2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192830

A Strategic Data Protection Plan for the Healthcare Industry- A Review

1st Aritra Mitra
Symbiosis Institute of Digital and Telecom Management,
Symbiosis International (Deemed University)
Pune, India
aritra.mitra2022@sidtm.edu.in

2nd Saikat Gochhait
Symbiosis Institute of Digital and Telecom Management,
Symbiosis International (Deemed University), Pune , India /
Neurosciences Research Institute, Samara State Medical University, Russia
saikat.gochhait@sidtm.edu.in

3rd Ahmed J.Obaid
Faculty of Computer Science and Mathematics,
University of Kufa,
Baghdad, Iraq
ahmedj.alijanaby@uokufa.edu.iq

4th Mohammed Ayad Alkhafaji
National University of Science and Tech, Iraq
mohammed.alkhafaji@nust.edu.iq

***Abstract*—Since revolutionary digitization has taken hold in all industries and companies, the excessive growth of data is overtaking the world around us. With this explosion of data comes an increased responsibility to protect it from external threats, exploitation and misuse of information. The healthcare industry is expanding its horizons with the latest cutting-edge technologies such as robotic process automation, cloud transformation and digitization, generating several zettabytes of data every year. With this excessive data growth, the responsibility to protect the data from external threats, exploitation and information misuse is also increasing. The steep rise in data breaches, disclosure of important public and corporate data, fraudulent activities such as threatening phone calls, false insurance claims, and even illegal monetary claims have rocked the world. This in turn increases the urgency and need for an advanced, standardized data protection strategy. In this research study, the Scopus database has been used as a source for a bibliometric analysis to discuss recent research activities on big data protection. The expected outcome of this research is a broader understanding of how organizations operating in the healthcare sector are addressing overall data management by shaping existing organizational policies and adapting new security standards.**

Keywords: Robotic process automation, Cloud transformation, Digitization, Data governance, Healthcare

I. Introduction

There has been a paradigm shift in healthcare due to recent developments in the digitalization of healthcare, which has enabled the transition to an electronic mode and workflow for storing patient records. This will greatly increase the amount of electronic data available, especially in the medical field. At the same time, difficulty, variety, and speed are increasing sharply, resulting in trillions of pieces of data. These hold the promise of supporting a variety of exceptional events, opportunities, and use cases, such as the following leading examples: driven by authoritative criteria, as well as the ability to improve treatment, save lives, save money, and support clinical decision making. For diseases that affect multiple organ systems, health insurance, disease surveillance, public health monitoring, adverse event tracking, and treatment refinement are important. State-of-the-art innovations are being introduced in the healthcare industry [1]. Healthcare must simplify the problem that their security teams face in order to protect themselves against cyber threats. It is imperative to achieve this goal by integrating security functionality and centralizing the monitoring and management of security systems [34]. Thus, physician mobility and wireless networking, patient data collection, and cloud computing are all signs of current developments in healthcare, and concerns about information security of sensitive data and privacy are growing every year [2]. To ensure that data is consistent, trustworthy, and not misused, the implementation of effective data governance is essential. The importance of data analytics to optimize operations and support business decisions is increasing, while organizations must contend with new privacy regulations. Every healthcare organization needs to adopt a constructive, preventive strategy and measures with respect to future protection and privacy requirements to avoid the loss of confidential information and other security-related incidents [3].

Researchers around the world often encounter a particular conflict, namely the belief that privacy is more appropriate for developed countries. Developing countries do not need aggressive data protection. There are many technology-based data protection solutions such as data loss prevention (DLP), storage with built-in data protection, firewalls, encryption and endpoint protection to ensure data security. According to a 2020 survey by the United Nations Conference on Trade and Development (UNCTAD), 81 percent of countries worldwide have electronic transaction legislation in place [4]. The European Union has the highest percentage (98%), followed by North and South America (91%). Africa has the lowest percentage (61%). Although 79% of countries have laws in place to combat cybercrime, the percentage varies greatly by region, with Europe having the highest percentage (89%) and Africa the lowest (47%). The global percentage of online consumer protection is 56%. However, adoption rates vary widely, ranging from 73% in Europe to 72% in the Americas and 46% in Africa. Sixty-six percent of countries have data and privacy protection laws. The figure is 96% in Europe, 69% in the United States, 57% in the Americas, 50% in Asia and the Pacific, and 50% in Africa.

