[CHAPTER] – THE IMPORTANCE OF DATA SHARING IN MANAGING PUBLIC HEALTH CRISES[1]

Authors: Rahul Patil, Inder Gopal
Centre for Society and Policy, Indian Institute of Science, Bangalore

BOOK: Global Pandemic and Human Security: Technology and Development Perspective

Abstract

The COVID-19 crisis has emphasized the importance of the free flow of information and data-driven applications in the management of public health crises. This chapter examines the potential benefits, concerns, and solutions related to sustainable and secure access to public health data. We study some of the data-driven solutions in action worldwide and present them as replicable use cases. We also examine why a large volume of data from public and private sources never reaches the desks of decision-makers and suggest technical and policy solutions to eliminate these sources of "data friction".

Keywords

Data sharing, personal data, non-personal data, public health crises, human security

1. Introduction

The unprecedented COVID-19 crisis has taught us many lessons on improving the management of global public health. The outbreak disrupted many existing solution approaches and shook societal, technological, and regulatory norms. The critical resource underlying many proposed new solutions and approaches is data. We have seen the practical benefits of sharing accurate and trusted public-health non-personal data, namely data that does not violate any individual's privacy. To maximize global benefit, such data should be freely available for all, without unnecessary national or international legal restrictions on access or usage. The rapid availability of data enabled the global community to identify the spread of COVID-19 in a matter of two weeks. In comparison, it took several weeks to months in the previous outbreaks of Severe Acute Respiratory Syndrome in 2002, Ebola in 2013, and Zika virus in 2014. Researchers were able to deposit the first novel coronavirus genome sequences in the GeneBank database at record speed, i.e. the sequence with accession number MN908947. The accelerated and open access to this sequence data aided countries in designing the diagnostic kits on a war footing at breakneck speed.

Such cross-border flows of data are indeed praiseworthy. However, there is a need to strengthen the international frameworks and to harmonize the legal instruments that facilitate this data flow. To their credit, international bodies have recognized this requirement. In 2001, the World Health Organisation (WHO) brought the attention of its member states to the flow of information and

---

[1] This work is supported by a grant from Omidyar Network. We gratefully acknowledge the discussions held with the Omidyar Network team and other academicians in shaping this paper.

surveillance data of urgent international importance by passing a resolution called Global Health Security: Epidemic Alert and Response (WHA, 2001). These efforts have tried to blunt major human insecurity caused through epidemics and bioterrorism with trans-national approaches that are data-centric, people-centred, context-specific, and prevention-oriented (UNTFHS, 2016). But, there remain many other ways in which the flow of data can be encouraged.

This chapter highlights the potential benefits, issues, and solutions related to sustainable and secure access to public health data. We show how data from official and governmental sources can complement data from non-governmental and private sources to enable a vast new set of public health applications. These new data-driven applications will correlate data, apply sophisticated analytics and inferencing, and will be able to proactively address public health problems and deliver new value to citizens. We examine a series of use cases *(Section 2)* where such applications were able to better manage pandemics and improve chances of avoiding one. While public agencies will create some applications for the public good, it is crucial to create a climate where private companies are also incentivised to create innovative new data-driven solutions for commercial benefit.

We also explore the reasons *(Section 3)* why a large volume of data from public and private sources never reaches the desks of decision-makers. We suggest technical and policy solutions (Section 4) to eliminate the various inhibitors in sharing data and show how eliminating this data friction is a necessary (but not sufficient) step in taking full advantage of data. International agencies particularly advise governments globally to strengthen data governance for the effective application of open data and big data analytics in the COVID-19 crisis (Yao and Park, 2020). We make some recommendations on the public IT infrastructure and public data policy frameworks that governments will need to put in place to ensure an appropriate level of public health data sharing.

2. Examples of data-driven applications contributing to public health

We examine a series of examples (or use-cases) where data-driven applications have shown the way to better manage disease outbreaks, improve chances of not catching a disease, or prevent outbreaks from turning into pandemics. We divided these examples into several categories based on the scope of application. For each category, we describe an example use-case typifying the benefits of data sharing within that category. We map a few similar use-cases in the Figure 1 to gather insights through comparative assessment. The details for each use-case are consolidated based on the publicly available data from literature, official press releases, and media reports. The use cases are compiled for demonstration purposes only and are not intended to be exhaustive. More use cases can also be proposed and considered to demonstrate the importance of data sharing, governance and application, and comparative analyses in a particular domain.

2.1 Use-case categories and examples:

Use-case 1: Dashboard-driven decision-making for epidemic surveillance & management (Example studied: BlueDot)

BlueDot provides a software-as-a-service platform predicting high-exposure epidemic areas worldwide. It analyses complex correlations of streaming datasets across demographic, migration, health, and weather dynamics sourced from government and international agencies, researchers, companies, and media. They can potentially detect severe outbreaks as much as ten days in advance, as seen in the COVID-19 outbreak (Allam, Z., et al., 2020).

BlueDot's solutions require structured and non-personal datasets. The dynamicity and timeliness of the sourced data vary across datasets. Some of the datasets like anonymized geotagged locations of cell phones or airport ticket data might be dynamic with real-time or near-real-time streaming. They also feed in static datasets like census and demographic statistics or simulated data like environmental predictors facilitating disease transmissions. Obtaining a wider variety of data from diverse sources is difficult as the government and international agencies' reporting is scarce, not readily available, non-real-time, or brings limitations. A lack of access to reliable local data and reporting delays pose major concerns during rapidly evolving disease outbreaks (Grubaugh, N. D., et. al., 2019). Though enhanced data sharing activities are crucial in such situations, efficient data privacy and security protocols must be enforced across data sharing networks and systems (Allam, Z., et al., 2020).

Disease outbreak monitoring, mapping, and management are crucial to curtailing the disease spread from becoming an epidemic outbreak. Characteristics of other dashboards, that of Johns Hopkins University (Dong, E., et. al., 2020) and System for Opioid Overdose Surveillance (SOS) (Ballesteros A, et. al. 2020) can be seen in the Figure 1.

Use-case 2: Phylodynamics tool for tracking pathogen evolution (Exmple studied: Nextstrain)

Nextstrain is a platform for tracking pathogen transmission and evolutionary patterns to retrieve epidemic history from genomic data (Hadfield, J., et al. 2018). The platform equips researchers to perform phylodynamic exercises using bio-sequences sourced from public repositories like NCBI, GISAID, ViPR, or GitHub. Nextstrain accelerated the process of sharing, mapping, and visualising the genomic data midst coronavirus pandemic as close to two days to weeks. It took a year to complete these operations in Ebola or former outbreaks. Nextstrain's phylogenetic charts and family trees helped epidemiologists, policymakers, and medical practitioners at the earliest. Nextstrain is open-sourced under the strongest copyleft GNU AGP license and hosted as a web application at nextstrain.org (Nextstrain: Copyright license, 2016).

