Impact of Big Data on the Privacy of Mental Health Patients

J. Robert Langlois
Indiana University Bloomington, School of Informatics and Computing
langloir@umail.iu.edu

ABSTRACT

Society has experienced a lot of benefits with the introduction of technology. Today, one of the essential functions of technology is the collection, storage, processing, and transmission of data. The healthcare industry, including mental health services, are huge benefactors of these advances in technology. From birth, medical facilities start collecting information about all individuals; they do so even up to the point of death and all points in between. Over a lifetime, that is an abundance of information about an individual. The question that must be answered is, how is that data be protected to ensure patients' privacy rights? The more information collected on individuals, the more responsibility is assumed by those who collect data; methods for how the data is collected, used and shared must ensure the protection of patients' privacy rights. This challenge is one that needs to be navigated and addressed by medical professionals and facilities, policymakers, and the individuals whose data is collected. Specifically in the mental health field, by resolving patients' privacy concerns, policymakers and researchers can transform the field by introducing more cost effective strategies, ensuring patients' sense of security, and establishing new and more appropriate norms to communicate sensitive health information.

KEYWORDS

Big Data, Mental Health and Privacy i523

1 INTRODUCTION

We live in an era where data is constantly being produced; data exists everywhere in large quantities. The advances in technology have opened the door for businesses to collect inconceivable amounts of information on individuals via emails, smart-phones, sensors, and other technology devices. The 21st century has witnessed a data explosion; many fields have experienced a data deluge that can contribute to boast the economy via data analysis, make new discoveries based on existing data, respond to health problems in a quicker manner, and so forth. While it is worth celebrating the rapid innovations in technology and the presence of huge amounts of data, it is also crucial to consider the number of barriers and risks that come with the increased availability of data; often refers to as big data. One of the barriers that big data faces is privacy. In the healthcare industry, for example, there are protocols to accessing data that can cause financial burdens and can be time-consuming. The cost of collecting, disseminating, and organizing patient information, along with the time it takes to handle the information are some of the challenges. There are also very serious concerns regarding who can have access to what kind of patient information. Policymakers have a very important role in establishing more up-to-date policies and parameters that address the massive amounts of information available and the appropriate ways to collect, share, and house the data. "When considering the risks that big data poses to individual privacy, policymakers should be mindful of its sizable benefits"[7]. While it is important to address the numerous advantages of big data, it remains relevant to figure out ways to prevent data leakage, and to protect the privacy of individuals. This paper showcases the advantages of big data and the ways to overcome the individual privacy concerns.

2 THE ADVANTAGES OF BIG DATA

Big data analysis presents numerous advantages. For instance, it helps businesses to increase their productivity. This has done through a process of analyzing raw data that produces information that identifies trends and patterns that will help businesses make cost effective decisions. It is also helpful in aiding government agencies to improve public sector administration, and assists global organizations in analyzing information that has wide-reaching impact on the world. The information produced by big data can help medical professionals to detect diseases in earlier stages. Some other advantages of big data analysis is present in many different areas, such as: smart grids, which monitor and control electricity use; traffic management systems, which provide information about transportation infrastructure likes roads and highways, mass transit, construction, and traffic congestion; retail by studying customer purchasing behavior to improve store layout and marketing; payment processing by helping to detect fraudulent activity, etc.[7].

Certain research studies have supported the idea that big data allows for real time tracking of diseases and the development, prediction of outbreaks, and facilitates the development of personalized healthcare. Big data can also be used to maximize profits in many disciplines, including healthcare if harnessed properly.[8]. As indicates in [2] "by harnessing big data, businesses gain many advantages, including increased operational efficiency, informed strategic direction, improved customer service, new products, and new customers and markets." While data exists in huge quantities in many fields, including the health care field, individual privacy concerns remain a big problem that policymakers have to tackle to meet current trends in data collection. Improved methods of protecting very personal, private and sensitive health information is needed in order to allow for safe, necessary and adequate access to protected health information within the health care industry. Without proper policies related to data use, access, and protection, this big data potential can not be realized [4]. What are the barriers to big data in heathcare?

