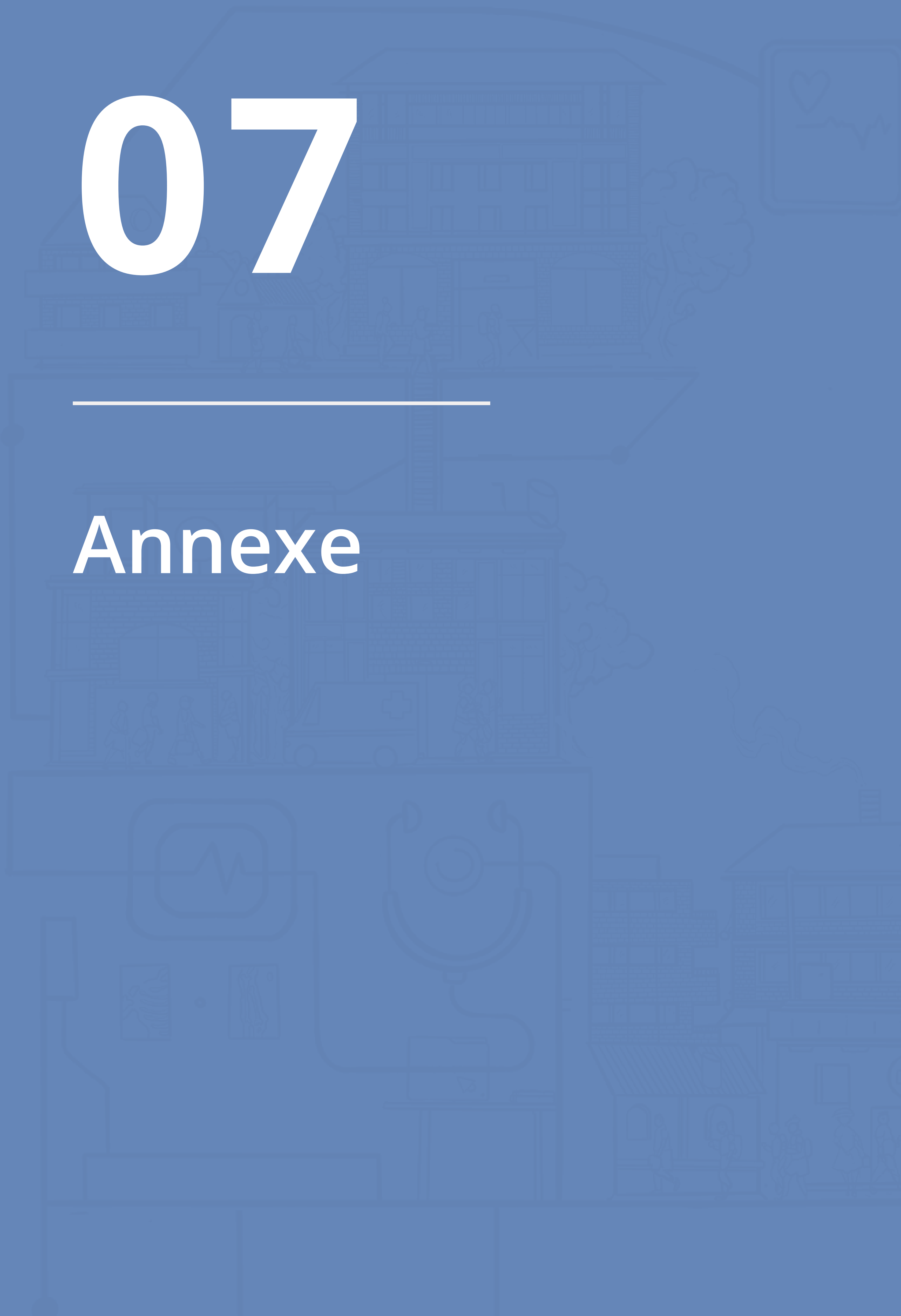
- Expand and strengthen role of community; engage patient groups, consult with advocacy groups and clinicians.
- Create inclusive decision-making structures for data sharing; build and provide tools for community participation in data governance.
- Publish or prescribe standards (ex: FHIR, Snow MED, ICD) for health data interoperability, to reduce transaction costs.
- Engage community, NPOs, academia to negotiate terms and purposes for secondary use of health data by data requestors.
- Leverage publicly accessible datasets – by taking advantage of sandbox, regulatory incentives, open innovation by organisations.
- Active and periodic communication channels with local communities; create feedback loops for building inputs to public consultations and represent community interests to influence governance decisions.

***Figure 13:** Instantiating stewardship in the health sector requires action at two levels: top-down policy interventions by the state and bottom-up community engagement.*

This study is best viewed as a journey towards stewardship, beginning with the problematisation of the current asymmetric health data economy where benefits from aggregated health information accrue overwhelmingly to providers alone. Stewardship presents a compelling alternative to this hegemonic paradigm, proposing actions for resolution that are grounded in accountability and fairness. Building and empowering communities of concern is the first step to creating equitable frameworks for health data governance and ensuring data is employed for public welfare.

07

Annexe



List of experts interviewed as a part of this study

Functional stewards:

Ernst Haffen
MiData

John Wilbanks and Megan Doerr
Sage Bionetworks

Rob Owen
RapidSOS

Javier Creus
Saluscoop

Wendy Charles and Rob Lubeck
BurstIQ

Scott Kahn
LunaDNA

Kaja Wasik
Variant Bio

Deven McGraw
Ciitizen

Adi Bergerzon
Healthcare Israel

Carrie Walter
Clue

Juan V Dura
Valencia Data

Ajay Nair
Swasth

Taunton Paine and Cheryl Jacobs
National Institute of Health

Graham Mecredy
IC/ES (formerly Institute for Clinical and Evaluative Studies)

Regina Estuar
FASSSTER

Moritz Kraemer and team
Global.health

Antti Piirainen
Findata

Mad Price Ball and Bastiaan Tzovaras
Open Humans

Jan Leindals
Diabetes.services

Ecosystem enablers:

Isaac Holeman
Medic Mobile

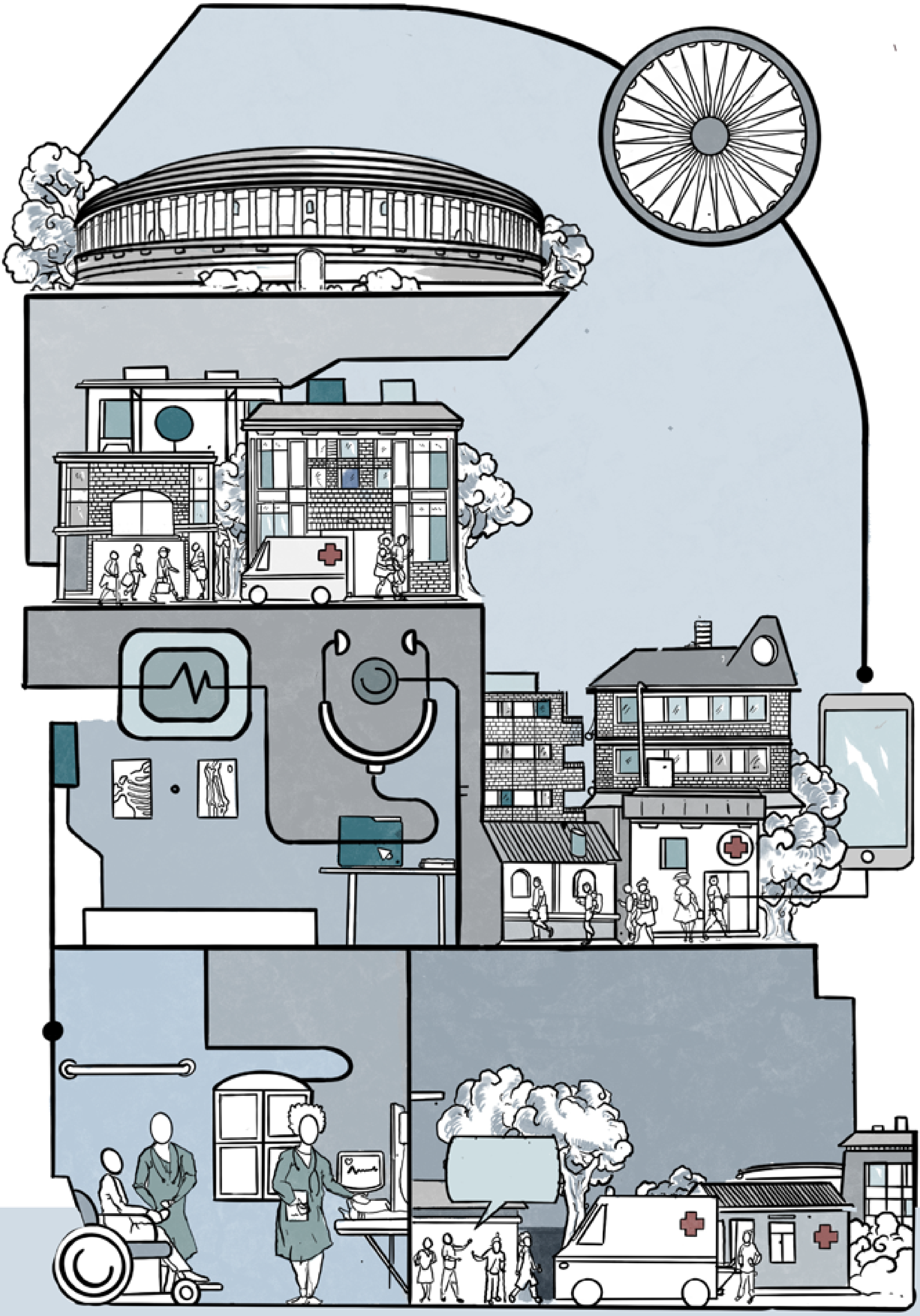
Neal Lesh
Dimagi

Atanu Garai
iKure Technologies

Kanishka Katara
PATH

Shiv Kumar
Swasti / The Catalyst Group

aapti institute



India: Health Data Stewardship Landscape and Recommendations

India: Health Data Stewardship Landscape and Recommendations

aapti institute

Soujanya Sridharan, Siddharth Manohar
and Astha Kapoor

July 2021

Acknowledgements

This report also draws from the expertise of numerous academic and industry experts and practitioners. We are grateful for their input during interviews and rigorous feedback sessions. We also thank Kevin O'Neil and Priyank Hirani of The Rockefeller Foundation for supporting this work.

Editor: Yana Bey | **Report design:** Meher Rajpal | **Illustration:** Siddhartha Iyer

Contents

Glossary

(CONCEPTUAL AND ABBREVIATIONS)

Data	Any representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means. (IT Act, PDP Bill)
Consent	Given by the individual to an entity for processing of their data; must be free, informed, specific, clear and capable of being withdrawn. (PDP Bill)
Data principal / user	Persons, both natural and legal, to whom any data relates. (PDP Bill)
Data fiduciary	An entity, company, or any individual who alone or in conjunction with others determines the purpose and means of processing of personal data. (PDP Bill)
Data trustee	As an organization, either a Government organization or a non-profit Private organization (Section 8 company / Society / Trust), that is responsible for the creation, maintenance, data-sharing of
Data trustee	High-value Datasets in India. (NPD Report) The data custodian is an entity that undertakes the collection, storage, processing, use, etc. of data. Typically, it is the data custodian that has a relationship with the consumer from whom data is collected. (NPD Report)
High Value Dataset	An HVD is a dataset that is beneficial to the community at large and shared as a public good, subject to certain guidelines pertaining to the management of an HVD and data sharing. (NPD Report)
PDP Bill	The Personal Data Protection Bill, 2019
NDHM	National Digital Health Mission, 2020
NPD Report	Revised Report by the Committee of Experts on Non-personal Data Governance Framework , 2020

01

Introduction

> India has made remarkable strides in enhancing health outcomes over the past three decades – life expectancy has increased by 12 years from 57.8 years in 1990 to 69.6 years in 2019,[1] infant mortality has reduced to 28.3 deaths in 2019 from 88.9 deaths per 1,000 live births in 1990,[2] and the maternal mortality ratio has declined from 370 deaths per 100,000 live births in 2000 to 145 deaths in 2017.[3]

However, the overall effectiveness of health systems is undermined by certain infrastructural gaps – availability of physicians is considered to be alarmingly low, with only one qualified doctor for every 1,511 members of the population, way below the World Health Organisation (WHO) prescribed norm of 1:1000 physician:population ratio. Similarly, India faces an acute shortage of nurses, with a nurse-to-population ratio of 1:670, while the corresponding WHO recommendation is 1:300.[4]

This highlights an essential paradox in India's development story – on the one hand, the nation's growth is considered a success, with impressive economic gains achieved in the post-liberalisation period. But another, more insidious tangent to this growth is the rise in inequalities[5] and persistent cuts in public investment in health.[6] Adoption of digital health solutions presents a new tool to leapfrog India's healthcare evolution.[7] Consequently, such solutions rely on instituting new data pipelines to deliver services.

