

Review Article

Ownership, privacy, and value of health-care data: Perspectives and future direction

Anthony Vipin Das¹, Madhu Viswanathan²

¹Department of eyeSmart EMR & AEye, LV Prasad Eye Institute, ²Institute of Data Science, Indian School of Business, Hyderabad, Telangana, India.



***Corresponding author:**

Anthony Vipin Das,
Associate Director, LV Prasad
Eye Institute, Hyderabad,
Telangana, India.

vipin@lvpei.org

Received : 16 November 2022

Accepted : 10 January 2023

Published : 06 May 2023

DOI

10.25259/IHOPEJO_33_2022

Quick Response Code:



ABSTRACT

In this ever expanding explosion of data in the world, we are at a crucial juncture to balance quality and quantity. In healthcare, there is a need to analyze voluminous datasets for the benefit of the patients while respecting their privacy and ownership. There is a need to understand the fundamental framework of co-creation of the data between the health-care provider and the patient. There is no more opportune time such as this to harness the potential of large datasets in healthcare to catalyze value-based care for the population.

Keywords: Data ownership, Healthcare, Electronic medical records, Big data, India

WHAT IS DATA AND ITS CHARACTERISTICS?

The past two decades have been defined by the staggering explosion in the amount of data generated by us. It is estimated that in 2021, people created 2.5 quintillion bytes of data every day. In addition, by 2022, 70% of the globe's gross domestic product (GDP) will have undergone digitization.^[1] The volume of data/information created, captured, copied, and consumed worldwide has increased from 2 zettabytes in 2010 to over 97 zettabytes in 2020, with a projected forecast of 181 zettabytes in 2025 [Figure 1].^[2] The generation of data has also shown exponential growth as can be evidenced by the fact that over 90% of the data of the world has been generated in the last 2 years.

A popular term in this space to describe a large volume of data that can be analyzed computationally to reveal trends, associations, and patterns is "Big Data." Specifically, big data are identified by the 5 "V" s which are volume, velocity, veracity, variety, and value [Figure 2].^[3] Volume is characterized by the sheer number of records or transactions being generated across time. Velocity is characterized by the real-time creation or streaming of data at a constant rate. Veracity is characterized by trustworthiness and authenticity of the data being generated. Variety is characterized by structured, unstructured, and multimodal data sets. Value is finally dependent on all of the above attributes and provides insights to drive growth and impact.

Health-care data are unique in its diversity across various populations and geographies. While the challenges of standardization of the delivery of care are compounded by the partial digitization of its capture across the world, there is a great opportunity to work on the generation of quality health-care data to realize the vision of value-based care. While various other sectors including social data, machine data, and transactional data have progressed into billion, trillion to quintillion bytes of data, structured health-care data that can be mined are lagging on the scale.

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, transform, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

©2023 Published by Scientific Scholar on behalf of IHOPE Journal of Ophthalmology

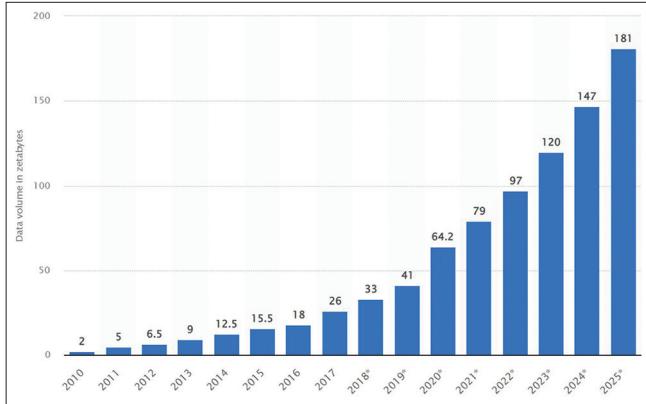


Figure 1: Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 (in zettabytes).

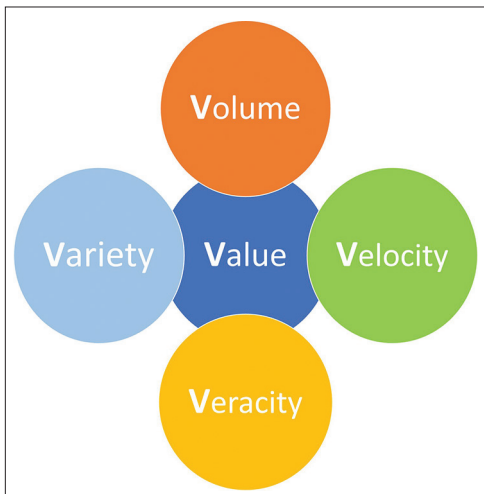


Figure 2: Five “V”s of Big Data.