3 THE BARRIERS TO BIG DATA IN HEALTH-CARE

One of the barriers faced by big data analysts in healthcare, including mental health services, is privacy. Regardless of the efforts

policymakers try to establish, the different strategies in place to protect individual health information can pose serious challenges that scientists have to wrestle with when it comes to big data analytics. One of the most notable efforts that policymakers have introduced to secure health information, is the creation of the Health Insurance Portability and Accountability Act (HIPAA) in 1996. HIPAA has established norms for data privacy and has mandated security provisions for safeguarding medical and mental health information. Every provider in the healthcare industry must comply with HIPAA privacy laws if they want their practices to remain up and running. The HIPAA laws prohibit providers from sharing patients' information without their consent. The challenge for big data analysts is that a lot of times, patients refuse to share their personal information for research purposes due to fears that the health issue will be the cause of being ostracized, discriminated against, marginalized, etc. "The unintended release of a person's health information into the public realm has huge potential to undermine personal dignity and cause embarrassment and financial harm"[8]. While the healthcare field is faced with a huge increase in health information, individual privacy concern remains a huge conundrum for big data analysis. What can policymakers do to overcome individual privacy concerns, but still allow for the sharing of information that would be for the better good of society at large?

4 WAYS TO OVERCOME PRIVACY CONCERN

4.0.1 Data Anonymization. One way policymakers can protect individual privacy is by making the data anonymous. Researchers have identified three types of data: personal and proprietary data that is controlled by individuals; government-controlled data, which government agencies can restrict access to; and, open data commons, which means that the data is centrally located and available to all. Big data analysts and researchers have advocated for linking data together that can help to improve health care planning at both the patient and population levels. They also argued for an increase in the amount of information that is available in open data commons. Although the anonymization of data appears to be a great technique that policymakers could espouse to address privacy concerns, other studies have indicated that some data can be traced back to their respective individual; thus, destroying the argument for anonymity.[8]. " Every copy of data increases the risk of unintended disclosure. To reduce this risk, data should be anonymized before transfer; upon receipt, the recipient will have no choice but anonymize it at rest…And re-identification is by design, in order to ensure accountability, reconciliation and audit." If proper norms are established for data analysis, this can potentially contribute to improvements in the health care industry.

Still, there are others that have advocated for data de-identification and data minimization. The term de-identification is the process by which the data is made anonymous. The proponents of this process explain that this protective measure is valid under security and accountability principles, but admonish that policymakers should think about other ways to protect patients' privacy. The term data minimization, describes the extent to which organizations can limit the collection of personal data. It is worth noting that data minimization is contrary to big data analysis because data minimization encourages deleting data that is no longer in use in order to protect privacy; whereas, big data analysts would prefer to archive the data for ulterior usage. While this technique can help protect privacy, it is antithetical to big data analysis because it contributes to reducing the amount of data collection that could be used in data analysis to make new discoveries, respond to crises, and maximize profits [7].

As found in [1], privacy principles should be introduced during the process of data architecture; privacy should be incorporated into the design and operational procedures. In so doing, personal health care data will be protected against malicious hackers who try to access individuals' personal health information for the purposes of stealing individuals' identity. Another type of data that has been introduced to the healthcare industry is concept quantified self data. It can be understood as the data produced by individuals that engage in self-tracking of personal health information, such as heart rate, weight, energy levels, sleep quality, cognitive performance, etc. These individuals use devices like smart-phones, watches, and wearable technology sensors in the collection of their personal data and biometrics. It has been shown that 60 percent of U.S. adults are tracking their weight, diet or exercise routines, while 33 percent are monitoring their blood sugar, blood pressure, sleep patterns, etc. This indicates that there is a vast amount of health information that has been produced by individuals. What is done with all of this data? This massive supply demonstrates the need to develop policies and protocols that involve individual patient consent to share their collected data; this data can be critical to the advancement of health-care with the support of data analysis. Before that can be done, however, we must first establish the proper norm to use this type of data so that the privacy of individuals can be protected; this ought to be the primary action to take. [6]. In the healthcare industry, Patients often do not want their health information to fall in the hand of other entities without their consent; however, with proper informed consent, patients seemed to become willing to share their personal health information.As agencies work with patients to disclose the purposes of collecting certain, sometimes sensitive, health information, they can empower patients to make informed decisions about their personal health information, thus engaging patients in the process. This can then serve to increase and improve the set of personal health information utilized for clinical research purposes, and subsequently improve people's lives [5]. "Privacy concerns exist wherever personally identifiable information or other sensitive information is collected and stored in any form"[3]. Thus, to protect privacy, other techniques, like encryption, authentication, and data masking may be utilized to ensure that the information is available only to authorized users.

