Joseph Annand

DSE 6003

Assignment 02

Businesses and researchers struggle with balancing data security and data usage. While companies across the globe seek to acquire and utilize high-quality data that can be analyzed to guide important business decisions and yield great financial and social benefit, the legal and ethical requirements set by government agencies to protect private personal data and the increasing threat of malicious attackers trying to steal that data inhibits totally free data access and sharing. Companies that handle people’s personal data must employ a variety of security measures and de-identification techniques while ensuring the data they have remains useful for business and analytical purposes. To understand if a balance between security and usage can be achieved, the benefits and costs of handling personally identifiable information (PII) and, more specifically, personal health information (PHI) must be investigated. Initially, the financial and social benefits of data usage and sharing are discussed, followed by the costs of data security breaches. Finally, some current methods of securing data while retaining utility are discussed to determine if and how a balance can be reached.

Data sharing and accessibility has the power to generate significant revenue for companies and society, and Deming argues that society needs more data sharing because the benefits of publicly available data often outweigh the costs (2021). Public data contributed $3 trillion in 2013 to seven different sectors of the global economy, including between $300 and $450 billion in the healthcare industry (Manyika, et. al., 2013). The Humane Genome project is a prime example of the financial and social benefits of data usage and sharing. A comparison between the publicly shared Human Genome Project and the similar privately developed project showed that work done by the private company led to fewer innovations and commercial products than the data that was made public immediately (Deming, 2021). The private company required licensing agreements and high monetary fees before allowing users to access their data. Meanwhile, one study estimates that the $3.8 billion public investment into the Human Genome Project yielded $796 billion in benefits in a single year, as well as 310,000 new jobs (Deming, 2021). Lack of data can hurt agencies and companies as seen with the distribution of funds for COVID-19 small business relief: a lack of data of small businesses financial well-being led to poor choosing of recipients for aid and high costs for the short program (Deming, 2021). Better data yielding better business is seen across many private companies. Highly data-driven organizations are three times more likely to report significant improvements in decision-making. Data-driven approaches lead to more confident decision-making and more proactive actions. Investing in big data can improve operational efficiency and reduce business costs (Stobierski, 2019). Companies that were more data driven in their decision-making were, on average, 5% more productive and 6% more profitable than their competitors. (McAfee, & Brynjolfsson, 2012). Having a dedicated team to reviewing and ensuring appropriate security measures in place through tools like tools like differential privacy can manage the risk of disclosure while releasing usable, beneficial data (Deming, 2021).

Data breaches at the hands of attackers is a significant risk in data usage and sharing not only to the individuals whose information is disclosed but also to the companies responsible for the data’s security. Healthcare is among the worst affected industries regarding data breaches. Between 2005 and 2019, the number of individuals affected by healthcare data breaches was 249.09 million (Seh, et. al., 2020). Most of these breaches, 157.40 million, occurred in the last five years of that range. The average cost of a data breach in 2019 was $3.92 million, but a breach in the healthcare industry costs more at $6.45 million (Seh, et. al., 2020). The cost of a breached record in the healthcare sector is $429 and increased 19.4% between 2014 and 2019, more than any other industry (Seh, et. al., 2020). The greatest number of breach attacks on health records was initiated by Hacking. Between 2015 and 2019, hacking incidents exposed more than 92% of records, and the number of incidents increased by 73.4% from 2018 to 2019 (Seh, et. al., 2020). The main locations from where confidential healthcare data were breached over the last four years were email and network servers. The number of HIPAA breaches are increasing due to the increasing digitalization of healthcare data and the increasing sophistication of cyberattacks. Healthcare data is more valuable on the black market than any other type of data because it takes longer for healthcare fraud to be discovered and stolen data can be used for longer compared to other instances of fraud that can be stopped once a breach is discovered, like credit card fraud (Murray-Watson, 2023). HIPPA penalties can be hundreds of thousands to even millions of dollars with the average penalty being $479,500 in 2023 (Murray-Watson, 2023).

