HOW TO DE-IDENTIFY PERSONAL DATA IN SOUTH KOREA: AN EVOLUTIONARY TALE

Haksoo Ko
Seoul National University School of Law
hsk@snu.ac.kr (corresponding author)

Sangchul Park
University of Chicago Law School
scpark@uchicago.edu

Abstract

In early 2020, South Korea's legislature made amendments to major laws in the area of data protection in order to, among others, promote the utilization of pseudonymised personal data. With these amendments, pseudonymised personal data can be processed, without consent from data subjects, for archiving purposes, scientific research purposes, or statistical purposes. Arguably, these amendments are largely inspired by the relevant provisions contained in the EU GDPR, although details differ between GDPR and South Korea's amended statutes. One unique aspect of South Korea's amended statutes is that they introduce a scheme under which designated agencies carry out the task of combining pseudonymised data that different entities possess.

Keywords

De-identification, pseudonymisation, pseudonymised data, Korean data protection law, Personal Information Protection Act

HOW TO DE-IDENTIFY PERSONAL DATA IN SOUTH KOREA: AN EVOLUTIONARY TALE

Haksoo Ko
Seoul National University School of Law
hsk@snu.ac.kr (corresponding author)

Sangchul Park
University of Chicago Law School
scpark@uchicago.edu

I. Introduction

In early 2020, South Korea's legislature made amendments to major laws in the area of data protection, in order to, among others, promote the utilization of pseudonymised personal data by allowing processing of the data for archiving, scientific research or statistical purposes without consent from data subjects. It is a latest attempt in Korea in its efforts to strike a balance between properly safeguarding personal data and fostering data analytics and other types of utilization. The statutes that were amended are as follows: (i) the Personal Information Protection Act (PIPA), (ii) the Act on the Protection and Utilization of Credit Information (Credit Information Act), and (iii) the Act on the Promotion of Information and Communications Network Utilization and Information Protection (IC Network Act). They were all amended as of 4 February 2020 and effective as of 5 August 2020 in South Korea (collectively, 2020 Amendments).

This Comment discusses the main factors that drove recent major amendments; debates preceding the 2020 Amendments; and details regarding the new rules concerning the use of pseudonymised data. Paying particular attention to the influence that the EU General Data Protection Regulation (GDPR)[1] had on the amendment discussions in Korea, this Comment aims to provide additional insights into promising prospects and daunting challenges for the harmonisation of data protection laws on a global level.

II. Legal Context

Statutory framework and regulatory governance

South Korea has a stringent legal regime regarding personal data protection. The primary body of legislation is the PIPA, which was enacted in 2011. Several other data protection laws cover specific areas. The IC Network Act largely governs personal data collected online. The 2020 Amendments, however, consolidated relevant provisions into the PIPA and, as such, the role of the IC Network Act as personal data protection law has been drastically reduced. The Credit Information Act covers issues in financial privacy and contains provisions governing the credit information industry.[2] Also, the Act on the Protection and Utilization of Location Information (Location Information Act) is a statute with a jurisdiction over issues in the location information collected from various sources.

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

[2] Pursuant to the Credit Information Act, the Korea Credit Information Services was established in 2016, succeeding the tasks that had been conducted mostly by several credit registries.

Thus, while the PIPA is in principle expected to serve as a general statute governing data protection matters, there are several other relevant statutes as well. Due to this, although these statutes would not have overlapping subject matter jurisdictions in theory, difficult jurisdictional questions may well be raised in practice. If these practical aspects are considered, Korea's data protection regime can be called a hybrid regime between the EU model of omnibus data protection legislation and the sectoral approach of the U.S. legal regime.[3] With the 2020 Amendments, some of the overlaps have been eliminated, in particular between the PIPA and the IC Network Act.

The data protection laws are enforced by different data protection agencies. Prior to the 2020 Amendments, the Korea Communications Commission (KCC) had the power to govern online data protection rules under the IC Network Act, and in the case of offline data protection rules under the PIPA, the role of policymaking and coordination resided with the Personal Information Protection Commission (PIPC), while the enforcement powers belonged to the Ministry of the Interior and Safety (MOIS). The 2020 Amendments, for the most part, transferred the regulatory power over online data from the KCC to the PIPC. Also, the enforcement authority of the PIPA was transferred from the MOIS to the PIPC. Thus, the PIPC's regulatory and enforcement authority has been reinforced extensively. The Credit Information Act and Location Information Act will continue to be enforced by the Financial Services Commission (FSC) and the KCC, respectively. Table A outlines the changes brought about by the 2020 Amendments.

