

A Study on the Policy Trends for the Revitalization of Medical Big Data Industry

Hyejin Kim¹, Myongho Yi^{2*}

¹Chairman, Global Digital healthcare Policy Research Institute,

²Associate Professor, Sangmyung University

의료 빅데이터 산업 활성화를 위한 정책 동향 고찰

김혜진¹, 이명호^{2*}

¹글로벌 디지털 헬스케어 정책 연구소 대표, ²상명대학교, 부교수

Abstract Today's rapidly developing health technology is accumulating vast amounts of data through medical devices based on the Internet of Things in addition to data generated in hospitals. The collected data is a raw material that can create a variety of values, but our society lacks legal and institutional mechanisms to support medical Big Data. Therefore, in this study, we looked at four major factors that hinder the use of medical Big Data to find ways to enhance use of the Big Data based healthcare industry, and also derived implications for expanding domestic medical Big Data by identifying foreign policies and technological trends. As a result of the study, it was concluded that it is necessary to improve the regulatory system that satisfies the security and usability of healthcare Big Data as well as establish Big Data governance. For this, it is proposed to refer to the Big Data De-identification Guidelines adopted by the United States and the United Kingdom to reorganize the regulatory system. In the future, it is expected that it will be necessary to have a study that has measures of the conclusions and implications of this study and to supplement the institutional needs to play a positive role in the use of medical Big Data.

Key Words : Healthare Big Data, Data Governance, Big Data Security, De-identification Guideline, Healthcare Industry Promotion.

요약 오늘날 비약적으로 발달한 의료 기술(Health Technology)은 병원에서 생성되는 데이터 외에도 사물 인터넷 기반의 의료기기를 통해 방대한 양의 데이터를 추적하고 있다. 수집된 데이터는 다양한 가치를 창출할 수 있는 원료가 되지만 우리 사회에는 의료 빅데이터를 활용하는데 근거가 되는 법적·제도적 장치가 미비한 상태다. 이에 본 연구에서는 빅데이터 기반 의료 산업의 활성화 방안을 모색하기 위해 의료 빅데이터의 활용을 저해하는 4가지 주요 요인을 살펴보고 그 외 국외 정책 및 기술적 동향을 파악해 국내 의료 빅데이터 활성화를 위한 시사점을 도출하였다. 연구 결과 의료 빅데이터의 보안과 활용성 강화를 동시에 만족시키는 규제 체계의 개선 및 빅데이터 거버넌스의 구축이 필요하다는 결론이 도출되었으며 이를 위해 미국과 영국이 채택하고 있는 빅데이터 비식별화 가이드라인을 참고해 규제 체계를 정비할 것을 제안하였다. 향후 본 연구에서 도출한 결론 및 시사점의 구체적 활용 방안을 다룬 연구가 필요할 것으로 보이며 본 연구를 참고해 제도적 미비점을 보완한다면 의료 빅데이터를 유용하게 활용하는데 긍정적인 역할을 할 것으로 기대된다

주제어 : 의료 빅데이터, 데이터 거버넌스, 빅데이터 보안, 비식별화 가이드라인, 의료산업 활성화

*Corresponding Author : Yi Myongho(josephlee@smu.ac.kr)

Received February 18, 2020

Accepted April 20, 2020

Revised March 16, 2020

Published April 28, 2020

1. Introduction

In recent years, the foundation of Big Data is constantly expanding due to the development of ICT (Information and Communication Technology) that is represented by smart devices, SNS (Social Network Service), or IoT (internet of Things). Ginni Rometty from IBM has defined Big Data as oil of 21st century and it is also assessed, depending on whether or not it is utilized, as an asset that could create a new value that was never existed in the past [1]. In the health and medical industry, the potential value and utilization of Big Data is significant and it is focusing on the eliminating factors of the 4th industry [2]. The liquidity expansion of opening for medical Big Data is forecasted to change health and medical industry[3]. Hence, UK is integrating dispersed data to support linking and analyzing the data and established to operate HSCIC (Health & Social Care Information Center) to safely protect personal medical records. This is a meaningful implication that we urgently need to improve to invigorate medical Big Data industry [4].

Currently, medical Big Data in Korea, unlike the cases of other countries such as the UK and other developed countries abroad, the connection between data is not smooth and sometimes impossible. Use of medical Big Data is under-utilized for various reasons, and this can be said to be a major obstacle to the expansion of the medical Big Data industry as the nature of Big Data it increases in value when each data is connected. The main reason for the difficulty in connecting data is excessive regulation of privacy policy and to solve this problem, the 'Data 3 Act', which is the legal basis for the use of Big Data in Korea, has been proposed to the National Assembly, but it is currently pending due to the protection of personal information [5]. On the other hand, major countries around the world, such as the United States, Europe, Japan,

and China have adopt policies that focus on the balance of regulation and support to encourage the Big Data industry while preventing the abuse of personal information by providing information for predictive medical diagnosis with the use of AI technology [6], reducing treatment time and contributing to the development of new drugs [7]. Therefore, through the expansion of the domestic medical Big Data industry, we need to look into the regulations and support measures of major foreign countries. In addition, we will look into the main factors that are currently preventing the expansion of the medical Big Data in Korea and examine the requirements of integrated governance on how to securely utilize medical Big Data [6].

2. Theoretical Background

Value of the Big Data in recent years are not merely simple creation of services that were not available in the past. Experts see that when the Big Data is efficiently utilized, when it can forecast potential risks and when if we can make strategy to respond to those risks, it is truly valuable [8]. If we can apply this value to medical Big Data, based on personal medical information, it is true that we can perform customized medical service, prediction and prevention of diseases [9].

But abusing of medical Big Data could lead to violation of personal rights protected by constitution, it is time to aggressively search for safe utilization of medical Big Data[10]. There is a global movement on the effort to balance between utilization of medical Big Data and privacy protection in order to accomplish ethical use of medical Big Data[11]. On the other hand, there is no efforts on the safe usage of medical Big Data. The purpose of this study is to identify obstacles on invigorating domestic medical Big Data and compare and analyze policies on

medical Big Data of global major countries. This study will examine factors to obstruct utilization of medical Big Data in 4 prospective and propose measurement on effective and ethical use of the medical Big Data.