However, reliance on technological innovation also throws up several regulatory and ethical issues. From a regulatory standpoint, India lacks the necessary legal protections to ensure that data, particularly sensitive health information, is used in responsible ways to improve health services. This spawns ethical considerations for citizens whose data is collected, processed and shared through mechanisms that may compromise their privacy and agency.[8]

Recognising that data is not a mere economic resource but is intrinsically tied to individuals and communities who help generate this information is

---

[1] Life expectancy at birth (total years) - India, World Bank, 2019

[2] Mortality rate, infant (per 1000 live births) - India, World Bank, 2019

[3] Modelled estimate maternal mortality ratio in India from 2000 - 2017, Statista

[4] "'1 doctor for 1,511 people, 1 nurse for 670 — Covid exposes India's healthcare 'fault lines'", Abantika Ghosh, The Print (February 11, 2021) retrieved June 30, 2021 from https://theprint.in/health/1-doctor-for-1511-people-1-nurse-for-670-covid-exposes-indias-healthcare-fault-lines/602784/

[5] "Health inequalities research in India: a review of trends and themes in the literature since the 1990s"

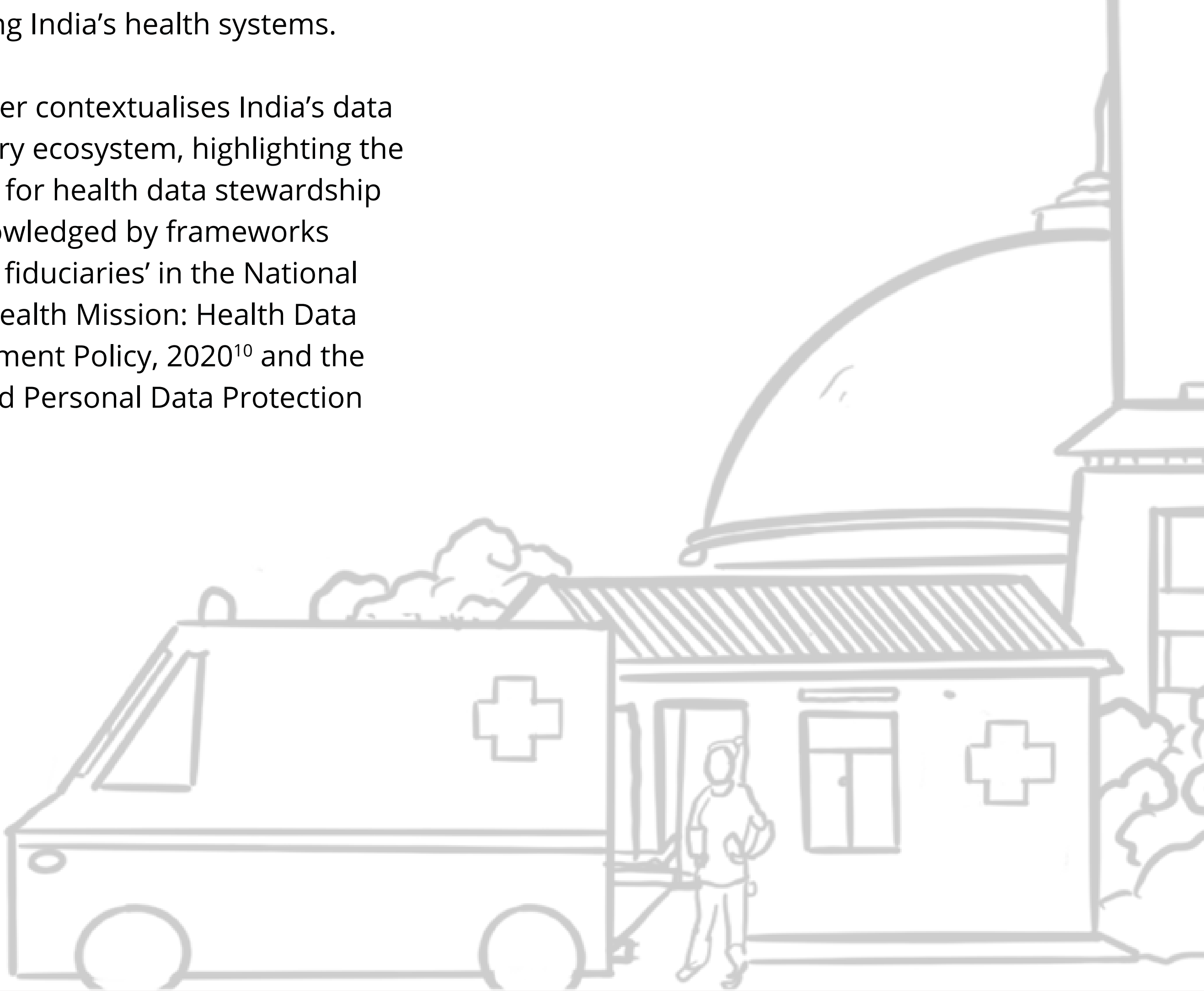
[6] "India's economy needs big dose of health spending", Mehra, Livemint (April 08, 2020), retrieved June 30, 2021 from https://www.livemint.com/news/india/india-s-economy-needs-big-dose-of-health-spending-11586365603651.html

[7] "Leapfrogging Health Outcomes in India", Vijayavargiya et al, Boston Consulting Group (September 03, 2019), retrieved June 30, 2021 from https://image-src.bcg.com/Images/BCG-Leapfrogging-Health-Outcomes-Report_tcm9-227963.pdf

[8] "Comments to the National Digital Health Mission: Health Data Management Policy", Mohandas et al, The Centre for Internet and Society (September 21, 2021), retrieved June 30, 2021 from https://cis-india.org/internet-governance/comments-to-national-digital-health-mission-health-data-management-policy-pdf

critical to confer data rights that form the basis of human-centric technical innovation. It is equally important that community data be used in pursuit of its members' empowerment and to enhance quality of life by delivering better health services. Adopting data stewardship[9] – a mechanism for data governance grounded in accountability that ensures meaningful participation of citizens in data decisions – is a vital response to harness data's transformative power for radically advancing India's health systems.

This paper contextualises India's data regulatory ecosystem, highlighting the avenues for health data stewardship as acknowledged by frameworks for 'data fiduciaries' in the National Digital Health Mission: Health Data Management Policy, 2020[10] and the proposed Personal Data Protection Bill, 2019[11], as also the prescribed role of 'data trustees' in the Report by the Committee of Experts on Non-personal Data Governance Framework, 2020.[12] Additionally, it surfaces the varying impetus for adoption of stewardship and provides a roadmap for its implementation within India's regulatory landscape. The paper concludes by outlining the benefits of stewardship and its potential for steering policy in favour of empowering citizens to have control over their data.



---

[9] Data Stewardship: A Taxonomy, Aapti Institute (June, 2020), retrieved June 10, 2021 from https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/

[10] Chapter III - Consent Frameworks, National Digital Health Mission: Health Data Management Policy, 2020

[11] Chapter II - Obligations of Data Fiduciaries, The Personal Data Protection Bill, 2019

[12] Section 7.7 - Data Trustees, Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

02

Regulatory Overview

India's health data ecosystem finds itself straddling a decisive watershed in its history: on the one hand, there is increased regulatory attention to health data protection and on the other, a nascent attempt to unlock the value of data for research and innovation through layers of intermediaries such as data trustees. This follows the many regulations, at varying stages of development, under consideration, affecting how health data is processed and shared across India's jurisdiction.

The Personal Data Protection Bill, 2019,[13] (PDP Bill) is a focal legislation, albeit pending parliamentary approval, that dictates terms of ownership, localisation, consent and use of personal data, including individual health data defined as sensitive information. The proposed Non-personal Data Governance Framework, 2020,[14] (NPD Report) directs sharing of de-identified health data to the extent that it creates provisions for mandatory disclosure of data by data custodians who include healthcare providers, pharmacies, health tech platforms and related service applications. Moreover, the umbrella frameworks of a Data Protection Authority[15] (mentioned in the PDP Bill) and Non-personal Data Authority[16] (mentioned in the NPD Report) are pan-sectoral regulatory institutions that affect data sharing within healthcare, among other sectors.

Both regulations – the PDP Bill and the NPD Report – are crucial for their postulation of the fiduciary relationship[17] between data principals (the original persons from whom data is derived) and data controllers (persons and entities concerned with processing

---

[13] The Personal Data Protection Bill, 2019, retrieved June 10, 2021 from http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

[14] Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020, retrieved from https://static.mygov.in/rest/s3fs-public/mygov_160922880751553221.pdf

[15] Chapter IX - Data Protection Authority of India, The Personal Data Protection Bill, 2019

[16] Section 7.10, Revised Report on Non-personal Data Governance Framework, 2020

[17] "Fiduciary relationships as a means to protect privacy", Bailey et al, Data Governance Network (November, 2019) retrieved June 10, 2020 from https://datagovernance.org/files/research/NIPFP_Rishab_Trishee_fiduciaries_-_Paper_4.pdf

data) that forms the foundation of data stewardship mechanisms.[18] Specifically, the NPD Report calls for the institutionalisation of data trustees[19] to exercise community rights and distribute value from data more broadly.

In addition to the above mentioned laws, there are certain sector-specific regulations such as the National Digital Health Mission, 2020 (NDHM)[20] which provides the base for digital health infrastructure in India. The NDHM lays down principles for the creation of a voluntary National Health ID for individuals and entities, a federated and decentralised architecture for health data sharing among interoperable information systems and consent managers (bearing fiduciary responsibilities) to protect individuals from privacy harms. Additionally, it posits the creation of certain 'building blocks' under the Sandbox Guidelines[21], attempting to foster controlled pilots and testing for National Health IDs, DigiDoctor, Health Facility Registry e-pharmacy, health clouds and telemedicine. Lastly, the NDHM is governed by the National Health Authority (NHA) which sets conditions for its implementation.[22]

However, the abundance of regulations underpins an essentially fragmented and scattered approach to data governance that has been characteristic of India's experience. First, the PDP Bill is pending approval by Parliament[23] and as a result, India has no comprehensive protections available for personally identifiable information, including health and financial data. Second, the NDHM has been piloted across six Union Territories[24] ahead of the PDP Bill, opening sensitive health data to the possibility of misuse by data controllers and downstream users. Significantly, the NDHM policy accords sweeping powers to the NHA in order to obtain access to anonymised health information[25] without stipulating conditions and checks against abuse of power by state authorities. TThird, the NPD Report recognises collective claims over data[26], without providing necessary procedures and structures to realise the rights. Similarly, the Report imposes mandatory data sharing by data

---

[18] Data Stewardship: A Taxonomy, Aapti Institute (June, 2020), retrieved June 10, 2021 from https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/

[19] Section 7.7, Revised Report on Non-personal Data Governance Framework, 2020

[20] National Digital Health Mission: Health Data Management Policy, 2020, retrieved June 10, 2021 from https://ndhm.gov.in/health_management_policy

[21] NDHM Sandbox (August, 2020), retrieved June 10, 2021 from https://ndhm.gov.in/documents/sandbox_guidelines

