Exploring the Impact of Big Data in Healthcare and Techniques in Preserving Patients' Privacy

Justice Asare-Frempong1[†] and Manoj Jayabalan2[††],

Asia Pacific University of Technology & Innovation, Kuala Lumpur, Malaysia

Summary

The era of big data brings several benefits and opportunities to the healthcare industry by way of providing timely patient care services, proactive disease detection, real-time monitoring among others. The adoption of big data in healthcare with its accompanying plethora of advantages has been a revelation, nonetheless, activities of data poachers and other adversaries has left the downside of big data much to be desired. The major challenge faced by healthcare providers on safeguarding the privacy and security of patients' data. This study explores technologies in preserving patients' privacy in the healthcare industry with emphasis on data stored in Electronic Health Records (EHRs). The techniques explored are anonymization techniques, cryptographic, and data management framework. In as much as the formerly mentioned techniques (anonymization and cryptography) possess some enviable merits, studies have proven ambiguities in their usage. The utilization of data management framework largely absorbs the deficiencies inherent in the former techniques, hence provides a more reliable mechanism in preserving privacy and security.

Keywords:

Anonymization, Big Data, Cryptography, Data Management, Electronic Health Record, Privacy, Security

1. Introduction

The need to digitize patients' records has brought in its wake a pressing demand for new techniques, methodologies as well as technologies beyond the existing traditional methods of handling and managing data [1]. With the recent accumulation of data in various and diverse forms spanning from medical databases comprising medical checkups, diagnosis, test reports, x-ray as well sensor data and other administrative records. In the healthcare industry, big data emerges as the new paradigm with the capability of handling the dynamic nature of these sources of data [2]. As reported that in 2012, data amassed in the healthcare industry reached the domain of 500 petabytes by estimate. This alarming growth in data notwithstanding, experts still predict over a 50% increase in this rate i.e. gearing towards 25 Exabyte's in the year 2020 [3]. The transformation of this massive quantity of data into meaningful knowledge, which is crucial for determining patient needs as well as augmenting the accuracy and rapidity of medical practitioners cannot be over emphasized. These dimensions provide the stimulus to harness the power of big data. According to Mckinsey Global Institute, an estimated annual increment in a profit of $100 billion could be made in the healthcare industry provided the dynamics of big data are fully utilized [4]. The new wave for real-time health analytics, diagnostic analytics as well as predictive analytics based on the patient-centric model has started to take stand in healthcare information technology with EHR being the foundation of this novelty [4]. As patients stand out as the fulcrum around which EHRs revolve, there comes with it crucial questions regarding the privacy and security of patients' record. This paper discusses the benefits and challenges of big data on patients' records in the healthcare industry coupled with techniques in preserving patients' data. This research discusses three major techniques in privacy preservation which are anonymization, cryptography, and data management framework. Mechanisms of information security comprising such methods as authentication, access control, log analysis among others are left beyond the scope of this research.

The rest of this paper is organized as Big Data in Healthcare comprising data sources, benefits of big data, challenges of big data in Section 2, Privacy Protection Techniques which includes anonymity, cryptographic techniques, and data management framework presented in Section 3. The key issues in the research are captured under conclusion.

2. Big Data in Healthcare

2.1 Data Sources

Technological advances in the use of devices have brought with its innovations, numerous avenues in generating and collecting data in the healthcare industry. Most data repositories in healthcare like other industries and organizations integrate data from heterogeneous sources such as EHR, social media, medical devices among others.

A) Electronic Health Records

Manuscript received August 5, 2017
Manuscript revised August 20, 2017

In a more organized dimension, data in a healthcare facility is housed in an EHRs with the core reason of providing a multifaceted outlook of patients' records [5]. The EHR is a repository of electronic healthcare data that primarily concerns patients, medical records, and healthcare administrative records. A classic EHR structure comprises and integrates sub-schemes like admission, discharge and transfer (ADT) of patients' schema, repetitive engagements, and planning schema, the procedure of entering prescription schema and notes on routine medical checks among others [6].

B) Social Media

Social media also presents another significant source of data in the health care industry. Data is collected on patient behavior and sentiment data on the recovery reactions of patients. In addition, social media posts, comprising Twitter feeds, blogs, status updates on Facebook as well as other platforms, and web pages can reveal and provide an indication of a person's health, mood, and state of mind that is of key relevance to health professionals.