Nextstrain respects researchers' intent to publicly or privately share the data with the options to store it in client-side computers or their servers. Nextstrain aims to maintain ~1hr turnarounds to channelise publicly available data into the live transmission and evolution tracking (Bedford, T., 2020). The platform's functionality depends upon the genome sequence submissions by researchers globally. The Open Science Prize felicitated Nextstrain's contributions with the grants of $230 000 for prototype development (OSP, 2017).

It can be inferred that micro-level assessments with bioinformatics tools can potentially fine-tune the decision-making in disease outbreaks management. Other tools such as Coronapp (Mercatelli, D., et. al., 2020) equip researchers to study the pathogen evolutions by monitoring nucleotide and polypeptide mutations (refer Figure 1).

Use-case 3: Virtual repositories of biospecimens and associated metadata (Example studied: UK Biobank)

UK Biobank is a resource centre providing genotypic and phenotypic data from large population-scale studies. This data is sourced from the genetic and clinical analyses of the medical samples donated by ~500,000 participants recruited across the United Kingdom from 2006 to 2010 (Bycroft, C. et. al., 2018). These participants also provided consent to their electronic medical records, EMRs. Any bonafide researcher worldwide from the academic, charity, public, and commercial entity can use this open-access resource for pre-approved research purposes only (Sudlow C, et al. 2015). The Biobank shares the de-identified data on ratifying material transfer agreements and undertakings about keeping data secure without re-identification attempts. The users of these datasets have to mandatorily publish their results along with derived data and return them to the Biobank. UK Biobank hosts terms and conditions backed by robust data sharing policies and non-discriminatory, transparent access protocols. UK Biobank is planning to avail the resource on its own cloud-based Data Analysis Platform since summer 2021 to facilitate access to the researchers without storage and processing capabilities (UK Biobank News, 2020a). The database is estimated to grow by 15 petabytes by 2025.

Similarly, we also examined EuroBioBank's role and data sharing characteristics and mapped in the matrix, Figure 1 (Mora, M., et al., 2014).

U*se case 4: Integrated Command and Control Centres (ICCC) (Example studied: Indian Smart Cities)*

Integrated Command and Control Centres (ICCCs) established in India's Smart Cities Missions are repurposed into COVID-19 War Rooms to tackle coronavirus outbreak. Total 47 such centres took the responsibility to provide city-level emergency responses around the clock and employed GIS-based COVID tracker dashboards (PIB Press Release, 2020a). They source data from city hospitals and tools such as cellular or IoT devices, CCTV surveillance cameras, or interactive voice response system based toll-free helpdesks (Deloitte Report, 2020). Access to these resources is restricted to the city administrators and authorised personnel. Some of the dashboard features are made available to the public on the respective municipality's websites.

ICCCs retrieve both personal and non-personal data. These datasets primarily include details of COVID-19 diagnostic tests, coronavirus hotspots, disease heat maps, geotags of home-quarantined people, lane closures and traffic data, audio files of helplines, surveillance camera footage, mobile questionnaires, etc. (PIB Press Release, 2020b). Some of these dynamic datasets are either real-time or updated daily. Functionalities of ICCC's are based on the heavy stakeholder coordination between several municipalities and services.

Similar crisis command centres are found in other parts of the world including the one at the State of Utah, USA (Baird, R. P., 2020).

Use case 5: Smart Public Infrastructure (Example studied: South Korean Smart Bus Shelter)

The smart bus shelter is a pilot initiative adopted as part of the policy initiative for enhanced convenience to Seoul's travellers. Such shelters are well-equipped with IoT sensors and interconnected smart devices capable of collecting, transmitting, and delivering data remotely (Seoul Press Release, 2020). COVID-19 oriented data-driven technologies auto-regulate their functionalities at the smart bus shelters, such as automated thermal-imaging cameras allowing persons with temperature below normal body temperatures, air-conditioning and sterilizing systems, sanitizer dispensers, bus arrival-departure schedules, automatic detection of crimes and fires (Park, M., 2020). The district administration restricts the data and shares it with police or fire authorities in crimes or fires (France-Presse, A., 2020).

These sensors-based systems use de-identified dynamic data to enhance convenience for travellers. These systems also interact with the users in real-time, e.g., aids voice assistance to those who require it, provides suggestions on wearing a mask in a disease outbreak. Such initiatives require extensive collaboration between administration and technology/solution providers (Park Han-na., 2020).

This use case facilitates citizens' comfort through IoT-enabled healthcare repones and explores opportunities of digital epidemiological surveillance using interconnected IoT devices. Similar, example (Figure 1) can be seen in the case of IoT-based pedestrian monitoring project for residents at Liverpool's City Centre, England (Verstaevel, N. et al., 2020).

2.2 Summary

The table shown in Figure 1provides a bird-eye-view of all the use cases and maps the characteristics of the underlying data sharing approaches. We map approaches such as data sourcing (including data origin, size, format, nature), delivery (including data end-user, format of data delivery, type of beneficiary), and intellectual property aspects (including data ownership and data access). We also assess the type of data transacted, including – variety (nature of data, e.g., types of data being used or delivered); velocity (timeliness of data, e.g., real-time, near-real-time, non-real-time); variability (change in data, e.g., dynamic or static); veracity (robustness: completeness or accuracy of shared data); and volume (size of data transferred). We also consider stakeholder stewardship as an essential factor in the ease of data sharing.

Comparative assessment of all the use cases provides essential insights into the preferences of varied organisations. Organisations prefer utilising non-personal data over personal data for both internal and external monetisation of data assets. Many organisations face the issue of mixed data where personal and non-personal data are inextricably linked. All the organisations prefer multisourced, dynamic, structured, and semi-structured datasets. Organisations use real-time, near real-time, non-real-time data, or their combinations depending upon the need of the desired results. Most of the time, organisations face limitations and inconsistencies in the datasets sourced from the government and international agencies. Organisations with public funding opt to develop open source applications and share raw data or derived insights with full public access considering the fact of the sensitivity of the data.