[22] "About National Health Authority", Ministry of Health and Family Welfare, Govt. of India, retrieved June 10, 2021 from https://ndhm.gov.in/home/nha

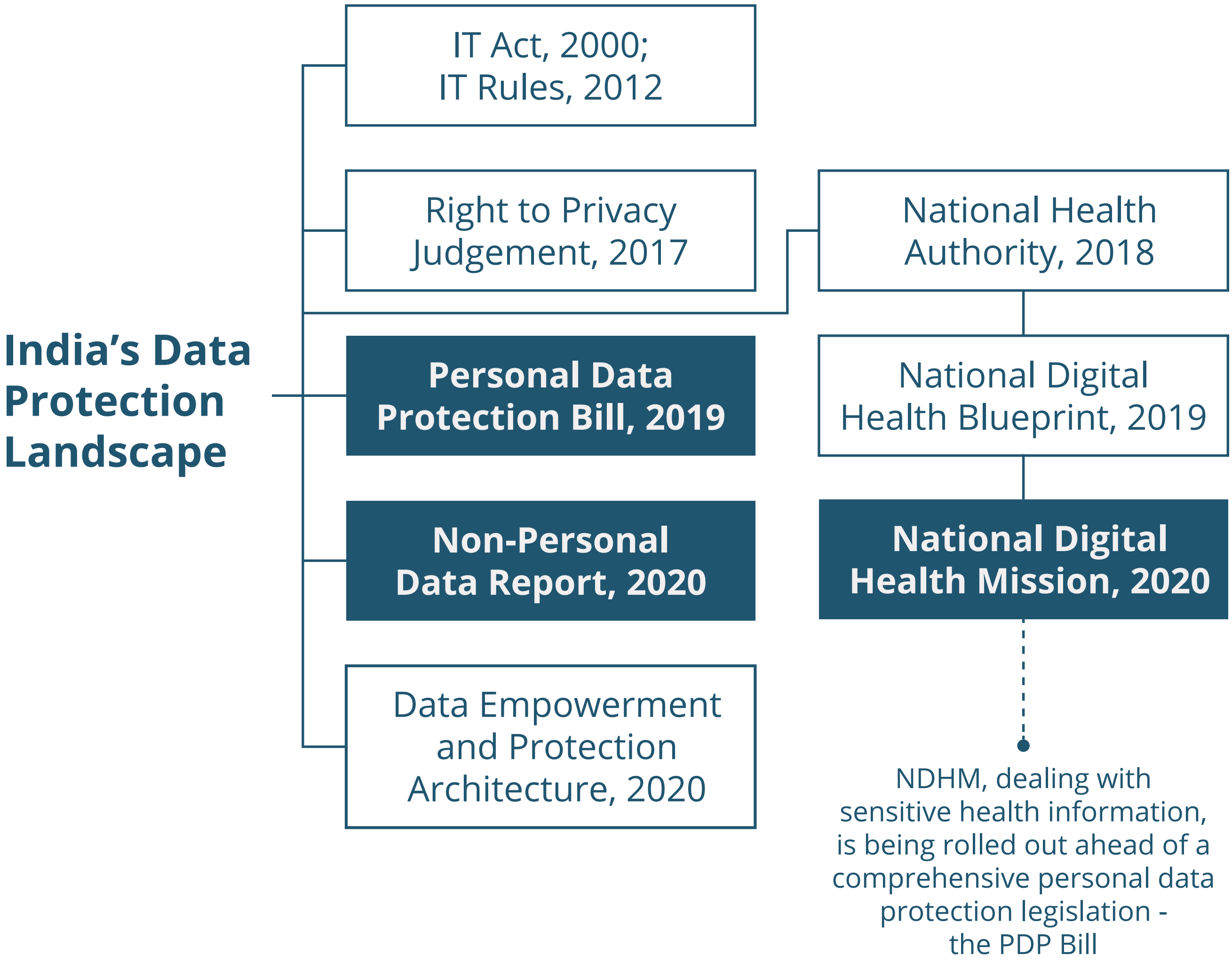
[23] "JPC on Data Protection Bill given time till winter session to submit long-pending report", Special Correspondent, The Hindu, retrieved July 26, 2021 from https://www.thehindu.com/news/national/jpc-on-data-protection-bill-given-time-till-winter-session-to-submit-long-pending-report/article35488448.ece

[24] "National Digital Health Mission: NHA invites suggestions ahead of roll out", FE Online, Financial Express (January 22, 2021) retrieved June 10, 2021 from https://www.financialexpress.com/lifestyle/health/national-digital-health-mission-nha-invites-suggestions-ahead-of-ndhm-rollout/2276327

[25] Section 29.1 – Health Data Management Policy, NDHM, 2020

[26] Section 7.1 – Non-personal Data Roles – Community, Revised Report on Non-personal Data Governance Framework, 2020

controllers which leads to centralisation of power while simultaneously disincentivising sharing of data for public good that is essential to create broad-based societal benefit.



NDHM, dealing with sensitive health information, is being rolled out ahead of a comprehensive personal data protection legislation - the PDP Bill

Figure 1: *India has multiple regulations for use, access and sharing of data. The PDP Bill, NPD Report and NDHM are relevant to this study.*

Consent is another salient feature along which the aforementioned legislation differs significantly. Meaningful consent clauses necessarily presume that data principals would authorise use of their data for purposes that provide value to them and their communities.[27]

Specificity is a crucial artifice of consent frameworks, implying that personal information such as health data should be collected, processed and shared according to explicit and stated purposes only.

---

[27] "Health data sharing being set up to fail by a regulatory mashup: roles of the NDHM, PDPB, consent managers and the NHA", Aapti Institute (June 16, 2021), retrieved June 20, 2021 from https://thedataeconomylab.com/2021/06/10/health-data-sharing/

To this end, the PDP Bill is guided by the principle of data minimisation wherein personal information sharing is consent-bound and purpose-specific; this is crucial to limit harms to individual privacy. The NPD Report, in its current form, requires consent for anonymisation of data only and data principals reserve the option to withdraw consent prior to anonymisation.[28] However, there is no requirement to collect consent for sharing of anonymised data, under the NPDR. In contrast, the provisions on consent collection in the NDHM Health Data Management Policy differ along two dimensions: one, consent is not collected for anonymisation of sensitive health information[29] and, two, for sharing of anonymised health data according conditions set forth by the NDHM implementing agency[30], upon request. Consequently, the NDHM provisions are not only in violation of the PDP Bill's data minimisation imperatives but also the NPD Report's requirement of consent for anonymisation itself.

So far, India's approach to data governance demonstrates the shortcomings of a narrow "privacy model"[31] that is overwhelmingly anchored in consent provisioning alone. In such a scenario, data principals' participation in processing is rendered meaningless as they lack the requisite awareness around data decisions to provide informed consent. A data fiduciary's role is limited to collection of consent, as is envisaged for "data fiduciaries" in the NDHM policy[32] and PDP Bill[33], without stipulating any mechanisms to hold the fiduciary itself accountable. Moreover, the possibility of "consent fatigue" emerging due to long-drawn processes (e.g. cookie notices and explanations) for collecting authorisation combined with high compliance burden makes the "privacy model" undesirable.

Alternatively, we propose an approach to data governance – data stewardship – whose roots lie in the "accountability / harms"' model. According to the "accountability model", a data fiduciary is accountable for harms to data principals arising from its own actions as well as other harms to privacy of individuals and communities. Similarly, communities can exercise claims over inferred data derived from their personal and non-personal information through their respective data fiduciaries. Consequently, not only are individual agency and privacy protected under stewardship, the value of data is distributed more widely through a framework of community rights.

---

[28] Section 5.4 (iii) - Consent for Anonymised Data, Revised Report on Non-personal Data Governance Framework, 2020

[29] Section 2.2.6 – Health Data Anonymisation and Aggregation,NDHM Strategy Overview, 2020

[30] Section 29.1 – Health Data Management Policy, NDHM, 2020

[31] "Rethinking personal data regulation in India", Amar Patnaik, The Indian Express (February, 2021), retrieved June 20, 2021 from https://bit.ly/3jEGPMz

[32] Section 4 (g) – definition of a "data fiduciary", NDHM Health Data Management Policy, 2020 retrieved from https://ndhm.gov.in/health_management_policy

[33] Chapter II – Obligations of Data Fiduciary, The Personal Data Protection Bill, 2019

03

Impetus for Health Data Stewardship

> Indian regulators have made preliminary attempts to carve out a role for data intermediaries within the PDP Bill, NPD Report and the NDHM policy. Both the PDP Bill[34] and NDHM policy[35] call for the appointment of 'data fiduciaries' who would determine the purposes and means for processing an individual's personal data, including sensitive health information.

In a similar vein, the NPD Report adopts the framework of trustees to provide for delegated representation of communities in the process of collecting and sharing de-identified data. This potpourri of intermediaries bears certain duties of loyalty and care towards communities and is obligated to steward data in accordance with their interests.

However, the regulations suffer from a narrow focus on privacy protection and blanket consent frameworks. The involvement of individuals and communities begins and ends with consent-provisioning, such that principals have no avenues to understand and engage with downstream use of their data. Communities cannot exercise control over how their data will be used in the future and by whom. Similarly, the role of fiduciaries is limited to ensuring compliance, by determining whether grounds for processing of personal and non-personal data by health service providers are in accordance with applicable regulations. Most crucially, the regulations fail to consider perspectives of communal harm, going beyond the current preoccupation with privacy as the only harm arising from data sharing, to effect real damage in the nature of discrimination and denial of access to services.[36] Lastly, the regulations fail to establish comprehensive and accessible mechanisms for grievance redressal by which data principals can hold both fiduciaries and data controllers accountable.[37]

Framing data rights to include the right to re-use information is a critical first step towards protection of community interests in the process of data sharing. This is important to

---

[34] Chapter II - Obligations of Data Fiduciary, The Personal Data Protection Bill, 2019

[35] Section 4 (g) - definition of a "data fiduciary", NDHM Health Data Management Policy, 2020 retrieved from https://ndhm.gov.in/health_management_policy

[36] "Six ways (and counting) how big data systems are harming society", Redden, The Conversation (2017), retrieved June 20, 2021 from https://theconversation.com/six-ways-and-counting-that-big-data-systems-are-harming-society-88660

[37] Comments on the Revised Report by the Committee of Experts on Non-personal Data Governance Framework, Aapti Institute (February 03, 2021), retrieved June 23, 2021 from https://thedataeconomylab.com/2021/02/03/comment-on-the-revised-report-by-the-committee-of-experts-on-non-personal-data-governance-framework/

ensure that fairness, transparency and accountability are enshrined within data protection legislation as non-negotiable first principles to enable safe data sharing within healthcare.

This paper presents one such framework for data governance – stewardship – that can upend the current monopolistic control of data by corporations and healthcare providers to put data back in the hands of communities to harness its underlying social value. Data stewardship's inherent potential lies in its ability to replace existing patterns of storing and sharing data in the health sector to distribute benefits more equitably through the data economy. Our research examines instantiation of stewardship in healthcare and the value it brings to stakeholders.[38]

At the individual and community levels, stewardship enables data principals to exercise control over their data, enhances accountability, trust and transparency in the process of sharing and facilitates collective bargaining with data requesters and fiduciaries in pursuit of group enfranchisement. For instance, India's regulations – particularly the NPR Report and NDHM – could benefit from reflecting on the experiences of organisations such as MiData[39] and SalusCoop[40] – both data cooperatives that integrate and share members' health data for citizen-driven research. Sharing decisions are authorised through a general body of members (data principals) who enjoy voting rights, thereby creating democratic and accountable stewardship models for the health sector.

For data holders or controllers, stewardship provides opportunities to collaborate with multiple actors in the ecosystem to solve pressing public issues. More importantly, stewardship creates mechanisms for participatory decision-making involving communities in data governance, increasing stakeholder faith in the process of health data sharing. Since data holders/controllers carry fiduciary duties, they are obligated to act in the best interests of the communities they serve and in the process, uphold their claims and preferences regarding data. To this end, the NPD Report should encourage bottom-up stewardship instruments such as data unions, collaboratives and trusts to actualise meaningful structures for the community participation it purports. Variant Bio[41] is one such data collaborative that works with historically marginalised populations to facilitate people-driven therapeutics.

