INTERNATIONAL JOURNAL OF MATHEMATICS AND COMPUTERS IN SIMULATION
DOI: 10.46300/9102.2021.15.11 Volume 15, 2021

Methodology for Measuring the Impact of the Privacy Protection Law on the Use of Big Data

Oh Kyu-cheol
Information Security Group
Korea Internet & Security Agency, KISA
Seoul, Korea
kcoh@kisa.or.kr

Received: June 15, 2019. Revised: July 22, 2021. Accepted: September 3, 2021. Published: November 26, 2021.

The ICT revolution triggered by the emergence of smart devices, typically represented by the iPhone and the iPad, is migrating into the new domain of 'big data' after passing the turning point of 'SNS Life,' which is represented by Twitter and FaceBook among others. These developments have brought significant changes in all areas of politics, economy and culture. The stock prices of Apple, Samsung Electronics, FaceBook and Google fluctuate depending on who takes the hegemony in the changes. Meanwhile, such a reform of the ICT sector has generated some new undesirable side-effects, including online disclosure of personal information, malicious comments, Smishing or other forms of financial scams. As we cannot abandon either big data or privacy protection, it is critical to find a compromise. It seems both evident and self-explanatory that the use of big data, which is attributable to technical innovation, conflicts with privacy protection based on the idea that individuals should be allowed to determine the disclosure or not of their personal information. Yet, the problem here is that the discussion of countermeasures remains at the level of catching the wind with a net. Therefore, this paper intends to present a framework that can objectively verify what impact the enhanced legal regulation concerning privacy protection has on the use of big data as the first step in exploring a compromise between the use of big data and privacy protection.

Keywords— big data, privacy protection, trade-off point, impact measurement, 5V Framework, large volume attribute, adaptability, real-time attribute

I. Introduction

'Big data' have become a popular topic that represents the recent trend of the ICT industry. The incumbent Korean government has specified in its agenda that it will establish strategies for the nation's future development and respond to crises by using big data.[1] Big data have emerged as a core of ICT, as the OECD treats 'measurement of the economic value of big data' as one of its essential agenda.

On the other hand, privacy protection has definitely emerged as the most important issue in the information security service industry. In particular, privacy protection has emerged as the greatest headache of ICT workers of businesses in the wake of a local law court's ruling that SK Communications should pay damages for leaking personal information. On 15 February 2013, a panel of the Seoul Western District Court issued a ruling that SK Communications should pay 200,000 won per claimant in acknowledgement of its negligence in leaking personal information. Privacy protection is then an issue related to the survival of a business when we consider the fact that the number of potential victims of the privacy leakage case is estimated to reach 35 million. (For example, if 200,000 won is paid out to 35 million people, the total damages could amount to seven trillion won, which is at least 35 times the amount of SK Communications' annual sales revenue of 197 billion won). Of course, SK Communications appealed immediately after the court ruling. Meanwhile, as of June 2013, seven bills, all of which are designed to enhance the law on privacy protection [See Table 1.], have been either passed by, or are pending at, the National Assembly.

TABLE I. Bills for amendments that have either been passed or are pending at the National Assembly as of June 2013

Bills for amendments	Date of introduction	Key Contents
Bills (3) for amendments to the Act on the Promotion of the Information Network Service and the Protection of Information	22 August 2012	Heavier punishment of business operators that leak personal information
	12 February 2013	Reinforcement of users' right to determine disclosure of their personal information
	4 June 2013	Establishment of criteria or guidelines for settling disputes concerning infringements of personal information
Bills (4) for amendment to the Act on the Protection of Personal Information	27 August 2012	Measure for ensuring the security of personal information
	29 August 2012	Mitigation of requirements for collective litigation
	31 January 2013	Restriction of handling or processing of resident registration numbers
	12 April 2013	Reinforcement of requirements for notifying leaks of personal information

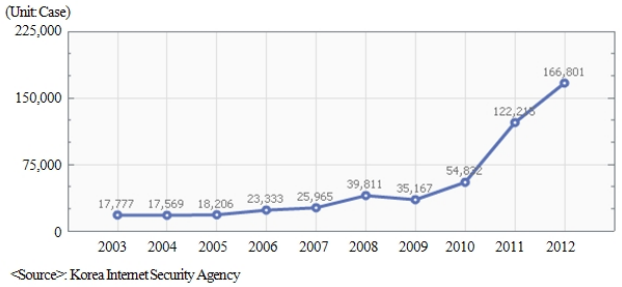
So far, the prevailing view has been that promotion of the ICT industry and the implementation of information security have existed in a trade-off relationship, in other words, that information security is a cost factor in the promotion of ICT business. Rightly or wrongly, such an understanding is linked

to the perceived conflict between the big data-based ICT revolution and the protection of personal information. As big data use is an unavoidable major trend in ICT services, while enhanced protection of personal information is also a demand of the times, the breakthrough should eventually come with the discovery of an optimum trade-off point.