		Use case categories	Dashboard-driven decision-making for pandemic surveillance and management			Phylodynamics or Informatics of Pathogen Evolution		Virtual biobanks		Crisis Command Centres		IoT-enabled healthcare response	
Use Case Characteristics	Use Case Dimensions	Use case examples I/O Data Characteristics	John Hopkins' Dashboard	BlueDot	The System for Opioid Overdose Surveillance	Nextstrain	Coronapp	UK Biobank	EuroBioBank	ICCCs – Indian Smart Cities	Command Centres at State of Utah	Pedestrian monitoring – Liverpool's City Centre	IoT-based antivirus bus shelters – South Korea
PRIVACY (Influx of data-type based on user and system identifiability)	Sensitivity	Personal data			✓								
		Non-personal data	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Mixed data			✓			✓	✓	✓	✓		
VARIETY (Use of data based on technical requirements)	Content style	Structured data	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
		Unstructured data	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Data Sources	Singular											
		Multiple	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VELOCITY (Need of real-time updates)	Data timeliness	Real-time	✓	✓		✓			✓	✓	✓	✓	✓
		Near real-time	✓	✓	✓	✓	✓					✓	
		Non real-time		✓				✓	✓				
VARIABILITY (Format of data used)	Rate of change	Dynamic	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
		Static						✓	✓				
INTELLECTUAL PROPRTY (Ownership and access dimensions of information and information system)	Data access type	Full public access	✓			✓	✓						
		Limited public access			✓	✓		✓	✓	✓	✓	✓	
		Restricted access		✓	✓	✓				✓	✓	✓	✓
	Application ownership	Open source	✓			✓	✓	✓	✓				
		Proprietary		✓	✓					✓	✓		
	Data ownership	Public data	✓			✓	✓			✓	✓	✓	✓
		Private data			✓	✓		✓	✓	✓	✓		
DATA SOURCE (What are the sources of information?)	Data Origin	Instruments								✓		✓	✓
		Internet of Things								✓		✓	✓
		Web	✓	✓									
		Surveys	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
		Commercial activity		✓							✓		
		Simulations	✓	✓	✓					✓	✓		✓
DATA DELIVERY (Audiences and data users for products and services)	Data Destination	Multipurpose usage	✓										
		Policymakers & Administrators		✓	✓	✓				✓	✓	✓	✓
		Research Community			✓	✓	✓	✓	✓				
		Public/Private Services			✓					✓	✓	✓	✓
DATA STEWARDSHIP (Stakeholders engaged across value chain from funding to value delivery)	Stakeholders involved	Private entity		✓				✓			✓		✓
		Research orgs	✓	✓	✓	✓	✓	✓	✓				
		Government orgs	✓	✓	✓			✓	✓	✓	✓	✓	✓

Figure 1: At-a-glance Mapping of Data Sharing Characteristics at Organisations of Interest – Use Case Matrix

Organisations adopt multiple data sharing approaches if data owners seek special provisions, as in Nextstrain. Almost all the organisations under study collected data by some sort of survey. Some adopted multiple strategies to procure the relevant data. These organisations preferably target a particular audience due to the specificity and granularity of datasets and insights achieved in their respective domains. We noticed that all the organisations emphasise multi-stakeholder engagement and collaborate with government entities or research institutions in particular.

3. Inhibitors in sharing data

The previous section shows how much value can be created by sharing of data. Yet, the use cases capture range of inhibitors. A plethora of data from public and private sources never reaches the desks of decision-makers. The public health data available for sharing remains restricted to government sources and is just a small fraction of the universe of public health data. The generic term, *data friction*, describes the various inhibitors that create barriers in sharing data. Eliminating data friction becomes a necessary (but not sufficient) step in taking full advantage of data.

There are four major causes of data friction.

- The data is not available or not of high enough quality.
- The data cannot be found by those who might need it.
- The data cannot be accessed and understood by those who try to use it.
- There are real or perceived security and privacy issues which inhibit sharing.

We examine each cause in more detail.

1. High-quality data is not available:

The biggest inhibitor in sharing data is that the data is simply not available or of poor quality. Several critical health metrics are simply not being measured, or the measurements are sporadic and often inaccurate. Some of the reasons are organizational, the cost of measurement, lack of appreciation of the value of data, and lack of access due to large numbers and constant mobility.

There are many strong organizational reasons why data is not shared. City administrators see no reason to share data and often view data sharing as an invitation to criticism. Why share data when it can be used against me? Broader engagement in data sharing can be dampened by the concerns about competition and loss of market share in for-profit organisations compared to the non-profit organisations as seen in the reports by Adler-Milstein J, et al. (2011).

The cost of data collection and storage is another significant inhibitor. For example, is estimated that the overhead for data collection (West, D. R. et. al., 2012) in a popular diabetes quality improvement program the estimated per-practice cost of implementation and first-year maintenance for the data collection and reporting was approximately $26,000 per practice. This is a significant amount for something which is often perceived as peripheral to patient care.

Poor and inefficient communication channels are also a frequent cause. The COVID-19 crisis has spawned many creative ways to communicate data-ranging from WhatsApp texts, scanned scraps of paper, Memes, short videos to telephone calls. These methods, however, need to scale as well as be standardized.

Yet another reason for poor data quality is that the data often remains unused and there is no feedback to the data provider. In the social sciences (and physics and experimental physics), the *observer's paradox* is a situation in which the phenomenon being observed is unwittingly influenced by the presence of the observer/investigator. In the case of data collection, the observer's paradox manifests itself in a positive fashion, namely in improvements in data quality as data providers receive feedback if the data is not usable. As an example, in the IUDX project (DE Framework, 2020), it is reported that installed Air Quality Data sensors often lost their calibration, and as a result the collected data was almost useless. The act of making data available through IUDX made the data providers cognizant of quality issues, and data quality improved rapidly even without any overt actions.

2. Finding pertinent data:

If the data universe expands to a multiplicity of sources, how does a prospective data consumer locate pertinent data? Currently, no catalogue or directory of data exists for users to identify and describe data. Such a searchable catalogue becomes essential to create a useful data ecosystem.

For example, in many Indian cities, it is reported (IUDX, 2021) that different city departments are unaware of data being collected by each other. Traffic police, Fire department and Hospitals collect critical data that can be of help in responding to a natural disaster or terrorist attack. As a result, the different departments are simply not aware of the data being collected by the others. In times of emergency, there is lack of coordination in emergency response.

A related issue is the lack of coherency in health data retention and archival policies. For health-related data, this implies that authorities should mandate longer minimum maintenance period for the clinical information (Corn, M., 2009; Dong, L., 2015). Recent guidelines issued by the National Digital Health Mission (NDHM) of India assigned the responsibility of long-term storage of health records of hospitals, diagnostic centres, and clinics to Health Repository Providers (HRPs), who will also provide them NDHM compliant software. On the other hand, Health Lockers are particularly assigned to provide software and long-term health records storage to individuals (NDHM, 2020).

3. Understanding the data:

Even if the data source gets identified, the same type of data is often represented differently from different sources. This may be as simple as different measurement units (Celsius versus

Fahrenheit, for temperature) or something more sophisticated as the object model for a data object.

Standards such as ETSI's NGSI-LD and the Smart Data Models from TM-Forum provide standardized ways of representing and accessing of public data. More specifically, the Need for standard-oriented interoperability of health IT systems is well recognized in a variety of instances. For example, it is well known that portability of medical records across public-private entities and states will reduce costs and save time (Balsari, S., 2014). Portability might necessarily arise due to the multiplicity of healthcare providers and healthcare facilities and for maintaining the patient care continuum (Radhakrishna K, et. al., 2014). On average, 5% of healthcare service providers such as hospitals or healthcare professionals change their IT solutions provider. Recent final rules of the 21st Century Cures Act availing the EHI export functionality might increase the portability of EHI and decelerate the rate of switching IT systems (ONC, 2020). Even though Beacon Communities achieved higher-than-average in EHRs adoption, lack of interoperability led to complicated information exchange (UC-NORC, 2014).

