Advanced Science and Technology Letters
Vol.136 (ITCS 2016), pp.54-58
http://dx.doi.org/10.14257/astl.2016.136.14

A Study on Personal Health Information De-identification Status for Big Data

Young-Chul Chung[1], Ya-Ri Lee[2], Jung-Sook Kim[3*1], Ho-Kyun Park[4]

[1]Information Technology Convergence Unit of KIHASA, Republic of Korea
[2]Personal Information Protection Center of SSIS, Republic of Korea
[3*]Division of Computer Science, Sahmyook University, Republic of Korea,
[4]School of IT Convergence Engineering, Shinhan University, Republic of Korea
cyc@kihasa.re.kr, i_lyaree@ssis.or.kr, kimjs@syu.ac.kr, hkpark@shinhan.ac.kr

Abstract. One of the most promising areas of application of Big Data is the field of medicine. In order to apply Big Data generally, it is necessary to prevent infringement of privacy, and to minimize the personal information misuse, to define and set the boundaries for the terms used with regards to medical information and personal health information. Thus, through an agreement between the various parties set for a range of personal health information de-identifying action it is required. Thus, through an agreement between the various parties set for a range of personal health information de-identifying action it is required.

Keywords: Personal Health Information, De-identification, Personal Information Protection, Guideline

1 Introduction

While the interest in big data explosion, the Government of the Republic of Korea makes its main keywords in the government's policy to utilize the excavations “Big Data”. The government has enacted “Act on Provision and Active Use of Public Data” in 2013 and enforced to actively express the will of it. With this groundwork, there are approximately 16,000 cases of public data registered at the Public Data Portal(www.data.go.kr) as of May 9, 2016 and according to a study done by the OECD in 2015, South Korea has an openness index of 0.98, making it the top country on the list[1].

The leverage big data in the medical field from a variety of areas such as Seoul National University in Korea in the middle are receiving much attention hospitals, Health Insurance Review Agency, Health Insurance, the National Human Resources Bank, variety, etc. medicinal sources health information Big Data use case is it is being introduced[2]. One of the issues that frequently arise in the usage of Big Data is the problem of infringement of personal information. In the current status of Korea's Big Data policy implementation [3], the issue of protecting personal information has

[*1] paper presenter

ISSN: 2287-1233 ASTL

been put forth as a limiting factor in the application of Big Data. In this sense, the application of Big Data and the protection of personal information appear to be two sides of the same token. Whereas there are high expectations for the application of Big Data in the medical field, it is totally necessary to hold discussions and come up with measures against the violation of personal information as Big Data contains a huge amount of sensitive personal information. Moreover, the rise of Big Data signifies the creation of new values in the integration of data in various domains. For example, one should be alarmed by the fact that personal health information is being unwittingly collected and used by means of elaborate data algorithms from day-to-day information collection involving credit cards, e-shopping, etc. in addition to medical services[4]. Meanwhile, personal identification information by which the individual is identified or at least has the potential of being identified in most countries [5]. A measure of de-identification which makes it impossible to identify a certain individual based on his personal information is known as one of the measures to minimize the risk of infringement of personal information.

This study describes and makes relevant suggestions for the current status of the de-identification process as a protective measure for personal information for the safe application of Big Data in the field of medical services.

2 Related Works

Among the personal information name, phone number, unique identification information (Social Security numbers, driver's license number, alien registration number, passport number), a biometric information (fingerprints, iris, DNA information), institutions and organizations of the user account information (registration number, account number) is the information that can identify individuals by themselves[6]. Health information is labeled using many terms such as health and medical information, medical information, medicine information, diagnostic information, health information, healthcare information, etc. without a clear distinction among them in many studies. In this study, the term "health information" is used, and the term "personal health information" zeroes in on the information among health information that has qualities which can be used to identify the individual. Eliminating identifying characteristics from personal health information and "de-identifying" it is currently in the spotlight as an optimum means of decreasing the risk of the violation of personal information. Thus, this study examines the de-identification regulations regarding personal health information in the HIPAA Privacy Rule, which is representative of de-identification regulations.

2.1 HIPAA Privacy Rule

The HIPAA (the Health Insurance Portability and Accountability Act) is a US federal law enacted in 1996, which standardizes the administration of and the electronic transfer of financial data with regards to medical services [7]. In conjunction with this law, the HIPAA Privacy Rule has been set up to protect personal health information, and it contains detailed regulations that deal with personal information protection in medically-related fields. The HIPAA Privacy Rule stipulates what constitutes PHI (Protected Health Information) which falls under the umbrella of protection by the relevant law, and it also establishes certain complete and partial exceptions to the rule to exempt certain health information from protection so it can be used and provided freely.

First, information that is protected is any information that can be used to identify the individual or gives reasonable clues as to the identity of the individual. On the other hand, health information that has been de-identified and cannot be used to identify the individual is clearly stated to be exempt from the Privacy Rule and can be freely used and provided by any party. This makes it clear that personal health information de-identification is an important means in the utilization of health information [7].

Meanwhile, additional explanation can be provided by the expert determination method as a de-identification method for complete rule exemption and the safe harbor method which removes 18 identifying factors as shown in Table 1[7].