C) Medical Equipment's

Medical devices and sensors such as pulse oximeters, glucose monitors, blood pressure monitors etc. generate enormous amounts of data that can provide some valuable insights about patients' health conditions [7].

The eruption of the Internet of Things (IoT) coupled with its capacity to give rapid access to medical needs is one of the driving elements for embracing big data in healthcare [8]. What is more, the presentation of Body Sensor Networks (BSN) and its uninterrupted application to healthcare, will equip healthcare providers with the capacity to screen indispensable parameters, and accurately anticipate impending medical dangers such as epidemics and pandemics by extension [4]. Body sensors create huge information, and connecting such medical data from varied sources is key for driving medical analytics.

2.2 Benefits of Big Data

With the incorporation of technology at the heart of its operation, big data brings many desired and productive benefits into the healthcare industry. For instance, IoT as an ally of big data enables medical staff and individuals to monitor heart rate, weight, blood pressure, varying levels of stress among others on patients. Recently, the convenience of mobile apps on smartphones facilitate the monitoring of users' exercise regimen and intensity as well as the duration and appropriateness of sleep. Examples of these apps are Pebble time, AliveCor heart Monitor, and MyfitnessPal [4]. Another merit of big data in the domain of healthcare is that it facilitates the ability to use the genetic blueprint to effectively detect beforehand, diseases cropping from patients' lifestyles [9]. A further probe into this ensures the determination of treatments and medications for patients on a personal basis. To a large extent, this proactive step helps to ensure the reduction of unforeseen costs of treatment and the eventual elimination of the possibility of chronic diseases with an effective usage of big data.

The 'omics theory' stemming from genomics, proteomics, and metabolomics, big data has again facilitated scientific studies into Epigenomics, Transcriptomics, Pharmacogenomics, Immunogenomics which have elevated knowledge into human physiology [10].

Individualized medication, otherwise considered 'precision medicine' endeavors to extract insights obtained from the omics theory and key contemporary data analytics techniques like Artificial Neural Networks to facilitate a satisfactory degree of rendering medical care.

In the same vein, data available in Healthcare Administrative Records/Repositories (HAR) is leveraged to identify and monitor patients' health progress regarding acute cases like type 2 diabetes and further helps to undertake health trend analysis [6].

Text mining as a component of data mining, as well as clustering, are some of the big data statistical techniques applied in ensuring quality health trend analytics. When one considers the perspective of EHR data as the interest of this study, big data has contributed immensely to leveraging patients' data. Drug potency, otherwise known as drug efficacy has been well boosted if not made possible in the healthcare industry through the adoption of big data. In a study conducted into the cost effectiveness of using Random Control Trials (RCTs) and EHR to determine the results of cardiovascular examinations by the University of Pennsylvania, School of Medicine, it was revealed that RCTs required higher costs relative to EHRs [11]. In addendum, healthcare suppliers employ EHRs in tracking their patients more acutely and updating insurers on related charges. These records accumulate a patient's private information, health and family history, genomic structures, vaccination and treatment information, payer information, and other useful collections concerning the patient [10].

2.3 Challenges in Big Data

On the hind side of the technological advancements and benefits following the adoption of big data in the healthcare industry lies a number of risk factors and serious challenges. Studies by different scholars have posited different views regarding the challenges of big data in the healthcare. For instance, Patel & Patel (2016) revealed aggregation as one of the menaces of big data in the healthcare industry.

More prominently was a high ranking conference in the United States which comprised key stakeholders in healthcare and other related domains. Issues regarding the challenges posed by big data particularly in health care revealed among other matters, the lack of standard terminology and ontology of warehoused information, a lack of access to required technologies and computational architecture, the vulnerability of data storage houses and cyber architectures [12] [8]. Outstanding among the factors, however, was the issue of patients' privacy and security connected to EHRs.