Health data needs to be represented in a standard-based fashion but standardized metadata is also essential. Metadata are the additional descriptors essential to describe the characteristics of data. Edward et al. (2011) reports metadata as a source of friction for impeding data sharing, particularly in scientific communications. Metadata is considered a fixed and highly structured information product, e.g., sets of descriptors, links, XML tags, catalogues, etc. Scientific communities need to consume additional resources to create, use, and manage such data leading to metadata friction (Edwards PN, 2010). Various metadata databases, catalogues, or standards are being employed to reduce this friction and increase the replicability of scientific outcomes. These include the Ecological Metadata Language (EML) (EML, 2011), Biological Data Profile (BDP) (FGDC, 1999), Content Standard for Digital Geospatial Metadata (CSDGM) (FGDC, 1998), Geographic Markup Language (GML) (GML, 2011), ISO Geospatial Metadata (ISO 19115) (ISO 19115-1, 2014), etc.

4. Privacy and security policies for health data sharing:

Personal health data about an individual should *never* be shared without the consent of the individual. These considerations are of paramount importance and there must be clearly understood opt-in and/or opt-out consents of individuals regarding collection, sharing, and deletion of data.

However, these concerns are often painted with too broad brush and used as an excuse for a blanket ban on sharing data. Personal data can sometimes be shared in aggregate or with appropriate anonymization without violating privacy. Moreover, while non-personal data is usually not subject to privacy considerations, privacy considerations are often invoked because of a lack of clarity in the interpretation.

Government regulations often limit the ability to share data for reasons that arise from security concerns. An example of this are laws with localisation provisions require domestic (within the national territory) installation of data servers storing personal or other data. Experts notice that localisation requirements in regulations might potentially hinder the free flow of information, data sharing, and reuse (OECD, 2018). It is debatable how useful these localization laws are and many suggest they are based more in xenophobia rather than genuine security considerations. Policies and regulations need to balance securing access to the citizens' data and free-flow of the data mindful of privacy and security concerns.

4. **Enablers for sharing data,**

Lowering data friction facilitates data sharing. Reducing data friction in a multi-faceted task that requires a combination of technical and non-technical approaches. We have organized these approaches into several broad categories as outlined below.

a. **Data sharing platforms:**

Data Sharing Platforms are a basic prerequisite for receiving data from many sources, consistently managing privacy and security, and normalizing all the data through a common set of Application Programming Interfaces (API's) and Data Models. These platforms also ensure control by data providers on who gets access to their data and enable them to obtain monetary compensation if appropriate. The data shared through these platform includes all types of data related to health care such as suitably anonymized patent and treatment outcome information, hospital statistics, aggregated health metrics, etc. In addition many environmental and other factor that relate to health of citizens are shared such as air pollution, traffic density, flooding information, water quality, etc.

There are three basic categories of data sharing platforms:

1. **Open Data Platforms:** The simplest are "Open Data Platforms" which aggregate data and allow access to all comers. Examples include national platforms such as Open Government Data (OGD) Platform India data.gov.in. The advantage of an open data platform is that anyone can get free access. The disadvantage of an open data platform can be loss of possible control of who gets access to data. This reduces dramatically the amount and type of data that will be shared as the data providers are concerned about losing control. Open data platforms are popular because they offer the opportunity to rapidly open up and share, but they are limited in their flexibility and attractiveness to data providers.

2. **Data Warehouses or Lakes:** An approach very common in corporate settings is a data warehouse (large warehouses as referred to as Data Lakes), which aggregates all data into a common place, using a common way of representing data, and under a common

set of access control mechanisms. Mechanism such as Extract, Transform and Load (ETL) are used to extract data from their existing repositories to the common data warehouse. Data warehouses have support for a wide variety of data analytics. The approach is often used by hospital systems to share data amongst their different departments. As an example, Geisinger built the Keystone community data warehouse as a freestanding repository where data is pooled from Geisinger and non-Geisinger community hospitals. (UC-NORC, 2014; Allen C., et al., 2014; McCarthy, D.B., et. al., 2014). Where personal data that is subject to governmental piracy mandates (such as HIPPA (HIPPA, 1996) or Personal Data Protection Bill (PDP Bill, 2019)), Data Warehouses are often the best approach as they offer tight and predictable protection since all the data is stored and managed in a single place. However, this approach does not scale to a broad deployment where health data is obtained from a wide variety of sources, public and private.

3. The third category of data sharing platform is termed **"data exchanges"**. Many cities such as Copenhagen (CDE, 2018), Columbus (smartcolumbusos.com) and Manchester (https://urbandata.exchange) have taken ownership of their data assets by creating such *data exchanges*. The objective of a data exchange is to *'interconnect data silos'* and not to aggregate data as in a data warehouse. An important idea behind a data exchange is that data silos are actually not a bad thing, as each silo often represents a domain-optimized service that performs that function very well. Instead of breaking silos or moving data en masse into a central repository, this approach chooses to interconnect the disparate and distributed entities. This provides a way for accessing data in a unified, common format, allowing for sharing of data between different departments in a city, as well as opening up data for third party developers to create innovative new applications and citizen services. In addition, there is an opportunity for third party providers of data, or third-party providers of data analytics or data annotation, to participate in what becomes a data marketplace.
   An example of note which encompasses many cities, the India Urban Data Exchange (IUDX), is described in the IUDX Discussion Paper (DE Framework. 2020). IUDX is completely open source, based on an underlying framework of open Application Programming Interfaces, data models, and the security, privacy and accounting mechanisms that will facilitate, easy and efficient exchange of data among disparate urban data silos. IUDX draws on ideas and, where feasible, code, from best-of-breed global projects such as CityVerve (cityverve.org.uk) and Fiware (fiware.org). It is based on important emerging standards in the space including NGSI_LD from European Telecommunications Standards Institute – ETSI (ETSI Press Release, 2019), and Smart Data Models created by the Telecom Management Forum (GitHub/Smart Data Models, 2020). IUDX is an example of a system well-tuned to the ecosystem of the country, including cultural norms, city nomenclature, payment and identification systems, etc.

b. Establishing the system of data providers:

The value of sharing data is only as useful as the availability of data. Thus establishing an organized network of data providers is essential. An example is the US-based eHealth Exchange. It is a healthcare information exchange network for sharing clinical information nationwide, electronically; vendor-independent, largest query-based network; incorporating - both federal and private institutions; 75% of all US hospitals, doctor offices, healthcare professionals; and facilitating more than 30 EHR technologies – Now: independent, non-profit health information network, supported by The Sequoia Project (formerly Healtheway) – Previously: began as the Nationwide Health Information Network (NHIN or NwHIN), the federal program under the Office of the National Coordinator for Health IT (ONC) (eHealth Exchange, 2018; FMPHPCE, 2014).

c. Development of an application ecosystem:

The iPhone did not become a success until a critical mass of independent application developers built innovative applications and services for the platform. A similar application ecosystem will be required to make a data platform successful. The application developers will require nurturing and encouragement and will need clear business models for managing interactions and dependencies, establishing and operating infrastructure systems, monetising data and associated services, and developing long-term expertise. Other significant factors for a cohesive, comprehensive ecosystem would include tools and support along with setting up the basic values cum philosophy of the ecosystem.