[Insert Table A here]

[3] See Paul M. Schwarz, 'Preemption and Privacy' [2009] 118 Yale L J 902, 908–15.

Consent principle

A major principle which is commonly observed in Korea's data protection statutes is the consent principle. That is, to collect statutorily defined personal data from a data subject for processing or for transferring to a third party, it is required to obtain prior informed consent from the data subject.[4] About the concept of personal data, identifiability of an individual is used as a key component in all relevant statutes.[5] Regarding giving notice to the data subject, each statute contains provisions with specific requirements.[6]

There are stipulated exceptions to the consent principle, which are specifically listed in these statutes. For instance, under the PIPA and the Credit Information Act, personal data may be collected and used without the data subject's consent in the following limited circumstances:

---

[4] Under the PIPA, consent should be obtained after disclosing each notice item so as to be clearly recognizable by the data subject, (i) with certain important notice items (concerning promotion, sensitive personal data, unique identifiers, retention period and the recipient and his or her purpose of use) being made more conspicuous if consent is obtained in writing or electronic document, and (ii) with consent-requiring items being separated from non-consent-requiring items (arts 22, 15(1)(i), 17(1)(i)). Under the Credit Information Act, for the disclosure of personal credit data, prior consent should be obtained per each case in writing or through public key certificate, accredited digital signature, personal password, voice recording, or other secure and creditworthy methods, while such formality does not apply to the collection and processing of personal credit data (arts 15(2), 32(1)(2)). The Location Information Act also provides for consent requirement (arts 18, 19).

[5] Under the PIPA, personal data is defined as any information relating to an individual which can be used to identify the individual alone or when easily combined with other information (art 2(i)). The Credit Information Act uses almost the same terminology to define personal credit data (art 2(ii)). The Location Information Act defines personal location data as 'location data relating to a specific individual (including data which can be used to identify the location of the specific individual when easily combined with other data even if the location cannot be identified by the data alone)' (art 2(ii)).

[6] Under the PIPA, the data subject must, before giving consent to collection, be given notice of (i) the purpose of collection and use, (ii) the items of data collected, (iii) retention and use period and (iv) (unless data is collected online) the data subject's right to refuse consent and disadvantages, if any, from the refusal (arts 15(2), 39-3(1)). The data subject must, before giving consent to disclosure, be given notice of the recipient and similar items as above (art 17(2)). Under the Credit Information Act, the data subject must, before giving consent to disclosure, be given notice of (i) the recipient, (ii) the recipient's purpose of use, (iii) the items of data disclosed and (iv) the recipient's retention and use period (art 32(1)), and under the amended Credit Information Act, the data subject must, before giving consent to collection, be given notice of the same items as set forth under the PIPA (art 34-2). The Location Information Act requires the disclosure of certain items in the standard terms of use before obtaining consent to the collection, use, or disclosure (arts 18, 19).

when permitted by law or necessary for compliance with a legal obligation; when required to exercise legal authority vested in the public agency; when necessary to enter into or perform a contract to which the data subject is party; when evidently necessary in order to urgently protect the vital, bodily, or pecuniary interests of the data subject or another person (when the data subject or her legal custodian is legally incapable, the address is unknown, or otherwise obtaining prior consent is infeasible); or when necessary for the purposes of the legitimate interests pursued by the data controller,[7] which evidently override the rights of the data subject.[8] These exceptions have similarities to the lawful basis for processing under Article 6(1)(b) through (f) GDPR, except among others, that the public interest alone does not constitute a lawful basis (see Article 6(1)(e) GDPR) and that not only the protection of vital interests but also the evident and urgent protection of bodily or pecuniary interests constitutes a lawful basis (see Article 6(1)(d) GDPR). Also, the Credit Information Act includes publicly available data among the exceptions (art 15(2)(ii)). Further, the PIPA, when the 2020 amendments were made, added an additional list of exceptions to the personal data collected by online service providers (art 39-3(2)).[9]

In practice, these exceptions appear to be rarely applied, in particular in the business context. For instance, although consent is not required when the data controller's legitimate interests evidently override the data subject's interests, it is not clear who is authorized to make the requisite decision and what the applicable criteria would be. In fact, its permissible scope is

[7] The PIPA defines a data controller as a 'public agency, entity, organization, or individual which processes personal data alone or through another person to administer personal data files for work purposes' (art 2(v)). In fact, the PIPA uses a term literally translated into 'personal data processor,' but it does not distinguish between data controller and data processor as in the GDPR.