---

[38] For a detailed analysis of the value created by health data stewardship, please refer to a corollary of this report, "Health data stewardship: Bottom-up stakeholder engagement"

[39] For more information, please visit https://www.midata.coop/en/home/

[40] For more information, please visit https://www.saluscoop.org/

[41] For more information, please visit https://www.variantbio.com/

Communities are engaged prior to the beginning of research projects, their data is then collected and used within a framework that ethically focalises community concerns.

Stewardship offers remarkable incentives to data requesters by unlocking previously inaccessible datasets in sensitive domains such as healthcare for further utilisation. Such datasets are made available in shareable and usable formats. Among its many functions, a steward provides certain value-addition services such as data curation, standardisation and interoperability. Moreover, the position of stewards as fiduciary agents demonstrates the reliability and provenance of data exchanged in the process of sharing. Born of concerted efforts by academics and technologists to address the pandemic, Global.health[42] is a unique civil society-driven data repository which provides access to real-time, anonymised epidemiological data. This is essential to bolster public health responses which are informed by open and trusted health data.

At the level of the public, stewardship abounds in possibilities to expand social value derived from data. It protects and empowers data principals to negotiate data decisions, while concomitantly offering expertise and guidance as a fiduciary. Subsequently, stewards forge partnerships with other stakeholders – providers, civil society organisations and academia – to channel data for productive, communitarian ends. Indian policymakers could refer to the Finnish model of health data sharing – FinData[43] – for pioneering a mechanism of making de-identified health and social data available for secondary use in research, policymaking and development interventions. Health and social data are collated from siloed sources across public agencies and authorised for sharing through data permits (to gain access to granular datasets) or data requests (to obtain statistical-level information and other insights from data).

A summary of the arguments presented in this section has been provided in Figure 2.

---

[42] For more information, please visit https://global.health/about/

[43] For more information, please visit https://findata.fi/en/what-is-findata/

Data Principals / Subjects

- Exercise more control over data.
- Receive transparency and accountability with regard to data use and access - increase trust in the process of sharing.
- Negotiate and collectively bargain with data requestors and fiduciaries.

Data Holders

- Safeguard data principal rights on data in the process of sharing – earn stakeholder trust in the process.
- Enable harnessing public value of data through trusted mechanisms of sharing.
- Collaborate with multiple actors to solve public questions.

Data Requestors

- Access large data sets other in non-shareable domains and formats, at reduced cost and increased efficiency.
- Trust the process of sharing and provenance of data.
- Collaborate with multiple actors to solve public questions.

Figure 2: *Stewardship can replace existing patterns of storing and sharing data to distribute value across stakeholders.*

04

Instantiating Health Data Stewardship in India

Effective and responsible data sharing in healthcare requires the deployment of varied models of stewardship[44] that make quality data available, while being mindful of the associated risks that may arise during the process. Consequently, sharing health information presents several institutional, ethical and technical challenges that can be resolved by instantiating stewardship. However, operationalising this within India's health data ecosystem hinges on the calibration and alignment of three correlated factors:

1

Legislative reforms

This follows the recognition of data rights of principals and prevention of harms to their privacy and autonomy. Establishment of data rights of individuals and communities (data principals) over raw and derived data, though recognised by the PDP Bill and NPD Report, respectively, is pending operationalisation as statutory framework. Harms included within these proposed pieces of legislation have to be reimagined in view of the potential of data processing to inflict individual and collective damage.[45] Similarly, the PDP Bill – the primary data protection legislation – should lay out mechanisms for grievance redressal to hold data processors and data fiduciaries accountable for their actions. Lastly, the NPD Report and NDHM policy should delineate clear purposes and processes for unlocking health information, in a manner that respects the interests of individuals and communities generating the data.

2

Regulatory frameworks

Such frameworks are necessary to facilitate stakeholder collaboration, serving as a space for communities and data requesters to negotiate fair terms for re-use of their health data To this end, evolving avenues for community organisations by instituting mechanisms for public consultation is important for communities to exercise choices over their collective data. In its current form, the NPD Report recognises community rights to economic value derived from own data.[46] However, it fails to outline comprehensive structures and processes for participatory data governance that is crucial to devolve power to individual data principals and communities. This recommendation is also particularly salient for the NDHM policy whose governance architecture – the National Digital Health Blueprint – calls for the establishment of decentralised leadership and decision-

---

[44] For more information on the different models of stewardship, please refer to 'Data Stewardship: A Taxonomy' by Aapti Institute, available at https://thedataeconomylab.com/2020/06/24/data-stewardship-a-taxonomy/

[45] "The Data Delusion: Protecting individual data isn't enough when the harm is collective", Tisne, Stanford Cyber Policy Centre (July, 2020) retrieved June 30, 2021 from https://cyber.fsi.stanford.edu/publication/data-delusion

[46] Section 7.1 - Non-personal Data Roles, Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

making as part of its implementation.[47] Consequently, the resultant burden on state agencies seeking to share and request data is reduced by minimising the role of top-down regulatory mechanisms to mediate data processing. Finally, articulating clear sharing norms, grounded in consent and purpose-driven principals, is indispensable to facilitating secure data exchange.

3

Enabling environment

This relates to an ecosystem that opens data for technical and regulatory innovation. Creating terms for safe testing for low-risk experiments involving health data, using anonymised and synthetic datasets. The NDHM Sandbox guidelines deliver on this recommendation,[48] but are limited to pilots involving technical innovation. Expanding the scope to include regulatory experiments is significant for building accountable legislation, informed by bottom-up community perspectives. Additionally, Indian regulators could look to constitute an industry working group to understand the specific needs of health information providers and explore ways to reduce transaction costs incurred on transitioning to prescribed data sharing formats such as ICD-10/11, FHIR-R4, LOINC and SNOMED-CT under the NDHM policy.[49] State support is imperative to ease compliance burden for the period of testing and beyond.

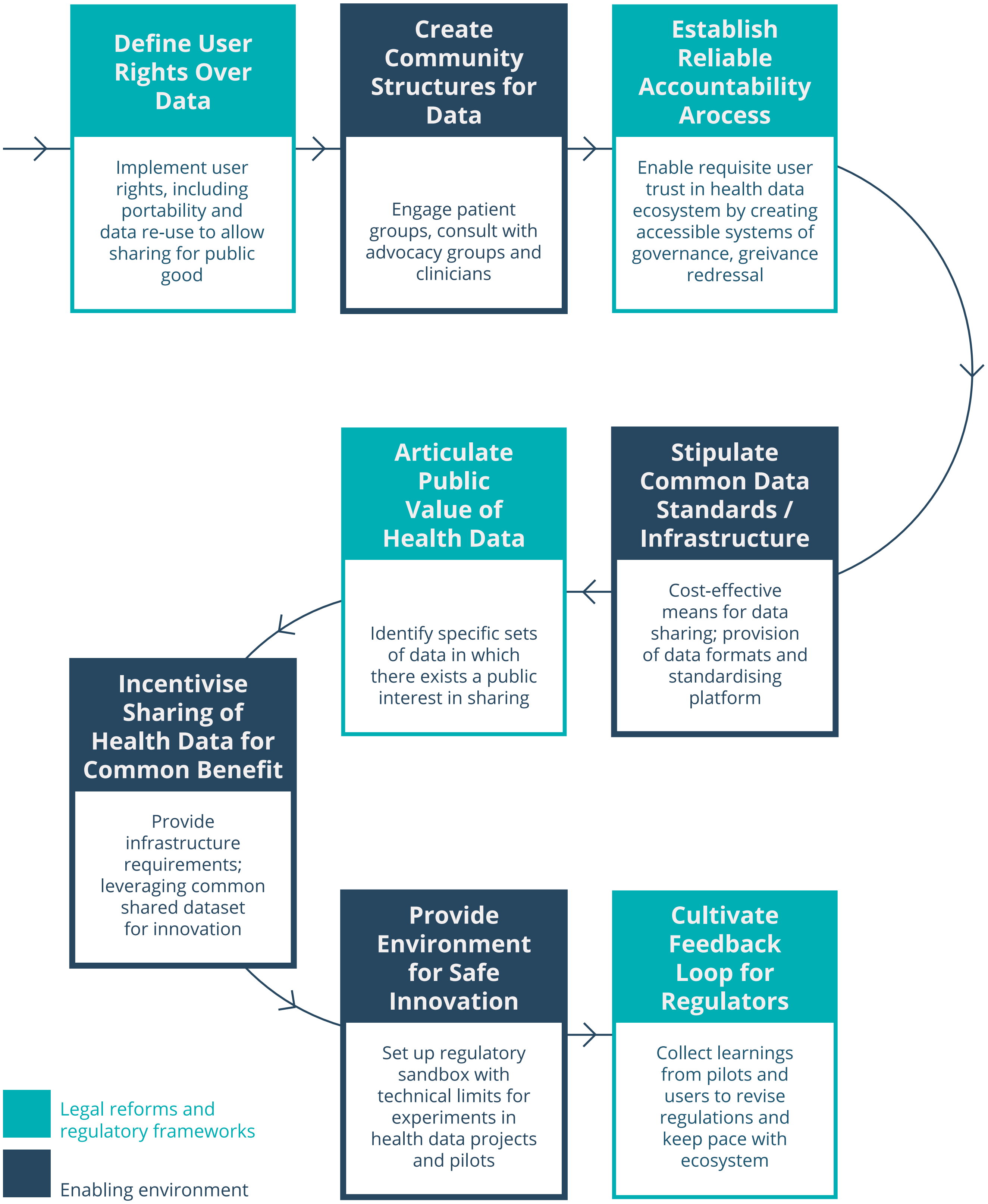
The factors explained above function as forerunners to a roadmap (see Figure 3) for implementing health data stewardship. Read together, the roadmap and its constituent elements, i.e. legal reforms, regulatory frameworks and enabling environment, provide a series of successive milestones that should be taken up by policymakers at the national level to build a thriving and secure health data economy.

---

[47] Section 4.2 - Essential Elements of the National Digital Health Mission, National Digital Health Blueprint, 2019. Retrieved June 30, 2020 from https://www.nhp.gov.in/NHPfiles/National_Digital_Health_Blueprint_Report_comments_invited.pdf

[48] NDHM Sandbox Guidelines (August, 2018), retrieved June 30, 2020 from https://ndhm.gov.in/documents/sandbox_guidelines

[49] Chapter 2.2.5 - Formats and Adoption of Standards for Health Data, NDHM Strategy Overview



Define User Rights Over Data

Implement user rights, including portability and data re-use to allow sharing for public good

Create Community Structures for Data

Engage patient groups, consult with advocacy groups and clinicians

Establish Reliable Accountability Arocess

Enable requisite user trust in health data ecosystem by creating accessible systems of governance, greivance redressal

Stipulate Common Data Standards / Infrastructure

Cost-effective means for data sharing; provision of data formats and standardising platform

Articulate Public Value of Health Data

Identify specific sets of data in which there exists a public interest in sharing

Incentivise Sharing of Health Data for Common Benefit

Provide infrastructure requirements; leveraging common shared dataset for innovation

Provide Environment for Safe Innovation

Set up regulatory sandbox with technical limits for experiments in health data projects and pilots

Cultivate Feedback Loop for Regulators

Collect learnings from pilots and users to revise regulations and keep pace with ecosystem

Figure 3: *A robust ecosystem for health data sharing depends on state action to create infrastructure for stewardship.*

This roadmap surfaces opportunities that governments can leverage to create a flexible, long-term strategy comprising regulations and action-oriented initiatives to instantiate health data stewardship. While certain milestones – defining user rights, creating community structures and establishing processes for accountability – have been included within the PDP Bill and NPD Report, the NDHM lags far behind on coherently assimilating these provisions. Stewardship can help solve for these inconsistencies by purposefully centring a paradigm that protects data rights and privacy of principals while concurrently encouraging secure sharing for public welfare.