WHERE IS DATA GENERATED?

Health-care data are growing with the increasing integration of technology into the delivery of care around the world. Today, the health-care industry generates approximately 30% of the world’s data by volume.^[4] By 2025, the compound annual growth rate of data for healthcare will reach 36%. This growth rate is 6% faster than manufacturing, 10% faster than financial services, and 11% faster than media and entertainment [Figure 3]. The various sources that generate big data in the health-care industry include hospital records, medical records of patients, results of medical examinations, biomedical research, and devices that are a part of internet of things (IoT). A health-care data point is created with the first interaction of the patient with the care provider through the patient registration that captures the most important aspect of the sociodemographic information of the patient. Age, gender, socioeconomic status, and geographic location form one of the most crucial signals that guide the identification of vulnerable cohorts to suffer from a particular disease. The clinical history

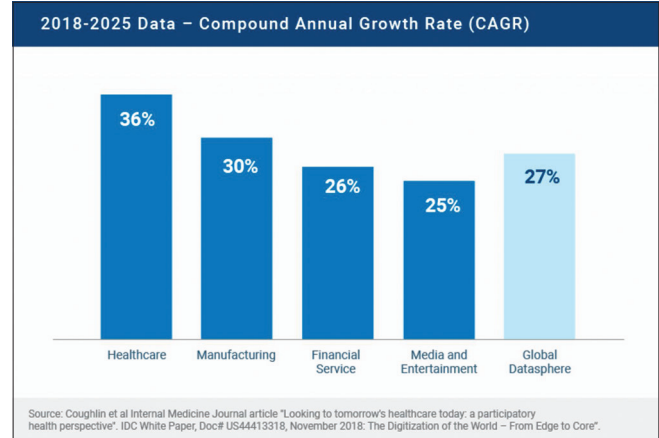


Figure 3: Global compound annual growth rate of health-care data.

and symptoms elicited from the patient and the documentation of the signs from the examination further adds to the deduction of the diagnosis that determines the next steps of medical or surgical care. The data generated from investigations form the bulk of the data generated in healthcare that ranges from various media such as clinical imaging [X-ray, computed tomography Scan (CT Scan), and Magnetic Resonance Imaging (MRI)], clinical photographs, audio, and video. There has been a rise in the use of personalized IoT devices globally. From the smart phone device in our hand to the wearable device on our wrist, today there is a vast amount of digital health information being captured from the users. It is expected that our digital device interactions will increase from 1400 interactions/person/day by the end of 2020 to 4909 interactions/person/day by 2025. There are large health-care datasets that exist across different regions, notably the National Health Service (NHS) digital,^[5] Intelligent Research in Sight (IRIS) Registry,^[6] European Health Data Space,^[7] Global Health Observatory Data Repository,^[8] and eyeSmart EMR.^[9] Big data techniques allow us to look at patterns in the presentation of disease, treatment outcomes, and the possibility to offer personalized templates to deliver care for the patients. There has been an evolution of the size of the patient sample from the hundred, thousand to the million and beyond^[10] over the past decades [Figure 4] and there is an increasingly evolving debate on the insights that are derived at each stage.^[11]

Another challenge that is often confronted with the data that are being collected is that about 80% of all of the health-care data are unstructured. This requires different ways of data collection, storage and analysis from that done with traditionally structured datasets. One of the big impetus for big data techniques in healthcare is the caution surrounding analysis and inferences built on smaller datasets^[12] that might not be representative of the population and synthetic datasets^[13] that are not a substitute for real-world data (RWD). In fact, there has been an emphasis on RWD and real-world evidence (RWE) for health-care decisions. RWE is

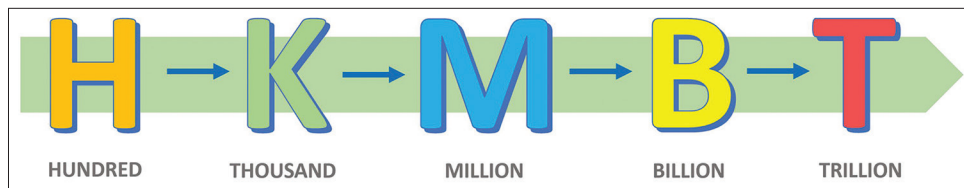


Figure 4: The evolution of the sample size in health-care research.

clinical evidence regarding the usage and potential benefits, or risks of a medical product derived from analysis of RWD and represents a more realistic picture of the health outcomes. For instance, the U.S Food and Drug Administration uses RWD and RWE to monitor post-market safety and adverse events and to make regulatory decisions. The health-care community uses the data to support coverage decisions and to develop guidelines and decision support tools for use in clinical practice. Medical product developers also use RWD and RWE to support clinical trial designs (e.g., large simple trials and pragmatic clinical trials) and observational studies to generate innovative, new treatment approaches. The RWD can come from several sources such as electronic health records, claims and billing activities, product and disease registries, and patient-generated data including in home-use settings and data gathered from other sources that can inform on health status, such as mobile devices.