To vindicate the consensus of the US stakeholders conference held, Kasier Permanente, one of the top-ranking non-profit making healthcare suppliers in the US in the year 2013 sent a notification to over 49,000 of its patients informing them of the compromise of their medical data which supposedly was saved on an unencrypted flash drive [4]. In a similar case, according to the 2012 Verizon's data examination, a compilation of 47,000 security cases emphasized 621 of them being cases of data compromise [13]. In another event, a research on the privacy of patients and the security of data resulted in the revelation that on the minimum, about 94% of healthcare facilities have suffered a singular case of security and for that matter, data compromise [14].

With the dynamics of society and advancements in technology, privacy and security threats are expected to climb up, hence the need to take a recursive look into the best practices and mechanisms to secure and sustain the overall privacy and security of patients' data with the EHR being the cardinal data repository in focus.

3. Privacy Protection Techniques

The successes that big data brings to the healthcare industry can only be grounded with the conscious acknowledgment of the privacy and security concerns of patients. In this regard, this section discusses the different techniques that can be employed for privacy in healthcare such as anonymization, cryptography, and data management framework. The incessant intrusion upon patients' privacy partly as a result of the compromise of their information stored in EHRs has attracted attention in the domain of big data in healthcare. This section reviews some of the most domineering models and techniques in protecting the privacy of patients' data particularly those stored in EHRs.

3.1 Anonymization

The k-anonymity model presumes that every record represents a different individual. If quite a few records in a table denote the same record holder, a collection of k records may denote fewer than k-1 record holders, and the record holder may be under protected [15]. To achieve k-anonymity, there is the need for generalization and compression of data. Generally, a table is k-anonymous if the quasi-identifier of its individual instances (tuple) remain parallel to those of k-1 other instances [16]. As for an example, a healthcare facility may provide patients' data for purposes of research into discovering features and patterns of diseases as observed in the tables below.

TABLE 1: Micro Data

ID	***Attributes***			
	Age	***Sex***	***Zip code***	***Disease***
1	28	F	63474	Syphilis
2	43	M	63571	HIV
3	32	M	63657	STD
4	29	F	63594	Ulcer

TABLE 2: Registration List

ID	***Attributes***			
	Name	***Age***	***Sex***	***Zip code***
1	Kelvin	28	F	63474
2	Dane	43	M	63571
3	Mabel	32	M	63657
4	Joan	29	F	63594

TABLE 3: 2-Anonymous Table

ID	***Attributes***			
	Age	***Sex***	***Zip Code***	***Disease***
1	2*	F	634**	Syphilis
2	4*	M	635**	HIV
3	3*	*	636**	STD
4	2*	*	635**	Ulcer

From Table 1 above, a healthcare facility releases the data about its patients while withholding their names. Regardless of this effort, an intruder with access to say the voter's register in Table 2 can easily uncover the identities of the patients by matching the two tables on Age, Sex and Zip code. Table 3 represents a 2-anonymous generalization of Table 1. With this, an ammunition of the voter's register can only enable an intruder to deduce that Kelvin might be related to the instances in Table 3 or similarly, the disease of Kelvin is uncovered with only a 50% surety (a split chance or accuracy or error). In summary, k-anonymity ensures the association of a tuple with a chance of 1/k. An associate technique that falls under k-anonymity is the perturbation technique.

Data perturbation to privacy preservation is a renowned approach in privacy-preserving data mining. This technique ensures the quality of data as well as safeguarding information stored. It's protective ability to add more profound regarding hiding of delicate information such as patients' health data from the adversary.

The perturbation technique comes in two broad dimensions namely, probability distribution and fixed data perturbation

[17]. With an emphasis on the latter, the delicate portions of patients such as diagnosis of deadly diseases like HIV AIDs which are saved under said patient's id is converted into '0s' in the health facility's database (eg EHRs). This way, confidentiality of patients' information is upheld. As might have been hinted earlier, the most suitable format of data used by this technique are numeric and nominal data types. By extension, this technique manages well, data types in the characters, boolean, and integers. By privacy tradition, the cryptographic technique that comprises encryption and decryption strategies are applied shortly after data perturbation is performed. This ensures higher protection to safeguard patients' data [16]. An eminent shortcoming of the perturbation technique lies in the fact that patients' (clients') data are metamorphosed to the point that they fail to correspond to the original data scheme and might therefore not make meaning to the final recipients of the data.