In the health case space, the need for creating an application ecosystem is particularly pronounced. While applications in, say, gaming or social networking can be readily launched with limited initial investment, health care applications require significantly more domain knowledge and expertise. Often, they involve careful and painstaking compliance with legal mandates and other regulations. For certain categories of applications, they may require extensive testing and possibly human trials. Thus, the capabilities of application vendors and qualities providing expected sophistication and quicker solutions needs to be carefully curated.

In the case of the State of Utah and Iowa, application vendors established digital infrastructure in a few days with necessary sophistication to manage COVID-19 pandemic versus EHR vendors for the Greater Tulsa Health Access Network Beacon Community took longer than anticipated, about seven to eight months, for setting information exchange platforms (Domo Press Release, 2020; UC-NORC, 2014).

d. Public investments in data sharing infrastructure or platforms

The capabilities of data sharing infrastructures and facilities available for secure and private data exchange may impact the overall success of large-scale healthcare digitisation programs. Appropriate stimulus through public investments in data sharing infrastructure can potentially galvanise the transition to sustainable adoption of data sharing technologies.

The HITECH Act of 2009 took forward steps to promote health information technologies and incentivising their use. More particularly, HITECH Act provisions tried to overcome financial obstacles by availing extra payments to the eligible healthcare professionals with meaningful use of EHRs and bonuses or extra payments-per-discharge of Medicare patients to the hospital with meaningful use of EHRs (Blumenthal, D., 2010).

To address the lack of infrastructure to exchange health information among stakeholders, the US Federal government channelled over $560 million in funds to states being the custodians of Medicaid and public health data. The US government also accelerated the incubation of Nationwide Health Information Network, a health information exchange infrastructure that is now an independent eHealth Exchange. Through these investments nurtured heavier adoption of digital EHRs, various other multi-dimensional barriers need to be addressed at the same time apart from technical barriers such as business barriers, complex privacy laws, misaligned procedural and incentivising mechanisms, etc. (HITPC, 2015).

e. Strengthening a culture of data sharing

The most pernicious factor that inhibits data sharing is the view amongst officials that they are better off hoarding data. Building a culture of data sharing becomes imperative. Diverse data types, different subject domains, multiple locations, and host institutions highlight the broad range of existing agencies and capacities that require them to come together for effective use of public health data. The default behaviour of parties that have public health data should be to share unless there are privacy reasons not to do so. Research demonstrates a change in overall culture (versus agency behaviour) has long term positive implications for policy design and implementation. Culture change would include an environment of trust, common data sharing values and norms, commonality in rules, and institutionalization of interest of all parties.

This cultural shift must be strongly advocated by the administration and leadership. Simply writing down a set of rules in a personnel handbook will not suffice. In a recent study, it was found that program officers were convinced that their data policy at the National Science Foundation, ensured publication of publicly funded data, even though the majority of NSF-funded Principal Investigators displayed near-complete ignorance of this policy and were reluctant to share data. The answer in these and other cases is not stricter standards, but a successful effort to integrate understandings of the working culture and practices of health care professionals into the design and implementation of those standards (Edwards, P. N., et. al., 2011).

Building a broad-based data driven health mission at a national or state level is essential to make this data cultural shift. As an example, the US HITECH Act as a part of the American Recovery and Reinvestment Act 2009, emphasized data driven health care. The HITECH Act provided $35 billion in grants and incentive payments to adopt and use EHRs at hospitals by health professionals. This legislation backed electronic Health Information Exchange (HIE), pushed digitisation of all patients' health records, mandated data standards facilitating nationwide data reuse by service providers, and incentivised hospitals to adopt certified EHR technology (Reisman M., 2017; Henry J, et. al., 2016; Mennemeyer, S. T., et. al., 2015). The HITECH Act of 2009 took several innovative steps to promote health information technologies and incentivise their use. The act's provisions tried to overcome financial obstacles by availing extra payments to the eligible healthcare professionals with meaningful use of EHRs and bonuses or extra payments-per-discharge of Medicare patients to the hospital with meaningful use of EHRs (Blumenthal, D., 2010).

f. Economic models incentivising data sharing or reward systems

Creating a broad-based data ecosystem is difficult if the impetus for sharing data is entirely top-down driven. Public and Private parties will share data if needed in response to an emergency or government mandate. However, routine sharing in normal times becomes far more likely if there are economic benefits for all parties involved. Public-private partnerships and data management ecosystems where government and private share for monitoring, supporting, and managing for the public good.

There are many examples of such data monetization. These are broadly divided into three categories:

1) Direct monetization: This type of monetisation primarily involves direct buying and selling of data assets considering economic and other benefits. Data providers and users adopt various modes of exchange of data assets which include either direct peer-to-peer transactions; use of trusted and managed third party data exchange ecosystems (Gilad-Bachrach, R., 2019) (examples, https://catalogue.iudx.org.in, https://www.quadrant.io, https://aws.amazon.com/data-exchange); or multi-party decentralised blockchain based data exchanges (ITU-T, 2019; Sikeridis, D., et. al., 2020) (examples, https://datapace.io, https://www.dawex.com/en/blockchain-data-exchange, or https://streamr.network).

2) Monetization of derived analytics: The power of data analytics is a breakthrough for variety of institutions. Data analytics services are being traded as a key commodity (Jiao, Y., et al., 2018). Derived data analytics potentially equips organisations to

achieve higher rate of performance by delivering innovative insights for decision-making. As seen in the case of BlueDot, they keep their clients well-equipped for epidemic surveillance & management by sharing real-time insights on emergence of disease outbreak.

3) New business models: The data-driven digital space is witnessing disruption with the emerging business models. One of the prime examples include monetization of personal DNA data to assess genetic ancestry and personalised genetic histories. Companies such as ancestry.com are offering direct-to-consumer genetic ancestry testing and use consumer's DNA data to compare with a vast library of DNA data (Royal, C. D., et. al., 2010). Results data further supplements historical documents for researching genealogy (Shriver & Kittles, 2004). Nextstrain's use case *(Section 2.2)* also captures an innovative model mapping pathogens' evolution using biosequences sourced from huge public repositories.