[8] PIPA, art 15(1)(ii) through (vi); Credit Information Act, art 15(2)(i).

[9] Exceptions are as follows: when it is remarkably difficult, due to economic or technical reasons, to obtain consent whereas personal data is required in order to perform contract duties in the context of online service; when personal data is needed to settle online service fees; or when other laws permit.

even narrower than the scope of the 'legitimate interests' basis for processing under Article 6(1)(f) GDPR, as it requires the legitimate interests to 'evidently' override the data subject's interests. Thus, in a business context, obtaining consent is commonly perceived to be the only legal way of collecting personal data from data subjects.

Reconsideration of the consent principle

The advent of artificial intelligence and the growing use of big data spurred repeated debates in Korea, and attempts were made to find legitimate ways to utilize data without obtaining consent from data subjects. Early attempts in this context were culminated in the 2016 by the publication of a Personal Data De-Identification Guideline by the government (2016 Guideline). The 2016 Guideline excludes data from the application of the PIPA and other data protection laws once certain requirements are met: personal data is de-identified, undergoes adequacy assessment, and is subject to follow-up control for the prevention of re-identification.

Even with the publication of the 2016 Guideline, however, uncertainty remained and controversies continued. With further discussions, the 2020 Amendments tried to bring greater clarity and introduced the concept of pseudonymisation explicitly into the relevant statutes. Once the 2020 Amendments come into effect, personal data under the PIPA will be recategorized into (i) non-pseudonymised personal data (subject to the consent principle) (art 2(i)(*ga*)(*na*)),[10] (ii) pseudonymised personal data (subject to *ex post* control for the prevention

[10] This category includes any information relating to an individual which can be used to identify the individual alone or when easily combined with other information (to determine whether it can be 'easily combined,' the time, expense, technology, etc. needed to identify an individual, including availability of other information, should be considered at a reasonable level) (art 2(i)(*ga*)(*na*)).

of re-identification) (art 2(i)(*da*)),[11] and (iii) anonymised or anonymous data (which is not legally deemed personal data) (art 58-2).[12] The 2020 Amendments also allow the use of personal data without consent if doing so does not impair the interest of a data subject and if appropriate safeguards such as encryption of data are in place, within the scope reasonably related to the purpose for which the personal data is initially collected.[13] The following section expounds on the debates surrounding the concept and mechanics of de-identification of personal data, which resulted in the 2020 Amendments.

III. Facts

Debates on the de-identification of personal data prior to the 2020 Amendments

Need for de-identification

From the perspective of regulatory compliance, it would be best if consent could easily be obtained from every data subject. For obvious practical reasons, however, obtaining consent from all relevant data subjects would be an exceedingly cumbersome process in most

[11] Pseudonymised data is explicitly defined as the information which has been pseudonymised so as not to be used to identify an individual without the use of and combination with additional information for the reconstruction of the original data (art 2(i)(*da*)). Pseudonymisation is defined as 'the processing of personal data in such a manner that a specific individual becomes not identifiable without the use of additional information, rendered through deletion of a part of the data or substitution of all or a part of the data' (art 2(i-2)).

[12] The term of anonymised or anonymous data is not explicitly referenced in the PIPA, but the amended PIPA provides that the law does not apply to the data when identification cannot take place even through combination with other information and that, in making the decision about identifiability, reasonableness of the time, cost and technology should be considered (art 58-2).

[13] PIPA, arts 15(3), 17(4). Credit Information Act, art 32(6)(ix-4) also provides for a similar exception. These new broad exceptions are analogous to the concept of 'compatibility with the purpose for which the personal data are initially collected' under Article 6(4) GDPR, although it remains unclear how this exception will be interpreted and applied in practice.

circumstances. In particular, if the data to be utilized involve a large number of data subjects, obtaining consent from each of these data subjects would be all but impossible.