3.2 Cryptographic Techniques

Hitherto, institutions used myriad methods of disguised identification techniques (de-identification) as a way to ensuring data security and privacy of their clients as well as their business operations [18], [19]. Authentication and access control provides security throughout active application level. Whereas, cryptographic techniques protect the patient data while the data-in-rest and data-in-motion [20]. A more progressive and novel resolution in using cryptography techniques incorporates Advanced Encryption Standard (AES), Rivest-Shamir-Adleman (RSA), Elliptical Curve Cryptography (ECC) among others. Contemporary exposures portray that the National Security Administration (NSA) might have discovered channels to breach encryption systems. What is more, virtual fences such as firewalls, secure sockets layer and transport layer security are mapped to impede entry to databases. These technologies can, however, be breached, thus, requiring the need to be constantly monitored.

Emerging as a formidable technique in handling privacy in EHRs, the cryptographic technique is employed to chiper data which are being disseminated along several heterogeneous systems [16]. This technique usually reinforces the security of databases and parallel sites after the extensive deployment of perturbation methods. With this, the perturbated records are manipulated using encryption and decryption techniques [17]. Encryption as a component of a cryptographic technique involves the conversion of data into a code with the sole rationale of regulating unauthorized access to databases. Data encrypted is called 'cipher text' whereas an unencrypted data is called 'plain text'. An invader of encrypted EHR of a health facility requires a key or password to gain access to the contents of the records stored.

Broadly, encryption algorithms fall under Symmetric and Asymmetric [21]. Encryption technique as a crucial component in a data protection methodology ensures a Personally Identifiable Information (PII) in the circumstance of a security breach. Generally, cryptographic techniques are used to perform authentication, non-repudiation, confidentiality as well as data integrity functions. The detailed exploration of these techniques is not captured in this review. Interested readers can refer article [14], [22].

3.3 Data Management Framework

Data Management (DM) relates to an institution's management of informational data/resources for secured and regulated access as well as safe keeping [23]. Generally, DM tasks include among others, creating frameworks for governing data to enhance analytical purposes, managing database integration systems, security of data, identification of data sources, segregation purposes and etc. [24]–[26]. The DM comprises various strategies geared towards the effective control of data ranging from creating (the architecture), processing, making the best utilization and finally the deletion stage [10], [27]. The implementation of DM is achieved through a robust architecture of technological and a governing roadmap (governance) which spells out the processes to be used all through the sequence of the data manipulation.

In a close alignment to Data management is Master Data. It is the central data of an institution which is used recurrently across applications and organizational processes [28], [29]. In the perspective of the healthcare, master data includes among others, information about patients, health plans, services, locations, other related units, and the relationships amid them. Consolidating this data produces a single version of the 'truth' on which a healthcare organization can base its analysis and planning. It enables us to see relations among cardinal organizational entities which facilitates the delivery of rapid and reliable insight [30]. In the arena of healthcare, master data comes in two categories; Identity Data which comprises patient, provider and location identifiers. In the next category is Health Level Seven (HL7) with its shared model, Reference Information Model (RIM) is founded on the core concept of recognition of external events as well as trigger events by computer applications. By complying with medical standard codes, HL7 becomes mainly relevant in terms of communication adaptability in meeting specific protocol requirements, allowing for international collaboration, ensuring the possibility of using varied codes and vocabularies in its messages [31]. Master Data Management (MDM) therefore encapsulates the practice of cleaning, streamlining and integrating data into an entity-wide 'system of record' for fundamental organizational activities [32]. It functions as a discipline used to bring compliance, accountability, and integrity to data. Wrongly, matching patients' data is bound to result in undesirable repercussions hence it becomes imperative to send the right identifiers of patients across systems especially liaising it with niche systems like Lab Info Systems (LIS) and Health Info Exchanges (HIE).

