White Paper

Centre for
The Digital Future

Data Sharing
for Effective Public
Health Management

FEBRUARY 2021

Background

India has made digital health the focal point of its public health strategy. The National Digital Health Mission is an important step in this direction. It (The NDHM) aims to develop the backbone necessary to support the integrated digital health infrastructure of the country (and) will bridge the existing gap amongst different stakeholders of the healthcare ecosystem through digital highways.[1]

Health data is universally acknowledged to be critical, sensitive and personal. The draft PDP (Personal Data Protection) Bill classifies health data as "sensitive personal data" with attendant restrictions on its collection, processing, storage and usage. The Draft Health Data Management Policy, released by NHA, accords very high priority to individual consent pertaining to the collection and usage of data.

Having said the above, the use of quality data made available in a timely fashion is crucial not only for an individual's health but also for public health. Data collected as part of a public health surveillance system can be used to estimate the magnitude of a problem, identify groups at higher risk of having poorer outcomes, examine relationships between risk factors and outcomes, develop interventions and with continued monitoring assess the effectiveness of the interventions to modify the complications or outcomes. The results of analyses from surveillance data can be useful for many purposes including health care and patient advocacy, providing a basis for priority setting and allocation of health care resources, ensuring the availability of better data on population health and supporting medical care quality assurance and quality improvement efforts.[2]

Introduction

The National Digital Health Blueprint (NDHB) keeps the overall vision of NHP 2017 in perspective and recommends a pragmatic agenda to start with, adopting the principle of 'Think Big, Start Small, Scale Fast'[3].

The philosophical angle of the NDHB is the adoption of an affiliated architecture, in several aspects, but more so in terms of how data is managed, which means that data resides at a point of collection or at a trusted source from where it is collected. This health data is accessible only to authorized persons through links. This avoids centralization, privacy concerns and management issues. Another key principle which is underlined is that the citizen should be in control of his/her data. This also serves as the essence of PDP Bill or DIPA (Data Empowerment and Protection Architecture).

The blue print also takes an ecosystem rather than a systems approach, as there are several players and levels that are addressed, involved and participating in the health sector – public/private/start-up; primary/secondary/tertiary and Centre/State/local.

The Centre for The Digital Future at the India Development Foundation, in collaboration with other researchers and institutions, aims to explore the use of real-time and other novel datasets to address research questions raised by experts. This project is supported by Facebook's Data for Good initiative. Under the aegis of this initiative, a number of Round

---

[1] https://ndhm.gov.in/home/ndhm
[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4515757/
[3] https://www.nhp.gov.in/NHPfiles/National_Digital_Health_Blueprint_Report_comments_invited.pdf

Tables have been and are being organized, one of them being: **Data sharing for effective public health management**.

This paper is based on the discussions during this particular Round Table, which sought to bring together a small group of experts from diverse backgrounds relevant to the topic under discussion. The aim was to have an in-depth discussion on the role of data in public health and the need and limits for data sharing in the context of public health.

An indicative list of questions that were addressed at the Round Table are:

a) What are the kinds of data that are needed for effective public health management and research?
b) What are the possible sources of such data?
c) What kinds of data should or should not be shared?
d) What policies and arrangements are needed to facilitate and enable such data to be shared?
e) How can data sharing be incentivized so that institutions (including the Government) that possess data share it with others who need that data, subject to laid down privacy and security stipulations?

Different kinds of data that are needed for effective public health management and research

Before understanding the different kinds of data needed for effective public health management and research, it is important to understand what is meant by public health? One of the definitions presented during the discussion was that from the Institute of Medicine which describes public health simply as: what society does collectively to assure the conditions for people to be healthy.[4]

Another aspect that was presented was that public health identifies and influences the determinants of health at the population level, in order to impact upon health at the individual level. Therefore, it identifies all determinants and the risk factors, and operates through policy, system and programmatic interventions, also looking at community engagement.

Once a broad picture of what public health can be defined as is clear, it is important to consider the next question: What is public health surveillance? An ongoing, systematic collection, analysis and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice. Surveillance is undertaken to inform disease prevention and control measures.[5]

Why is surveillance needed?

- To serve as an early warning system, identify public health emergencies,
- To guide public health policy and strategies,
- To document impact of an intervention or progress towards specified public health targets/goals,

---

[4] https://www.publichealth.columbia.edu/public-health-now/news/what-public-health-introduction
[5] https://www.who.int/immunization/monitoring_surveillance/burden/vpd/en/

- To understand/monitor the epidemiology of a condition to set priorities and guide public health policy and strategies.[5]

Further to the above, there are three kinds of data that need to be considered – PII (Personally Identifiable Information, which attracts stringent privacy and security provisions of the proposed PDP law), anonymized data (which is related to specific individuals, but without the possibility of the individual concerned being identified) and aggregate data, which is cumulated at some level and contains no individual information. These three forms of data characterize both health and non-health data. Further, there exists today a market for data. Sharing of and access to data required for public health purposes, therefore, needs to be suitably regulated and enabled in the eco-system without any compromise on privacy or security rights and obligations.

