Computer Science Journal of Moldova, vol.**24**, no.2(71), 2016

# Digital Health Data: A Comprehensive Review of Privacy and Security Risks and Some Recommendations [*]

Shahidul Islam Khan, Abu Sayed Md. Latiful Hoque

**Abstract**

In todays world, health data are being produced in ever-increasing amounts due to extensive use of medical devices generating data in digital form. These data are stored in diverse formats at different health information systems. Medical practitioners and researchers can be benefited significantly if these massive heterogeneous data could be integrated and made accessible through a common platform. On the other hand, digital health data containing protected health information (PHI) are the main target of the cybercriminals. In this paper, we have provided a state of the art review of the security threats in the integrated healthcare information systems. According to our analysis, healthcare data servers are leading target of the hackers because of monetary value. At present, attacks on healthcare organizations' data are 1.25 times higher compared to five years ago. We have provided some important recommendations to minimize the risk of attacks and to reduce the chance of compromising patients' privacy after any successful attack.

**Keywords:** Health Data, Privacy, Security, Data Breach, PHI

## 1 Introduction

Health data refers to pieces of information collected to use in the diagnosis of a health condition. Health Information is collected about

[*] This research is supported by the ICT Division, Ministry of Posts, Telecommunication and Information Technology, Government of the People's Republic of Bangladesh.

a patient, his/ her family, often during creating of a nursing history for the patient. A health record may include multiple types of health data such as various notes entered by health care professionals over time, recording observations and administration of drugs, test results, x-rays, reports, etc. Digital health data are health data generated by medical devices in digital form e.g., fasting plasma glucose test (FGT) result, or other patient health related information e.g., height, weight, blood group etc stored in digital form at computers, laptops, or in database of health information systems [1]–[3].

At present, enormous quantity of digital health data are generated daily by healthcare providers. Medical records of patients are increasingly digital, in the form of Electronic Health Record (EHR). These EHRs are more useful than paper records for better healthcare and medical research because electronic data can be stored easily and manipulated by software. These precious data are stored in various health information systems (HIS) in hospitals, research centers and diagnostic laboratories. Many of these data fall in the category of protected health information.

Protected health information (PHI) is defined as personally identifiable health information collected from an individual, and covered under federal or international data breach disclosure laws [4]. PHI of an Individual is information which relates to:

a. the individuals past, present, or future physical or mental health or condition,

b. the provision of health care to the individual,

c. the past, present, or future payment for the provision of health care to the individual, and that identifies the individual or for which there is a reasonable basis to believe that the information could be used to identify the individual.

PHI includes many common identifiers such as name, date of birth, address, National ID / Social Security Number, telephone and fax numbers, E-mail addresses etc. when they can be associated with the health information listed above [5].Laboratory reports, medical records, and hospital bills are examples of PHI because each document contains a patient's name and/or other identifying information associated with the health data content.

Security of a HIS deals with protecting medical data from intruders, malwares, and frauds. It retains confidentiality and integrity of healthcare data. Privacy concerns exist wherever personally identifiable information or other sensitive information is collected and stored in any form. A major challenge in health data privacy is to share data among medical practitioners while protecting personally identifiable information. Information privacy may be applied in numerous ways, including encryption, authentication and data masking – each attempting to ensure that information is available only to authorized persons [6], [7].

Nowadays, hacking PHI by cybercriminals is observed as a growing trend. Hackers goal is to take advantage of personal information of the patients. Average sell value of a complete medical record varies from \$10 to \$1,000 in black market. Although privacy of a patient can be compromised with paper based medical records, it alarmingly increased along with digitized record keeping by the healthcare providers [8], [9].

It is obvious that developing a national health data warehouse (NHDW), where integrated data from all the diverse HIS will be made available for better health delivery and medical research, is very much essential for every country [10]–[16]. However NHDW raises high risk to data security and privacy of individuals. Before integration to NHDW, sensitive and private data of patients reside to a single organization such as a hospital or a diagnostic center. Only that particular organization is responsible by law to protect the data privately. Now the situation is far different in the case of national warehouse. So proper measures have to be taken to safeguard privacy of patients in the NHDW.

In this paper we have presented a comprehensive review of security and privacy risks of digital health data and integrated health information systems. We have exposed the statistics of high rise of security threads in healthcare data servers. In addition, we have provided some general recommendations to reduce risks of PHI breaches and some specific recommendations for developing national scale integrated health information systems.