With the rise of the concept of Master Patient Index (MPI) which is used to manage patient data by way of assigning unique identifiers to every individual patient. The unique identifier acts like a linkage to other systems and applications to refer to a patient. An equally robust and innovative technology named integrated Rule-Oriented Data (iRODS) is believed to be a reliable means of ensuring privacy and security in EHRs [33]. This iRODS has been mapped and intended at absorbing the inadequacies inherent in the anonymization and cryptographic techniques of privacy preservation. It will allow healthcare organization to setup and install solutions that manage and releases data that is relevant to the healthcare agencies such as the Health Insurance Portability and Accountability Act (HIPAA) in the USA requirements, patient consent and etc. [34]. Among other important features, iRODs features federated data grids (intelligent clouds), a distributed rules engine, 'iCAT' meta sequence, a storage access platform that permits entry, a super blend of GUIs as well as APIs for interaction with an iRODS data grid [34]. iRODs employs a series of data management technologies hence has been adopted by numerous organizations across the globe [35]. Entity Resolution as a procedure ensures the efficient consolidation of data from multiple sources into a well-structured while accounting for key security metrics. This interwoven-multifaceted database is crucial for references determination [36]. While subjecting restrictions on the accessibility to sensitive data, the storage system has to make room for impromptu occurrences. To remedy this situation, healthcare systems incorporate such techniques as "break-the-glass" (BTG) which has the propensity to allow users access to sidestep control protocols during emergency circumstances [37].

A crucial component of data management framework is emphasizing patient consent to the storage and processing of personal clinical data. Legislation in many countries take serious measures to ensures data privacy include the European Data Protection Act, 1998 in Europe, the National Standards to Protect Patients' Personal Medical Records, in the United States and Canada (through a variety of provincial laws) [38]. In the United Kingdom, for instance, articles of privacy legislation include Data Protection Act (1984, updated in 1998) (hereafter DPA) and the Human Rights Act (1998) (hereafter HRA). Whereas the HRA endorses reverence for the individual's right to privacy, the DPA reinforces previous legislation regarding the processing of personal data [38].

It is worthwhile to note that, patients consent can be effectively utilized in the real-time auditing process to detect deviations. Auditing of quality and safety measures during routine daily work can detect and quantify a varied collection of errors and systems glitches in a short period of time [39], [40]. Safety audits help to identify clinical errors and safety problems that lead staff to make immediate changes to improve performance.

4. Conclusion

This study discussed the adoption of privacy preserving techniques and related use of big data in the healthcare industry. With this, the benefits of big data in health care were explored as well as some of its challenges as enshrined in other related literature. Emphasis was put on how to safeguard the privacy and security of patients' records, especially in EHRs. In the light of this, some techniques in protecting the privacy of patients were discussed. They comprise the anonymization technique, cryptographic technique, and the data management framework. The study revealed that in as much as the former techniques endeavor to protect and ensure privacy, which does not consider patient consent and dynamic nature of healthcare operation. Consolidating data as a core principle of a Data Management Framework (DMF) leads to producing a unilateral perspective of the 'truth' on which a healthcare institution can use as a platform for its analysis as well as planning procedures. There are several pros that a good DMF brings to the table among, which are elevating the quality of healthcare delivery services by providing consistency in facility, patient and procedural information. With this, healthcare institutions can attain deeper insights into metrics such as patient populations and also enable the identification and replication of desirable practices from in-network physicians. Supporting patient growth emerges as another benefit of a DMF. The framework as a solution helps to reduce out-of-network referrals by presenting relations between patients, physicians and referral conduct to remove points of referral leaks. As a force of remedy, the iRODS technique absorbs these weaknesses and ensures a more reliable avenue to protecting as well as ensuring the privacy of patients.

In future works, we envisage to explore conceptual model integrating dynamic rule engine in access control for EHRs in accommodating Big Data.

References

[1] M. Jayabalan and T. O'Daniel, "Access control and privilege management in electronic health record: a

systematic literature review," J. Med. Syst., vol. 40, no. 12, p. 261, Dec. 2016.

[2] M. Hajirahimova, "The Big Data Era in Healthcare: Promises and Challenges," Probl. Inf. Technol., vol. 8, no. 1, pp. 64–72, 2017.

[3] M.-T. Kechadi, "Healthcare Big Data: Challenges and Opportunities," in Proceedings of the International Conference on Big Data and Advanced Wireless Technologies - BDAW '16, 2016.

[4] H. Kupwade Patil and R. Seshadri, "Big Data Security and Privacy Issues in Healthcare," in 2014 IEEE International Congress on Big Data, 2014, pp. 762–765.