One way to solve this conundrum is to de-identify personal data. Personal data, once properly de-identified, would no longer be considered personal data under the PIPA or other statutes. As such, once de-identification is conducted, at least in theory, various stringent statutory requirements for personal data protection would not need to be complied with. Then, de-identified data can be used without having to consider the consent requirement or the purpose for which the consent was initially obtained. De-identified data may also be disclosed to a third party.

Prior to the 2020 Amendments, there was already a provision in the PIPA which allowed for the utilization of de-identified personal data. That is, Article 18(2)(iv) of the PIPA permitted the use of personal data for purposes other than the initial purposes notified to the data subjects prior to consent or its disclosure to a third party, without having to obtain additional consent, to the extent that the following safeguards are satisfied: (1) personal data at issue is de-identified and (2) utilization of the personal data is limited to statistical purposes or academic research purposes.

Thus, there were two possible avenues for statutory interpretation related to de-identified data for data utilization. First, de-identified data would no longer be considered to fall under the legal definition of personal data in all relevant statutes. Second, personal data, once de-identified, can satisfy an exception under Article 18(2)(iv) of the PIPA. Either way, for de-identified data, there would be room for manoeuvre and utilization.

The 2016 Guideline

The 2016 Guideline was a result of the joint efforts made by multiple government agencies to balance between utilization and protection. The 2016 Guideline did not have the power or authority of a statute because it simply is a government-issued guideline and not even an administrative ordinance. Nonetheless, since it was issued by multiple government agencies, it carried a heavy *de facto* authority.

The 2016 Guideline offered an explanation on various de-identification techniques and suggested utilizing these techniques appropriately as needed. Specifically, it established a four-step approach for de-identifying personal data. First, it should be examined if the given data falls under the legal definition of personal data. Obviously, if deemed personal data, such data should be de-identified prior to engaging in analytics or other utilization. Second, actual de-identification is conducted. For de-identification, identifiers should be removed in a given dataset and, as a general rule, attributes should also be eliminated to the extent that they are not needed for the purposes of the proposed analytics. With the remaining data elements, various statistical methods may be applied in order to prevent linkage and other attacks for re-identification. Most notably, the concept of *k*-anonymity is proffered as an important concept to be applied.[14] Further, if needed, the concepts of *l*-diversity[15] and/or *t*-closeness[16] could be applied as well. The 2016 Guideline provided illustrations as to how various statistical methods can be applied such as pseudonymisation, aggregation, data reduction, data suppression and

[14] A release of data is said to satisfy *k*-anonymity if each equivalence class contains at least *k* records (in other words, each person's information is indistinguishable from at least *k* - 1 individuals' information in the release) (Pierangela Samarati and Latanya Sweeney, 'Protecting Privacy When Disclosing Information: k-Anonymity and its Enforcement through Generalization and Suppression' [1998] Proc IEEE Symp Security and Privacy 4–5).

[15] A release of data is said to satisfy *l*-diversity if each equivalence class has at least *l* well-represented values for each sensitive attribute (Ninghui Li, Tiancheng Li and Suresh Venkatasubramanian, 't-Closeness: Privacy beyond k-Anonymity and i-Diversity' [2007] IEEE 23rd Intl Conf Data Eng 107–8).

[16] A release of data is said to satisfy *t*-closeness if the distribution of a sensitive attribute in any equivalence class is close (within a threshold *t*) to the distribution of the attribute in the table (Ibid 109–11).

data masking. Third, after a de-identification process is finished, an assessment should be made by a panel of experts regarding the adequacy of de-identification that took place. An important criterion in the assessment process is again the concept of *k*-anonymity.[17] Fourth, if de-identification is considered to be adequate, de-identified data can then be utilized without consent. At the same time, the data controller is required to assume a duty to prevent illegitimate use of data including re-identification.

After the 2016 Guideline was issued, several public agencies were designated,[18] with mandates for assisting the process of de-identifying personal data and for carrying out tasks of linking or combining datasets that are de-identified. Within a year, pursuant to the 2016 Guideline, certain de-identification and data combination projects were reported to have been carried out involving multiple organizations, including mobile carriers, insurers, and credit card companies.[19] Through these projects, companies de-identified part of their data and asked one of the designated agencies to link the data, to confirm the de-identification status, and to return the consolidated dataset. Amid controversies over the validity and legality of carrying out these projects, in November 2017, eleven NGOs[20] joined forces and filed criminal complaints

[17] The expert assessment panel thus determines the reference or threshold value of *k*, after taking into consideration various factors such as the number of (quasi-)identifiers, possibilities of accruing linkable data, possibilities of re-identification attempts, potential damages from re-identification and purposes of data utilization. Once the reference value of *k* is determined, among other tasks, the expert panel compares the value of *k*-anonymity achieved through de-identification with the reference value of *k*. In principle, if the level of *k* achieved through de-identification is higher than the reference *k*, de-identification would be considered successful.