05

Deep Dives: Milestones in the Roadmap

India, being in the early stages of development of a health data ecosystem, is trammelled by a myriad of capacity and policy constraints. Therefore, ushering in a new regime for health data sharing should follow a diagnosis of these constraints, prioritisation of data protection, and lastly, creation of the necessary technical infrastructure to facilitate secure information exchange.

Accordingly, in this section, the analysis examines each of the eight milestones defined in our proposed roadmap for instantiating health data stewardship in India. This begins with problematising India's policy position regarding each category of milestone, followed by submissions on how the positions can be revised bearing in mind the key tenets of stewardship[50] and concluding with recommendations for lawmakers to modify appropriate policies.

---

[50] For more information on key tenets of data stewardship, please refer to an allied report, "Health data stewardship: Bottom-up stakeholder engagement"

MILESTONE 1

Define user rights over data

Institutionally-backed data rights constitute the basis of participation of individuals and communities in the data economy. Further, meaningful engagement of data principals in the process of sharing is crucial to the vision of stewardship. However, the PDP Bill (primary data protection legislation) and NDHM policy (the basis for India's health data ecosystem) are marred by inconsistencies, particularly with regard to the role envisaged for the National Health Authority.

Where India stands:

Individual user (data principal) rights are set to be established under the PDP Bill. This includes rights to data access, erasure, portability, and the right to be forgotten.[51] Processing of data under the PDP Bill is subject to purpose and collection limitations, in pursuance of data minimisation principles.[52] Significantly, the right to portability is key to enabling re-use of data and to allow data principals to authorise use of their data for collective benefit. However, the NHA's expansive powers to compel mandatory data disclosure, as prescribed by the NDHM policy, is diametrically opposed to the principle of minimisation inherent in the PDP Bill.[53]

How India should proceed:

The NDHM policy provisions ought to be harmonised with the sector-agnostic, primary data protection legislation that is the PDP Bill. The NHA's ability to mandate access to de-identified health data should be curtailed by including provisions for consent-driven anonymisation, as purported by the PDP Bill. Additionally, Indian regulators would benefit from looking at the **European Union's General Data Protection Regulation, 2016**[54] **and Digital Services Act, 2020**[55] that enable safe re-use of data for research purposes through consent-driven anonymisation and sharing.

---

[51] Chapter V - Data Principal Rights, The Personal Data Protection Bill, 2019

[52] Chapter II - Obligations of Data Fiduciary, The Personal Data Protection Bill, 2019

[53] "Health data sharing being set up to fail by a regulatory mashup: roles of the NDHM, PDPB, consent managers and the NHA", Aapti Institute (June 16, 2021), retrieved June 23, 2021 from https://thedataeconomylab.com/2021/06/10/health-data-sharing/

[54] "Viewing the GDPR Through a De-Identification Lens: A Tool for Clarification and Compliance", Mike Hintze (October 31, 2016), retrieved June 23, 2021 from https://fpf.org/wp-content/uploads/2016/11/M-Hintze-GDPR-Through-the-De-Identification-Lens-31-Oct-2016-002.pdf

[55] "Platform access is a lynchpin of EU Digital Services Act", Engler, Brookings (January 15, 2021) retrieved June 23, 2021 from https://www.brookings.edu/blog/techtank/2021/01/15/platform-data-access-is-a-lynchpin-of-the-eus-digital-services-act/

The **US's Health Insurance Portability and Accountability Act** (HIPAA), **1996**[56] performs functions similar to the GDPR. The HIPAA establishes conditions for use and disclosure of data, subject to patient authorisation. But, it leaves all initiative and costs of processing data to be borne by entities concerned with such activities.[57] The HIPAA presents a cost-effective mode of data processing that can be reflected in Indian health data protection policies.

Further, the **Personal Health Information Protection Act, 2004,** of Ontario, Canada, requires consent for collection, use and disclosure of personal health information. Additionally, users (data generators) reserve the right to expressly prohibit the use of their personal health information for purposes specified by them.[58] This adds a layer of nuance to data re-use, attempting to truly comply with individual interests and preferences. Indian regulators must imbibe this principle to deliver patient-centric health services.

How can India get there:

Indian regulators can aim to resolve inconsistencies by adopting certain principles:

- **Data rights should be institutionalized :** User agency over data must be concretised through national legislation. As matters stand, the PDP Bill is pending approval by Parliament while the NDHM policy has been rolled out across large swathes of the country, putting sensitive health information at risk of misuse without substantial safeguards in the form of data protection regulation. Accelerating the process of implementing the PDP Bill is critical to preserving user privacy and agency.

- **Capacity-building is necessary for enforcement :** Building state capacity to facilitate effective data governance is an essential prelude to carrying out data sharing activities. This requires investment on three fronts: human capital, administrative infrastructure and technical foundations. The need to build secure information systems for data interoperability is magnified in the case of the health sector where most information transacted is potentially sensitive in nature.

- **Limiting state power over processing:** Reducing the scope for mass surveillance and privacy harms by limiting exemptions granted to government agencies such as the NHA is key to preventing abuse of state power.[59]

---

[56] Health Insurance Portability and Accountability Act, 1996, retrieved from https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html

[57] Summary of the HIPAA Security Rule, retrieved from https://www.hhs.gov/hipaa/for-professionals/privacy/index.html

[58] Personal Health Information Protection Act, 2004, retrieved from https://www.ontario.ca/laws/statute/04p03

[59] "Took Covid vaccine using Aadhaar? Your National Health ID has been created without your permission", Dogra, India Today (May 24, 2021) retrieved June 23, 2021 from https://bit.ly/3yB2TMB

MILESTONE 2

Create community structures for data governance

While India recognises community rights over data as part of the NPD Report, this is not backed by coherent structures or processes for realisation. Developing a consultative, rights-first approach to data governance is a necessary precondition to put power back in the hands of communities which are the original generators of data. Looking inwards to adapt principles for data governance from other domains such as forest rights and local self-government (city or village democratic councils) is one under-studied mechanism for actualising collective claims. Similarly, the experiences of organisations operating as stewards in the health sector hold critical insights for Indian regulators.

Where India stands:

The NPD Report attempts to implement community governance but remains deliberately vague in articulating a concrete definition of communities. The terms of association between an individual data principal and the broader data community are also unclear.[60] Additionally, the Report omits outlining grievance redressal processes when communities are failed by their trustee (data fiduciary), undermining transparency of decision-making mechanisms.[61]

Further, the NDHM policy fails to include any conception of community governance in its framework, decisions are centralised by the state without checks for unfettered sovereign access to health data and consent by data principals. The Data Empowerment and Protection Architecture,[62] an ancillary data sharing framework for the financial sector, wholly disregards the role of communities in data governance.

How India should proceed:

The Committee of Experts deliberating a framework for non-personal data governance should adapt perspectives from existing laws and utilise

---

[60] Comments on the Revised Report by the Committee of Experts on Non-personal Data Governance Framework, Aapti Institute (February 03, 2021), retrieved June 23, 2021 from https://thedataeconomylab.com/2021/02/03/comment-on-the-revised-report-by-the-committee-of-experts-on-non-personal-data-governance-framework/

[61] "#NAMA: Issues With Definition Of Communities, Public Good, And Unabated Sovereign Access To Non-Personal Data", Barik, Medianama (January 22, 2021), retrieved June 23, 2021 from https://www.medianama.com/2021/01/223-nama-issues-with-definition-of-communities-public-good-and-unabated-sovereign-access-to-non-personal-data/

[62] The Data Empowerment and Protection Architecture (draft), 2020, retrieved fromhttps://niti.gov.in/sites/default/files/2020-09/DEPA-Book_0.pdf

established collective structures to meaningfully realise community rights over anonymised data:

- **Principles from Forest Rights Act, 2006:** Rights of communities should include the right to ownership, access, collection and use of data.[63] Further, communities have claims to value created or derived from use of data, including community data used to build intangible assets.[64]

- **Bottom-up participation:** Leverage the lowest levels of established and widely acknowledged (self) governance (panchayats[65], gram sabhas[66], ward committees[67]) to serve as decision-making structures for collective data governance.

How can India get there:

Community engagement can be facilitated by incorporating best practices from organisations stewarding health data:

- **Variant Bio[68]:** An American for-profit data collaborative in which clinical research projects, priorities and purposes are designed in partnership with communities who generate the data. The company has outlined a long-term benefit-sharing strategy[69] as part of its effort to build bottom-up governance structures and distribute value from data – both monetary and epistemic – more broadly among the populations and groups it works with.

- **ICES[70]:** The Ontario not-for-profit data repository follows ownership, control, access and possession (OCAP)[71] principles for stewarding indigenous data. Communities identify use-cases and themes for inquiry in collaboration with researchers, in a bid to dictate how data originating from First Nations is collected, shared and used.

- **Swasti[72]:** The Indian civil society organisation does not collect data directly, but produces it as a by-product of community health projects. However, this data is not shared externally. Only local health workers have access in order to ascertain any observed health concerns within the community.

---

[63] Forest Rights Act, 2006, Chapter II - Section 3 (1(c)), retrieved https://tribal.nic.in/downloads/FRA/FRAActnRulesBook.pdf

[64] Forest Rights Act, 2006, Chapter II - Section 3 (1(k))

[65] The Constitution (Seventy-third Amendment) Act, 1992, Section 243(d) retrieved https://tnrd.gov.in/constitutionalprovision.html

[66] The Provisions of the Panchayats (Extension to the Scheduled Areas) Act, 1996, Section 4(c), retrieved https://legislative.gov.in/sites/default/files/A1996-40.pdf

[67] The Constitution (Seventy-fourth Amendment) Act, 1992, Section 3(d), retrieved from https://www.india.gov.in/my-government/constitution-india/amendments/constitution-india-seventy-fourth-amendment-act-1992

[68] For more information, please visit https://www.variantbio.com/

[69] 'Long-term benefit sharing pledge', Variant Bio (202), retrieved June 25, 2021 from https://www.variantbio.com/pdfs/vb_benefit_pledge_11_2020.pdf

[70] https://www.ices.on.ca/Research/About-ICES-Research

[71] The First Nation Principles of OCAP (n.d.), retrieved June 25, 2021 from https://www.afn.ca/uploads/files/nihbforum/info_and_privacy_doc-ocap.pdf

[72] For more information, please visit https://swastihc.org/

MILESTONE 3

Establish reliable accountability process

Accountability constitutes the basis for intelligible data protection regimes. It is influential in building trust of data principals in the process of sharing. Creating accessible fora for grievance redressal marks the first step towards building a responsive and participatory data governance framework. To this effect, Indian laws such as the PDP Bill, NPD Report and NDHM policy make vague commitments to accountability that should be revised in view of proposed submissions on stewardship.

The PDP Bill proposes requirements for large data fiduciaries to appoint data protection officers[73] who would monitor and guide processing activities. However, the Bill falls short of defining a comprehensive and dependable grievance redressal mechanism. It is left to individual data protection officers housed within significant data fiduciaries and equivalent officers in other organisations[74] to draft rules as part of internal frameworks for accountability.

According to the NPD Report, communities are required to approach a Section 8 not-for-profit company[75] to engage with the designated regulatory authority – the Non-personal Data Authority – to raise complaints. This gives rise to several issues, the most grave of which is that the definition of "communities" is far from clear in this Report. Similarly, there is a dichotomy between the role of the data trustee and that of the Section 8 company set to be established under this framework.[76] As a result, accountability processes remain mired within regulatory obfuscations.

Lastly, the NDHM policy requires the appointment of a data protection officer (DPO) by fiduciaries and publishing of details such as name, processes and format for filing complaints on their respective websites. It also calls for the establishment of an NDHM Grievance Redressal Officer for health, to allow

[73] Chapter VII - Data Protection Officer (Sec. 30), The Personal Data Protection Bill, 2019

[74] Chapter VI - Grievance redressal by data fiduciary (Sec. 32), The Personal Data Protection Bill, 2019

[75] Establishing Rights over Non-personal Data (Sec. 7.2(ii))

[76] Comments on the Revised Report by the Committee of Experts on Non-personal Data Governance Framework, Aapti Institute (February 03, 2021), retrieved June 25, 2021 from https://thedataeconomylab.com/2021/02/03/comment-on-the-revised-report-by-the-committee-of-experts-on-non-personal-data-governance-framework/

for escalation of complaints unresolved by the DPO. Lastly, the policy allows formal litigation as an appeals mechanism.[77] However, the increasingly technocratic approach does little by way of accounting for data literacy levels among the broader populace. The cumulative effect of complex grievance redressal mechanisms is that the systems of accountability remain inaccessible for data principals.