HOW IS DATA GENERATED?

There has been an increasing debate on the ownership of patient data over the past few years and it is important to understand how the data are being generated to understand the arguments in this debate. In the hospital lifecycle of a patient, data are generated at multiple points both clinically, surgically, and financially. The first touch point for the patient involves providing the sociodemographic details such as national ID, age, gender, location, socioeconomic status, and ethnicity among others. Then, the health-care provider (clinical or paraclinical) elicits the symptoms, for which the patient has presented to the clinic. The health-care provider with the consent of the patient then proceeds to elicit signs from the patient and correlates them with the symptoms mentioned by the patient. The next course of action can involve ordering a set of diagnostic tests (invasive or non-invasive) to gather additional information. Once this data is in place, a combination of self-reported information by the patient coupled with the examination and expertise of the healthcare provider is analyzed to arrive at a diagnosis. It is reasonable to make the argument that the data are CO-CREATED as it contains not only information from the patient but also the healthcare provider who together generate this Protected Health Information (PHI). The notion that the entire PHI belongs to the patient alone merits debate. The perspective that patients own their entire health-

care data, and they decide whom to share with different health-care providers and still maintain continuity of care has not really taken off at scale. There are reasons that are not necessarily privacy-centered, such as lack of awareness of benefits and cost of implementation of such services that could explain the lack of take-off. Further, a lot of benefits from big data come from aggregation across several individuals and accrue to society and public health as such which may not be internalized by individuals. Collaborative creation of data ensures that both parties have equal access to the data they have co-created. It is the accountability of the health-care provider to ensure the privacy of the health-care data that they have created with the patient according to the legal regulations.

Health-care providers must ensure a culture of trust and transparency in the delivery of care to patients in accordance with data privacy and legal regulations. There has been a paradigm shift in the way personal data that are being perceived in the hands of service providers across various domains. The United Nations Conference on Trade and Development reports that 137 out of 194 countries have put in place legislation to secure the protection of data and privacy. About 71% of countries have legislation in place, 9% of countries have draft legislation, 15% of countries do not have any legislation, and 5% of countries have no data.^[14] The Health Insurance and Portability and Accountability Act (HIPAA) enacted in 1996 in the United States determines the data privacy and security requirements of PHI and all stakeholders handling this data must ensure compliance.^[15] The various covered entities include health-care providers, health plans, health-care clearinghouses, and business associates. Most of the privacy laws around the world have defined precise geolocation data, ethnic origin, genetic data, biometric information for purposes of identification, and health information as sensitive personal information. Informed consent is a key component in medicine and is the explicit documented approval given by the patient to receive the medical intervention after having reflected on the pros and cons of the same.^[16] Making informed consent an informed choice in the language known to the patient is of utmost importance to prevent the misuse of information collected by the health-care provider. In most LMIC countries, general practitioners do not maintain a standard consent form. It is either implied or risks being

coercive and is usually collected during registration or before any intervention.^[17] Broad blanket consents collected by the industry are to be discouraged that may run the risk of misuse of the health-care data, but, on the other hand, may prevent the realization of the true potential of the application of machine learning or artificial intelligence algorithms on large datasets gathered from the population.

WHO IS COLLECTING THE DATA?

A very important aspect to understand is the motivation behind the collection of data and who benefits from data. There are many stakeholders in the ecosystem and the final value that is generated depends significantly on the comprehensive definition of the data points that are being collected. From the perspective of the clinicians, the data must be comprehensive in the sociodemographics to understand the distribution of the diseases. The medical and surgical documentation at the presenting visit helps to understand the severity of the disease at presentation. The temporal trends also help to quantify the treatment outcomes for the patients.

From a health-care organization perspective, the operational efficiency, and costs of the delivery of health-care services, the reimbursements from the public and private sectors are some of the most important key performance indicators that need to be tracked at regular intervals. The insurance industry mandates the collection of basic information for claims verification and reimbursements, but it lacks the granularity of the clinical documentation including outcomes that will help define value-based health-care services. Patient engagement groups would want to collect quality of life data, patient apprehensions, confidence, and trust levels in the system.