The possible sources of data needed for effective public health management and research

Having said the above, there are different types of data required for effective research in public health management; hence, the sources of these data are also varied. Let us take the case of the current epidemic. The Government of India collects epidemiological information periodically through the Health Management Information System (HMIS) and other varied survey sources such as National Family Health Survey (NFHS), District Level Household Survey (DLHS)[6]. Much of the epidemiological data regarding infectious and communicable disease in India is captured by the integrated disease surveillance program. The ICMR also runs a network of 26 disease specific institutes across India, such as the Antimicrobial Resistance Surveillance Network, or the Rotavirus Surveillance Network, which collate data on clinical, epidemiological, and virological information, which is used to devise evidence-based treatment guidelines to improve care and monitor and evaluate transmission-modifying interventions such as vaccines.[6] These are just some examples.

However, sources of data will obviously vary for other uses for e.g. The National Family Health Survey (NFHS), the National Sample Survey Office (NSSO) census survey, the National Aids Control Organization (NACO), and other such serve specific diseases.

The nature and purpose of data are like inseparable twins. ISO/TS 14265:2011 defines a set of high-level categories of purposes for which personal health information can be processed[7]. These categories include clinical care provisions, emergency care provisions, public health emergency, research and others.

The challenge, however, remains in unlocking the potential of data in assisting public health management and research in a manner that is responsible? The difficult part lies in collating data from diverse stakeholders like for example private and public hospitals, all of whom may not be willing to share; besides this, data protection comes into the forefront too; however, if patients are given adequate information on the cause for which their data is being used, and how it is going to benefit him/her and society, the "Good Samaritan" element may come into play.

---

[6]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5930391/#:~:text=The government of India collects health information and Sample Registration

[7] https://www.iso.org/standard/54547.html

Another aspect is: can bulk, anonymized individual data be made available on the public domain in a timely fashion, and in a format in which it can it be readily used, meeting the end purpose effectively? Are we just collecting data, time and again, repeating ourselves, duplicating, and, hence, wasting resources or we focussing on what is, actually, needed, for a specific cause to be addressed? How successful would a non-unique personal health indicator be (which is, currently, being tested, on a pilot phase, in six geographies)?

What kinds of data should or should not be shared?

According to The Personal Data Protection Bill, 2019: Personal data is data which pertains to characteristics, traits or attributes of identity, which can be used to identify an individual[8].

A data fiduciary is an entity or individual who decides the means and purpose of processing personal data. Such processing will be subject to certain purpose, collection and storage limitations. For instance, personal data can be processed only for specific, clear and lawful purpose. Additionally, all data fiduciaries must undertake certain transparency and accountability measures such as: (i) implementing security safeguards (such as data encryption and preventing misuse of data), and (ii) instituting grievance redressal mechanisms to address complaints of individuals. They must also institute mechanisms for age verification and parental consent when processing sensitive personal data of children.

The Bill allows processing of data by fiduciaries only if consent is provided by the individual. However, in certain circumstances, personal data can be processed without consent. These include:
   i.   if required by the State for providing benefits to the individual,
   ii.   legal proceedings,
   iii. to respond to a medical emergency

The central government may direct data fiduciaries to provide it with any: (i) non-personal data and (ii) anonymised personal data (where it is not possible to identify data principal) for better targeting of services.[8]

The point, however, remains that though data is available on the public domain, it does not necessarily mean that it can be accessed by anyone, anywhere; mechanisms have to be and are in place.

Data sets collected by large government or government authorized surveys are made available on request but they are quite often uploaded on the public domain delayed. Initially, the delay was because surveys were conducted at wide time intervals, an aspect which has changed now, resulting in data being uploaded in a timelier fashion. A challenge that also applies to data emerging from a large number of demographic surveillance sites, research studies, cross sectoral surveys, cohort studies conducted by agencies/researchers and such is that it is seldom placed in the public domain either due to the policy of the funding agency or/and because a lead of usually a year is given during which the output is expected to be published. The above, however, critically impacts policy making as the data

---

[8] https://www.prsindia.org/billtrack/personal-data-protection-bill-2019

on which research (for formulation of policy) is based, is outdated, caused by the lapse of time.

What policies and arrangements are needed to facilitate and enable data to be shared?