How India should proceed:

The preliminary requirement for forging accessible systems of accountability is establishing comprehensive processes and specific contact points for filing and escalation of grievances. To do this, India's lawmakers can follow precedents from the following jurisdictions:

- **General Data Protection Regulation, 2016:** The regulation recognises the right to file complaints with supervisory authorities appointed by each member state of the European Union. Additionally, it creates provisions for escalating proceedings to courts within the country where the concerned supervisory authority is located. Delegated representation of data subjects' grievances through established NPOs is another salient feature of this regulation.[78] Integrating similar clauses within the NDHM, particularly delegated representation, would go a long way in designing approachable grievance resolution mechanisms.

- **Health Insurance Portability and Accountability Act, 1996:** The law establishes a process for reporting complaints to the Office of Health Secretary and the Office for Civil Rights[79] which are empowered to enforce HIPAA standards and achieve compliance. Specifically, complaints can be lodged through an online portal associated with the OCR as well as written mail, fax or email.[80]

- **Personal Health Information Protection Act, 2004:** The Information and Privacy Commissioner of Ontario is charged with protection of user rights over health data. Such a complaint should be filed within a year of the reported violation. The Act provides detailed procedures for reviewing complaints and determining necessary penalties or directives on the subject matter of the complaint.[81]

---

[77] Grievance Redressal (Secs. 32(1), 32(2) and 32(4)), National Digital Health Mission: Health Data Management Policy, 2020

[78] Chapter 8 - Art. 77, 78, 80, General Data Protection Regulation, EU 2016/279

[79] 45 CFR Part §160.304, Privacy Rules, Health Insurance Portability and Accountability Act, 1996

[80] "How to file an information privacy or security complaint", Dept. of Health and Human Services, retrieved June 25, 2021 from https://www.hhs.gov/hipaa/filing-a-complaint/complaint-process/index.html

[81] Part IV - Administration and Enforcement (Complaints, reviews and inspections), Personal health Information Protection Act, 2004

MILESTONE 4

Stipulate common data standards and infrastructure

Consent management – a critical corollary to the establishment of common data standards – remains inconsistent across policies. This poses a potent threat that can be used to undermine rights of data principals, in the absence of uniform and crystallised rules for processing health data. Moreover, other important stipulations on standards for interoperability of electronic health information fail to account for low levels of uptake among service providers chiefly charged with its implementation. Covering for the aforementioned inconsistencies while subsidising adoption of new technical standards is indispensable for creating policies that are in harmony.

Where India stands:

Under the PDP Bill, consent is required for processing and sharing of personal data. Consent is considered meaningful only if it is free, informed, specific, clear and capable of being withdrawn. Importantly, the PDP Bill stipulates exemptions to mandatory requirements of consent collection in cases of health emergencies, among others.[82] However, according to both the NPD Report[83] and NDHM policy[84], consent is not required for collection of anonymised health data. Consequently, data principals are excluded from data decisions on sharing of de-identified health information.

The NDHM Strategy Overview prescribes adoption of ICD-10/11, FHIR-R4, LOINC and SNOMED-CT as recommended standards for maintenance of digital health information. However, the policy itself acknowledges the abysmal levels of adoption of the standards across India's healthcare sector.[85] Lastly, the Electronic Consent Framework[86] puts forth machine-level standards for consent management that have been issued by the Reserve Bank of India for processing of financial data. However, this framework is yet to be adopted across other sectors such as health.

---

[82] Chapter III - Grounds for Processing Data without Consent (Section 12(e))

[83] Comments on the Revised Report by the Committee of Experts on Non-personal Data Governance Framework, Aapti Institute (February 03, 2021), retrieved June 23, 2021 from https://thedataeconomylab.com/2021/02/03/comment-on-the-revised-report-by-the-committee-of-experts-on-non-personal-data-governance-framework/

[84] "Health data sharing being set up to fail by a regulatory mashup: roles of the NDHM, PDPB, consent managers and the NHA", Aapti Institute (June 16, 2021), retrieved June 23, 2021 from https://thedataeconomylab.com/2021/06/10/health-data-sharing/

[85] Chapter 2.2.5 - Formats and Adoption of Standards for Health Data, NDHM Strategy Overview

[86] Electronic Consent Framework, 2020, retrieved from http://dla.gov.in/sites/default/files/pdf/MeitY-Consent-Tech-Framework%20v1.1.pdf

How India should proceed:

The first step to creating an enabling environment for health data stewardship is to harmonise consent standards across legislation. Specifically, the NDHM policy should follow the lead of the PDP Bill to revise consent clauses as health data is sensitive in nature. Regulators should also include clauses on consent-led sharing of anonymised data within the NPD Report. Consequently, the Report should be revised to engage data principals in the process of collection and sharing of de-identified data, moving beyond consent for mere anonymisation. This is important to ensure participation of principals at every step of the data value chain.

Building momentum for uptake of new standards relies on state subsidies to facilitate transition from offline to digital modes of maintaining health information. Accordingly, the state should cover for the cost of adopting renewed standards such as FHIR, SNOMED-CT, etc. This will not only reduce compliance costs for stakeholders but also encourage quicker transition to prescribed formats. Similar approaches to standards setting include the Findability, Accessibility, Interoperability and Reuse (FAIR) principles[87] employed globally in scientific research to manage data with minimal human intervention.

Lastly, evolving sector-specific guidelines on the nature of legislation on secondary uses of health data for public benefit, along the lines of Finland's FinData[88] enabling framework – Act on the Secondary Use of Health and Social Data – would go a long way in clarifying the premises and modalities to unlock health data for socially productive ends.

How can India get there:

Cost-effective, efficient management of consent can be facilitated through insights from a few use-cases interviewed as a part of this study:

- **BurstIQ:** A for-profit data exchange that provides blockchain-based health solutions which offer users dynamic consent controls. Users retain granular rights to authorise use of their data for specific health purposes and withdraw whenever required.[89]

- **Ciitizen:** This is a data collaborative working with individual patients and broader advocacy groups to facilitate user-driven health research. Ciitizen follows a comprehensive, process-based consent management system that allows users to opt in and opt out of specific health research projects. Authorisation for participation in the projects is largely digitally mediated.[90]

---

[87] "What are the FAIR Data Principles?", August C. Long Health Sciences Library, Columbia University (January 31, 2021) retrieved June 23, 2021 from https://library.cumc.columbia.edu/insight/what-are-fair-data-principles

[88] For more information, please visit https://findata.fi/en/what-is-findata/

[89] For more information, please visit https://www.burstiq.com/

[90] For more information, please visit https://www.ciitizen.com/

MILESTONE 5

Articulate public value of health data

Articulation of the value of health data is critical to garner public support for stewardship. This is because emergency response, pandemic preparedness and broader public health outcomes stand to benefit from greater efforts at sharing data.[91] However, Indian regulations, specifically the NPD Report, view data as sources of commercial and sovereign leverage, in a clear repudiation of its inherent social value in effecting positive gains in health systems.

Where India stands:

According to the NPD Report, the value of data is perceived as collective and to be stewarded by the state. While it acknowledges intent to generate benefits from the social value of data[92], it remains unclear on how this value is measured and distributed within and across communities.

In the non-data context, the Forest Rights Act, 2006 uses principles of vesting to ensure rights of Scheduled Tribes and other forest-dwelling communities over resources extracted from forests they inhabit.[93]

An important point to consider is that the NDHM policy is remarkably silent on the public value of health data due to its constricted focus on digitally mediating access to health services, disregarding the regulatory implications of such a glaring policy gap.

How India should proceed:

India's policymakers should reflect on the following principles to better articulate and distribute the value inherent in health information:

- **Publicly-managed health data banks:** Setting up and managing open databases containing health data from public agencies is prescient to articulating its welfare potential. The National Institute of Health (NIH) is one such database that acts as an open repository of anonymised health information intended for research. Moreover, NIH-funded research outputs

---

[91] "Understanding Value in Health Data Ecosystems", Marjanovic et al, RAND Health Quarterly (2018), retrived June 25, 2021 from https://www.rand.org/pubs/periodicals/health-quarterly/issues/v7/n2/03.html

[92] Public good purpose (Section 8.2(ii))

[93] Forest Rights Act, 2006, Chapter II - Section 3 (1(k))

are shared on an open access basis. India's pan-sector equivalent open data repository – the Open Government Data Platform India[94] – is plagued by quality issues. Data is made available in formats which are not usable or shareable for further analysis. Insufficient standardisation and slow transition to e-governance are often cited for the platform's failure.[95]

- **Value from use of data:** Borrowing from the Forest Rights Act, 2006[96], the Indian data ecosystem should recognise communities as originators of data and accord them the right to the value derived from its use, including community data used to build intangible assets.

---

[94] For more information, please visit https://data.gov.in/

[95] Report on Open Government Data in India, Wright et al, Centre for Internet and Society (n.d.), retrieved June 25, 2021 from https://cis-india.org/openness/publications/ogd-report

[96] Forest Rights Act, 2006, Chapter II - Section 3 (1(k))

MILESTONE 6

Incentivising sharing of health data for common benefit

Academic and private entities are often incentivised to keep data exclusive for personal and monetary gain. These challenges can be countered with ecosystem-enabled voluntary sharing that can incentivise data release, while also permitting its use in research and innovation. One way of achieving this is to support data pooling and similar collaborative models which can offset narrow private interests to unlock data for public benefit.

Where India stands:

Indian policy has thus far attempted mandating data sharing rather than incentivising outcomes. Policymakers have systematically deprioritised data protection in their quest to push for sharing.

The sole feature of the NPD Report that contributes to this factor is the limitation of mandatory data sharing to “high value datasets”.[97] However, the Report fails to outline clear standards and intermediaries to reduce cost of sharing. In fact, basic administrative principles for establishing legible processes for data sharing are overlooked within this framework.[98]

How India should proceed:

Regulators can benefit from recognising and remedying the commercial, technical and procedural barriers to health data sharing. This can be achieved by:

- **Providing financial support:** Subsidising infrastructural costs to address issues arising from commercial and technical barriers to health data sharing. Estonia’s X-Road[99] is a data exchange layer that enables information exchange among public agencies, private entities, not-for-profit organisations and citizens using state-sponsored data infrastructure. Specific to the health sector, the Central Health Information Systems

---

[97] Section 7.8, Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

[98] Comments on the Revised Report by the Committee of Experts on Non-personal Data Governance Framework, Aapti Institute (February 03, 2021), retrieved June 25, 2021 from https://thedataeconomylab.com/2021/02/03/comment-on-the-revised-report-by-the-committee-of-experts-on-non-personal-data-governance-framework/

[99] Estonian data exchange layer for information systems (X-Road), retrieved June 25, 2021 from https://scoop4c.eu/index.php/cases/estonian-data-exchange-layer-information-systems-x-road

and Patient Portal[100], supported by laws mandating user rights to access and re-use data, allows for low-cost access maintenance of patient records.

- **Creating structures for community representation:** Empowering community organisations to negotiate with third parties for community data sharing is an effective tool to overcome the procedural barriers to data sharing. Representative organisations must bear a duty of care and loyalty towards their communities, always prioritising their preferences over monetary or instrumental prerogatives, e.g. the NPD Report attempts this through appointment of NPOs as trustees.[101] Extending similar provisions to the NDHM policy would enable effective negotiation on terms for sharing sensitive and valuable health information.