II. Conflict between Big Data Technologies and Privacy Protection Law

Before exploring this trade-off point, let's further examine the meaning of the conflict arising between these two hot ICT issues – the use of big data and the protection of personal information. The social consensus concerning the importance of privacy protection has gradually increased, resulting in a qualitative enhancement of the related regulations and a quantitative increase in the number of complaints concerning infringements of personal information [Figure 1].

Fig. 1. Growth Trend of Counseling Cases concerning Privacy Infringement



<Source>: Korea Internet Security Agency

99.2% of individual Internet users agree that privacy protection is critical and believe that the most serious undesired side-effects of ICT are the infringement of personal information or privacy and the damages caused by illegal spasm [5]. Conversely, only 35.4% of the businesses that collect personal information allocate budget funds to the protection of personal information, showing a gap in their understanding when compared with users [6].

This gap in understanding has resulted in enhanced legal regulations for privacy protection. Let's review specific cases to find out what practical problems are caused by such enhancements of the regulations. The current Act on the Protection of Personal Information (Article 21) stipulates that personal information should be destroyed without delay when it is no longer required, including cases where the purpose of their processing or handling has been accomplished, and any personal information in either electronic or digital formats should be permanently deleted using a means that makes it impossible to retrieve the information in question (Article 16). Superficially, it is highly reasonable to delete personal information whose purpose of collection and use has been accomplished. However, this is not a simple problem from the perspective of the engineers who have to handle it.

For example, what specific level of deletion is meant by permanent deletion using an irrecoverable means? In litigation proceedings, it is extremely difficult to mount a legal defense by describing the technical difficulty. A literarily exact compliance with the provisions of the law is more important than anything else. Therefore, a penalty of 30 million won or less may be imposed if personal information that has been destroyed immediately when no longer required can be recovered from a hard disk, as it constitutes a breach of the law.[7]

According to the Guidelines on Degaussing or Disabling the Storage Media of Information Systems [8] presented by the National Intelligence Service in this regard, the storage media should be incinerated/destroyed/melted; stored data should be deleted using a degausser [9]; or the storage media should be fully formatted three times in order to delete personal information. However, is it really possible for business employees in charge of personal information to delete personal information using such methods 'without delay' in an actual work environment? In particular, if tens or hundreds of thousands of pieces of personal information are simultaneously processed in a second, should personal information be recorded on several thousand newly purchased hard disks each and every minute?

Furthermore, big data entails the constant accumulation, without deletion, of even unnecessary data for the analysis of possible future events. Article 16 or 21 of the Privacy Protection Act mentioned above may be interpreted to mean that acts of processing or handling big data including the slightest piece of personal information automatically constitutes a breach of the law in Korea.

To approach these problems more systematically than can hardly be asserted by anyone at present, common criteria are required to align the viewpoints of the engineers who handle big data and the legal experts who handle the accompanying legal issues.

Therefore, this paper presents a framework for calculating the value of big data, known as 5V serving, as such a common criterion.

III. Framework for Calculating Value Based on the Charateristics of Big Data

First, let's examine the characteristics of big data and what is meant by such characteristics in order to determine what impact the protection of personal information has on the use of big data.

Generally, the characteristics of big data that distinguish them from conventional data processing are expressed with 3Vs Gartner suggested. The '3Vs' represent Volume, Variety, and Velocity. New values created by big data are determined depending on how the 3Vs are defined and treated.

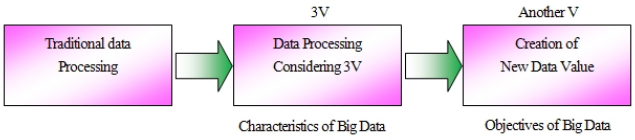
[Table 2] shows the characteristics of big data and specific examples of added values:

The processes or flow by which added values are created by big data may be seen when 1V (the objective of big data) is added to the 3Vs (the characteristics of big data). This process of added value creation (value chain) is expressed as 4V below.

TABLE II. CHARACTERISTICS OF BIG DATA AND TYPES OF ADDED VALUES

Classification	V	Examples
Characteristics of big data	Volume	o Large volume attribute of data o Large volume attribute of processing o Large-volume sharing or utilization (issue of data consistency)
	Variety	o Variety of data types o Variety of data processing methods o Variety of data sources o Variety of data times
	Velocity	o Real-time data acquisition o Real-time data processing o Real-time data analysis o Real-time data understanding (or representation)
Added values of big data	Value	o Cost saving o Advanced decision making o Rapid capture of customer preferences o Enhanced accuracy of future forecast o Discovery of meaningful patterns

Fig. 2. 4V: Value chain of big data



It should be remembered here that new values are not created by the 3Vs (the characteristics of big data); rather, the way in which 3V is defined and responded to is the factor that creates new added values.