## 2 Data Breaches of Health Information Systems

A health data breach or leakage is defined as an event that involves the loss or exposure of personal health records. Personal health records are data containing privileged health related information about an individual that cannot be readily obtained through other public means, which information is only known by an individual or by an organization under the terms of a confidentiality agreement [17]. For example, leakage of a health insurer's record of the policyholder with doctor and payment information will be treated as a health data breach. According to the research by IBM and Ponemon Institute in 2015 where 350 companies in 11 countries were interviewed extensively, more than 18 thousand records were breached on an average in each breached incident [18]. This is presented in Fig. 1.

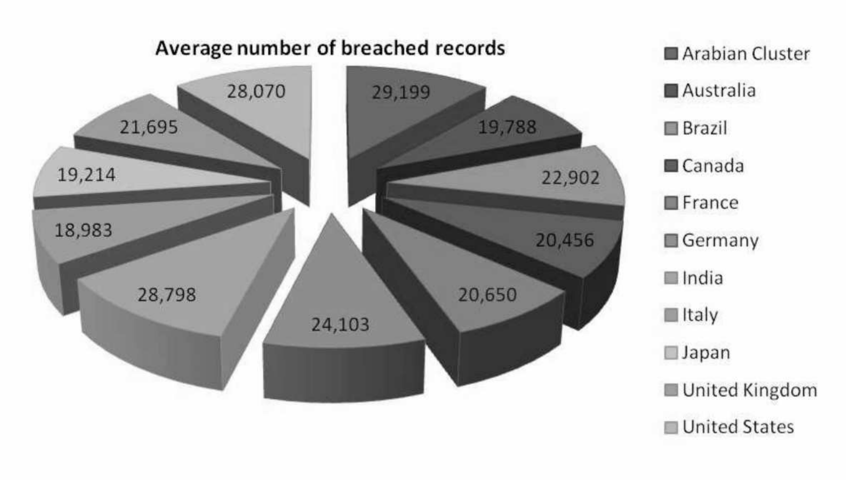

Figure 1. Average number of breached records in a data breach incident

The costs of a data breach can vary according to the cause and the protections in place at the time of the breach. Direct costs refer to the direct expense spent to carry out a given activity such as hiring forensic experts and law firm or offering identity protection services to the victims. Indirect costs include the time, effort and other organizational resources spent during the data breach resolution. Indirect costs also include the loss of goodwill and customer churn. In 2015, the average cost of data breach per lost or stolen record was 154USD but in case of a breach of healthcare organization, the average cost was 363USD [18]. This is shown in Fig. 2.

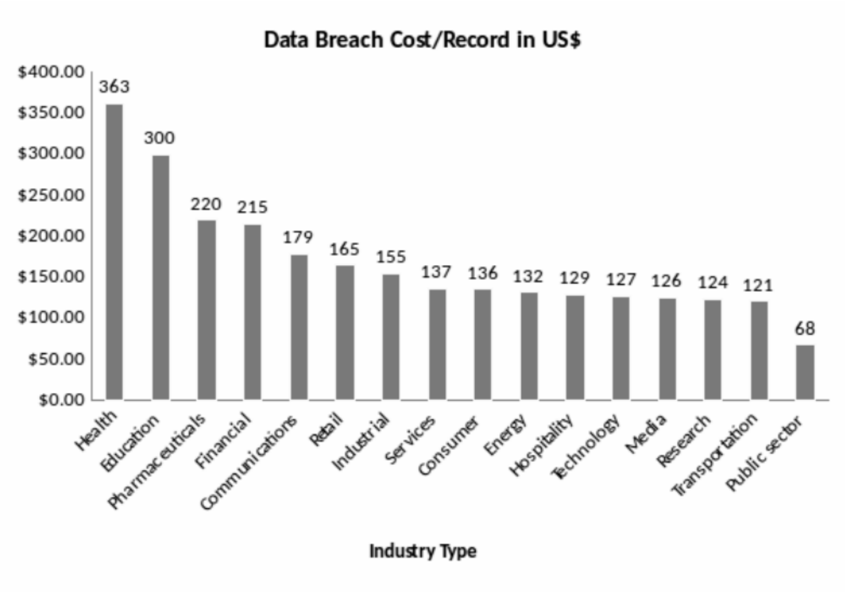

Figure 2. Cost of each breached record in different sector. The cost is maximum for the healthcare industry.