How can India get there:

India needs to create spaces for innovation with projects relating to data use and sharing. This can be done by formulating sandboxes for health projects, as proposed for the NDHM policy.[102] However, the scope of the NDHM Sandbox Guidelines suffers from a restricted focus on testing for technical innovation only. Revising it to introspect proposals for regulatory innovation, such as health data stewardship, is critical to build ecosystem incentives for sharing.

Similarly, communication of value to communities will also help enhance possibilities of greater health data sharing. LunaDNA[103] and MiData[104] are data cooperatives that invest resources in communicating the public value of their health information to their members.

---

[100] Estonian Central Health Information System and Patient Portal, CEF Digital, retrieved June 25, 2021 from https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/2019/07/26/Estonian+Central+Health+Information+System+and+Patient+Portal

[101] Section 7.7, Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

[102] NDHM Sandbox Guidelines, Ministry of Health and Family Affairs (August 18, 2020), retrieved June 25, 2021 from https://ndhm.gov.in/documents/sandbox_guidelines#:~:text=NDHM%20Sandbox%20is%20a%20framework,market%20reactions%20to%20the%20same

[103] For more information, please visit https://www.lunadna.com/

[104] For more information, please visit https://www.midata.coop/en/home/

MILESTONE 7

Providing environments for safe innovation

The health data ecosystem in India is ingenious for its use of the sandbox approach for controlled testing of technical services and products. However, there remain concerns about the NDHM policy, stemming from weak consent protections that allow the possibility of function creep in use of health data. Piloting stewardship is one way to resolve the concerns as it allows for dynamic and granular consent management.

Where India stands:

The NPD Report suggests the health sector[105] as a pilot to test the high value datasets and mandatory data sharing framework. In the absence of the protections afforded by the PDP Bill, this is undesirable due to the highly sensitive nature of information under consideration.

Another important provision to be noted is the framework of consent managers who constitute the potential vehicles for stewardship within India's regulatory ecosystem and find mention in the PDP Bill[106] and NDHM policy[107]. This framework prescribes machine-level tools for identity and access management of individuals. But there have not been any requests for pilot proposals as yet.

Lastly, the 2019 version of the PDP Bill introduces a regulatory sandbox for AI/ML start-ups[108] with associated relaxations on data processing, consent collection and localisation. However, the Data Protection Authority (DPA) is vested with power to decide which fiduciaries can be included in the sandbox and on what terms. Conditions of "public interest" and "innovative use of data" determined by the DPA create arbitrary, undefined parameters for participation of start-ups.[109]

How India should proceed:

Indian regulators should forge a unified coherent vision for use of health data. This is underpinned by a first principles-based approach[110] for use of health data. Subsequently, regulations should impose reasonable limits to powers of

---

[105] Section 7.2(iv), Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

[106] Chapter IV - Rights of Data Principals (Art 21.1), The Personal Data Protection Bill, 2019

[107] Section 3.5.3, National Digital Health Mission: Health Data Management Policy, 2020

[108] Chapter VIII - Exemptions (Art. 40.1), The Personal Data Protection Bill, 2019

[109] "Regulatory Governance Under The PDP Bill: A Powerful Ship With An Unchecked Captain?", Smriti Parsheera, Medianama (January 7, 2020), retrieved June 26, 2021 from https://www.medianama.com/2020/01/223-pdp-bill-2019-data-protection-authority/

[110] "Heralding a new health data regime in India", Deo et al, Observer Research Foundation (July 06, 2020), retrieved June 26, 2021 from https://www.orfonline.org/expert-speak/heralding-a-new-health-data-regime-in-india-69259/

the state to mandate data sharing and focalise social welfare as the objective of health data sharing.

An immediate imperative is to define clear and narrow policy on use of data for public health actions. India's lawmakers can look towards the proposed European Health Data Space[111] framework whose clauses are harmonised with provisions of the GDPR. Similarly, the NDHM policy must be reconciled with the PDP Bill and specifically, do away with expansive powers accorded to the NHA.

Lastly, policymakers should facilitate testing for regulatory innovation in the health sector. In their current form, the NDHM Sandbox Guidelines are limited to technical innovation. Expanding the scope to facilitate governance and regulatory experiments such as stewardship is best to protect user autonomy while opening data for social good.

How can India get there:

Indian regulators should reflect on the following suggestions to create a safe environment for innovation using health data:

- **Incorporating community health pilots:** At the outset, the NDHM Sandbox can look to support pilots by civil society organisations working in the health sector. This is because NPOs play a critical role in filling gaps in healthcare delivery in under-resourced contexts.

- **Gleaning insights from global health data pilots:** India's regulators can explore FinData, which was essentially guided by e-Prescriptions pilots[112] that began in the country in 2001. Similarly, the proposed European Health Data Space framework is based on insights from the SUSTAINS pilot.[113]

- **Providing playbooks and technical knowledge resources:** The Government of Ontario is developing a digital health exchange[114] and Playbook[115] in consultation with health practitioners and technical professionals. Similar tools and guides for stakeholders in India's health data ecosystem can be taken up by nodal agencies concerned with implementing data governance regulations.

---

[111] For more information, please visit https://ec.europa.eu/health/ehealth/dataspace_en

[112] "The ePrescriptions System in Finland: A case study", Dovancescu et al (n.d), retrieved June 26, 2021 from https://www.wi.rwth-aachen.de/wi/theses/ePrescription.pdf

[113] "Access to your health records - Results of SUSTAINS pilot", European Commission (n.d.), retrieved June 26, 2021 from https://ec.europa.eu/digital-single-market/en/news/results-sustains-pilot-patients-accessing-their-health-data

[114] Digital Health Information Exchange Policy (draft), Govt. of Ontario (April, 2020), retrieved June 26, 2020 from https://www.ontariocanada.com/registry/showAttachment.do?postingId=32590&attachmentId=43697#:~:text=The%20purpose%20of%20this%20policy,and%20provincial%20digital%20health%20tools

[115] Digital Health Playbook, Government of Ontario (December, 2019), retrieved June 26, 2020 from https://health.gov.on.ca/en/pro/programs/connectedcare/oht/docs/dig_health_playbook_en.pdf

Based on the above analysis, we have postulated what a few health data stewardship pilots for India might look like as part of Figure 4.

Pilot 1:

Community organisation as "data trustee" under the Framework for Non-personal Data Governance.

'X' is a not-for-profit enterprise involved in providing telemedicine consultations for primary healthcare across seven rural districts in southern Karnataka. As a part of its work, the enterprise has built an application to collate and track data on incidence of diarrhoea among children below the age of five in its operative districts. X wants to share this data with a start-up working to create low-cost nutritional supplements for children below the age of 10.

Under the current framework of the NPD Report, X qualifies as a 'data trustee'[116] which is stewarding children's data relating to vital health stats. Upon consultation with the Non-personal Data Authority, X has identified a high value dataset (HVD) containing data on childhood malnutrition in southern Karnataka. However, before proceeding to share the HVD with the concerned start-up, X holds processes for community consultation with the guardians / parents of children whose data it stewards. It defines clear clauses for purpose limitation such that data sharing is driven to protect community interests.

Consequently, X moves beyond its role as a 'data trustee' stipulated by the NPD Report, to become an active data steward with community participation at the heart of its governance framework. The concerned start-up can now request access to this HVD for use in the development of its nutritional supplement and X is obligated to share the HVD under the NPD Report's mandatory data sharing clause. Results from this pilot can potentially aid innovation and bring value addition to health products, one of the aims stipulated in the NPD Report for use of HVDs to create public good.[117]

---

[116] Section 7.7 (ii) - Non-personal Data Roles: High Value Datasets and Data Trustee, Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

[117] Section 8.2 - Public Good Purposes, Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

Pilot 2:

Community organisation as 'data trustee' working with a social enterprise, as a part of the NDHM Sandbox.

'Y' is a social enterprise which provides genetic diagnostics services through its take-home DNA kits in Bengaluru, Mumbai and Pune. The organisation collects and collates data, thus becoming a 'data custodian'[118] under the NPD Report. Further, as part of its informed consent practice, the enterprise seeks consent to use aggregated data to generate insights. A community health organisation – 'Z' – requests Y to disclose data on prevalence of Type-II diabetes to create an HVD on prevalence of Type-II diabetes in Bengaluru. As a result, Z becomes the data trustee[119] of this HVD on diabetes.

'AV' is a start-up incubated in the NDHM sandbox working to provide e-pharmacy solutions, specifically aimed at diabetics in India. AV requests the relevant data trustee – in this case, Z – for access to its HVD on Type-II diabetes for its pre-approved e-pharmacy sandbox experiment. The health start-up is then able to utilise the HVD to generate insights for its product development. This arrangement is negotiated by the trustee, either to give the community a degree of access to or control over how the derived data is used, or negotiate on other terms.

The innovation incubated in the sandbox is able to promise secure and beneficial outputs to the data trustee which is entrusted to negotiate for the community. In turn, the trustee and custodian are able to utilise these insights for the priorities of the community.

Figure 4: *Pilots can differ depending on context, type of data and purpose – all of these suggestions will be the subject of innovation by start-ups and social enterprises, once afforded the opportunity to experiment with health data.*

---

[118] Section 7.4 - Non-personal Data Roles: Data Custodian, Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

[119] Section 7.7 (ii) - Non-personal Data Roles: High Value Datasets and Data Trustee, Revised Report by the Committee of Experts on Non-personal Data Governance Framework, 2020

MILESTONE 8

Cultivate feedback loops for regulators

Laws and regulations for an ever-evolving landscape of the digital economy must be informed by learnings and insights from pilots to facilitate evidence-based policymaking. This requires policymakers to go beyond limited prerogatives of drafting legislation by attempting to actively engage with impacted stakeholders – communities, providers and NPOs – to build a cohesive sharing environment for health data.

Where India stands:

Contemporary history of the Indian policymaking process displays a lack of a consultative element. The public comment period for the NDHM policy proved perfunctory, granting only a week for stakeholder consultation.[120] This demonstrates that the Pre-legislative Consultation Policy[121] which mandates a minimum period of 30 days for public feedback is frequently flouted by state agencies.[122]

How India should proceed:

Data governance regulations can draw from a range of existing mechanisms to engage with stakeholders:

- **Embed a comprehensive consultation process:** Work with community representatives to increase awareness on data issues and enable decision-makers to represent community interests.

- **Panchayats (Extension to Scheduled Areas) Act, 1996:** According to this law, local communities in tribal-majority areas must collectively assent, through Gram Sabhas[123], to any use of their resources by the state or third parties. Adapting this principle, as suited, may be explored for realising community data rights.

---

[120] "NHA confirms that the consultation period for Draft Implementation Strategy of NDHM is in contravention of Pre-Legislative Consultation Policy", Software Freedom Law Centre (May 17, 2021), retrieved June 27, 2021 from https://sflc.in/nha-confirms-consultation-period-draft-implementation-strategy-ndhm-contravention-prelegislative

[121] Pre-legislative Consultation Policy, Ministry of Law and Justice (2014) retrieved from https://legislative.gov.in/documents/pre-legislative-consultation-policy

[122] "When Bills get passed without following Pre-legislative Consultation Policy", Patel et al (August 18, 2019), LiveLaw, retrieved June 27, 2021 from https://www.livelaw.in/columns/when-bills-get-passed-without-following-pre-legislative-consultation-policy-147258

[123] The Provisions of the Panchayats (Extension to the Scheduled Areas) Act, 1996, Section 4(c), retrieved from https://legislative.gov.in/sites/default/files/A1996-40.pdf

- **Mine and Minerals Act, 1956:** According to the amendment of 2015, the Act creates a statutory body, the District Mineral Foundation (DMF)[124], to protect the interests of impacted communities. The DMF is funded by a fixed rate of returns from the income accruing to mining leaseholders. Similar provisions can be drawn up to measure and disburse compensation to communities for current and future use of their data.

How can India get there:

Collection of feedback on pilots is a useful method of assessing successes and weaknesses of policies and projects. Successful pilots originating in the EU provide a starting point for cultivation of feedback loops. Feedback from the Finnish e-Prescriptions pilot was incorporated into law while building a national platform for health records called Kanta services.[125] Further, the EU-funded SUSTAINS pilot was successful in granting patients access to electronic health records and other digital health services. Learnings from this pilot were factored into the European Health Data Space proposal.[126] Similarly, the results of the NDHM Sandbox should guide incremental revisions of health data ecosystem policies in India.