Among many other challenges, governments in all countries have reflected on concerns about data loss, particularly with respect to the health care system. Several countries have tightened their privacy policies, while many have adapted for the first time. Policies such as HIPPA (Patient Safety and Quality Improvement Act), (General Data Protection Regulations (GDPR), Data Protection Directive, Personal Information Protection and Electronic Documents Act (PIPEDA), Data Protection Act (DPA) have been widely adopted in countries such as the US, Europe, Canada, UK and many others.

India is not far behind in this field [6]. A national health data repository by the Indian government based on ID along with an underlying vast IT network to achieve a centralized approach to data access and mitigate security threats is the need of the hour.

Conceptual Background and Methodology

Data protection is a top priority for almost all industries. An industry like healthcare must develop a concrete strategy to protect patient data. Privacy mechanisms are at the forefront of every healthcare department. Scopus is the most widely used source of bibliometric data [33]. Retention of electronic patient records, personal data, and financial data is a top priority for medical institutions [7].

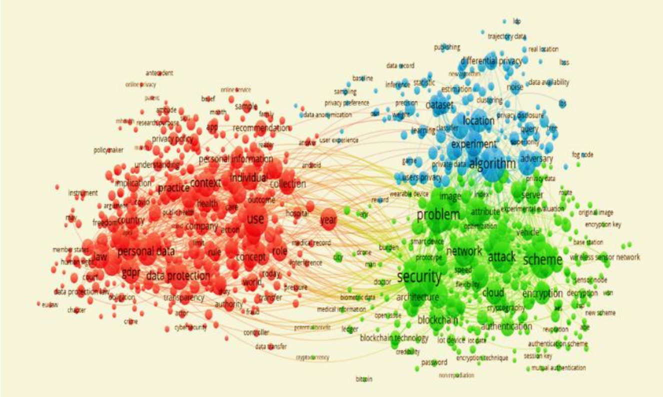

Fig 1: VOS viewer plot for co-occurrence

Co-occurrence matrix social network map is shown in figure 1. Using this approach, it is possible to intuitively identify the relationship between the data protection strategy and healthcare. This study helped me to understand the research outcome and the potential research gap in this very field.

Table 1: SUMMARY OF FINDINGS FROM VOS VIEWER

Topic Mining i.e., Co-occurrence Clustering	We have received 3 topic clusters for 1981-2021. Few topics with very high link strength and co-occurrence value are listed below. 1. Personal Data in Healthcare 2. Data Protection Strategy 3. Cloud Computing for Electronic Data 4. Security of Patient Data 5. Accuracy of Information & Criticality 6. Algorithm protecting healthcare information	Pluralistic topics; some topics weakened, or deviated, or merged to a slice of other topics, some themes were evolving

II. Literature Review

A. Importance of Personal Data in Healthcare

Personal data is information about an individual natural person that allows or could allow that person to be identified when performing certain functions. It's necessary to distinguish between personally identifiable data (even if encrypted) and completely anonymized data, as the Regulation applies only to the former. The data must be collected and used by another person (a natural or legal person) to be protected by the Regulation, regardless of whether it's personal or professional data.

The constitutional freedom of patients to have their medical information protected poses a significant problem in several respects, including medical treatment, including eHealth and cross-border medical services, and in science (clinical trials, clinical investigations, epidemiological tests, and patient registries are all examples of clinical trials. etc.). Clinical trials such as clinical research, clinical investigations, epidemiological tests, and patient registries are considered "important data" under EU law and are subject to additional safeguards [9]. Unauthorized disclosure of confidential health information can adversely affect patients' personal and professional lives.

B. Security of Patient Healthcare Records

Healthcare facilities are collecting, storing, and sharing more patient data every year as a result of the proliferation of advanced medical devices, mobile technologies for clinicians, and the evolving Internet of Things across industries. An abundance of data is rewarding for hackers. And gradually, their attack tactics are becoming more sophisticated and refined [16]. A data breach can be very costly. Patients must be notified and the breach must be reported to the authorities, which negatively impacts the organization's image and can result in large fines [17].

Companies must now bear the burden of cybersecurity programs. Every penny and time spent on data security must be sustained by the economics of such a facility [18]. Hospitals must implement strategies that are both reliable and profitable to protect patient data without breaking IT budgets. Implementing effective security technologies and devices requires a multi-layered approach that considers both regional and remote patient environments to keep pace with evolving information technology systems and cyber threats [19].