## 2.1 Health data breaches

According to 2015 Fifth Annual Benchmark Study on Privacy and Security of Healthcare Data which covered 90 healthcare organizations in USA, more than 90% of healthcare service providers had a data breach, and 40% had more than five data breaches over the past two years [19]. The following chart of Fig. 3 shows the total numbers of health data breaches in USA in last five years till February 26, 2016. We have calculated the data from [20].

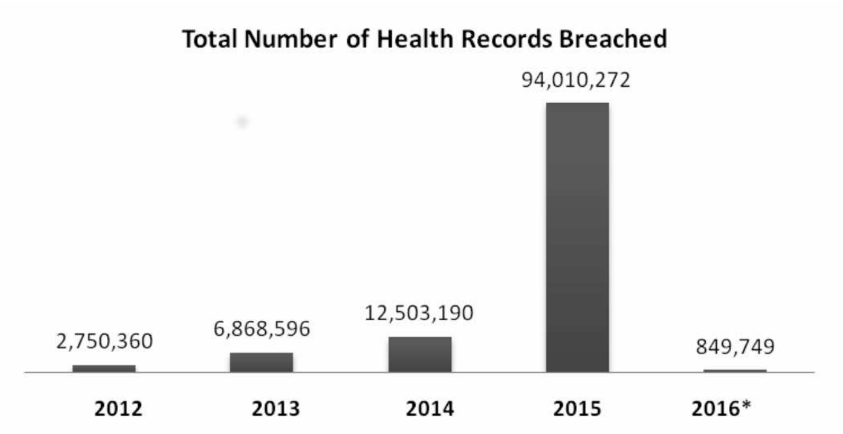

Figure 3. Total number of health records breached in USA

According to the report [19], for the first time, criminal attacks are the number one cause of healthcare data breaches. Criminal attacks on healthcare organizations are 1.25 times higher compared to five years ago. The main causes of data breach in healthcare sectors are illustrated in Fig. 4.

Some recent attacks on health information centers are listed below:

- Hackers have shut down the internal computer system at a Hollywood Presbyterian Medical Center for more than a week for a payoff of 9,000 bitcoins, or almost USD 3.7 million [21]. It is due to a malicious software called ransomware that encrypts sensitive data until it can only be decrypted with a code.

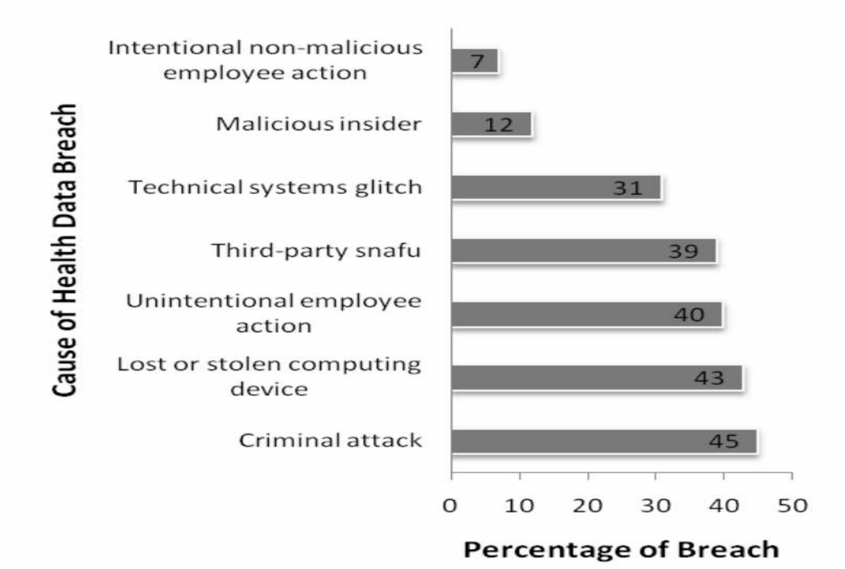

Figure 4. Main causes of data breach in the healthcare industry

- In February 2016, Jackson Health System discovered that a hospital employee have stolen confidential PHI of patients including names, birthdates, social security numbers and home addresses around 24,000 patient records over the last five years [22].