The submissions suggested as a part of this section are born of a quest to reimagine the position of the state in mediating the relationship between communities and entities concerned with processing data. India's data governance landscape is primed for novel interventions that can subvert the digital divide which has thus far rendered communities powerless and unable to advocate for their own data rights. While layers of complexities abound when considering possible avenues for collective governance of data, stewardship can reliably address these concerns to spur citizen-led efforts for health data sharing.

---

[124] Section 9B, Mines and Minerals (Development and Regulation) Amendment Act, 2015, retrieved from https://www.mines.gov.in/writereaddata/UploadFile/The_Mines-and-Minerals_Amendment_Act,2015.pdf

[125] For more information, please visit https://www.kanta.fi/en/what-are-kanta-services

[126] "The European Health Data Space: an opportunity for the public health community", European Public Health Alliance (June 2, 2021), retrieved June 27, 2021 from https://epha.org/the-european-health-data-space-an-opportunity-for-the-public-health-community/

1. **Enact data protection legislation for greater transparency in use of health data:** Set up an enforcement agency with regulatory capacity to audit practices of companies involved in health data management and sharing.

2. **Legislate on rights of users over their personal information:** Define rights of users to access, re-use, modify, delete their information held by third parties, including public agencies.

3. **Establish community-level rights over data:** Use principles from existing legislation such as the Forest Rights Act and the Panchayats (Extension to Scheduled Areas) Act to establish rights of communities over shared resources (data) when taken over by companies and the state.

4. **Adapt localised, existing structures for community governance:** Account for existing bodies of collectivised representation, including village-level and municipality-level assemblies (73rd and 74th Constitutional Amendments) and associations organised around identities.

5. **Formulate sector-specific taxonomy and rules for health data maintenance and sharing:** Define the extent of jurisdiction for health data regulation; relevant types of data, institutions including clinics, hospitals; roles and responsibilities of relevant individuals, patient groups, health workers, clinicians.

6. **Harmonise health data policies with primary data protection legislation:** Furnish a narrower, clearer definition of data sharing requirements within the NDHM, in accordance with the data minimisation principle of the PDP Bill.

Figure 5: *This is a snapshot of recommendations presented in this section. They are at the level of policy and legislation, owing to the lack of uniform standards in health data governance in India. Creating a sharing environment requires empowerment of communities and a common data ontology – both of which can be achieved by instantiating stewardship.*

06

Regulatory Value of Health Data Stewardship

The imperatives for Indian regulators to instantiate stewardship are layered and compelling. In a milieu where private healthcare providers account for 62% of India's health infrastructure[127], the power to collect data and exercise control over its use is overwhelmingly concentrated in the hands of commercial entities. Insights derived from data accrue to large corporate hospitals primarily, systematically invisibilising the role of individuals and communities as primary generators of health information. Responsible stewardship is a manifestation of the need to challenge this hegemony, presenting persuasive evidence of an accountable, agency enhancing mechanism for data governance.

Subsequently, this research has identified four categories of imperatives to embed stewardship in the Indian regulatory ecosystem:

1. Constitutional imperative

The Indian state is required to act in alignment with the Directive Principles of State Policy[128], contained within its Constitution. This requires development of healthcare to be a primary objective of policymaking, along with redistribution of material resources to prevent its concentration and directs its use for fulfilment of "common good".[129] Data stewardship helps communities to wrest control of their data from private healthcare providers and distribute its value broadly among stakeholders in the data economy. Moreover, progressive investments in health data stewardship can enable timely access to up-to-date health information that can enhance disease and treatment management, promote innovation in services through research and reduce costs of accessing healthcare.

2. Financial sustainability of the system

Data-driven care has increased the preventive scope of health services, averting adverse outcomes as well as reducing costs of treatment measures. One study by Nesta[130] estimates a total savings of £4.4 billion on public spending when patients and communities are involved in their own care. Stewardship's central tenet is grounded in enhancing participation of data principals in governance. As a result, patients and caregivers are empowered with necessary data to make informed decisions about their health.

---

[127] "Private Healthcare in India: Boons and Banes", Jaffrelot, et al, Institut Montaigne (November, 2020), retrieved June 20, 2020 from https://www.institutmontaigne.org/en/blog/private-healthcare-india-boons-and-banes

[128] Part IV: Directive Principles of State Policy, retrieved June 21, 2021 from http://www.mea.gov.in/Images/pdf1/Part4.pdf

[129] Article 39 (b) and (c)

[130] "Public Value: How can it be measured, managed and grown?", Mulgan, et al., Nesta (May, 2019), retrieved June 21, 2021 from https://media.nesta.org.uk/documents/Public_Value_WEB.pdf

3. Multiplier effect in data economy

Government spending in health and education follows the fiscal multiplier pattern, demonstrating the long-ranging economic benefits of such investments. A report by WHO[131] offers a compelling case for investing in public health, producing approximately four-fold returns on every dollar spent on healthcare. Another more recent estimate by researchers in Europe[132] indicates that multipliers are particularly abundant for social policies, bringing in benefits of €3 or more per euro spent in health system overhaul. In the Indian context, investments in health tech and regulatory initiatives such as stewardship can help scale a nascent data ecosystem and increase accuracy of healthcare delivery.

4. Social imperative

Stewardship will result in stronger collaborations to facilitate faster and socially beneficial open innovation, as in the case of COVID-19 vaccine development. Chinese researchers shared the genome sequence of Sars-CoV-2[133] in early January 2020 on an open access basis which enabled the production of vaccines in record time. Similarly, public data repositories maintained by the EU and other governments have led to rapid, cost-effective and safe data sharing to advance outbreak research.[134] Providing a wider set of tools for care also increases the chances of greater self- and peer-care options, while decentralising access to care itself in the process. To this end, a study by Nesta estimates annual savings of £950 million per year (for health systems alone) from targeted peer support and self-management education for specific health conditions.[135] Indian regulators could benefit from a paradigm like stewardship that places individual patients, their caregivers and related patient advocacy groups at the heart of the health data ecosystem, thereby granting agency over the process of delivering care to communities.

---

[131] "The Case for Investing in Public Health", World Health Organisation - Regional Office for Europe (2014), retrieved June 21, 2021 from http://apps.who.int/iris/bitstream/handle/10665/170471/Case-Investing-Public-Health.pdf?sequence=1&isAllowed=y

[132] "Social and economic multiplier: What they are and why they are important for health policy in Europe?", Stuckler, et al. Scandinavian Journal of Public Health (August 29, 2017), retrieved June 21, 2021 from https://journals.sagepub.com/doi/full/10.1177/1403494817707124

[133] "Genetic sequencing takes lead in Covid-19 vaccine development", Hindustan Times (April 02, 2020), retrieved June 22, 2021 from https://www.hindustantimes.com/india-news/genetic-sequencing-takes-lead-in-covid-19-vaccine-development/story-A4mjutfeefKjkSeGeM8NCL.html

[134] "Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research", Ian le Guillo, Horizon: EU Research and Innovation Magazine (April 2020), retrieved June 22, 2021 from https://horizon-magazine.eu/article/covid-19-how-unprecedented-data-sharing-has-led-faster-ever-outbreak-research.html

[135] "Public Value: How can it be measured, managed and grown?", Mulgan, et al., Nesta (May, 2019), retrieved June 21, 2021 from https://media.nesta.org.uk/documents/Public_Value_WEB.pdf

The pandemic has renewed interest in the ability of data-driven solutions to improve the effectiveness of public health interventions. Stewardship in healthcare is one such data-driven regulatory innovation that offers credible advantages to a cross section of stakeholders: for the public exchequer, it improves the financial sustainability of health systems and creates positive externalities for the wider population; for service providers and health technology start-ups, stewardship provides opportunities to increase cost efficiencies; for individuals and communities, stewardship can channel their data in safe and effective ways to meet prioritised health outcomes.

07

Way Forward

> Unlocking siloed data hitherto processed through limited bilateral and multilateral arrangements is important to address the pressing challenges of our times. The need for wider and safe health data sharing is particularly magnified in the context of India's experience of managing the COVID-19 pandemic.

The second wave of the pandemic placed exorbitant strain on the country's already fragile health infrastructure where timely data sharing and analytics proved critical in saving lives. Communities[136], private entities, civil society[137], local city and village governments[138] came together to build collaborative data platforms that attempted to provide real-time information on availability of hospital beds, ventilators, life-saving medication such as Remdesivir and, most sadly, slots in crematoria to cremate the thousands of Indians who became victims of the pandemic.

These efforts served to surface a more fundamental point: apathy towards health systems governance can no longer remain a glossed-over status quo. Stepping up public investment, strengthening health ecosystems for emergency response and building capacity of health workers have emerged as paramount to India's fight against COVID-19. Further, perspectives on collective impact have renewed the focus on health interventions, necessitating structured and coordinated action by the state, providers and related entities to forge a common agenda that delivers the best quality of care for India's communities.

For instance, patients of COVID-19 and their caregivers benefitted from data sharing initiatives, such as Bangalore's municipal corporation-run control rooms which provided real-time information on availability of hospital beds, ventilators and oxygen tanks.[139] The BBMP acted as a steward of community data to deliver greater access to health services during the second wave of the pandemic in Bangalore. Guided by an essentially human-centric approach to data governance, such initiatives place creation of public benefit as the primary goal of health data sharing.

Data is at the heart of all these efforts, providing indispensable insights

---

[136] "India COVID warriors", Al Jazeera (June 24, 2021), retrieved July 05, 2021 from https://www.aljazeera.com/program/101-east/2021/6/24/indias-covid-warriors [Video]

[137] "Responding to the second wave of COVID-19 in rural India", Rapid Rural Community Response to COVID-19, India Development Review (May 05, 2021), retrieved July 05, 2021 from https://idronline.org/responding-to-the-second-wave-of-covid-19-in-rural-india/

[138] "Bengaluru's Index app for COVID-19 management recognised by the Centre", The New Indian Express (June 21, 2021), retrieved July 05, 2021 from https://www.newindianexpress.com/cities/bengaluru/2021/jun/26/bengalurus-index-app-for-covid-19-management-recognised-by-centre-2321796.html

[139] "BBMP launches COVID-19 helpline for citizens", Special Correspondent, The Hindu (April 16, 2021), retrieved July 01, 2021 from https://www.thehindu.com/news/cities/bangalore/bbmp-launches-covid-19-helpline-for-citizens/article34330111.ece

into problem areas and gaps in healthcare access and delivery. All data decisions should be directed towards empowering communities to participate in health systems. To achieve this, the current study has outlined an eight-point roadmap with successive milestones to help India's regulators instantiate stewardship in healthcare. Additionally, the research furnishes pointed recommendations on how Indian regulators can revise and reposition policy to accommodate an agency-enhancing, transparent mechanism for data governance provided by stewardship.

Fundamentally, data stewards function as responsible intermediaries between data principals and data controllers / requesters. The Indian state is in a powerful position to mediate this relationship by creating a robust policy ecosystem for instantiating stewardship within the health sector, supplying frameworks and standards to guide interactions between communities, providers and other stakeholders in the health data ecosystem.[140] Community-centric purpose is central to this study's conception of stewardship as generation of social value is a core attribute identified through conversations with practitioners and organisations operating as stewards in the health sector. Additionally, stewardship is essential to create a paradigm in which data is used in the service of public well-being through clinical research and innovation in healthcare delivery. Embedding this framework of stewardship within India's data regulatory landscape is pivotal to empowering patients and their caregivers to control their data and structure data flows in ways that deliver better health outcomes for the larger Indian populace.

---

[140] For more information on stakeholders in the health data ecosystem, please refer to a corollary of this study, "Health data stewardship: Bottom-up stakeholder engagement"

aapti institute 2021

Aapti is a public research institute that works at the intersection of technology and society. It examines the ways in which people interact and negotiate with technology both offline and online.

✉ contact@aapti.in

www.aapti.in | www.thedataeconomylab.com