5. Conclusion

This book chapter shows the critical role of data and specifically data sharing in improving public health management. Global policymakers must understand how important the globalisation of a data-driven public health system is to maintain human security, and strengthen and harmonize the legal and policy frameworks that facilitate the required transnational data. We have highlighted the practical benefits of sharing accurate and trusted public health non-personal data. A comparative assessment of use cases of data-driven applications has provided some insights to better healthcare outcomes. We also emphasized the importance of breaking data silos and the need for decision-makers need to complement government data with non-government and private sources. Private companies also need to be included and incentivised to create novel and innovative data-driven solutions. We have suggested technical and policy solutions to overcome data friction and its impact and believe these approaches can be valuable in global management of disease outbreaks.

6. References

(Adler-Milstein J, et. al. 2017) Adler-Milstein J, DesRoches CM, Jha AK. Health information exchange among US hospitals. Am J Manag Care. 2011 Nov; 17 (11): 761-8. PMID: 22084896. https://pubmed.ncbi.nlm.nih.gov/22084896/

(Allam, Z., et al., 2020) Allam, Z., Dey, G., & Jones, D. S. (2020). Artificial Intelligence (AI) Provided Early Detection of the Coronavirus (COVID-19) in China and Will Influence Future Urban Health Policy Internationally. AI, 1(2), 156–165. doi:10.3390/ai1020009 https://www.mdpi.com/2673-2688/1/2/9 (last accessed on 14 December 2020)

(Allen C., et al., 2014) Allen, C., Des Jardins, T.R., Heider, A., Lyman, K.A., McWilliams, L., Rein, A.L., Schachter, A.A., Singh, R., Sorondo, B., Topper, J. and Turske, S.A., 2014. Data Governance and Data Sharing Agreements for Community-Wide Health Information Exchange: Lessons from the Beacon Communities. eGEMs (Generating Evidence & Methods to improve patient outcomes), 2(1), p.5. DOI: http://doi.org/10.13063/2327-9214.1057

(Baird, R. P., 2020) Baird, R. P. (2020). How Utah's Tech Industry Tried to Disrupt Coronavirus Testing. The Newyorker. https://www.newyorker.com/tech/annals-of-technology/how-utahs-tech-industry-tried-to-disrupt-coronavirus-testing (last accessed on 14 December 2020)

(Ballesteros A, et. al. 2020) Ballesteros A, Roche J, Totty Z. (2020). Expansion of the Michigan system for opioid overdose surveillance (SOS) to inform data-driven community responses. Injury Prevention; 26: A22. DOI: 10.1136/injuryprev-2020-savir.55 https://injuryprevention.bmj.com/content/26/Suppl_1/A22.2 (last accessed on 14 December 2020)

(Balsari, S., 2014) Balsari, Satchit. (2014). A digital push for healthcare: Portability of data will cut costs, save time. Indian Express. https://indianexpress.com/article/opinion/columns/india-health-sector-data-ecosystem-niti-aayog-national-health-stack-bluebook-6553690/

(Bedford, T., 2020) Bedford, T. (2020). Early warnings of novel coronavirus from genomic epidemiology and the global open scientific response. Blog at bedford.io. https://bedford.io/blog/genomic-epi-for-ncov-response/ (last accessed on 14 December 2020)

(BlueDot: Press Release, 2019) Press Release by BlueDot. (2019). Digital Health Company BlueDot Secures $7M USD Series A Financing. Inside BlueDot. Toronto. August 7th, 2019. https://bluedot.global/digital-health-company-bluedot-secures-7m-usd-series-a-financing/ (last accessed on 14 December 2020)

(Blumenthal, D., 2010) Blumenthal, D. (2010). Launching HITECH. New England Journal of Medicine, 362(5), 382–385. DOI: 10.1056/nejmp0912825 https://www.nejm.org/doi/full/10.1056/NEJMp0912825

(Bycroft, C. et. al., 2018) Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. Nature, 562(7726), 203–209. doi:10.1038/s41586-018-0579-z https://www.nature.com/articles/s41586-018-0579-z (last accessed on 14 December 2020)

(CDE, 2018) City Data Exchange. 2018. City Data Exchange - Lessons Learned from a Private/Public Data Collaboration. Municipality of Copenhagen and Capital Region of Denmark. https://cphsolutionslab.dk/media/pages/projekter/data-platforms/city-data-exchange/1837671186-1612174620/city-data-exchange-cde-lessons-learned-from-a-public-private-data-collaboration.pdf

(Corn, M., 2009) Corn, M. (2009). Archiving the Phenome: Clinical Records Deserve Long-term Preservation. Journal of the American Medical Informatics Association, 16(1), 1–6. doi:10.1197/jamia.m2925 https://academic.oup.com/jamia/article/16/1/1/864394

(DE Framework, 2020) DE Framework. (2020). Data Exchange Framework for Indian Smart Cities. Discussion Paper. Smart Cities Mission, Ministry of Housing and Urban Affairs (MoHUA), Government of India. http://www.iudx.org.in/archives/Discussion_Paper-Data_Exchange_Framework_for_Indian_Smart_Cities.pdf

(Deloitte Report, 2020) India Smart Cities COVID 19 Response: Case Studies. (2020). Deloitte. https://www2.deloitte.com/content/dam/Deloitte/in/Documents/public-sector/in-gps-IndiaSmartCitiesCOVID19Response_11Jun_Print-noexp.pdf (last accessed on 14 December 2020)

(Domo Press Release, 2020) Domo Press Release. 2020. Domo Launches the COVID-19 Crisis Command Center for the State of Iowa https://www.domo.com/news/press/domo-launches-the-covid-19-crisis-command-center-for-the-state-of-iowa

(Dong, E., et. al., 2020) Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases. doi:10.1016/s1473-3099(20)30120-1 https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/fulltext (last accessed on 14 December 2020)

(Dong, L., 2015) Dong, L. (2015). Taking the long view of medical records preservation and archives. Journal of Documentation, 71(2), 387–400. doi:10.1108/jd-11-2013-0141 https://www.emerald.com/insight/content/doi/10.1108/JD-11-2013-0141/full/html

(Edwards PN, 2010) Edwards PN (2010) A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming. Cambridge, MA: MIT Press. https://mitpress.mit.edu/books/vast-machine

(Edwards, P. N., et. al., 2011) Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. Social Studies of Science, 41(5), 667–690. doi:10.1177/0306312711413314 https://journals.sagepub.com/doi/abs/10.1177/0306312711413314 (eHealth Exchange, 2018) eHealth Exchange overview. 2018. https://ehealthexchange.org/what-we-do/ or https://s3.amazonaws.com/seqprojectehex/wp-content/uploads/2018/09/12051759/eHealth-Exchange-Overview-2-20-18.pdf

(EML, 2011) Ecological Metadata Language (EML) 2011 https://www.dcc.ac.uk/resources/metadata-standards/eml-ecological-metadata-language