- The Washington State HCA reported, in February 2016, that an employee error resulted in a healthcare data breach compromising 91,000 Medicaid patient files. The information affected includes clients social security numbers, dates of birth, Apple Health client ID numbers and private health information [23].

- Six hard drives containing personal and health information on clients of health insurance company Centene Corp were lost which contained Social Security numbers, birthdates, health data, names, addresses, and insurance identification numbers for 950,000 patients who received laboratory services between 2009-2015 [24].

- Premera Blue Cross was targeted with a sophisticated cyber attack after hackers gained access to the financial and medical information of 11 million members in January 2015. Hackers swiped Social Security numbers, financial information, medical claims data, addresses, email addresses, names and dates of birth [25].

- Health insurer Anthem Inc. has suffered a massive data breach on March 3, 2015 after hackers gained access to a corporate database reportedly containing personal information on around 80 million of the health insurer's current and former USA customers and employees [26].

- In last ten years at least 18 health breaches reported in Europe affected minimum 9,337,197 individual records [17]. The health records include details on the patients conditions, names, home addresses and dates of birth. The health networks and servers containing integrated health records are in high risk of cyber attacks all over the world.

### 2.2 Data breaches of healthcare servers

From 2014, hackings on healthcare servers increased terrifyingly. The attackers motivation is to get huge PHI in a single successful hack. Table 1 presents last 12 big criminal attacks on integrated health records in USA within last 12 months. We have summarized these data from [20].

We have analyzed the data provided by U.S. Department of Health and Human Services and found that hackers are increasingly targeted healthcare servers which is very alarming to national level health information system development. Table 2 and Fig. 5 illustrate the fact clearly.

### 2.3 Other impacts of health data breaches

There are other impacts of health data breaches. They are discussed below:

Table 1. Latest 12 big breaches in USA on Health Data Servers

| Sl. | Name of Healthcare Org. | Affected Individuals | Breach Date | Type of Breach |
|---|---|---|---|---|
| 1 | Alliance Health Networks, LLC | 42372 | 2/15/2016 | Hacking/IT Incident |
| 2 | OH Muhlenberg, LLC | 84681 | 11/13/2015 | Hacking/IT Incident |
| 3 | Excellus Health Plan, Inc. | 10000000 | 9/9/2015 | Hacking/IT Incident |
| 4 | Medical Informatics Engineering | 3900000 | 7/23/2015 | Hacking/IT Incident |
| 5 | University of California, Los Angeles Health | 4500000 | 7/17/2015 | Hacking/IT Incident |
| 6 | CareFirst BlueCross BlueShield | 1100000 | 5/20/2015 | Hacking/IT Incident |
| 7 | Freelancers Insurance Company | 43068 | 3/24/2015 | Hacking/IT Incident |
| 8 | ATnT Group Health Plan | 50000 | 3/23/2015 | Hacking/IT Incident |
| 9 | Premera Blue Cross | 11000000 | 3/17/2015 | Hacking/IT Incident |
| 10 | Anthem, Inc. Affiliated Covered Entity | 78800000 | 3/13/2015 | Hacking/IT Incident |
| 11 | Virginia (VA-DMAS) | 697586 | 3/12/2015 | Hacking/IT Incident |
| 12 | Georgia Department of Community Health | 912906 | 3/2/2015 | Hacking/IT Incident |

Table 2. Statistics of Healthcare server attack compared to total healthcare breach

| Reporting Year | Total Health Data Breach affecting 500 or more individuals | Healthcare Server Attach |
|---|---|---|
| January 1, 2011 to December 31, 2011 | 194 | 27 |
| January 1, 2012 to December 31, 2012 | 202 | 25 |
| January 1, 2013 to December 31, 2013 | 263 | 35 |
| January 1, 2014 to December 31, 2014 | 290 | 55 |
| January 1, 2015 to December 31, 2015 | 265 | 50 |