2.1 Need of utilization on medical Big Data under changes of medical environment due to 4th industry

2.1.1 Understanding on value creation for definition and utilization of the medical Big Data

In health and medical industry Big Data is basic prerequisite of precision medicine as a data of personal medical treatment obtained through self-tracking technology[12]. Precision medicine not only blood test, DNA test or interview but also daily data such as heart rate, blood oxygen saturation, temperature and blood pressure. This is because the purpose of precision medicine is to improve behaviors or habits to maintain good health by understanding and analyzing personal physiological response. Thus, we call the data used in precision medicine as life-logging or quantified self[13]. The reasons of the concerns on abusing medical Big Data and its harm is because it contains daily physiological information. But as a nature of the Big Data, when generic, biomedical and behavior information is connected, the completeness of the Big Data, compared to single data, is enhanced. This means it is imperative to draw agreement and participation of individuals towards opening and sharing of personal medical information in order to collaborate and utilize collected data [14]. In other words, it is required to establish systematic and safe system for security of privacy and data protection as well as public utilization of medical Big Data to balance protection and utilization. Under the objective of 'All the citizens must receive medical services regardless of economic situation', Since 1948 UK is operating NHS (National Health Service) based

on taxation and also operates Care.data program to improve the quality of medical service. Care.data program opens collected medical Big Data after DE identification process to various health and medical agencies [15] and UK government unfolds policy to induce citizens' participation and choice on opening and sharing of medical data [16]. In case of the United States, Office of the National Coordinator for Health Information Technology has established ONC-HIT Certification and detail progress stages that are commonly applied to collecting, analyzing and utilizing of medical data. When the target progress and the corresponding verification is satisfied, it is recognized as Meaningful Use and the service agency is provided with incentives [17] since 2009. On the other hand, Hermon and Williams[18] see that the Big Data can be used in administration, services, supporting on the decision of clinician, clinical information, individuals and consumers and Krumholz[19] proposed the Big Data must be utilized to draw development of study for improving success rate of the forecast, improving the performance and comparing the efficiency. Also Roski, Bo-Linn, and Andrews [20] assessed that the Big Data can contribute in personal diagnosis and treatment of detail risks and utilization of clinical determination system and it is also highly effective in the health analysis and prevention of population group.

Opening and sharing as well as active connection of data must be accomplished to improve effectiveness of the medical Big Data and on the other hand, privacy and data protection must be safely secured. In other words, value creation for the utilization of the medical Big Data can only be triggered when each agency shares and connects the data and the decision makings based on the result is expected to put the medical service one step above[14].

2.1.2 Business trend and prospective of the medical Big Data

In recent years, mobile related technology shows significant improvement and smart phones are extensively used to increase the use of smart health care Apps based on IoT. This influenced significant increase of the medical Big Data in real time and as the development of AI recognition and process of atypical data, it accelerates growth of the medical Big Data based service industry. The US based General Electric announced investment and plans for technology development in the medical analysis and diagnosis technology under the motto of "Make the Invisible Visible" to become leader in medical innovation by utilizing AI and cutting edge Big Data fusion technology. Also the US based venture company Enlitic currently applied the medical Big Data analysis and diagnosis technology to deep learning technology to provide information for forecasted medical diagnosis by adopting machine learning based AI to medical diagnosis video[6]. In case of China, a platform to connect doctors and patients through application has been established and the system enables simple treatments and prescriptions through online. Through this system, upon agreement of the patients, the doctors can receive medical history and prescription history of the patients to be helpful in saving time and cost of the treatment and development of new medicines[7]. On the other hand, as the development of ICT technology, the medical Big Data industry will significantly fill in the gaps of treatment cost and level of the medical team[21]. In order to accomplish this, collection and utilization of the medical Big Data must be prerequisite and it is forecasted that if the collected Data is utilized with AI, the level of medical treatment will be improved to above average and at the same time medical treatment cost will be lowered[22]. In case of Korea, 'Digital Healthcare Partners' is established to

professionally incubate startups in digital healthcare industry who utilize the Big Data in 2016. Lunit is one of the leading startups whose technology is globally recognized. Looking at the trend of current medical Big Data based industry, mostly the businesses utilizing accumulated data from hospital are growing and domestic general hospitals established in-house Big Data Center to search for utilization of the medical Big Data [23].

2.1.3 Domestic utilization of the medical Big Data and its limitations

When we look into the domestic utilization of the medical Big Data, it is obvious to find that the level of utilization is significantly low compared to the quality and quantity of accumulated Data. Rapid development of IT industry enabled massive establishment of various public and private data in Korea. Especially, in case of Medical Insurance Data that contains medical information of the whole population and medical facilities, the scale and diversity of the data is assessed to be the best in the world (1.57 trillion cases, if piled in papers the height is 24,841 times higher than Mount Everest) but it has never been meaningfully utilized due to absence of policy and legal foundations that enable the integration and connection among the agencies [24]. Recently many countries in the world are utilizing generic data to analyze fundamental mechanism of the disease but in the aspect of completeness of the information by the scale and connection of the data among the agencies, Korea has competitive disadvantage. The reason is because it is unable to integrate and connect data from different agencies and without foundation of integrated Big Data, the value of utilizing single data is highly limited[6]. Currently in the evaluation of policy and competitiveness on the health and medical industry, Korea is in mid to low ranking and in 2016, US ITIF (Information Technology &

Innovation Foundation) has rank Korea on 29th in the report on 'Ranking of biomedical policies contributing to global innovation of biomedicine industry among 56 countries'. Considering Taiwan and Singapore are ranked on 3rd and 4th, it shows the competitiveness of Korea is significantly low[2]. The biggest reason for the limited utilization of the medical Big Data is the conflict on the legal structure between collection and protection of the personal data. Also there are lack of detailed provisions on the utilization of the medical Big Data and application of strict privacy protection law. In addition, positive regulation system is another reason that prevents services utilizing the Big Data[22].