[18] They include Korea Internet & Security Agency (KISA), National Information Society Agency (NIA), Korea Credit Information Service (KCIS), Korea Financial Security Institute (KFSI), Social Security Information Service, Korea Education & Research Information Service.

[19] It was disclosed in October 2017 by Assemblyperson Hyeseon Chu, who obtained the records of de-identification projects from the four designated public agencies in the course of the parliamentary inspection of the administration (Jaesup Kim, 'A Surge of Distribution of Pseudonymised Customer Personal Data – 340 Million Cases within a Year' *The Hankyoreh* (Seoul, 9 October 2017) <www.hani.co.kr/arti/economy/it/813776.html> accessed 16 June 2020).

[20] They include a lawyer group (Lawyers for a Democratic Society), four healthcare NGOs (Korean Federation of Medical Activist Groups for Health Rights, Association of Physicians for Humanism, Korean Pharmacists for Democratic Society and Association of Korea Doctors for Health Rights), a national trade union centre (Korea

against four of the designated public agencies and the companies which carried out these projects, for alleged violation of the PIPA, the Credit Information Act and the IC Network Act.[21] The position of these NGOs was that the 2016 Guideline was faulty and illegitimate in itself as it infringed data subjects' constitutional right to control personal data and that, as such, any de-identification work conducted following the 2016 Guideline was illegal.[22] In March 2019, Seoul Central Prosecutors' Office decided not to indict these public agencies and companies on the ground that the consolidated data do not fall under the definition of personal data under the PIPA as it would not be possible to reidentify data subjects from the data; that the accused acts are not punishable as they were conducted pursuant to the direction of relevant authorities; and that data combination was carried out for research purposes.[23] While the accused public agencies and companies were cleared from possibilities of criminal liability by this decision, the mere fact that a criminal complaint was filed had a chilling effect on the business community.

Pseudonymisation under the 2020 Amendments

Preparation for legislation

Confederation of Trade Unions) and five other NGOs (People's Solidarity for Participatory Democracy, Korean Progressive Network Centre, Citizens' Action Network , People's Coalition for Media Reform and Solidarity for Workers' Health).

[21] Taewoo Park, 'Does Common Customer Tendency Analysis Qualify for the Research Purpose – Concerns Raised over Retrogression in Data Protection' *The Hankyoreh* (Seoul, 28 March 2019) <www.hani.co.kr/arti/economy/it/887851.html> accessed 16 June 2020).

[22] Ibid.

[23] Ibid.

With the 2016 Guideline, a major contentious issue arose in relation to the mechanics which allowed for the combination of datasets from two or more sources. Under this scheme, in order for the combination to take place, organizations holding datasets should transfer their data to one of the designated public agencies, so that this public agency can combine the datasets. In that process, data would initially be de-identified, and the data combination would be carried out with such de-identified data. What was not entirely clear was whether such de-identified-and-combined data should legally be deemed personal data.

While controversies surrounding the 2016 Guideline ensued, some observers in Korea argued that pseudonymisation could provide a useful alternative to the data de-identification scheme contained in the 2016 Guideline. They took note of the pseudonymisation provisions contained in the GDPR. In particular, the Presidential Committee on the Fourth Industrial Revolution, which was established in 2017, served as an important venue for public debates. The Committee, among others, held two 'Hackathon' meetings in 2018 to discuss issues related to personal data.[24]

The first of these Hackathon meetings concluded that legal concepts related to personal data need to be streamlined and that the concept of pseudonymisation needs to be discussed further. At the second Hackathon meeting on personal data, the concept of pseudonymisation was included as a main discussion item and the GDPR's relevant provisions were more extensively discussed. The participants reached the following conclusion about pseudonymisation:[25]

[24] One of the authors participated in the Hackathons and related meetings. While called Hackathon, it was in practice akin to a lengthy townhall meeting among stakeholders and experts.