[5] R. Rahman and C. Reddy, "Electronic health records: a survey," Healthc. Data Anal., 2015.

[6] D. B. A. Saranga Jayawardena, "A Systematic Literature Review of Security , Privacy and Confidentiality of Patient Information in Electronic Health Information Systems," Sri Lanka J. Bio-Medical Informatics, vol. 4, no. 2, pp. 25–31, 2013.

[7] M. Rushanan, A. D. Rubin, D. F. Kune, and C. M. Swanson, "SoK: Security and Privacy in Implantable Medical Devices and Body Area Networks," in 2014 IEEE Symposium on Security and Privacy, 2014, pp. 524–539.

[8] S. Patel and A. Patel, "A Big Data Revolution in Health Care Sector: Opportunities, Challenges and Technological Advancements," Int. J. Inf. Sci. Tech., vol. 6, no. 1/2, pp. 155–162, Mar. 2016.

[9] S. M. Riazul Islam, Daehan Kwak, M. Humaun Kabir, M. Hossain, and Kyung-Sup Kwak, "The Internet of Things for Health Care: A Comprehensive Survey," IEEE Access, vol. 3, pp. 678–708, 2015.

[10] Federal Bureau of Investigation, "National and Transnational Security Implications of Big Data in the Life Sciences," 2014.

[11] M. Sandberg, J. Kristensson, P. Midlöv, and U. Jakobsson, "Effects on healthcare utilization of case management for frail older people: A randomized controlled trial (RCT)," Arch. Gerontol. Geriatr., vol. 60, no. 1, pp. 71–81, Jan. 2015.

[12] M. Kumar and S. Wambugu, "A Primer on the Privacy , Security , and Confidentiality of Electronic Health Records A Primer on the Privacy , Security , and of Electronic Health Records," pp. 1–13, 2016.

[13] F. H. R. France, "Security of health care records in Belgium," Int. J. Med. Inform., vol. 73, no. 3, pp. 235–238, Mar. 2004.

[14] A. Mahfuth, J. S. Dhillon, and S. M. Drus, "A Systematic Review on Data Security and Patient Privacy Issues in Electronic Medical Records," J. Theor. Appl. Inf. Tech Nol., vol. 90, no. 2, pp. 106–115, 2016.

[15] M. E. Rana, M. Jayabalan, and A. A. Mohung, "Privacy Preserving Anonymization Techniques for Patient Data: An Overview," in Third International Congress on Technology, Communication and Knowledge (ICTCK 2016), 2016.

[16] J. Wang, Y. Luo, Y. Zhao, and J. Le, "A Survey on Privacy Preserving Data Mining," in 2009 First International Workshop on Database Technology and Applications, 2009, pp. 111–114.

[17] S. Mekala and S. Sathappan, "Privacy Preserving In Patient Health Record Using Data Perturbation And Decision Tree," Int. J. Adv. Res. Basic Eng. Sci. Technol., vol. 2, no. 19, pp. 160–164, 2016.

[18] C. A. Kushida, D. A. Nichols, R. Jadrnicek, R. Miller, J. K. Walsh, and K. Griffin, "Strategies for De-identification and Anonymization of Electronic Health Record Data for Use in Multicenter Research Studies," Med. Care, vol. 50, no. 7, pp. S82–S101, Jul. 2012.

[19] W. Susilo and K. T. Win, "Security and Access of Health Research Data," J. Med. Syst., vol. 31, no. 2, pp. 103–107, Mar. 2007.

[20] M. Jayabalan and T. O. Daniel, "Continuous and Transparent Access Control Framework for Electronic Health Records : A Preliminary Study," in International Conference on Information Technology on Information Technology, Information Systems, and Electrical Engineering (ICITISEE 2017), 2017.

[21] S. M. Diesburg and A. A. Wang, "A survey of confidential data storage and deletion methods," ACM Comput. Surv., vol. 43, no. 1, pp. 1–37, Nov. 2010.

[22] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing," in 2010 Proceedings IEEE INFOCOM, 2010, pp. 1–9.

[23] J. Beel, B. Gipp, S. Langer, and M. Genzmehr, "Docear," in Proceeding of the 11th annual international ACM/IEEE joint conference on Digital libraries - JCDL '11, 2011, p. 465.