Public health researchers collect sample population data and basic scientists collect the genome data which are stored in disparate databases that do not interact with each other. Non-governmental organizations collect data from the field for evaluation of government schemes and to understand the social determinants of health. The government collects large amounts of data from the population due to the sheer magnitude of its presence across the population, but it does have the inherent limitation of not being granular enough.

COUNTRY REGULATIONS

As more data are being collected and generated several laws and regulations are being created across the world to structure compliance on issues surrounding ownership, protections, responsibilities, usage, and exchange of data. These compliance regulations can be categorized as those pertaining to all types of personally identifiable information and other special category information such as the General

Data Protection Regulation (GDPR)^[18] and compliance regulations specific to PHI such as HIPAA. While HIPAA is very specific to health-care data and applies to health-care providers, health plans, and health-care clearinghouses, and business associates, the broader laws apply to any data controller (entities that deal/process personal data) and data processors (entities that process data on behalf of others).^[19] Given the many different sources of data that are being gathered nowadays, it is imperative that we have a more general data compliance policy. To understand why, notice that nowadays a lot of health-care information is being provided by firms that operate/manufacture wearable smart health devices. The data collected by these firms would not be governed by laws having a strict definition of the health-care market. Both general data protection regulation (GDPR) and HIPAA require informed consent, right to access, update, and portability of records. While HIPAA had loopholes, newer laws such as GDPR and California Consumer Privacy Act^[20] require that data portability happens in a common format allowing for a certain degree of interoperability. While all laws require consent, there are exceptions built in as well. For instance, the Data Privacy Bill (DPB),^[21] 2021 that was tabled and later withdrawn in India allows personal data to be processed without obtaining consent under certain grounds such as, (i) *responding to any medical emergency*; and (ii) *for any provision or service to be received by an individual from the State*. In addition to consent all compliance regulations have rules regarding notifications of breaches and penalties that can be potentially levied in case of non-compliance. The regulations also specify appointments of data protection officers (as in GDPR) or privacy officers (as in HIPAA) to oversee implementation of privacy policy in line with regulations and to enforce compliance. Finally, rules also specify regular impact assessment in terms of data protection and compliance. The common thread that ties all these regulations, despite differences in definitions, language, and obligations, is the reliance on principles related to qualified consent and consumer/patient privacy rights.

EVOLVING ECOSYSTEM

Data and associated data science are seeing a golden era where advances in technology, data collection and data science are allowing us to leapfrog traditional challenges in healthcare. Advances in artificial intelligence (AI)/machine learning (ML) technologies have demonstrated remarkable progress in image-recognition tasks and have been particularly adept in assessments of radiographic characteristics and have impacted many different aspects of radiology and oncology treatment.^[22] While data science can be a force multiplier in the health-care context and provide efficiencies for individuals, organizations, and societies, they are also sensitive to abuse. The Ayushman Bharat

Digital Mission (“ABDM”)^[23] envisions a federated data architecture and an ecosystem that is based on technological interoperability between various entities, “consent” as well as “privacy and security of personal data.” This requires many building blocks such as personal health records, electronic medical records, data consent managers, and health information exchanges (HIEs). At present, there are no clear linkages between the data requirements envisioned in ABDM and the DPB, 2021. The DPB, 2021 includes “health data” as sensitive personal data and defines it as the data related to the state of physical or mental health of the data principal and includes records regarding the past, present or future state of the health of such data principal, *data collected in the course of registration for, or provision of health services, data associated with the data principal to the provision of specific health services*. It is not clear if data collected from ancillary services such as insurance providers would fall under this definition. Similarly, there are no clear rules that govern information sharing in HIEs and to whose benefit.^[24] In many several cases, organizational and public health concerns may override individual consent requirements.

FUTURE DIRECTIONS

Data are the new oil, but oil is of no use if it is not processed appropriately. In order to do so we need to evolve into a value-based approach, where the collection of health-care data points to aid in the understanding of the efficiency of the care being provided. This is possible through a holistic approach that combines a piece of everyone. The point of interaction with the patient where this framework is most conducive to being collected is in the hospital. We must encourage our health-care organizations on their digital transformation journeys and contribute to the value-based health-care pool of information. We need to be cautious of the inclusivity of the data being collected as well. The Vision and Eye Health Surveillance System report on the IRIS registry in the United States highlighted certain limitations such as lack of patient level data that prevents tracking outcomes at a granular level, not considered representative of the general population due to its current nature, does not include all the ophthalmology practices, cannot identify the payer-specific procedures and that the eye examination rates are not reported due to the lack of a suitable denominator.^[25] The ability to collate nationwide data is a great opportunity to understand the nature of the services being provided to the population, but we also need to complete the circle by quantifying the value delivered through the outcomes. This will enable us to prioritize the patients in need and to create a positive reinforcing mechanism for the health-care providers to focus on quality rather than the quantity of the care being provided to those in need.