2.2 Main obstacles on invigoration of the medical Big Data

2.2.1 Integrated restructuring of the laws and system of the medical Big Data

Currently the medical Big Data is seeking to be one pillar of developing national industrial strategy but meaningful accomplishments of the Big Data analysis along with its benefits are quite incomplete. This is because of the absence of systematic national strategy of the Big Data to be publically utilized and as a result data sharing among the agencies is only possible within the legal limit. Especially there is no systematic and physical foundation to utilize the high value added clinical data accumulated in nongovernment sector[14]. There are 4 stages that current legal structure is conflicting with utilization of the medical Big Data. First there could be violation in the laws including personal privacy protection, public privacy protection and medical law in collecting and storage · analysis stage. Next is Data value adding stage that there is conflict between remote medical treatment law and medical data sharing law. Lastly for the Data optimization stage, there are factors that hamper utilization of the medical Big Data such as

medical equipment law, national health protection law and integrated medical equipment certification law[21]. On the other hand, the current personal privacy protection law in the medical Big Data collection stage, there is prerequisite of unexpected new legal relations. As Yoon[9] claimed, even though the primary collector of the medical Big Data is public sector, as provider of the medical forecast service will be private sector, there is a new legal relations. In other words, when public sectors collect personal medical information, there is no requirement to get approval from each individuals but when the data that is collected by public sector is transferred to private sector, it is not dualistic legal relations that is prerequisite for personal privacy protection law any more but triple legal relations resulting removal of requirement of current personal privacy protection law. Also when private sector is provided with the medical Big Data from public sector, if there is relative grounds in the current law, there is no problem even if there is no agreement from the individuals. Even though there are numbers of different opinions, the revision of relevant law and system is far from execution and this is the biggest factor to hamper invigoration of the Big Data utilization [25].

2.2.2 Operating structure of the medical Big Data utilization by the government

The current health and medical system governed by the government targets establishment of health and medical infrastructure focused on prevention but what is current shared is limited to health service status provided by local government and location information. In order to overcome the limitations and provide customized health and medical information service for potential patients, it is possible when the medical Big Data that are stored in public agencies is efficiently utilized [26–28] Also in order to have smooth exchange between public agencies and

medical institutions, it is required to establish an intermediary system to analyze and deliver the medical Big Data stored in the public agencies to medical institutions. There are services provided by the public agencies and related companies but the utilization is very insignificant[29]. On the other hand, in the new born intensive care unit, status of new born is monitored in real time through the Big Data streaming technology. This system enables 24 hours faster recognition of any disease of the new born than the medical team. In order to utilize this technology, it is most important to analyze time series data and it is required to correct atypical data by using linear or polynomial regression analysis. The hospital is then required to store both original data and corrected data and this means complete establishment of governance for data protection and quality control[7]. In case of overseas, digitalization of the medical system such as EMR is only about 80% but we have more than 90% of the whole medical institutions have adopted EMR for digitalization. But compared to overseas case of 40% data exchange among the medical institutions, our treatment information data exchange among the medical institutions is less than 1% to show significant difference[30]. In case of foreign countries, medical systems such as EMR(Electronic Medical Record) was established in the mid-2000s and the digitalization ratio was only about 80% while more than 90% of all medical institutions in Korea have established EMR (Electronic Medical Record) in the mid-1990s resulting the highest level in the world in terms of digitalization. However, compared to overseas cases where the information exchange rate among medical institutions is close to 40%, the exchange of medical information among medical institutions in Korea is currently less than 1% showing significant difference [31]. There are various reasons on this significant difference but one of the fundamental reason is lack of deliberation

structure and preparation of operating structure that will support efficient data exchange of the medical Big Data. Currently the medical Big Data is managed and operated by not only Ministry of Health and Welfare but by Ministry of Employee and Labor, Ministry of Trade, Industry and Energy, Ministry of Science and ICT, Korea Centers for Disease Control and Prevention, Ministry of Food and Drug Safety, and National Statistics. Also significant amount of data is accumulated public agencies such as in National Health Insurance Corporation, and Health Insurance Review & Assessment Service. Additionally, University hospitals are individually operating Big Data collected by EMR. Diversity and scale of the Data is highest in standard but due to lack of main operator to connect and integrate the Data and lack of system for decision making prevent meaningful utilization of the Big Data[24]. For instance, Health Insurance Review & Assessment Service provide opening of their Big Data to privates and quality management of the Big Data is constantly executed but the Big Data from a single institution has limitation in leading invigoration of the medical industry based on the Data. Therefore, there is a proposal to expand the utilization of the Big Data that is stored in other public institutions such as the National Weather Service and the National Statistics but it requires the most sophisticated technology to integrate the medical Big Data and other Big Data [32]. For the invigoration of the medical Big Data, it is important to have a smooth Data exchange among the institutions but Ministry of Health and Welfare, Ministry of Trade, Industry and Energy, and Ministry of Science and ICT have different stance toward the utilization. It is also proposed to establish governance to build consistent policy for the utilization of the medical Big Data as well as requirement of establishing committee to integrate each public institutions and privates[33]. On the other hand, the Ministry of

Health and Welfare has established the Medical Big Data Policy Review Committee composed of the experts in academic, medical and municipal organizations to continue discussion on the utilization of the medical Big Data and the [34] Big Data platform has been established but the level of concerns of the municipal organization in the Medical Big Data Policy Review Committee is high[35]. This is concern of sensitive private information could be illegally abused and in order to overcome this concern and safe utilization of the social medical Big Data it is required to prepare the review and approval process on the using purpose and process method of the Data. In other words, It is needed to have the review and approval process on public use of the Data and secure safety during collection and sharing of the personal medical information[36]. Therefore it is imperative to establish the exchange platform to effectively utilize the medical Big Data and when the national medical Big Data sharing platform is established, it is possible to provide the medical service customized for the consumers[14].