C. Data Accuracy of Computerized Patient Records

Meanwhile, data integrity continues to be a concern for healthcare providers. Data integrity ensures that data are accurate and have not been manipulated in any way. Inaccurate data poses significant health risks to patients and significant liability to providers, leading to problems such as fraud, harassment, data loss, and incorrect or incomplete treatment. In the healthcare arena, electronic health records have tremendous potential. When used properly, electronic health records help patients receive better care from their physicians.

After considering all the potential sources of error, the question remains: how will the industry proceed to solve this problem? What can be done to address these issues while maintaining data integrity? When recommendations are proposed and developed, there are a variety of tools and services that providers can use to ensure that their data is reliable and represents them well. Symphony's mission is to provide healthcare organizations with leading technologies and resources to deliver the best care possible by ensuring the quality of clinical and administrative data. Managing electronic health records (EHRs) can be difficult, especially when there are so many potential sources of error [23]. With the right tools, these systems can be transformed into the assets they were designed to be, helping individuals achieve their goals and get the results you want.

D. Usage of Algorithms in Healthcare Software & Procuring Data

Algorithmic decision making based on Big Data and artificial intelligence (AI) is becoming more popular in healthcare than in other industries. The hope is for increased data collection and access, as well as algorithms for purposes as diverse as predictive policing, government efficiency assessment, scientific analysis, news reporting, and healthcare. To process daily data, modern healthcare organizations use a variety of tools. In the context of health, any information that pertains to a patient's health or the health of a group of people is considered health data. Health care professionals, insurance companies, and government organizations process this information using health information systems (HIS) and other technological tools. There is IP protection for this software. The machine design, applications, processes, and data have been disclosed. A software algorithm can be so complex that even experienced software gurus struggle to understand it, and it's almost impossible to be sure that it'll work correctly. Algorithms generated by the software and created by machine learning, genetic programming or other AI.

E. *Differential Privacy in Healthcare*

Differential privacy is a concept of data protection developed to ensure that statistical analysis does not compromise privacy. It ensures that an individual's data have a minimal impact on the overall outcome of the model. In other words, whether or not a particular person's data is included in the dataset, the result of the algorithm is essentially the same. Differential privacy is often implemented by adding statistical noise to the input (local differential privacy) or output (global differential privacy) of the model or statistical query. The contributions of individual users remain hidden, but thanks to the noise, insights into the entire population are gained without compromising privacy [30].

F. *General Data Protection Regulations (GDPR) Protecting Public Data*

The General Data Protection Regulation introduces a new data protection requirement. It affects any company that collects, processes or stores personal data of EU citizens, regardless of where that data is acquired, processed or stored. This gives the rule extraordinary reach, extending to countries outside the EU and affecting organizations in every industry around the world. For the healthcare industry, which requires a great deal of personal data? It's an opportunity to strengthen systems, rules, and practices to stay ahead of any potential threat to facility and patient data [26]. The GDPR establishes new rules for the collection, processing, and storage of personal data. Penalties and fines are among the tools used to enforce the new rules under the General Data Protection Regulation. Healthcare organizations are in a unique position because healthcare organizations process a wide variety of data, including financial records and health insurance information, as well as patient test results and biometric data. Some of these types of data are more sensitive than the data regularly collected by non-healthcare organizations. These data are inextricably linked to an individual and are, for the most part, immutable. For example, an individual may be able to change his or her email address, but not his or her medical history or dental records, which poses a major privacy risk if such information is stolen. The healthcare industry faces particular obstacles, but there are also excellent security solutions that can benefit an organization in the long run [30].

IV DISCUSSION, CONTRIBUTION, IMPLICATIONS & FUTURE DIRECTIONS

This study sought to analyze how healthcare organizations identify and address the potential gap for protecting their data and strive for smooth operations while maintaining the highest level of data security and information privacy. It also focuses on the privacy dimension of health records, developing strategic solutions, and assessing vulnerabilities while complying with specific policies and regulations. With the advent of hyper-automation, digital transformation, and the adoption of the digital mode of data storage and transmission, the attack surface for data breaches has become increasingly large. Sensitive personal data such as sexual orientation, diagnoses, social security numbers, credit card information, etc. is extremely important to protect from external and internal fraudsters. Providing accurate real-time health data is enabled by the digital transformation of healthcare [31]. Therefore, existing public health information policies need to be strengthened over time. In addition, regular monitoring of these policies is needed to ensure that the privacy strategy is effective enough to achieve its goal in the first place.