[25] Presidential Committee on the Fourth Industrial Revolution, *Press Release: Reviewing the Scope and Objective of Utilisation of Pseudonymised Data, the Reform of the Information Rating System for Facilitating the Use of Cloud Computing, and the Alleviation of Difficulties that the Drone Industry Faces* (5 April 2018) 3 <4th-ir.go.kr/hackathon/detail/7?num=03> accessed 16 April 2020.

- Pseudonymised data can be utilized for a purpose other than the initial purpose of collection or can be disclosed to a third party to achieve the following purposes: (i) archiving purposes in the public interest, (ii) scientific and/or research purposes or (iii) statistical purposes;[26]
- During the relevant processes, appropriate safeguards such as technical and organizational measures should be employed;[27] and
- Scientific and/or research purposes may include an industrial research purpose, and the statistical purposes may include a commercial purpose.[28]

It was also agreed that the use of personal data for the purposes compatible with the initial purposes should be permitted to the extent that due considerations are made regarding various circumstances including pseudonymisation.[29]

2020 Amendments

A conclusion reached at the end of these Hackathon meetings on personal data was that suitable statutory amendments are needed. In response, government agencies undertook the task of preparing an amendment proposal. In November 2018, the government submitted to Korea's

[26] Although GDPR provisions were not mentioned in the public announcement, this is arguably modelled after Articles 5(1)(b) and 89(1) of GDPR, which exempts, from purpose limitation, the processing for archiving in the public interest, scientific or historical research, or statistical purposes.

[27] Although GDPR provisions were not mentioned in the public announcement, this is arguably modelled after Article and 89(1) of GDPR, which requires safeguards relating to processing for archiving in the public interest, scientific or historical research, or statistical purposes.

[28] Although GDPR provisions were not mentioned in the public announcement, this is arguably inspired by the broad interpretation of scientific research and statistical purposes under Recitals 159 and 162 of GDPR.

[29] Presidential Committee on the Fourth Industrial Revolution (n 25) 3. "EU GDPR" was explicitly referenced in the public announcement in the part where the concept of compatibility was mentioned.

legislature, the National Assembly, amendment proposals for the PIPA, the Credit Information Act and the IC Network Act.[30] The amendment proposals finally passed the legislature on 9 January 2020 and were promulgated on 4 February 2020, with the effective date being 5 August 2020.

The amendments to the PIPA contain several provisions on pseudonymisation. It defines pseudonymisation as 'the processing of personal data in such a manner that a specific individual becomes not identifiable without the use of additional information, rendered by removing a part of the data, replacing all or a part of the data, etc.' (art 2(i-2)). This definition is somewhat analogous to the definition of pseudonymisation under Article 4(5) GDPR.[31] It is unclear if the concept of pseudonymisation under Korea's PIPA carries the same meaning as the concept of pseudonymisation under GDPR. In terms of the relevant statutory text of the PIPA, the following appears to be noteworthy: unidentifiability, rather than unattributability, is a key element; certain methods of pseudonymisation (i.e., removal or replacement) are specifically mentioned, although these methods are not meant to be exhaustive; and the separation of additional information and technical and organisational measures are not embedded in the definition itself but instead constitute part of the legal obligations that should be complied with for the processing of pseudonymised data.

[30] These statutory amendment proposals were submitted by individual lawmakers, with the appearance that they were submitted by the lawmakers' own initiatives. They were, in all practicality, the government's amendment bills.

[31] Article 4(5) GDPR provides that "'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person".

The amendments make clear that pseudonymised data still falls under the definition of personal data (art 2(i)).[32] Pseudonymised data can be processed without consent from a data subject (i) for statistical purposes, (ii) for scientific research purposes and (iii) for archiving purposes in the public interest (art 28-2(1)). This is also similar to Article 89(1) GDPR, except that 'historical research purposes" are not explicitly included. The pseudonymised data so disclosed to a third party should not include identifiers (art 28-2(2)).