The growing use of spyware technology has put people’s right to privacy and sensitive data at an increased risk. Pegasus spyware has been used to spy on politicians, human rights activists, doctors, lawyers, academics, and many others. The manufacturer of Pegasus admitted that their customers target between 12,000 and 13,000 individuals every year (United Nations, General Assembly, 2022). The Pegasus software can turn smartphones into 24-hour surveillance devices through which an intruder can access all messages, geolocation data, and application information on the phone, meaning financial and health data could be at risk (United Nations, General Assembly, 2022). PHI can be used to create fake insurance claims, allowing for the purchase and resale of medical equipment or illegal access to prescriptions for use or resale (Center for Internet Security, n.d.). Real cases of breached medical data show the impact on the patients whose data was compromised. Victims report fear, embarrassment, and emotional stress after the breach of their PHI (Wairimu, & Fritsch, 2022). More severe effects on victims include anxiety, trauma leading to psychological effects and medication, blackmail, stigmatization, and stalking (Wairimu, & Fritsch, 2022). One study demonstrated that an exposure of the health data of client with a heart pacemaker could have resulted in an attacker hacking the device and cause the death of the patient (Wairimu, & Fritsch, 2022). The Human Rights Council has called upon countries around the world to encourage businesses to implement solutions to protect communications and transactions of confidential information, including measures of encryption, pseudonymization, and anonymization (2022).

Re-identification of PHI can result in financial or personal impacts for individuals, including denial of insurance, lack of job offers, stigmatization, and damage to personal reputation. Stricter privacy regulations imposed by HIPAA may lead to fewer research studies or less robust studies (Lane, & Schur, 2010). Balancing utility and usage can start with removing delays in data release that reduce data utility while not improving data protection (Lane, & Schur, 2010). Additionally, Lane & Schur recommend the use of remote access data enclaves to facilitate high-quality data usage while imposing appropriate data security measures (2010). According to Nikoloski, protecting privacy starts with establishing accountability by complying with data privacy regulations and being transparent about how data is used (2023). Privacy risks and potential areas of intrusion or attack should be considered. Echoing the suggestion of the Human Rights Council, Nikoloski emphasizes the importance of minimizing data collection and retention and implementing anonymization and pseudonymization techniques can significantly reduce the risk of privacy loss when analyzing or distributing data. Finally, technical or organizational measures, such as encryption or access control, can reduce the risk of data breaches.

Increase concern of personal data protection has increased pressure on business to prioritize data privacy. Balancing personalization marketing and data protection can be achieved through a “privacy-by-design” approach, in which privacy issues are considered from the start of the project rather than something that is squeezed in at the end of development (Theriot, 2023). This could include policies for limited data retention or secure data handling through robust security measures. Another privacy-by-design strategy is data minimization, or collecting only the data that is necessary for a specific purpose. Theriot proposes to protect personal data using pseudonymization, as many others recommend, or leverage nonpersonal data in creative ways to minimize the risk of privacy breaches (2023). Theriot suggest finding workarounds to using personal data in business activities, like tailoring ads to the content of a specific webpage rather than the individual while marketing or advertising. Possibly the best and most unique suggestion from Theriot is for companies to encourage individuals to share their data directly with your company to improve transparency (2023).

While numerous recommendations on preserving data privacy have been discussed, a determination of how well these techniques impact data utility is missing. Hotz, et. al. argues the development of new privacy protection strategies and technologies has not sufficiently considered the impact on data usability and its detriment to society (2022). Criticizing common security methods, Hotz, et. al. calls out the ability of synthetic datasets to actually reduce the risks of data disclosure into question, while also accusing differential privacy of not adequately evaluate the impact on data utility nor the absolute disclosure risk (focuses on the relative disclosure risk) (2022). Hotz, et. al. recommends the use of a disclosure risk criterion that evaluates the risk individuals are most concerned about as well as an exploration into new methods of privacy protection and disclosure risk criterion that evaluates the balance between usability and privacy protection (2022). Current methods may not be adequate for finding that balance; however, further research on alternative methods, like a cost-benefit analysis used in economics, could support the goals of companies using Big Data (Hotz, et. al., 2022).

No simple solution exists for securing personal data that companies and researchers desire to get their hands on for cutting costs, making smarter investments, or contributing to the betterment of society. Studies have shown that access to high-quality data can benefit companies by saving money, increasing productivity, and raising profits. Lack of high-quality data can lead to poor decision making and slow down businesses. Unfortunately, the costs of data breaches weigh heavily on the minds of organizational infrastructure managers, and it should worry functional managers, as well. Failure to have proper security measures for PHI can result in severe fines from the government, only one of which can cost a company tens of thousands of dollars. If individuals are more at risk of their data being stolen due to the advancement of hacking techniques and spyware, companies are more at risk of data breaches and the monetary consequences of those breaches. Government fines are not the only consequence of data breaches, though. When personal information is disclosed or stolen, real people suffer the negative effects of their private financial or medical information being revealed to the public. This could open the company up to further legal troubles through possible civil lawsuits from affected individuals and damage the reputation of the company.