Declaration of patient consent

Patient’s consent not required as there are no patients in this study.

Financial support and sponsorship

Nil.

Conflicts of interest

Anthony Vipin Das is on the editorial board of the Journal.

REFERENCES

1. Available from: <https://www.techjury.net/blog/how-much-data-is-created-every-day/#gref> [Last accessed on 2022 Oct 25].
2. Available from: <https://www.statista.com/statistics/871513/worldwide-data-created> [Last accessed on 2022 Oct 25].
3. Ishwarappa, Anuradha J. A brief introduction on Big data 5Vs characteristics and hadoop technology. *Procedia Comput Sci* 2015;48:319-24.
4. Coughlin S, Roberts D, O’Neill K, Brooks P. Looking to tomorrow’s healthcare today: A participatory health perspective. *Intern Med J* 2018;48:92-6.
5. Available from: <https://www.digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets> [Last accessed on 2022 Oct 27].
6. Available from: <https://www.aao.org/iris-registry/about> [Last accessed on 2022 Oct 27].
7. Available from: https://www.health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en [Last accessed on 2022 Oct 27].
8. Available from: <https://www.apps.who.int/gho/data/node.home> [Last accessed on 2022 Oct 27].
9. Das AV, Kammari P, Vadapalli R, Basu S. Big data and the eyeSmart electronic medical record system- An 8-year experience from a three-tier eye care network in India. *Indian J Ophthalmol* 2020;68:427-32.
10. Das AV. People to policy: The promise and challenges of big data for India. *Indian J Ophthalmol* 2021;69:3052-7.
11. Gumpili SP, Das AV. Sample size and its evolution in research. *IHOPE J Ophthalmol* 2022;1:9-13.
12. Available from: <https://www.kaggle.com/code/rafjaa/dealing-with-very-small-datasets/notebook> [Last accessed on 2022 Oct 29].
13. Kimmel R, Tai XC, editors. *Processing, Analyzing and Learning of Images, Shapes, and Forms: Part 1*. 1st Ed. Netherlands: Elsevier; 2018.
14. Available from: <https://www.unctad.org/page/data-protection-and-privacy-legislation-worldwide> [Last accessed on 2023 Jan 10].
15. Available from: <https://www.cdc.gov/phlp/publications/topic/hipaa.html> [Last accessed on 2022 Oct 29].
16. Available from: <https://www.ahrq.gov/health-literacy/professional-training/informed-choice.html> [Last accessed on 2022 Nov 02].
17. Saksena N, Matthan R, Bhan A, Balsari S. Rebooting consent

- in the digital age: A governance framework for health data exchange. *BMJ Global Health* 2021;6:e005057.
18. Available from: <https://www.gdpr-info.eu> [Last accessed on 2022 Nov 02].
 19. Tovino SA. The HIPAA privacy rule and the EU GDPR: Illustrative comparisons. *Seton Hall Law Rev* 2016;47:973-93.
 20. Available from: <https://www.oag.ca.gov/privacy/ccpa> [Last accessed on 2023 Jan 10].
 21. Available from: <https://www.mondaq.com/india/privacy-protection/1213494/a-guide-to-the-data-protection-bill-2021> [Last accessed on 2022 Nov 02].
 22. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. *Nat Rev Cancer* 2018;18:500-10.
 23. Available from: <https://www.nha.gov.in/PM-JAY#:~:text=Ayushman%20Bharat%20PM%2DJAY%20is,the%20bottom%2040%25%20of%20the> [Last accessed on 2022 Nov 02].
 24. Available from: <https://www.mondaq.com/india/privacy-protection/1150676/health-data-under-the-data-protection-bill-2021-and-recommendations-of-the-joint-parliamentary-committee-on-data-protection> [Last accessed on 2023 Jan 10].
 25. Available from: <https://www.cdc.gov/visionhealth/vehss/data/ehr-registries/iris.html> [Last accessed on 2022 Nov 02].

How to cite this article: Das AV, Viswanathan M. Ownership, privacy, and value of health-care data: Perspectives and future direction. *IHOPE J Ophthalmol* 2023;2:41-6.