[24] A. E. Youssef, "A Framework for Secure Healthcare Systems Based on Big Data Analytics in Mobile Cloud Computing Environments," Int. J. Ambient Syst. Appl., vol. 2, no. 2, pp. 1–11, Jun. 2014.

[25] S. S. Bhowmick, L. Gruenwald, M. Iwaihara, and S. Chatvichienchai, "PRIVATE-IYE: A Framework for Privacy Preserving Data Integration," in 22nd International Conference on Data Engineering Workshops (ICDEW'06), 2006, pp. 91–91.

[26] N. Chakraborty, V. Sharma, and J. Ranjan, "A perceptual study on factors of medical data security in Indian organizations," J. Theor. Appl. Inf. Technol., vol. 84, no. 1, pp. 59–78, 2016.

[27] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing," ACM Comput. Surv., vol. 42, no. 4, pp. 1–53, Jun. 2010.

[28] R. Fang, S. Pouyanfar, Y. Yang, S. Chen, and S. S. Iyengar, "Computational Health Informatics in the Big Data Age," ACM Comput. Surv., vol. 49, no. 1, pp. 1–36, Jun. 2016.

[29] W. Fan, J. Li, S. Ma, N. Tang, and W. Yu, "Towards certain fixes with editing rules and master data," VLDB J., vol. 21, no. 2, pp. 213–238, Apr. 2012.

[30] L. Goldman, G. Benjamin, W. Wong, and K. Permanente, "Advancing the Health of Communities and Populations A Vital Direction for Health and Health Care," Natl. Acad. Med., 2016.

[31] G. W. Beeler, "HL7 Version 3—An object-oriented methodology for collaborative standards development1Presented at the International Medical Informatics Association Working Group 16 Conference on Standardisation in Medical Informatics—Towards International Consensus and C," Int. J. Med. Inform., vol. 48, no. 1–3, pp. 151–161, Feb. 1998.

[32] E. Insight, I. Efficiency, and I. Care, “Healthcare Data Management for Providers,” 2013.

[33] A. Rajasekar, R. Moore, C.-Y. Hou, C. A. Lee, R. Marciano, A. de Torcy, M. Wan, W. Schroeder, S.-Y. Chen, L. Gilbert, P. Tooby, and B. Zhu, “iRODS Primer: Integrated Rule-Oriented Data System,” Synth. Lect. Inf. Concepts, Retrieval, Serv., vol. 2, no. 1, pp. 1–143, Jan. 2010.

[34] B. Matturdi, X. Zhou, S. Li, and F. Lin, “Big Data security and privacy: A review,” China Commun., vol. 11, no. 14, pp. 135–145, 2014.

[35] X. Zhang, C. Yang, S. Nepal, C. Liu, W. Dou, and J. Chen, “A MapReduce Based Approach of Scalable Multidimensional Anonymization for Big Data Privacy Preservation on Cloud,” in 2013 International Conference on Cloud and Green Computing, 2013, pp. 105–112.

[36] Hyunmo Kang, L. Getoor, B. Shneiderman, M. Bilgic, and L. Licamele, “Interactive Entity Resolution in Relational Data: A Visual Analytic Tool and Its Evaluation,” IEEE Trans. Vis. Comput. Graph., vol. 14, no. 5, pp. 999–1014, Sep. 2008.

[37] A. Adriansyah, B. F. Van Dongen, and N. Zannone, “Controlling Break-the-Glass through Alignment,” in 2013 International Conference on Social Computing, 2013, pp. 606–611.

[38] A. M. Clark and I. N. Findlay, “Attaining adequate consent for the use of electronic patient records: An opt-out strategy to reconcile individuals’ rights and public benefit,” Public Health, vol. 119, no. 11, pp. 1003–1010, Nov. 2005.

[39] R. Ursprung, “Real time patient safety audits: improving safety every day,” Qual. Saf. Heal. Care, vol. 14, no. 4, pp. 284–289, Aug. 2005.

[40] M. E. Rana, M. Kubbo, and M. Jayabalan, “Privacy and Security Challenges Towards Cloud Based Access Control in Electronic Health Records,” Asian J. Inf. Technol., vol. 16, no. 2, 2017.