Why should data be shared? Capturing, collating, analyzing and extrapolating data is an expensive affair.

There are many initiatives around the world supporting the sharing of medical data, leading the way to open science while still respecting the privacy rights of the patients: GIFT-Cloud, Personalized Consent Flow, the Sync for Science (S4S), 1+ Million Genomes Initiative, the Pan-Cancer Analysis of Whole Genomes (PCAWG). Data sharing goes beyond the academic world. Many public-private partnerships have been set up, in order to make sure that discoveries are not only published, but also applied in a product such as a medical device, a medicine or a computer program.[9]

Funding programmes, such as the Horizon 2020 Research and Innovation Programme of the European Union, very much stimulate data sharing with companies. In May 2016, it was announced that Deepmind, a company owned by Google and most famous for its innovative use of AI, was given access to the healthcare data of up to 1.6 million patients from three hospitals run by a major London NHS trust[9]. Of course, when sharing data with commercial parties, privacy needs to be taken into account.

Can sharing of data be based on just the 'Good Samaritan' principle? Or, should data be monetized, keeping in mind the cost of the resources plugged to collect, aggregate and extrapolate it? Or, should indirect incentivization be promoted, instead? Are people who have data connected with the very people who need it?

Currently, incentivisation, pertaining to collection of data, in India is mainly indirect or non-financial in nature, one such example citied during the discussion was: Swachh Survekshan report; The MoUD modified the scoring pattern for ranking cities on cleanliness (by transferring) 100 points previously allotted to "municipal documentation" to "citizen engagement". Municipal documentation refers to urban local bodies' (ULBs') assessment of their own infrastructure efforts on collection, transportation and disposal of solid waste, and strategies for open defecation-free towns. Citizen engagement consists of two parts: Online, telephonic and social media surveys, and the use of the Swachhata App for citizens to connect with their ULBs. The app (carries) 150 points, which accounts for 25% of citizen feedback points (600 points) and holds 7.5% weightage in the overall scoring of 2,000 points[10].

The reservation of a certain percentage of points for sharing of qualitative data (which serve social scientists well as these are subjective responses) serves as a good initiative for collection of data.

---

[9] Quinn, B. Google Given access to Healthcare Data of up to 1.6 Million Patients. 2016. Available online:
https://www.theguardian.com/technology/2016/may/04/google-deepmind-access-healthcare-datapatients (accessed on 25 April 2020).
[10]https://www.business-standard.com/article/current-affairs/swachhata-bharat-app-shows-citizen-are-driving-up-the-cleanliness-rankings-117051700278_1.html
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7246891/

Of course, before we into the length and breadth of incentivisation, the focus has to be on collection of data required for to meet an end cause in a timely manner, via a structured process, in a standardized format and not disjointed, while maintaining the privacy and safety of data.

Conclusion

Health-tech companies and start-ups are testing waters to gauge what their role in the bigger picture is going to be, especially since the former are, mostly, in a nascent stage. This also encourages them to think on lines of a business model as far as the sharing of data is concerned, especially since collection, aggregation and extrapolation of data is an expensive affair. A cause of anxiety for them would, hence, be forced sharing of data without any sensical business model attached. Data sharing on the other hand is paramount to enable improvisation of public health care management.

When we're dealing with public health research, wasted data can translate into shorter, less healthy lives. Improving data management so that data can be shared is a first step to reducing that waste. But it will not be enough. We need to change the incentives that pit the interests of individual researchers against the interests of public health, that pit institutional interests against the more rapid advancement of knowledge and understanding. Governments may hold micro-data back from international organizations, but there's no excuse for international organizations to limit access to the aggregate data that governments do provide[11].

Essential are discussions between all stakeholders to reach an amicable middle path; essential is leadership that will take this forward beyond just a statement of good intentions.

---

[11] https://www.who.int/bulletin/volumes/88/6/09-074393/en/

About IDF: IDF is powered by the belief that rigorous research should inform debates, discussions and decisions on matters of public policy. Over the past 16 years of IDF's journey, the organization has worked with over 80 partners including Governments, multilaterals, corporates and civil society organisations. IDF's research is based along four pillars: fostering an innovation economy; using technology for developmental objectives; enhancing India's human capital; and measuring the efficacy of development policies. More info on: www.idfresearch.org

About CDF: CDF was launched on 30th October, 2019, with a vision to conduct actionable research on the impact of digitization on the economy and society. The inquiries are analytical, without any pre-determined bias, multi-dimensional and evidence-based and provides policy and regulatory insights that enable the transition to an optimal digital economy and society. The Centre was established and incubated as an entity by the India Development Foundation (IDF), a private non-profit research organisation set up as a Trust in 2003. More info on: www.cdfresearch.org