(ETSI Press Release, 2019) ETSI Press Release. (2019). ETSI CIM Group Releases Full-Feature Specification for Context Information Exchange in Smart Cities https://www.etsi.org/newsroom/press-releases/1519-2019-01-etsi-cim-group-releases-full-feature-specification-for-context-information-exchange-in-smart-cities?jjj=1615695136274

(FGDC, 1998) Federal Geographic Data Committee (1998). Content Standard for Digital Geospatial Metadata. Metadata Ad Hoc Working Group, Federal Geographic Data Committee. FGDC-STD-001-1998. https://www.fgdc.gov/standards/projects/metadata/base-metadata/index_html

(FGDC, 1999) Biological Data Profile (1999). Part 1: Content Standard for Digital Geospatial Metadata 1998, Biological Data Working Group, Federal Geographic Data Committee and USGS Biological Resources Division. FGDC-STD-001.1-1999 https://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/biometadata/biodatap.pdf

(FMPHPCE, 2014) The Impacts of the Affordable Care Act on Preparedness Resources and Programs: Workshop Summary. Washington (DC): Forum on Medical and Public Health Preparedness for Catastrophic Events; Board on Health Sciences Policy; Board on Health Care Services; Institute of Medicine. National Academies Press (US); 2014 Aug 27. Available from: https://www.ncbi.nlm.nih.gov/books/NBK241386/ doi: 10.17226/18755

(France-Presse, A., 2020) France-Presse, A. (2020). South Korea installs anti-virus bus shelters with temperature sensors and UV lamps. The Guardian. https://www.theguardian.com/world/2020/aug/13/south-korea-installs-anti-virus-bus-shelters-with-temperature-sensors-and-uv-lamps (last accessed on 14 December 2020)

(Gilad-Bachrach, R., 2019) Gilad-Bachrach, R., Laine, K., Lauter, K., Rindal, P., and Rosulek, M. (2019). Secure Data Exchange: A Marketplace in the Cloud. In Proceedings of the 2019 ACM SIGSAC Conference on Cloud Computing Security Workshop (CCSW'19). Association for Computing Machinery, New York, NY, USA, 117–128. DOI: 10.1145/3338466.3358924 https://dl.acm.org/doi/abs/10.1145/3338466.3358924

(GitHub/Smart Data Models, 2020) GitHub: Smart Data Models, 2020. FIWARE Foundation, IUDX and TM Forum. https://github.com/smart-data-models/data-models/blob/e40715f2d2914d181146dabd89d66513e4a4a3ce/README.md

(GML, 2011) Geography Markup Language (GML). 2011. https://www.loc.gov/preservation/digital/formats/fdd/fdd000296.shtml

(Grubaugh, N. D., et. al., 2019) Grubaugh, N. D., Saraf, S., Gangavarapu, K., Watts, A., Tan, A. L., Oidtman, R. J., … Andersen, K. G. (2019). Travel Surveillance and Genomics Uncover a Hidden Zika Outbreak during the Waning Epidemic. Cell, 178(5), 1057–1071.e11. doi:10.1016/j.cell.2019.07.018 https://www.sciencedirect.com/science/article/pii/S0092867419307834 (last accessed on 14 December 2020)

(Hadfield, J., et al. 2018) Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R. A. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. doi:10.1093/bioinformatics/bty407 https://academic.oup.com/bioinformatics/article/34/23/4121/5001388 (last accessed on 14 December 2020)

(Henry J, et. al., 2016) Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of Electronic Health Record Systems Among U.S. Non-Federal Acute Care Hospitals: 2008–2015. Washington, DC: Office of the National Coordinator for Health Information Technology; May, 2016. https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php

(HIPPA, 1996) Health Insurance Portability and Accountability Act of 1996 (HIPAA). 1996. Public Law 104–191. 104th Congress, the United States of America. https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf

(HITPC, 2015) Challenges and Barriers to Interoperability. (2015). Report to Congress. Prepared by The Health Information Technology Policy Committee (a federal advisory committee to the Office of the National Coordinator for Health IT focused on health information technology policy). https://www.healthit.gov/sites/default/files/facas/HITPC_Final_ITF_Report_2015-12-16%20v3.pdf

(ISO 19115-1, 2014) ISO 19115-1:2014. 2014. Geographic information — Metadata — Part 1: Fundamentals. ISO/TC 211 Geographic information/Geomatics. https://www.iso.org/standard/53798.html

(ITU-T, 2019) ITU-T. (2019). Proposal to Advance Draft Technical Specification of “data exchange and sharing for supporting IoT and SC&C”. Blockchain-based data exchange and sharing for supporting IoT and SC&C: Technical Specification D3.6. Telecommunication Standardization Sector of ITU 2019. ITU-T Focus Group on Data Processing and Management to support IoT and Smart Cities & Communities. https://www.itu.int/dms_pub/itu-t/opb/fg/T-FG-DPM-2019-3.6-PDF-E.pdf

(IUDX, 2021) IUDX: India Urban Data Exchange. (2021). https://nudm.mohua.gov.in/wp-content/uploads/2021/02/IUDX-Booklet-FINAL.pdf

(Jiao, Y., et al., 2018) Jiao, Y., Wang, P., Feng, S., & Niyato, D. (2018). Profit Maximization Mechanism and Data Management for Data Analytics Services. IEEE Internet of Things Journal, 5(3), 2001–2014. doi:10.1109/jiot.2018.2819706 https://ieeexplore.ieee.org/abstract/document/8326475

(McCarthy, D.B., et. al., 2014) McCarthy, D.B., Propp, K., Cohen, A., Sabharwal, R., Schachter, A.A. and Rein, A.L., 2014. Learning from Health Information Exchange Technical Architecture and Implementation in Seven Beacon Communities. eGEMs (Generating Evidence & Methods to improve patient outcomes), 2(1), p.6. DOI: http://doi.org/10.13063/2327-9214.1060

(Mennemeyer, S. T., et. al., 2015) Mennemeyer, S. T., Menachemi, N., Rahurkar, S., & Ford, E. W. (2015). Impact of the HITECH Act on physicians' adoption of electronic health records. Journal of the American Medical Informatics Association, 23(2), 375–379. doi:10.1093/jamia/ocv103 https://academic.oup.com/jamia/article/23/2/375/2572425

(Mercatelli, D., et. al., 2020) Mercatelli, D., Triboli . L., Fornasari, E., Ray, F., Giorgi, F. M. (2020). Coronapp: A web application to annotate and monitor SARS-CoV-2 mutations. Journal of Medical Virology. doi: 10.1002/jmv.26678. https://onlinelibrary.wiley.com/doi/10.1002/jmv.26678 (last accessed on 14 December 2020)