2.2.3 Improve quality of the medical Big Data

Quality of the Big Data impacts derived forecast results as a result of the analysis and also reflected to the decision making, it is a most important factor in utilization of the Data. Therefore it is required to have systematic approach for the Data quality and the factors composing the Data quality are as follow[24]. First is superiority of the Data that is intrinsic quality such as accuracy, objectivity and authenticity. In more detail, it means the level of accurate containing of the Data towards the objective without any distortion. Second is accessibility quality that indicates level of environmental superiority for the Data access and it is asserted that the Data access should be executed under the condition of privacy protection and the Data security. Third is

contextual quality such as completeness and intemporality of the Data, it indicates suitability of users' targeted situation and enhances utilization value when the Data is used in timely manner. Forth is representational quality that is clarity of the Data such as analysis, conciseness, manageability, and consistency and the expression used for the Data connection or integration must be consistent and distinct as well as easy to understand [18] On the other hand, according to announcement of Health Insurance Review & Assessment Service (hereafter HIRAS), the Big Data stored in the HIRAS are graded as platinum class by Korea Data Agency but it is highly limited in terms of securing diversity to be internally restricted for the Data quality improvement[32].

2.2.4 Expandability of the Big Data based comparative effectiveness research

There is a view that utilization of the Big Data is to get economic benefits but we can find out that initial Big Data business started in the US from Comparative Effectiveness Research (CER) by National Health Institute and Agency for Health Research & Quality. In order to utilize the Big Data in clinical site, the results can be different by the level of expansion of the results derived by CER and level of how effectively utilized. In case of the US, in the initial stage when the Big Data utilization was limited, it was induced to share the Data by establishing the networks by multiple institutions through research of various topics. Similarly, we also need to induce establishment of the network to constantly connect and integrate dispersed Data[18] Big Data in the 4th industrial era is a strong future asset and especially the medical Bit Data has high potential for utilization as it can draw resolution to various social problems derived from extended life cycle of the modern people [37].

The major countries are establishing national strategy to utilize the medical Big Data and the

utilization plan of the Big Data is being pursued from the multi-directions in UK to provide prevention focused medical service instead of treatment. The UK government is enhancing managing and supervising structure to use the medical Big Data within the purpose of public benefits and secure safe usage of the Data [38,39]. On the other hand, in case of the US, the US government allows collection and utilization of the medical Big Data by private institutions as well as impose responsibility of the management to protect the Big Data. At the same time, Office of the National Coordinator (ONC) is established under the Department of Health and Human Service to support the medical Big Data business of private institutions for public utilization of the Big Data [40].

Core responsibility of ONC is to support standardization and commercialization of the medical Big Data and it is expected to reduce excessive medical expenses of the US by introduction of new private medical insurance structure through utilization of the medical Big Data. This is more effective when the Data accumulated by HER is efficiently utilized and the US government is encouraging expansion of HER by incentivizing the institutions that complies with meaningful use of EHR. Besides, Biomedical Translation Research Information System (BTRIS) integrates the medical Big Data of institutions under NIH and provide the Data and analysis tool to support profit making of the private institutions through utilization of the Big Data [41]. Singapore has also established National Electronic Health Record (NEHR) to integrate the Data accumulated through HER and built inter-hospital EMR sharing system to confirm the personal medical Data without distinction between public or private medical institutions. Singapore has accepted Personal Medical Responsibility that automatically deposits 6% of monthly income to the medical savings account and each individuals are

responsible for all the medical services except for vaccinations and public health care. This medical system caused various problems as demands for the medical attention increases due to the aging society. As the need for integrated medical raised, Singaporean Government is trying for the establishment of national management structure to resolve the problem of increased demand of medical attention and currently there is pilot integration towards partial chronic diseases [42].

2.3 Methods for the safe utilization of the medical Big Data

2.3.1 Increasing social embracing of privacy protection law for utilization of the medical Big Data and overseas cases

In Recommendation of OECD Council on health Data Governance, it is suggested to adopt process of asking for consent after enough explanation of the purpose of the medical Big Data utilization and this is because the damage is significantly high if the essentially sensitive medical Data is abused [43] [44]. Also OECD has pointed out that when the purpose of the medical Big Data utilization does not aligns with the public interest, there should be alternative plan or exceptions as well as separate protection measurement. It is also needed in Korea to obtain consent from the individuals when their medical Data is utilized and clear instruction of how the Data is utilized, how it is processed and protected must be provided[31]. On the other hand, in 2012, the UK Department of Health and Social Care asserted the importance of the medical Big Data and announced 'The Power of Information: Putting all of us in control of the health and care information we need'. In this announcement, the UK Department of Health and Social Care guaranteed that all the Data of the UK citizen will be protected and prepare sharing structure and standard that the collection and analysis of the medical Big Data will return

to benefits of the citizens. Besides, through digital care support, the UK Department of Health and Social Care announced to execute service to coincide with demands of the country and also support the most optimized technology and the utilization of the Data. Also the Department has announced policy for expanded accessibility of the health and medical Data[18] Privacy protection has both meanings of information security and information protection. Information security is to maintain the security of information from man-made leak or distortion while information protection is respond to protect information from the natural disasters or mistakes[24]. In order to protect information, confidentiality, integrity and availability must be guaranteed and confidentiality is guaranteed when information is accessed by approved personnel. Integrity is guaranteed when information is safely processed, not artificially manipulated such as changing or deleting the Data. Availability is guaranteed when the personnel who is approved to access the Data can operate the system when needed. In other words, availability is guaranteed when authorized personnel can freely access the Data and related assets when needed[45]. On the other hand, personal Data literally contains various Data about the person such as name and social security number to identify certain individual. In case of Korea, as each individuals are identified by the social security number, exchanges of the personal information can be made easily and due to development of information technology, personal Data can be easily collected from online which significantly increases the concern on invasion of privacy[46]. Also as there are many cases that personal Data is closely related to basic right guaranteed by the constitution, it must be deliberately handled[47] Personal medical Data contains doctor's diagnosis and prescriptions as well as all the information researched and analyzed by the information. In