5 CONCLUSION

We have seen that healthcare data exists in large quantities; however, privacy concerns are one of the biggest barriers and challenges that scientists face when it comes to utilization of healthcare data. Certain researchers have proposed data anonymization as a solution to privacy concerns, while others have proposed a minimization of the amount of data collected on individual patients, as well as authenticate the data so that it can only be accessed by intended users. Suggestion was also made to involve patients in the collection of health data, so that they can be more willing to share their information that can play a vital role in improving healthcare and mental

health research, reduce health care cost, maximize profits, etc. It is almost certain that scientists will always have to wrestle with privacy concern whenever they are dealing with personal health information; thus the importance for policymakers to continue to encourage dialogue among healthcare providers and patients, and develop policies and regulations on how to utilize healthcare data without compromising patients' privacy rights.

REFERENCES

[1] Ann Cavoukian and Jeff Jonas. 2012. *Privacy by design in the age of big data.* Information and Privacy Commissioner of Ontario, Canada.

[2] Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Zakira Inayat, Waleed Kamaleldin Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani. 2014. Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal* 2014 (2014).

[3] Shahidul Islam Khan and Abu Sayed Md Latiful Hoque. 2016. Digital Health Data: A Comprehensive Review of Privacy and Security Risks and Some Recommendations. *Computer Science Journal of Moldova* 24, 2 (2016).

[4] Joachim Roski, George W Bo-Linn, and Timothy A Andrews. 2014. Creating value in health care through big data: opportunities and policy implications. *Health affairs* 33, 7 (2014), 1115–1122.

[5] Robert H Shelton. 2011. Electronic consent channels: preserving patient privacy without handcuffing researchers. *Science translational medicine* 3, 69 (2011), 69cm4–69cm4.

[6] Melanie Swan. 2013. The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data* 1, 2 (2013), 85–99.

[7] Omer Tene and Jules Polonetsky. 2012. Big data for all: Privacy and user control in the age of analytics. *Nw. J. Tech. & Intell. Prop.* 11 (2012), xxvii.

[8] J Van Den Bos, K Rustagi, T Gray, M Halford, E Zeimkiewicz, and J Shreve. 2011. Health affairs: At the intersection of health, health care and policy. *Health Affairs* 30 (2011), 596–603.

6 BIBTEX ISSUES

Warning–empty address in cavoukian2012privacy

Warning–page numbers missing in both pages and numpages fields in khan2014big

Warning–page numbers missing in both pages and numpages fields in khan2016digital

(There were 3 warnings)

7 ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

7.1 Uncaught Bibliography Errors

Missing bibliography file generated by JabRef

Bibtex labels cannot have any spaces, _ or & in it

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

7.2 Formatting

Incorrect number of keywords and HID not included in the keywords

Separate keywords with comma

7.3 Writing Errors

Do not use the phrase *In this paper/report we show* instead use *We show.* It is not important if this is a paper or a report and does not need to be mentioned

7.4 Citation Issues and Plagiarism

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

Uncited Quotes

Need to paraphrase long quotations (whole sentences or longer)

Claims made without citations provided. e.g. in section 4

The citation mark should not be in the beginning of the sentence or paragraph, but in the end, before the period mark. example: ... a library called Message Passing Interface(MPI) [7].

Put a space between the citation mark and the previous word

7.5 Structural Issues

Abstract is unnecesarily long