For example, 'variety' is generally a factor that obstructs the creation of added values. It becomes more difficult to analyze data into consistent or 'meaningful' information or intelligence when the data that are to be processed by computers exhibit greater variety.

In other words, in order to create values with big data, you need to secure 'adaptability' enabling the processing of diverse data rather than by making efforts to enhance the variety of data.

These observations may be summarized as follows:

- Acquisition or securing of the large volume attribute in terms of volume;

- Acquisition or securing of adaptability in terms of variety;

- Acquisition or securing of the real-time attribute in terms of velocity (real-time acquisition and processing of data).

The co-relations among the large volume, adaptability and real-time attributes may be expressed more intuitively with a vector. F or example, when the large volume attribute (A) enhances adaptability (B) for a certain reason, the efficiency (C) of the added value newly created by them (A and B) will be higher.

Fig. 3. Cases where synergy is generated by elements creating added values

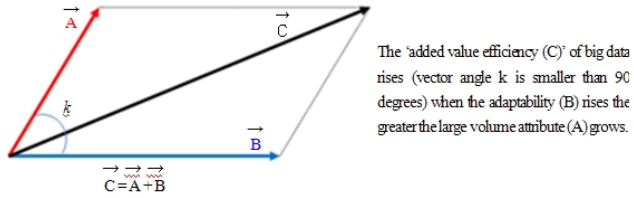
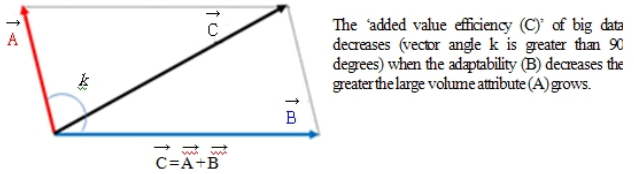


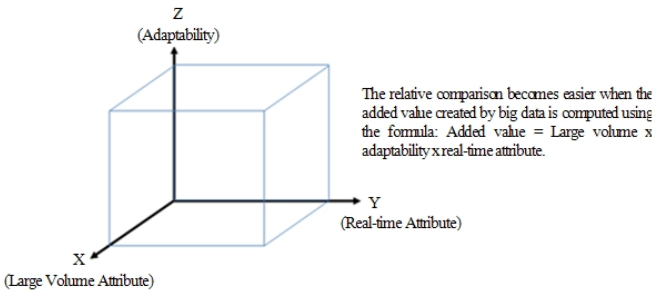
Fig. 4. Cases where conflicts or crashes arise among elements creating added values



However, as the mutual relevance among the large volume, adaptability and real-time attributes may vary significantly depending on the type of big data, the value of big data may be expressed simply by the area of a cu be, which is defined with x, y and z axes as shown in [Figure 5], if the three attributes meet each other orthogonally.

Let's refer to the big data factor that creates added values, which are expressed as a Vector, as 5V.

Fig. 5. 5V: Framework for measuring the added values of big data



IV. SIMULATED ANALYSIS OF THE IMPACT OF PERSONAL INFORMATION PROTECTION USING THE 5V FRAMEWORK

The following situation is assumed in order to simulate the real-life extent of the impact of the personal information protection-related law on big data-based business using 5V framework:

<Business service scenario>

- Business type: Customized shopping information is provided to customers using smart phones in connection with location information maintained by mobile service operators.

- Period of service provision: The contract is automatically terminated after providing services under a t wo-year agreement.

- Core competiveness: The nearest store is guided in real time that has the customer's most desired merchandise by analyzing big data concerning customer preferences, including age, occupation, gender, and past purchase history.

- Business growth potential: 300,000 customers subscribed to the service in the first year, and the number of subscribers increased by 5% each year for five years.

<Restrictions under laws concerning personal information protection>

- Maintenance of customer information: Personal information, including past purchase history and regularly visited stores (location information), should be deleted when the agreement period under Article 21 pa ragraph 1 of the Personal Information Protection Act expires.

- Method of subscriber recruitment: Customers originally subscribed to the service online via PCs or mobile appliances or by using an offline application or subscription form. Because of the problems encountered in securing users' consent to the collection and use of personal information in the case of offline applications, only online and mobile subscription is allowed currently (in connection with Article 15 or 22 of the Personal Information Protection Act).

※ In reality, the Korean Information Network Act and the Location Information Act should be considered as well. For this simulation, however, only the Personal Information Protection Act was applied.

Let's compute the large volume, adaptability and real-time attribute values under the above conditions in the current year (5th year).