other words, all the information collected during the treatment is regarded as personal medical Data such as diagnosis, treatments, results, all the medical records and financial records [48]. The medical Big Data, as the nature of the Data, it contains significantly sensitive and personal information such as genetic features, family relations, drug record, and daily habits. As this personal Data is required to medical team to treat the patient and could be utilized in research purposes or legal information, social members from various areas are utilizing the personal Data. In other words, discussion on invasion of privacy protection on the medical Big Data is not limited to information just stored in the information system[49]. Other than weak points related to information system, the Big Data has other weak points during collection, storage and processing such as pattern analysis, excessive information collecting and transfer of the Data. So the security related to the Big Data requires not only technical aspect but requires security by steps of collection and analysis, utilization and disposal and followings are cases of overseas countries [50]. Similar to our current situation, until mid-90, the US government could not completely protect the personal medical Data. In 1974, Privacy Act was established in the US but personal medical Data protection was not properly enforced resulting various side effects. For example, only 5% of the medical Data was utilized from the veteran's hospital and Viagra promotion mails were sent to those people who were treated for sexually transmitted diseases. The proper protection policy for the medical Data in the US started in 2003 and after 1 year grace period, from 2004, all the medical institutions followed the standard personal medical Data protection policy. On the other hand, in 1996, HIPPA law was established to be legal foundation for the protection of all the medical Data of US citizens[51]. The range of medical Data that is under the protection of

HIPPA (Health Insurance Portability and Accountability Act) is not only digitally written and saved but all the information delivered by documents or orally spoken. But development of Information Technology all the medical Data is digitally saved in the medical institutions and this means significant high risk of information breach. Therefore importance of the personal Data security and protection has become a social issue and brought up the need for establishment of the system to guarantee confidentiality, integrity and availability of the Data. The legal definition for privacy protection in the US is still developing and Professor Solove pointed out that accurate range and definition of privacy must be defined and continue to the discussion of personal Data protection[52]. There are many spectrum of audiences who can access personal medical Data including medical team, hospital administrator, insurance company or employer but there is the risk of illegal transfer or abuse of the personal medical Data is not completely protected. For instance, there could be a case when the patient never agreed but the patient's medical Data could be transferred to 3rd party. Therefore this has raised a necessity to have more severe punishment in the US for 3rd party privacy invasion through the personal medical Data and HIPAA has established regulation to support is idea. Also as the protection of the personal medical Data is very important factor between patient and doctor, there are higher concerns of risks of illegal invasion compared to other personal Data[53].

2.3.2 Establishment of a new governance for the national utilization of the Big Data

CMS has established ONC as an institution to expand interoperability in order to establish national Data sharing environment as well as inducing the medical Big Data exchange and integration and invigorate the Big Data research. CMS and ONC are collaborating for meaningful

use business of HER and this is regarded as an model case for establishment of national governance in the utilization of the Big Data[14]. The utilization value of the medical Big Data is determined by the Data integration of Data including various factors to determine personal health and proper timing of using the Data. In other words, it is very important to guarantee proper timing of using the medical Big Data and in the UK, in order to utilize required Data in timely manner, Care Data program was established [54],[4] Care Data program is operated by joint hosting of NHS and HSCIC and it is defined as information collecting system that collects the Big Data to one location. With the operation of Care Data, there has been changes on collection and analysis method of clinical Data in the UK and based on Health and Social Care Act 2012, HSCIC can request for personal confidential Data of the patient under special circumstance. Also all the UK citizens have right to opt-out their own and their family's treatment records and transfer only other personal Data to HSCIC[24].

Currently, Korean medical Data is dispersed in many public institutions such as National Health Insurance Corporation, Health Insurance Review and Assessment Service or National Cancer Center and these medical institutions have established their own Big Data platforms. Also as the personal Data is strictly managed, even the Data is collected from individual institution, it is realistically impossible to conduct Data exchanged analysis. Therefore it is necessary to establish the user portal and the Big Data platform in order to boost convenience of the medical Big Data users [55].

2.4 Big Data utilization platform and regulation trend of major countries

2.4.1 Policy of each country for the personal information protection and step by step conditions to protect the medical Big Data

Overseas countries are building the Big Data

platforms and providing services across the public and private sectors, including health care. For example the US ONC(Office of the National Coordinator) collects and organizes various medical Big Data to provide in standard form to act as an intermediate between platform provider and user[56].

In the US, there are policies that vary the approach depending on the extent and extent of the use of personal Data, and federal law shows that there are no regulations that impair privacy. In this regard, it is stated in the Federal Data Protection Act that Germany has a responsibility for the transfer of personal information to the transfer agency, and the UK and Australia have documented regulations for the collection and processing of sensitive information[57]. In the UK, HSCIC is in charge of the entire process of exchanging and utilizing the medical Big Data between institutions, and uses standardized open exchange technology when connecting the medical Big Data. In addition, an administrative information sharing center was established to establish a system for administrative information [58] As such, countries have introduced a variety of policies for the safe use of the medical Big Data, and looking at overseas cases, it is necessary to prepare for the risks that may occur in the entire process of collecting and processing the medical Big Data. Accordingly, information protection and security measures tailored to all stages of data collection, analysis, utilization, and disposal. First of all, in the Data collection stage, direct and indirect collections should be distinguished. In the former case, the Data owner's consent must be obtained.

In addition, the purpose of collecting Data should not be in violation of the owner's policy[59]. In the Data storage stage, it is most important to encrypt Data to maintain safety even when Data is leaked. Accordingly, in the analysis stage, the encrypted Data should be analyzed and in this time, de-identification

process should be performed on the Data that can identify individuals. In addition, it is required to manage the service provision process to prevent unauthorized leakage and invasion of personal Data in the utilization stage [60-62]. In the Data disposal stage, the Data should be destroyed immediately when the purpose of using the Data is achieved. If the Data is not destroyed, it is a serious violation of the privacy law [63,64] Meanwhile, European countries, including the UK and Germany, are currently pursuing amendments to the Personal Information Protection Act based on the EU's recommendations on privacy protection. In the case of Germany, in the past, the Personal Information Protection Law, which was established by dividing the public and private sectors, was proposed in the EU Recommendations. As a result, it was revised to the integrated privacy protection law[65].