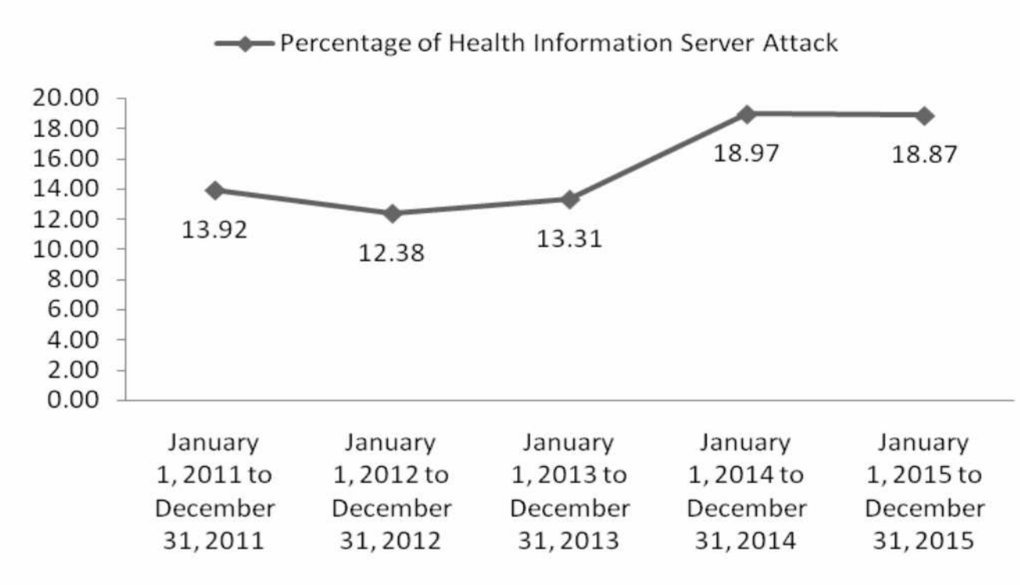

Figure 5. Criminal attack on Healthcare data servers are increasing high.

a. Breaches of PHI drastically effect on the goodwill of a healthcare organization. In a research report it is shown that, people are withholding their health information from healthcare providers because they are concerned that there could be a confidentiality breach of their records [27]. An unwillingness to fully disclose information could delay a diagnosis of a communicable disease. This is not only a potential issue for the treatment of a specific patient; there are potential public health implications.

b. Penalty of healthcare providers are imposed in two ways. They have to pay ransom to the hackers to get their breached data back or to restore their hacked system [21] and they also pay the government privacy penalty for failing to safeguard patient information [28].

## 3 Analysis of the risks related to Health Information Systems

If we analyze the increase trend of healthcare data breach around the globe, it becomes quite clear that the main reason of the breaches is the sell value of complete health records. What makes medical data so unique is that it often contains most of the information hackers are looking for such as credit card information, and Social Security and bank account numbers giving them a one-stop stealing strategy. Fraudsters use this data to create fake IDs to buy medical equipment or drugs that can be resold, or they combine a patient number with a false provider number and file made-up claims with insurers. Sometimes the cyber criminals use this data to blackmail a patient with good social status. For example, F1 racing legend Michael Schumachers and pop legend Michael Jacksons medical records were hacked.

If we look at Table 1, we can see that, all big breaches in healthcare servers are cause of hacking or IT incident though there are other causes available in the U.S. Govt. reporting form i.e., Theft, Unauthorized Access/Disclosure, Lossor unknown cause. So the owner of the healthcare servers should pay high attention to develop a secure framework to protect their health information servers from hacking or improper IT involvements.

Another important thing to notice is that, a healthcare company is looser in many ways after a successful breach. It has to pay money to both the hackers and the government. This situation will eventually increase healthcare cost and decrease better healthcare delivery. Policymaker should think about this.

If the stored health data are de-identified in every place from health information system software to backups and also in health data warehouses, then the risk of data breach can be significantly reduced. Because there is almost no sell value of de-identified health records. Another positive thing of de-identification is if a data breach occurs, privacy of individual patient will not be affected.