After all this, is there a balance to be found between data security and usage? There might be; however, companies are limited to the current methods of determining risk and utility. There are several commonly used and accepted methods for de-identification, such as anonymization and pseudonymization. Organizational Infrastructure managers should be careful that the pseudonymization method used meets the HIPAA requirements. Appropriate encryption techniques should be employed for the storage and transportation of PHI and other personal data. Differential privacy ca be a valuable technique in evaluating the risk of privacy loss; however, as mentioned above, it does not measure data utility after de-identification methods are performed on the data. While the data community makes current methods more robust or develop new methods for evaluating privacy loss and data utility, the scale between data security and data usage should rest on the side of security. Companies should implement practices accepted and recommended by HIPAA and other data security standards organizations. It is better to be safe when handling PII to protect the company’s assets and reputation.

References

Center for Internet Security. (n.d.). *Data Breaches: In the Healthcare Sector*. CIS Center for Internet Security. Retrieved October 8, 2023, from <https://www.cisecurity.org/insights/blog/data-breaches-in-the-healthcare-sector>

Deming, D. (2021, February 19). Balancing Privacy With Data Sharing for the Public Good (Published 2021). *The New York Times*. <https://www.nytimes.com/2021/02/19/business/privacy-open-data-public.html>

Hotz, V. J., Bollinger, C. R., Komarova, T., & Spencer, B. D. (2022, July 25). Balancing data privacy and usability in the federal statistic system. *PNAS*, *119*(31). <https://doi.org/10.1073/pnas.2104906119>

Lane, J., & Schur, C. (2010). Balancing access to health data and privacy: a review of the issues and approaches for the future. *Health services research*, *45*(5 Pt 2), 1456–1467. <https://doi.org/10.1111/j.1475-6773.2010.01141.x>

Manyika, J., Chui, M., Farrell, D., Van Kuiken, S., Groves, P., & Almashi Doshi, E. (2013, October 1). *Open data: Unlocking innovation and performance with liquid information*. McKinsey. Retrieved October 8, 2023, from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

McAfee, A., & Brynnjolfsson, E. (2012, October). *Big Data: The Management Revolution*. Harvard Business Review. Retrieved October 8, 2023, from https://hbr.org/2012/10/big-data-the-management-revolution

Murray-Watson, R. (2023). *Healthcare Data Breach Statistics*. The HIPAA Journal. Retrieved October 8, 2023, from https://www.hipaajournal.com/healthcare-data-breach-statistics/

Nikoloski, F. (2023, July 27). *Data Privacy and Analytics: Finding the Right Balance*. Cookie Law Info. Retrieved October 8, 2023, from <https://www.cookielawinfo.com/data-privacy-and-analytics/>

Seh, A. H., Zarour, M., Alenezi, M., Sarkar, A. K., Agrawal, A., Kumar, R., & Khan, R. A. (2020). Healthcare Data Breaches: Insights and Implications. *Healthcare (Basel, Switzerland)*, *8*(2), 133. <https://doi.org/10.3390/healthcare8020133>

Stobierski, T. (2019, August 26). *The Advantages of Data-Driven Decision-Making | HBS Online*. HBS Online. Retrieved October 8, 2023, from <https://online.hbs.edu/blog/post/data-driven-decision-making>

Theriot, K. (2019, March 9). *Marketing In The Age Of Privacy: Balancing Personalization And Data Protection*. Forbes. Retrieved October 8, 2023, from <https://www.forbes.com/sites/forbesagencycouncil/2023/08/09/marketing-in-the-age-of-privacy-balancing-personalization-and-data-protection/?sh=9676b194d090>

United Nations, General Assembly. (2022, August 4). *A/HRC/51/17: The right to privacy in the digital age*. ohchr. Retrieved October 8, 2023, from <https://www.ohchr.org/en/documents/thematic-reports/ahrc5117-right-privacy-digital-age>

Wairimu, S., & Fritsch, L. (2022). Modelling privacy harms of compromised personal medical data - beyond data breach. *In Proceedings of the 17th International Conference on Availability, Reliability and Security (ARES '22)*. Association for Computing Machinery, New York, NY, USA, Article 133, 1–9. <https://doi.org/10.1145/3538969.3544462>