Technological advancement is the most interesting and promising area of the privacy strategy. With the advances and synergies of technologies such as hybrid cloud computing, robotic process automation, remote servers and telemedicine, Big Data, and blockchain, the scope for this topic is expanding [32].

II. CONCLUSION

Bibliometric analysis presents a unique and fascinating topic for understanding privacy ideas and their applications in healthcare. Bibliometric analysis has been used to discover new insights in healthcare at the intersection of regulations, policies, frameworks, and technology and their applications in the areas of privacy, data security, confidentiality and availability of health data, innovation, and sustainability. This research provides a unique perspective on international privacy policy and the technology and security standards necessary to improve information security, transparency, cybersecurity, individual security, and the acquisition and retention of highly sensitive personal data. Focusing on scholarly articles, this bibliometric study identifies notable authors and academics who are implementing information security, digital storage, Big Data, and cloud security strategies. The research also found that it is a growing problem that will lead to more research around the world in the future. It is hoped that this research agenda will motivate researchers to continue research on data protection strategies for healthcare organizations.

REFERENCES

[1] J. L. Fernández-Alemán, Señor et al., “Security and privacy in electronic health records: A systematic literature review,” J. Biomed. Inform., vol. 46, no. 3, pp. 541-562, 2015. doi:10.1016/j.jbi.2012.12.003.

[2] P. S. Rao and S. Satyanarayana, “Privacy preserving data publishing based on sensitivity in context of Big Data using Hive,” J. Big Data, vol. 5, no. 1, 2018. doi:10.1186/s40537-018-0130-y.

[3] M. Shrivastava et al., "A Review on Digital Twin Technology in Healthcare," 2023 International Conference

on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India, 2023, pp. 741-745, doi: 10.1109/ICIDCA56705.2023.10099646.

[4] B. Jacobs and J. Popma, “Medical research, Big Data and the need for privacy by design”, Big Data & Society, vol. 6, no. 1, 2019. doi:10.1177/2053951718824352.

[5] Y. Mcdermott, “Conceptualising the right to data protection in an era of Big Data”, Big Data & Society, vol. 4, no. 1, 2017. doi:10.1177/2053951716686994
.

[6] D. V. Ford et al., “health research and evaluation,” October, 2009. doi:10.1186/1472- 6963-9-157.

[7] A. Appari and M. E. Johnson, “Information security and privacy in healthcare: Current state of research,” Int. J. Internet Enterpr. Manag., vol. 6, no. 4, p. 279, 2010. doi:10.1504/IJIEM.2010.035624.

[8] I. Keshta and A. Odeh, “Security and privacy of electronic health records: Concerns and challenges,” Egypt. Inform. J., vol. 22, no. 2, pp. 177-183, 2021. doi:10.1016/j.eij.2020.07.003.

[9] R. Ducato, “Data protection, scientific research, and the role of information. Computer law & security review,” The Int. J. Technol. Law Pract., vol. 37, p. 105412, 2020. doi:10.1016/j.clsr.2020.105412.

[10] S. Dash et al., “Big data in healthcare: Management, analysis and future prospects,” J. Big Data, vol. 6, no. 1, 2019. doi:10.1186/s40537-019-0217-0.

[11] B. C. M. Fung et al., “Privacy- preserving data publishing: A survey of recent developments,” ACM Comput. Surv., vol. 42, no. 4, 1-53, 2010. doi:10.1145/1749603.1749605.

[12] J. Han et al., “Research on electronic document management system based on cloud computing,”, Computers, Materials and Continua, vol. 66, no. 3, 2645-2654, 2021. doi:10.32604/cmc.2021.014371.

[13] M. Ahmadi and N. Aslani, “Capabilities and advantages of cloud computing in the implementation of electronic health record,”, Acta Inform. Med., vol. 26, no. 1, pp. 24-28, 2018. doi:10.5455/aim.2018.26.24-28.

[14] D. J. Nigrin, “When ‘hacktivists’ target your hospital,” N. Engl. J. Med., vol. 371, no. 5, pp. 393-395, 2014. doi:10.1056/NEJMp1407326.

[15] P. T. Jaeger et al., “Cloud Computing and Information Policy: Computing in a Policy Cloud?”, Journal of Information Technology & Politics, vol. 5, no. 3, 269-283, 2008. doi:10.1080/19331680802425479.