3. Discussions

These days, Big Data is expanding its foundation with the development of ICT technology represented by smart devices, SNS, Internet of Things, etc., and is regarded as a powerful future asset to create new value. In the field of health care, Big Data has great potential value and utility, and the use of Data is becoming more important as the medical paradigm of the 4th industry is changed from evidence-based medicine to precision medicine and customized medicine. On the other hand, currently the utilization of the medical Big Data in Korea is very limited because the connection between the Big Data is virtually impossible. Different views on the laws and systems surrounding the use of the medical Big Data have been raised, but the legal and institutional systems that encompass them is still far away and discussions and legal foundation need to be

established as soon as possible. On the other hand, in the case of overseas, the digitalization of medical systems such as EMR is only 80%, while in Korea, more than 90% of all medical institutions have reached the world's highest level in adopting EMR. However, compared to foreign cases where the exchange rate between medical institutions is close to 40%, the exchange of medical information between medical institutions in Korea is only less than 1%, which is very different. One of the fundamental reasons for this difference is the lack of a deliberation structure and operation system supporting the efficient exchange of the medical Big Data. In addition, due to the difficulty in exchange Data distributed among institutions, significant restrictions are placed on securing diversity, which implies inherent limitations in improving data quality. Currently, even the "Data 3 Act," which is the legal basis for utilizing Big Data in various fields, such as health care, is being severely restricted in the use of medical big data as legislation is delayed. In this regard, the Big Data related policies of the world's major countries, such as the US, the UK, Germany, and Japan, suggest that we have great implications for us. The case studies of major countries around the world show that the focus is on balancing protection and use of the medical Big Data. This shows what part of Korea, where excessively strong personal information protection is being applied, should have a policy for Big Data utilization. As it can be seen in the cases of overseas countries, when the Data accumulated in each institution is connected, various studies that could not be done by the data of individual institutions are possible, and the completeness becomes higher when various information is connected due to the characteristics of the medical Data. There is a need for an improved regulatory framework that addresses the need for the Data security to be secured through regulation, but focuses on

enhancing the use of the medical Big Data. In this regard, this study examines the regulatory trends of the medical Big Data in major countries of the world and draws implications applicable to Korea. First of all, EU countries such as the UK, Germany, and other countries have been amending their privacy laws to comply with the EU's recommended privacy guidelines. In particular, Germany, which used to separate the public and private sectors in the past, has also revised the Privacy Act, which integrates the private and public sectors. In the US, there are policies that vary the approach depending on the extent of the use of personal Data, and federal law shows that there is no regulation that impairs privacy. In this regard, it is stated in the Federal Data Protection Law that Germany has a responsibility for the transfer of personal data to the authorities, and the UK and Australia have documented regulations for the collection and processing of sensitive Data. Meanwhile, in Korea, the 'Data 3 Act', that is the legal basis for utilizing the Big Data in various areas, including the medical Big Data, has been proposed, but it is currently pending in the National Assembly. It is argued that the information handled can be easily provided to others. In addition, concerns about conflicts with the current medical law and the National Health Protection Act also act as obstacles to the legislation of the 'Data 3 Act'. In addition, the lack of an operating system for the medical Big Data utilization, the limitation of improving the quality of the medical Big Data, and the lack of scalability of the Big Data-based comparative effectiveness research were identified as major factors that hinder the invigoration of the medical Big Data. Therefore, this study proposes the de-identification guideline of the medical Big Data adopted by the US and the UK for the protection of personal Data when using the medical Big Data. First of all, in the US, the principle of self-regulation is established when dealing with non-identified

personal Data. The self-regulation principle is the principle that the user is responsible for re-identification of the items and scope of de-identification of personal Data according to the purpose of analysis. In the US, the rule of self-regulation is adopted, with the emphasis on the management to ensure prevention of re-identification. In the UK, in November 2012, the Anonymization Convention was announced in accordance with the Information Protection Act, and anonymization was conducted to protect the privacy of individuals and to ensure the usefulness of using the Big Data. As such, both the US and the UK have adopted policies that autonomously determine items on personal identification, and have subsequently placed greater importance on follow-up to prevent the risk of re-identification. In this regard, Korea has also demonstrated the possibility that medical institutions such as the National Cancer Center, National Health Insurance Corporation, and Asan Medical Center have developed their own de-identification technology to satisfy the need for the utilization and protection of the medical Big Data. After supporting the required areas with reference to the cases of each country and domestic institutions, the 'Data 3 Act', which is the legal basis for utilizing big data, will have a positive effect.

4. Conclusion

In the 4th industrial era, the modern medical paradigm is changing from evidence-based medicine to precision medicine and personalized medicine and accordingly the importance of utilizing Big Data is growing as a large amount of Big Data is generated and accumulated in medical devices based on hospitals and information and communications technologies (ICT). The medical big data collected through various channels has strong potential to create

new values that were never previously available but even today, even though we have the world's best medical Big Data, we are not able to utilize it meaningfully. Therefore, in this study, as a result of examining the main factors that hinder the effective use of medical Big Data, it is very difficult to connect the Big Data accumulated in individual institutions under the current legal systems and reached the conclusion that we should build a systematic and safe system for protecting personal information while utilizing the medical Big Data. This can be found in policies adopted by major foreign countries to balance the protection and utilization of medical Big Data. To satisfy the security and enhance usability of medical Big Data at the same time, first step is to improve the current regulatory system to secure medical Big Data and build Big Data governance for the safe public utilization. In addition, this study proposed to improve the regulatory system by referring to the guidelines for de-identification of Big Data adopted by the United States and the United Kingdom for the safe protection of personal information stored in Big Data when improving the regulatory system. In the future, based on the results and implications derived from this study, if the institutional defects are redeemed, it is expected that it will have a positive effect on the useful utilization of the medical Big Data.