## 4 Some general recommendations to reduce the chance of health data breaches

a. At the very least, healthcare companies should back up all their important health data regularly so that, in emergency situations, hard drives can be cleaned and restored to their previous states. PHIs in database backups must also be encrypted.

b. Internal HIS software should be screened for loopholes that could be way in of hackers. All third party software should be updated with latest patch and service packs. No free software from unknown or un-trusted source should ever be downloaded or installed.

c. Doctors and nurses should be more careful when handling PHI of patients. They should encrypt these records in their own laptops and pen drives. After working in the workstations, they must always sign out from their accounts when they have finished inputting patient information or viewed patients reports.

d. Health-care consumers should be smarter. The more the patients will query healthcare providers about how they are securing PHI, the more attention the providers will pay to enhance security and privacy of patients PHI.

e. It is more effective to integrate privacy and security into health apps, devices, and services from the start. For any piece of information collection and storage, the following should be considered:

   i. Minimize the amount of personal information collected
   ii. Decide how long the information needs to be stored
   iii. Encrypt information when possible
   iv. Delete the information earliest

f. Rather than spending a lot of money after breaches, the healthcare organizations should increase their budget for HIS security. Prevention is better than cure- this proverb should always be remembered.

g. Medical practitioners need to be more cautious of email attachments and shouldnt include health information in e-mail unless encryption is used. If encryption is not available, confidentiality statement needs to be included like below at the top of the e-mail:

> **Notice: Privacy & Confidentiality of Information**
> *This communication may contain non-public, confidential, or legally privileged information intended for the sole use of the designated recipients. If you are not the intended recipient, or have received this communication in error, please notify the sender immediately by reply email at xxx@xxx.xx or by telephone at +xxx-xxxxxxxx, and delete all copies of this communication, including attachments, without reading them or saving them to disk. If you are the intended recipient, you must secure the contents in accordance with all applicable state or federal requirements related to the privacy and confidentiality of information, including the HIPAA/ EU Data Protection Directive Privacy guidelines.*

# 5 Specific Recommendations for Deployment of National Health Data Warehouse

No information system can be assumed to be completely protected from all kind of criminal and cyber attacks. Security can be more vulnerable in the case of large scale, national level health information systems where Internet communication has to be maintained for the sake of easy data collection from far-most parts of the country. So integrated health information systems should be designed in such a way that:

- There is enough data to maintain record linkage so that doctors, researchers can get useful insight from the system.
- If data breach occurs, individual patients privacy will be safeguarded.

Record linkage is the process of identifying record pairs from different information systems which belong to the same real world entity. Given two repositories of records, the record-linkage process consists of determining all pairs that are similar to each other. Record linkage is essential when joining datasets based on entities that may or may not share a common identifier such as national id or social security number [29], [30]. For discovering effective knowledge such as correlations among diseases from medical dataset it is very essential to maintain record linkage. On the other hand, identifiable health data have high risk to patient privacy and make the health information systems security vulnerable to hackers [31], [32] For development of national level health data warehouse our recommendations from security and privacy point of view are:

1. No Medical record can be stored in any level, from diagnostic centers to National Health Data Warehouse, with personal identifiable attributes of the patients.
2. To facilitate knowledge discovery process of the Healthcare researchers, sufficient record-linkage data have to be kept in medical

records by replacing personal identifiable attributes with unique code using suitable computer cryptographic technique.

3. A data-protection strategy has to be implemented that will cover data everywhere it is stored, and at every stage, from creation and processing, to storage, backup and transmission.

4. Proper security measures have to be taken and tested before connecting the national health data warehouse with Internet.

5. Proper security measures have to be taken and tested before deploying the national health data warehouse in the public cloud.

We propose the following flow chart that will significantly reduce cyber attack in the national health data warehouse and also retain the privacy of the patients after any data breach incident shown in Fig. 6.

## 6 Conclusions

Widespread use of digital health data could bring positive changes to the healthcare system in a various ways, as these data are the foundational piece to softwares and technologies that could advance health care delivery radically. Having every patient's data stored digitally, in a national platform creating an easy transfer and comparison of data among providers, insurers, and researchers, will allow recognition of interesting medical patterns, development of personalized and predictive medicine, reductions in medical errors, better disease management, predicting and preventing disease outbreaks, elimination of insurance fraud, identification of low cost treatments and many more. However integration of protected health information has high risk to patients' privacy and makes such systems vulnerable to hackers. In this paper, we have provided a state of the art review of security and privacy risks of integrated healthcare information system. We have analyzed current security and privacy threats and provided some recommendations to reduce health data breaches. We have also provided some guidelines for developing national scale integrated health information systems.