[16] A. H. Seh et al., “Healthcare data breaches: Insights and implications,”, Healthcare (Basel), vol. 8, no. 2, 2020. doi:10.3390/healthcare8020133.

[17] C. S. Kruse et al., “Cybersecurity in healthcare: A systematic review of modern threats and trends,” Technol. Health Care, vol. 25, no. 1, pp. 1-10, 2017. doi:10.3233/THC-161263.

[18] B. L. Filkins et al., “Privacy and security in the era of digital health: What should translational researchers know and do about it?,” Am. J. Transl. Res., vol. 8, no. 3, pp. 1560-1580, 2016.

[19] P. Mehndiratta et al., A Model of Privacy and Security for Electronic, Mar., 2014. doi:10.1007/978-3-319-05693-7.

[20] K. Abouelmehdi et al., “Big healthcare data: Preserving security and privacy,” J. Big Data, vol. 5, no. 1, pp. 1-18, 2018. doi:10.1186/s40537-017-0110-7.

[21] N. Johnson et al., “PRACTICE OBSERVED surveillance: Comparative study of influenza data,” vol. 302, no. Mar., pp. 763-765, 1991.

[22] M. A. Habib et al., “Privacy-based medical data protection against internal security threats in heterogeneous Internet of Medical Things,”, International Journal of Distributed Sensor Networks, vol. 15, no. 9, 2019. doi:10.1177/1550147719875653.

[23] W. R. Hogan and M. M. Wagner, “Accuracy of data in computer-based patient records,” J. Am. Med. Inform. Assoc., vol. 4, no. 5, pp. 342-355, 1997. doi:10.1136/jamia.1997.0040342.

[24] G. Rong et al., “Artificial intelligence in healthcare: Review and prediction case studies,” Engineering, vol. 6, no. 3, pp. 291-301, 2020. doi:10.1016/j.eng.2019.08.015.

[25] J. Tohka and M. Van Gils, “Evaluation of machine learning algorithms for health and wellness applications: A tutorial,” Comput. Biol. Med., vol. 132, no. Mar., p. 104324, 2021. doi:10.1016/j.compbiomed.2021.104324.

[26] F. Fatehi et al., “General Data protection Regulation (GDPR) in healthcare: Hot topics and research fronts,”, Stud. Health Technol. Inform., Jul., 1118-1122, 2020. doi:10.3233/SHTI200336.

[27] J. Starkbaum and U. Felt, “Negotiating the reuse of health- data: Research, Big Data, and the European General Data Protection Regulation”, Big Data & Society, vol. 6, no. 2, 2019. doi:10.1177/2053951719862594.

[28] X. Larrucea et al., “Towards a GDPR compliant way to secure European cross border Healthcare Industry 4.0,” Comput. Stand. Interfaces, vol. 69, 2020. doi:10.1016/j.csi.2019.103408.

[29] A. Rawat and S. Gochhait, "Iot Enabled Mental Health Diagnostic System Leveraging Cognitive Behavioural Science," 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 2022, pp. 1401-1405, doi: 10.1109/DASA54658.2022.9765032.

[30] F. K. Dankar and K. El Emam, “Practicing differential privacy in health care: A review,” Trans. Data Privacy, vol. 6, no. 1, pp. 35-67, 2013.

[31] S. Gochhait et al., “Implementation of EHR using Digital Transformation: A study on Telemedicine,” International Conference for Emerging Technology (INCET),

vol. 2020, 2020, pp. 1-4. doi:10.1109/INCET49848.2020.9154146.

[32] M. Pathapati and S. Gochhait, “Intelligent data management to facilitate decision-making in healthcare,” International Conference on Decision Aid Sciences and Applications (DASA), vol. 2022, 2022, pp. 1-5. doi:10.1109/DASA54658.2022.9765260.

[33] S. Gochhait and A. Srivastava, (2023). “Security Threats in Healthcare Systems—A Bibliometric Study”.In: Gunjan, V.K., Zurada, J.M. (eds) Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications. Lecture Notes in Networks and Systems, vol 540, 2023 Springer, Singapore. https://doi.org/10.1007/978-981-19-6088-8_2

[34] K. Bajaj et al., “Risks and Regulation of Cryptocurrency during Pandemic: A Systematic Literature Review,” WSEAS Transactions on Environment and Development, Vol 18, 2022.