REFERENCES

- [1] S. W. Park. (2018). *Big Data Era and Data Integration*. http://m.kisdi.re.kr/mobile/repo/com_view.m?key1=14253&selectPage=
- [2] M. H. Ryu. (2017). *Suggestion to promote Bio Health Industry*. http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07113234&language=ko_KR
- [3] McKinsey & company. (2013). *The 'big data' revolution in health care*. 1-2.
- [4] Y. H. Lee. (2015). *The Services Strategies of the UK's Health and Social Care Information Centre*. Health

- and welfare policy forum.
- [5] J. S. Baek. (2019). *Data 3 Act halted by Chae Eebae Committee. Money Today.* <https://news.mt.co.kr/mtview.php?no=2019112915367649221>
- [6] I. J. Kim, & S. J. Kim. (2017). *A Study to Create an ISP That Taps into Convergent Medical Big Data.* Research Institute for Gangwon.
- [7] S. B. Ko. (2019). *A study on the harmonization of the use and protection of personal data in the age of big data.* Master dissertation, Konkuk, Seoul.
- [8] H. K. Ka. (2014). *(An)empirical study on the influencing factors for big data intended adoption.* Doctoral dissertation, Chung-Ang University, Seoul
- [9] S. J. Yoon. (2015). *Conflicts of Personal Information Protection and advantage of Big Data, the problem and the legislative policy issues. Focus on the Health & Medical Big Data.* Chung-Ang Law Association, 17(1), 7-47.
- [10] M. Y. Kim, & J. H. Lee (2016). *A Strategy for the Introduction of an Opt-out Policy of the Personal Data Protection Act based on Security Enhancements. The Journal of Korean Association of Security and Safety, 11(2), 113-139.*
- [11] H. S. Kim, & D. Y. Park. (2014). *A study on the harmonization of the use and protection of personal data in the age of big data.*
- [12] T. Sharon. (2017). *Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. Philosophy & Technology, 30(1), 93-121.*
- [13] M. Swan. (2012). *Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. Journal of Personalized Medicine, 2(3), 93-118.*
- [14] H. J. Kang. (2016). *National-Level Use of Health Care Big Data and Its Policy Implications. Health and welfare policy forum, 55-71.*
- [15] Korea Health Industry Development Institute. (2015). *2015-Medical and IT Infra Implementation.* <http://khiss.go.kr/board/view?pageNum=15&rowCnt=10&no1=317&linkId=64017&menuId=MENU00308&schType=0&schText=&boardStyle=&categoryId=&content=&schStartChar=&schEndChar=&country=>
- [16] A simple guide to Care.data (2014). <https://www.wired.co.uk/article/a-simple-guide-to-care-data>
- [17] S. J. Park, H. S. Ryu & J. E. Ahn. (2014). *New ICT Strategy for health and medicine Information Policy Research.* https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=66361&bcIdx=13469
- [18] R. Hermon & P. Williams. (2014). *Big data in healthcare: What is it used for?* Paper presented at the Australian eHealth Informatics and Security Conference.
- [19] H. M. Krumholz. (2014). *Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Affairs, 33(7), 1163-1170.*
- [20] J. Roski, G. B. L.G, & T.A. Andrews. (2014). *Creating value in health care through big data: opportunities and policy implications. Health Affairs, 33(7), 1115-1122.*
- [21] S. C. Hong. (2017). *Promotion of Healthcare Industry and Social Impact Korea Insurance Research Institute.* <http://inthenews3.mediaon.co.kr/mobile/article.html?no=7802>
- [22] M. J. Jung, & M. H. Lee. (2017). *Medical Big Data Activation Policy for Health Care Industry Development. Seoul.*
- [23] D. E. Lee & S. K. Kim. (2018). *New Trends in Digital Healthcare* <https://www.stepi.re.kr/module/publishDownFile.jsp?cateCd=A0502&ntNo=143>
- [24] H. J. Kang (2015). *Strategic Planning for promotion of healthcare big data. Korea Korea Institute for Health and Affairs*
- [25] S. Y. Lee. (2014). *A Comparative Law study on the Legislation of Big Data.* <http://www.ndsl.kr/ndsl/commons/util/ndslOriginalView.do?dbt=TRKO&cn=TRKO201500001396&rn=&url=&pageCode=PG18>
- [26] J. Manyika, M. Chui, & B. Brown. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
- [27] B. J. Park. (2013). *Scientific Evaluation Methods for Health Technology. Journal of Health Technology Assess, 1(1), 9-15.*
- [28] D. S. Heo. (2009). *Evidence-based Healthcare in Korea. Journal of Korean Medical Association(52), 934-935.*
- [29] H. J. Lee, & Y. H. Kim. (2017). *Open Platform for Improvement of e-Health Accessibility. Journal of Digital Contents Society, 18(7), 1341-1346.*
- [30] W. S. Cho, (2017). *Role of Governance to promote big data utilization.* https://www.kab.co.kr/kab/home/common/download_cnt.jsp?sMenuIdx=036015015000035021&sBoardIdx=045005125001039020127&sFileIdx=045005125001
- [31] H. J. Byun, (2018). *Issues in medical big data platform projects. Monthly Social Welfare, 23-30.*
- [32] T. J. Woo. (2017). *Customized Service using Medical Big Data Open Platform.* <https://www.kihasa.re.kr/common/filedown.do?seq=36589>