The large volume value steadily decreased to 42.9% in the fifth year along with the 'available data volume' under [Table 3] in comparison with the case where no restrictions are in place:

TABLE III. IMPACT ON YEARLY CUSTOMER DATA VOLUME

Year	New subscribers	Cumulative subscribers	Net subscribers (considering those who terminate contract)	Available data volume (compared to that when there are no restrictions)
1	300,000	300,000	300,000	100.0%
2	315,000	615,000	615,000	100.0%
3	330,750	945,750	645,750	68.3%
4	347,288	1,293,038	678,038	52.4%
5	364,652	1,657,689	711,939	42.9%

TABLE IV. IMPACT ON VARIETY OF SUBSCRIPTION METHODS

Subscription method	When a specific customer's consent is not required	When a specific customer's consent is required
Online	Allowed	Allowed
Mobile	Allowed	Allowed
Offline application	Allowed	Disallowed

It is assumed that the real-time attribute has no impact as it cannot be measured under the given scenario because of the statutory restrictions.

Combing the above outcome, the big data which a given business operator can utilize decreases to 28.6% of the original volume because of restrictions under the law related to personal information protection. {Big data value (28.6%) = Large volume (42.9%) * Adaptability (66.7%) * Real-time attribute (100%)}

The relevant business may enhance adaptability in terms of variety either by securing a l arge volume by applying anonymity technology [8], such as data reduction or data perturbation, to the personal information of its customers whose contract expiry is imminent or by improving off-line subscription procedures to secure the customers' consent to the collection and use of their personal information, in order to maintain the value of its big data. In particular, more value will be lost in terms of volume as time passes when the situation is left alone. As such, it is desirable to apply anonymity technology to personal information within two years from business commencement.

So far, we have simulated an example that derives a solution while understanding the trade-off or conflict point between the protection of personal information and the utilization of big data using the 5V Framework.

V. CONCLUSION

This paper has so far introduced the 5V Framework, which can measure the impact of the law related to personal information protection on big data utilization. Big data has a new value chain (4V) different from that of conventional data processing. This added-value creation factor will become a core asset in the future industry and economy of Korea. However, the law related to personal information protection is highly likely to have inherent restrictions on the accumulation, analysis or utilization of big data, possibly acting as a factor that impedes national competitiveness in the future. Furthermore, it will definitely affect Government 3.0 [9], which is being ambitiously prepared by the government, and the deployment of the social safety network, including the proactive prevention of crimes through the analysis of statistical information. Of course, as the protection of personal information is a critical value, we cannot or should not give it up based on simple economic logic. (In reality, the trend is towards the relevant laws being steadily strengthened.) In conclusion, we need to discover a trade-off point, which is a social consensus between the use of big data and the protection of personal information.

The problem here is that the position of the engineers who need to assign values by actually processing or utilizing big data can hardly be reflected, as discussion about the impact that personal information protection has on big data utilization has mostly been devoted to superficial legal interpretations. For example, as to the requirement (discussed under Chapter

II) that 'personal information that is no longer required should immediately be deleted permanently using an irrecoverable means', legal experts consider what constitutional value such deletion would have, while engineers labor to determine what technology should be applied practically. When it is believed that the deletion of personal information has a g reater or higher value than the commercial or profit-seeking activities of a business, the relevant fine or penalty provision is strengthened without considering the current technology level. When the value of big data cannot be measured quantitatively, a legal standard needs to be applied, and the above-mentioned trade-off point will incline toward legal grounds.

When the 5V framework is utilized, we can quantitatively measure the extent of the impact of personal information protection activities on the value of big data. This author expects that it will provide a lead or a starting point to the derivation of a trade-off point between the utilization of big data and the protection of personal information.

REFERENCES

[1] "140 Government Tasks of the Park, Guen-hye Administration", 18th Presidential Office Take-over Committee, Feb. 2013, pp. 209.

[2] "Measuring the economics of big data", Working Party on Indicators for the Information Society, DSTI/ICCP/IIS(2011)4, OECD, 2011. 6

[3] "Seoul West District Court Case No. 2011gahap11733."

[4] "2012 Information Protection Survey (Individuals)", Korea Internet Security Agency, Dec. 2012, .pp. 21.

[5] "2012 Information Protection Survey (Businesses)", Korea Internet Security Agency, Dec. 2012, .pp. 110.

[6] Article 75 (Penalty) paragraphs 2-4 of the Personal Information Protection Act: On those who fail to destroy personal information in violation of Article 21 paragraph 1, http://www.law.go.kr, as of June, 2013.

[7] National Intelligence Service, "Guidelines on Information Storage Media Sanitization."

[8] "Code of practice for data protection, ICO(Information Commissioner's Office, UK), 2012.

[9] "Government 3.0" , National Information Society Agency.