(Mora, M., et al., 2014) Mora, M., Angelini, C., Bignami, F. et al. (2014). The EuroBioBank Network: 10 years of hands-on experience of collaborative, transnational biobanking for rare diseases. European Journal of Human Genetics, 23(9), 1116–1123. doi:10.1038/ejhg.2014.272 https://www.nature.com/articles/ejhg2014272 (last accessed on 14 December 2020)

(NDHM, 2020) Guidelines for Health Information Providers, Health Repository Providers, Health Information Users and Health Lockers. (2020). National Digital Health Mission. https://ndhm.gov.in/documents/hip_hiu_Policy

(Nextstrain: Copyright license, 2016) Nextstrain/auspice license GNU Affero General Public License v3.0 (2016). https://github.com/nextstrain/auspice/blob/master/LICENSE.txt (last accessed on 14 December 2020)

(OECD, 2018) OECD Expert Workshop on Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-Use. 2018. Working Party on Security and Privacy in the Digital Economy. Danish Business Authority. Directorate for Science, Technology and Innovation: Committee on Digital Economy Policy. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/CDEP/SPDE(2018)4&docLanguage=En

(ONC, 2020) Final Rules of 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. 2020. Prepared by Office of the National Coordinator for Health Information Technology (ONC), the US Department of Health and Human Services (HHS). https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification

(OSP, 2017) Open Science Prize. (2017). nextstrain.org wins Open Science Prize! https://www.opensicenceprize.org/p/s/2023040/

(Park Han-na., 2020) Seoul unveils virus-blocking bus shelter. The Korea Herald. http://www.koreaherald.com/view.php?ud=20200806000634 (last accessed on 14 December 2020)

(Park, M., 2020) Park, M. (2020). Keeps out rain and COVID-19, Seoul tries smart bus shelter to fight virus. Reuters. https://uk.reuters.com/article/amp/idINKCN25A1U4 (last accessed on 14 December 2020)

(PDP Bill, 2019) Personal Data Protection Bill, 2019. The Republic of India. Bill No. 373 of 2019. http://164.100.47.4/BillsTexts/LSBillTexts/Asintroduced/373_2019_LS_Eng.pdf

(PIB Press Release, 2020a) 5th Anniversary of Urban Missions. Ministry of Housing & Urban Affairs. (2020). Press Information Bureau, Government of India. https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1634268 (last accessed on 14 December 2020)

(PIB Press Release, 2020b) Smart city missions integrated data dashboards being used at Command and Control Centres to monitor COVID-19. Ministry of Housing & Urban Affairs. (2020). Press Information Bureau, Government of India. https://pib.gov.in/PressReleasePage.aspx?PRID=1611592 (last accessed on 14 December 2020)

(Radhakrishna K, et. al., 2014) Radhakrishna K, Goud BR, Kasthuri A, Waghmare A, Raj T. (2014). Electronic health records and information portability: a pilot study in a rural primary healthcare center in India. Perspectives in Health Information Management. 2014; 11:1b. https://europepmc.org/article/pmc/pmc4142510

(Reisman M., 2017) Reisman M. EHRs: the challenge of making electronic data usable and interoperable. Pharmacy and Therapeutics. 2017; 42 (9): 572–5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5565131/

(Royal, C. D., et. al., 2010) Royal, C. D., Novembre, J., Fullerton, S. M., Goldstein, D. B., Long, J. C., Bamshad, M. J., & Clark, A. G. (2010). Inferring Genetic Ancestry: Opportunities, Challenges, and Implications. The American Journal of Human Genetics, 86(5), 661–673.

doi:10.1016/j.ajhg.2010.03.011 https://www.sciencedirect.com/science/article/pii/S0002929710001552

(Seoul Press Release, 2020) Seoul city, bus stop upgrades to the world's first 'Smart Shelter'. (2020). Smart City, South Korea. Ministry of Land, Infrastructure and Transport. https://smartcity.go.kr/en/2020/06/09/서울시-버스정류소-세계최초-스마트쉘터로-전면-업/ https://www.seoul.go.kr/news/news_report.do#view/318141 (last accessed on 14 December 2020)

(Shriver & Kittles, 2004) Shriver, M. D., & Kittles, R. A. (2004). Genetic ancestry and the search for personalized genetic histories. Nature Reviews Genetics, 5(8), 611–618. doi:10.1038/nrg1405 https://www.nature.com/articles/nrg1405

(Sikeridis, D., et. al., 2020) Sikeridis, D., Bidram, A., Devetsikiotis, M., & Reno, M. J. (2020). A blockchain-based mechanism for secure data exchange in smart grid protection systems. 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC). doi:10.1109/ccnc46108.2020.9045368 https://ieeexplore.ieee.org/document/9045368

(Sudlow C, et al. 2015) Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. (2015) UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med 12(3): e1001779. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001779 (last accessed on 14 December 2020)

(UC-NORC, 2014) NORC, University of Chicago. (2014). Data Sharing to Enable Clinical Transformation at the Community Level: IT Takes a Village. Evaluation of the Beacon Community Cooperative Agreement Program. https://www.healthit.gov/sites/default/files/beacondatasharingbrief062014.pdf

(UK Biobank News, 2020a) UK Biobank creates cloud-based health Data Analysis Platform to unleash the imaginations of the world's best scientific minds (2020) https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/uk-biobank-creates-cloud-based-health-data-analysis-platform-to-unleash-the-imaginations-of-the-world-s-best-scientific-minds

(UN, 2016). United Nations Trust Fund for Human Security. (2016). Human Security Handbook. https://www.un.org/humansecurity/wp-content/uploads/2017/10/h2.pdf

(Verstaevel, N. et al., 2020) Verstaevel, N. et al. (2020). Assessing the effects of mobility on air quality: The Liverpool Smart Pedestrian project. Transportation Research Procedia, 48, 2197–2206 https://www.sciencedirect.com/science/article/pii/S2352146520306967 (last accessed on 14 December 2020)

(West, D. R. et. al., 2012) West, D. R., Radcliff, T. A., Brown, T., Cote, M. J., Smith, P. C., & Dickinson, W. P. (2012). Costs Associated with Data Collection and Reporting for

Diabetes Quality Improvement in Primary Care Practices: A Report from SNOCAP-USA. The Journal of the American Board of Family Medicine, 25(3), 275–282. doi:10.3122/jabfm.2012.03.110049 https://europepmc.org/article/med/22570390

(WHA, 2001). World Health Assembly, 54. (2001). Global health security - epidemic alert and response: report by the Secretariat. World Health Organization. https://apps.who.int/iris/handle/10665/78718

(Yao and Park, 2020) Yao, K. and Park, M. K. (2020). Strengthening Data Governance for Effective Use of Open Data and Big Data Analytics for Combating COVID-19. UN/DESA Policy Brief 89. UN Project Office on Governance, Division for Public Institutions and Digital Government, UN Department of Economic and Social Affairs – DESA. https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-89-strengthening-data-governance-for-effective-use-of-open-data-and-big-data-analytics-for-combating-covid-19