Table 1. Identifying factors that have been eliminated using the safe harbor method

Category	Identifying factor.	Category	Identifying factor
1	Name	10	Account numbers
2	Address	11	Certificate/license numbers
3	Date Information	12	Vehicle identifiers and serial numbers
4	Telephone numbers	13	Device identifiers and serial numbers
5	Fax numbers	14	Web Universal Resource Locators
6	Email address	15	Internet Protocol addresses
7	Social security numbers	16	Biometric identifiers
8	Medical record numbers	17	Full-face photographs
9	Health plan beneficiary numbers	18	Any other unique identifying number

3 Korea’s Personal Health Information De-identification Rule

In Korea, guidelines, case study books, etc. on personal health information de-identification are being made available based on legislations in other. Personal information protection guidelines regarding the opening up and sharing of public information suggest items such as the de-identification rule, gradual de-identification measures, de-identification techniques, etc. through taking protective measures by de-identifying personal information in the handling and analyzing of public information.

"Personal Information De-identification Casebook for the Application of Big Data[9]" proposes specific principles for each technique such as de-identification measures for each step of Big Data application, de-identification techniques, etc. It also provides actual case studies of de-identification in Big Data application.

"Cases of Hands-on Applications in the Field of Medicine" presents actual cases of de-identification in items such as name, resident registration number, age, address, phone number, registration number, health insurance number, bank account number, license number, vehicle registration number, biological information(fingerprint, profile, iris, pulse, voice, handwriting recognition, etc.), genetic information, homepage member ID, employee identification number, personal identification number, sanatorium registration number, earnings, ID, diagnostic data, prescription date, diagnostic testing date, general testing date, etc.

"Guidelines for Big Data Personal Information Protection[10]" stipulates that information and communications service providers take de-identification measures to protect personal information when dealing with open information and usage record information. Moreover, it stipulates that one cannot collect or use previously open information that has been de-identified without the consent of the user.

"The Guide to Free Optimality Assessment for Personal Information De-identification[11]" includes the procedures for optimality assessment for personal information de-identification, detailed assessment methods, re-identification risk management measures, etc. It also suggests 18 identifying factors based on the example of the HIPAA Privacy Rule's 18 identifying factors, after having modified them to fit Korea's situation. Also, related legislations employ a four-step procedure for the measures to de-identify personal information.

First, when there is no basis for it, personal identifying factors are deleted by applying personal identifying factor elimination techniques and other detailed regulations.

Second, statisticians and other math-related experts examine the possibility of individual identification by the given information.

Third, the data is utilized.

Fourth, sustained checking and monitoring for the possibility of re-identification are carried out.

"Guidebook on Techniques for Personal Information De-identification for the Application of Big Data [12]" gives guidelines for personal identification information categories for each field, 18 de-identification technique application methods, and case studies of application in each field.

4 Conclusion and Expected Effects

On many de-identification regulations regarding personal information include de-identification principles, step-by-step de-identification measures, de-identification techniques and detailed application methods, case studies of de-identification application in the processing of Big Data in each field, optimality assessment procedures and detailed assessment methods, re-identification risk management measures, etc. Moreover, it would be appropriate to apply these regulations universally to the fields of medicine and medical information since they have as their basis the HIPAA Privacy Rule. However, Korea's medically-related legislations such as medical laws, national health insurance laws, etc. specify that medical information include not only biographical data and disease data but also organizational and administrative data. Therefore, personal health information must be limited in scope to information that has individual identifying characteristics, and not be applied to the whole domain of health information. In addition to a clear definition of health information and types of health information, a distinction must be made among general health information and personal information, information that identifies the individual by its own virtue and information that has become usable items by being combined with other information, and actual case studies. Furthermore, there must first be an agreement among the various agencies involved.

Acknowledgements. This paper is based in “Health and Welfare Policy Forum (2015), De-identification Policy of Personal Information and Tasks on Healthcare Big Data” first reported study.

References

1. Ministry of Government Administration and Home Affairs, "Korea is ranked No. 1 in public data openness index in a study by the OECD," Ministry of Government Administration and Home Affairs press release, (2015)
2. Kim, G.-H., Lee, J.-Y., Oh, A.-S.: “Convergence of healthcare IT and Big Data,” Korea Society of Computer and Information, Korea Computer and Information Science, Vol. 21, No. 2, pp.7-26, (2013)
3. Jointly by concerned ministries, "Big Data application expansion plan for the realization of a competent government," (2014)
4. WAJ Korea, "Medical data mining made elaborate, privacy infringement controversy," (2013)
5. Lee, I.-h.: "An interpretation theory regarding the concept of personal information as it relates to personal information protection laws," Korea Information Law Association, Vol. 19, No. 1, pp. 59~87, (2015)
6. Ministry of Security and Public Administration, "Guidelines on Personal Information Protection Regarding Personal Information Protection and the Opening up and Sharing of Public Information," (2013)
7. HHS OCR, "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule," (2012)
8. Ministry of Security and Public Administration, "Guidelines on Personal Information Protections Regarding the Opening up and Sharing of Public Information," (2013)
9. National Information Society Agency, "Personal Information De-identification Casebook for the Application of Big Data," Ministry of Science, ICT, and Future Planning, (2014).
10. Korea Communications Commission, "Guidelines for Big Data Personal Information Protection," (2014)
11. National Information Society Agency, "The Guide to Free Optimality Assessment for Personal Information De-identification," Ministry of Science, ICT, and Future Planning, (2015)
12. National Information Society Agency, "Guidebook on Techniques for Personal Information De-identification for the Application of Big Data," Ministry of Government Administration and Home Affairs, (2014)