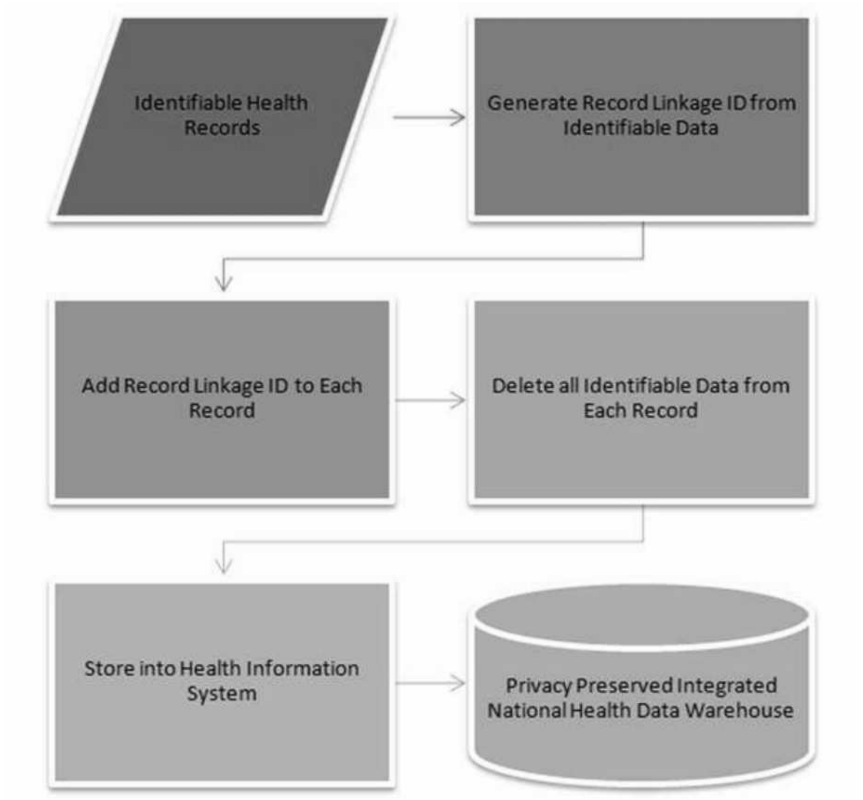
Identifiable Health Records
Generate Record Linkage ID from Identifiable Data
Add Record Linkage ID to Each Record
Delete all Identifiable Data from Each Record
Store into Health Information System
Privacy Preserved Integrated National Health Data Warehouse

Figure 6. Flow chart of security and privacy management of National health data Warehouse

## References

[1] C. K. Reddy and C. C. Aggarwal, *Healthcare data analysis.* CRC Press, 2015.

[2] Y. Zhang and C. Poon, "Editorial note on bio, medical and health informatics," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 543–545, 2010.

[3] M. L. Braunstein, *Practitioners Guide to Health Informatics.*

Springer, 2015.

[4] (2016, Feb.). [Online]. Available: http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

[5] (2016, Feb.) Protected health information: What does phi include? [Online]. Available: https://www.hipaa.com/hipaa-protected-health-information-what-does-phi-include

[6] F. T. Harold and K. Micki, *Information Security Management Handbook*, 6th ed. CRC Press, 2015, vol. 2.

[7] B. P. Robichau, *Healthcare Information Privacy and Security: Regulatory Compliance and Data Security in the Age of Electronic Health Records*, 1st ed. Apress, 2014.

[8] (2015, Sep.) Why hackers are targeting health data. [Online]. Available: http://www.databreachtoday.asia/hackers-are-targeting-health-data-a-7024

[9] (2015, Sep.) Your medical record is worth more to hackers than your credit card. [Online]. Available: http://www.reuters.com/article/2014/09/24/us-cybersecurity-hospitals-idUSKCN0HJ21I20140924

[10] Y. Zhang and C. Poon, "The development of health care data warehouses to support data mining," *Clin Lab Med.*, vol. 28(1), pp. 55–71, 2008.

[11] S. Nugawela, "Data warehousing model for integrating fragmented electronic health records from disparate and heterogeneous clinical data stores," M.Sc. Thesis, Queensland University of Technology, Australia, 2013.

[12] W. Kerr, E. Lau, G. Owens, and A. Treer, "The future of medical diagnostics: large digitized databases," *Yale J Biol Med*, vol. 85, no. 3, pp. 363–377, 2012.