- [33] M. H. Oh. (2019). *Use of Healthcare Big Data and Government Policy* Hospital News. <http://www.khanews.com/news/articleView.html?idxno=144061>
- [34] T. W. Kim. (2019). *Press release on Public Medical Big Data* <http://www.kmedinfo.co.kr/news/articleView.html?idxno=57246>
- [35] K. H. Ahn. (2019). *R&D and Market Big Jump* http://www.dt.co.kr/contents.html?article_no=2019112702101431650002
- [36] E. W. Lee, J. H. Lim, B. I. Oh, & Y. K. Jang. (2017). *Research on how to support on use of data integration. Personal Privacy Committee.* http://www.prism.go.kr/homepage/entire/retrieveEntireDetail.do?research_id=1079930-201700009
- [37] J. Bresnick. (2014). *Healthcare analytics market to expand at 25% CAGR to 2019.* <http://healthitanalytics.com/news/>
- [38] E. D. Hwang. (2014). *Promotions and Issues in Medial Big Data.* HIRA, 8(6).
- [39] T. H. Lee. (2019). *Roles and Responsibilities NHS in UK National Knowledge Information System, 13(1), 76-82.*
- [40] ONC (2014). *Federal Health IT Strategic Plan 2015-2020.* <http://HealthIT.govdashboard.healthit.gov/>
- [41] Institute of Medicine (2009). *Leadership Commitments to Improve Value in Healthcare: Finding common ground: Workshop Summary. Washington (DC): National Academies Press (US)*
- [42] McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competition, and productivity*
- [43] S. H. Wang. (2018). *Guidelines for Medical Data Open Data Governance Korea Legislation Research Institute.*
- [44] NIA (2019). *Artificial intelligence in Society.*
- [45] S. H. Kim. (2002). *Information Security Management and Policy.* Seoul, Korea: Life and Power Press.
- [46] D. S. Kim & M. S. Kim (2006). *E-health and Medical Data Issues and Strategies Informatization policy, 13(4), 128-148.*
- [47] S. G. Hong (2004). *Personal Informaton and Policy conflict - NEIS case study. Information Security Review, 1(2), 27-46.*
- [48] D. C. Kang. (2004). *IT trends and Privacy*
- [49] Y. M. Chae, (2005). *eHealth Policy Renovation*
- [50] H. S. Kim & D. Y. Park (2014). *A study on the harmonization of the use and protection of personal data in the age of big data.*
- [51] B. M. Seo (2006). *A Study on Private Information and HIPAA in the USA.*
- [52] J. Solove. (2007). *'I've Got Nothing to Hide' and Other Misunderstandings of Privacy. San Diego Law Review, 44, 745-760.*
- [53] S. Volkert. (2000). *Telemedicine: Rx for the Future of Health Care. MTTLR (Michigan Telecommunications and Technology Law Review), 6(1), 147-214.*
- [54] Department of Health. (2012) *The power of information.* <http://www.gov.uk/>
- [55] S. Y. Oh, J. H. Park, & J. W. Kim. (2019). *Protect Citizen using Open Medical Big Data.* <http://www.healthinnews.co.kr/news/articleView.html?idxno=11645>
- [56] I. Y. Jung & W. M. Koo. (2018). *Data integration for healthcare Ecosystem. Trends and Issues(46), 1-38.*
- [57] Y. H. Bang, H. S. Lee., & I. H. Lee. (2014). *A Comparative Study of Regional Medical Information Protection Act and Privacy Act. The Journal of the Korea Contents Association, 14(11), 164-174.*
- [58] S. Captain (2016). *Obama's Precision Medicine Initiative Is The Ultimate Big Data Project.* <http://www.fastcompany.com/3057177/obamas-precision-medicine-initiative-is-the-ultimate-big-data-project>
- [59] J. H. Kim, D. H. Son, K. J. Lee, H. S. Shin, & S. A. Lee. (2017). *Development of Educational Information and Statistics Policy Tasks for the Big Data Age.*
- [60] Korea Internet & Security Agency. (2009). *Technical and Managerial Security Guideline for Personal Information* https://www.kisa.or.kr/public/laws/laws3_View.jsp?cPage=7&mode=view&p_No=259&b_No=259&d_No=102&ST=T&SV=
- [61] Korea Communications Commission (2011). *Guideline for technical and managerial security for personal informaton*
- [62] J. Y. Lee (2005). *Database security System Implementation,* Doctoral dissertation, Dongguk University, Seoul
- [63] S. Y. Son. (2013). *Big Data, Online Marketing and Privacy. KISDI Premium Report, 13(1), 1-26.*
- [64] J. S. Lee. (2013). *Technologies to protect privacy in big data. Internet & Security Focus(3), 79-104.*
- [65] P. O. Kwon. (2018). *EU's GDPR and Countermeasures* https://news.kotra.or.kr/site/kotranews/upload/report/s/20180504153154_1525393651686.pdf

김혜진(Hye-jin Kim)

[상위권]



- 2006년 2월 : Teachers College, Columbia University 교육심리학 (임상심리학) (석사)
- 2007년 5월 : Teachers College, Columbia University 장애인 재활학 (석사)
- 2009년 5월 : Teachers College, Columbia University 철학 (석사)
- 2009년 5월 : Teachers College, Columbia University 특수교육학 박사 (Ph.D)
- 2012년 8월 : 국회 입법자문위원
- 2012년 2월 : 국회 전 5급비서관(도가니 사건, 나주사건, 헬스케어, 아동교육과 성폭력 사건 등 입법)
- 2010년 ~ 2017년 : 단국대, 외래교수 (성폭력법, 입법과 정책, 성폭력 피해자 인권, 디지털 성범죄, 장애인 성교육, 특수교육학 개론)
- 2019년 6월 ~ 현재 : 글로벌 디지털 헬스케어 정책연구소 대표(Global digital Healthcare Policy Research Institute)
- 관심분야 : Digital Healthcare Policy, Digital Sexual Crime, healthcare big data for early prediction of dementia. Senior Wellness IT.
- E-Mail : hk2188@caa.columbia.edu

이명호(Myongho Yi)

[상위권]



- 1995년 12월 : 텍사스주립대학교 컴퓨터학과 (학사)
- 1996년 12월 : 텍사스주립대학교 정보학과 (석사)
- 2006년 5월 : 플로리다 주립대학교 정보학과 (박사)
- 2012년 9월 ~ 현재 : 상명대학교 문헌정보학과 부교수
- 관심분야 : 빅데이터 분석, 온톨로지
- E-Mail : josephlee@smu.ac.kr