[13] S. I. Khan and A. S. M. L. Hoque, "Towards development of health data warehouse: Bangladesh perspective," in *Proc. 2nd International Conference onElectrical Engineering and Information Communication Technology (ICEEICT)*, May 2015, pp. 1–6.

[14] S. I. Khan and A. Hoque, "Towards development of national health data warehouse for knowledge discovery," in *Intelligent Systems Technologies and Applications*, ser. Advances in Intelligent Systems and Computing. Springer-Verlag, 2016, vol. 385, no. 2, pp. 413–421.

[15] S. I. Khan and A. S. M. L. Hoque, "Development of national health data warehouse for data mining," *Database Systems Journal*, vol. VI, no. 1, pp. 3–13, 2015.

[16] (2015, Jul.) A quiet revolution: Strengthening the routine health information system in bangladesh. [Online]. Available: http://health.bmz.de/good-practices/GHPC/A_Quiet_Revolution/HIS_Bangladesh_long_EN.pdf

[17] (2016, Feb.) Reported breaches of compromised personal records in europe. [Online]. Available: http://cmds.ceu.edu/sites/cmcs.ceu.hu/files/attachment/article/663/databreachesineurope.pdf

[18] IBM and P. Institute, "2015 cost of data breach study: Global analysis," IBM and Ponemon Institute, Research Report, 2015.

[19] P. Institute, "Fifth annual benchmark study on privacy & security of healthcare data," Ponemon Institute, Research Report, 2015.

[20] (2016, Feb.) Breach portal: Notice to the secretary of hhs breach of unsecured protected health information. [Online]. Available: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf

[21] (2016, Feb.) Hospital pays hackers 17,000 to unlock ehrs frozen in 'ransomware' attack. [Online]. Available: http://www.modernhealthcare.com/article/20160217/NEWS/

160219920/hospital-pays-hackers-17000-to-unlock-ehrs-frozen-in-ransomware

[22] (2016, Feb.) Jackson health: rogue employee suspected of stealing private patient information. [Online]. Available: http://www.miamiherald.com/news/health-care/article59339038.html

[23] (2016, Feb.) 91k patients data compromised in wa healthcare data breach. [Online]. Available: http://healthitsecurity.com/news/91k-patients-data-compromised-in-wa-healthcare-data-breach

[24] (2016, Feb.) Missing drives contained phi on 950k centene customers. [Online]. Available: http://www.scmagazine.com/missing-drives-contained-phi-on-950k-centene-customers/article/467860/

[25] (2015, Sep.) Premera blue cross breach exposes financial, medical records. [Online]. Available: http://krebsonsecurity.com/2015/03/premera-blue-cross-breach-exposes-financial-medical-records/

[26] (2016, Feb.) Anthem hit by massive data breach. [Online]. Available: http://www.healthcareinfosecurity.com/anthem-health-hit-by-massive-data-breach-a-7876

[27] Verizon, “Protected health information data breach report,” Verizon, Research Report, 2015.

[28] (2016, Jan.) Lincare ordered to pay 239,800 hipaa privacy penalty. [Online]. Available: http://www.modernhealthcare.com/article/20160209/NEWS/160209856/lincare-ordered-to-pay-239800-hipaa-privacy-penalty

[29] L. Jin, C. Li, and S. Mehrotra, “Efficient record linkage in large data sets,” in *Proc. Eighth International Conference on Database Systems for Advanced Applications (DASFAA 2003)*, Mar. 2003, pp. 137–146.

[30] E. Sauleau, J. Paumier, and A. Buemi, “Medical record linkage in health information systems by approximate string matching and

clustering," *BMC Med Inform Decision Making*, vol. 5, pp. 32–44, 2005.

[31] N. K. Abel, P. C. John, L. J. Kathryn *et al.*, "Design and implementation of a privacy preserving electronic health record linkage tool in chicago," *Journal of the American Medical Informatics Association*, pp. 1–9, 2015.

[32] S. I. Khan and A. Hoque, "Privacy and security problems of national health data warehouse: A convenient solution for developing countries," in *Proc. 2nd International Conference on Networking Systems and Security (NSysS)*, Jan. 2016, pp. 157–162.

Shahidul Islam Khan, Abu Sayed Md. Latiful Hoque

Received October 21, 2015

Revised April 5, 2016

Dept. of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000
E-mail: nayeemkh@gmail.com
asmlatifulhoque@cse.buet.ac.bd