The PIPA explicitly introduced a scheme for the consolidation of pseudonymised data. Only designated agencies are permitted to link or combine pseudonymised datasets sourced from different data controllers and to prepare a combined dataset (art 28-3). Also, appropriate technical, organizational and physical safeguards should be in place for the processing of pseudonymised data. This includes separate safeguarding of additional data which would be needed to reconstruct the original data from the pseudonymised data (art 28-4(1)). The proposed amendments to the Enforce Decree for the PIPA stipulate that the data controller who requested linkage or combination can analyse the consolidated data at a secured space within a designated agency and, in order to bring out the consolidated data from the designated agency, a separate approval would be needed from the designated agency (art 29-2(3)(4)). The data controller should also keep records regarding the history of processing of pseudonymised data, including its purposes and recipients (art 28-4(2)). The processing of pseudonymised data is exempted from not only the consent requirement but also various other PIPA rules such as the data subject's right to access, rectification and erasure and the data controller's obligation to notify data breaches (art 28-7).

[32] This is in line with Recital 26 GDPR, which provides that "[p]ersonal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person."

The amendments to the PIPA do not make an explicit reference to anonymous or anonymised data, unlike Recital 26 GDPR. It, however, contains a provision which can be interpreted to define anonymised data in an indirect way. That is, the PIPA provides that it does not apply to the data when identification through combination with other information cannot take place and that, in making the decision about identifiability, reasonableness of the time, cost and technology should be considered (art 58-2).

The amendments to the Credit Information Act also include provisions on pseudonymisation. It defines pseudonymisation as 'the processing of personal credit data in such a manner that a specific credit data subject becomes not identifiable without the use of additional information' (art 2(xv)). Similar to the amendments to the PIPA, the Credit Information Act provides that pseudonymised data may be processed for archiving purposes in the public interest, for research purposes or for statistical purposes (art 32(6)(ix-2)). This provision, different from the PIPA, explicitly declares that research purposes include industrial research purposes and that statistical purposes may include commercial purposes such as market survey. The amendments further stipulate that credit bureaus, credit registries, debt collection agencies, 'MyData' service providers,[33] and data furnishers and users (which include banks and other financial institutions) (collectively, Creditors) must keep additional data that were used for pseudonymisation separately or otherwise erase the data (art 40-2(1)). Appropriate technical, organizational and physical safeguards must also be in place for pseudonymised personal credit data. These safeguards include: preparing for an internal management plan or a retention system for log records (art 40-2(2)); and establishing a system to recall pseudonymised data, suspend its

[33] A 'MyData' service provider may gather and consolidate a customer's credit data from various sources; give the customer access to the consolidated data; analyse the customer's credit rating, financial risks and consumption patterns; provide customized financial consulting; and recommends financial products. Overall, this service is predicated upon the concept of data portability and upon the introduction of credit management business for individuals. Article 2 9-2 & 9-3 Credit Information Act.

processing and immediately erase identifier, if an individual becomes identifiable in the course of using pseudonymised data (art 40-2(3)).

The amendments to the Credit Information Act also include provisions on the linkage or combination of datasets held by the Creditors. Only a designated data agency can link or combine a dataset held by a Creditor with another dataset held by another party (art 17-2(1)). The Creditor can transfer personal credit data to the data agency for this purpose without obtaining consent from the data subjects (art 32(6)(ix-3)). The data agency must pseudonymise or anonymise the consolidated dataset before releasing it back to the Creditor or to a third party (art 17-2(2)).

A Creditor can further request the data agency to review the adequacy of anonymisation, and the data agency's finding of adequacy would be considered prima facie evidence that personal data has been rendered unidentifiable (art 40-2(3) to (5)).

IV. Analysis

General review

Korea does not have a long history of enforcing its legal regime for data protection. The PIPA was enacted in 2011 and, in particular, with regard to data de-identification, it has only been a few years since serious discussions began to take place. While the government published its 2016 Guideline, due to various legal and practical constraints, businesses and other potential users have not actively embraced the de-identification methodology contained in the 2016 Guideline.

In the meantime, the GDPR was enacted in the EU and, in Korea, active discussions took place seeking useful implications from the GDPR. As such, some of the concepts and provisions contained in the GDPR had a significant impact in the development of Korean data protection regime, in particular in the context of discussing the concepts of de-identification and pseudonymisation.

The 2020 Amendments, at the same time, have unique features such as provisions on data linkage or combination and on designating agencies for carrying out such linkage or combination.

Uncertainties that remain about the production, use and linking of pseudonymised data

Following on the 2020 Amendments, preparatory work for issuing subordinate regulations and guidelines are under way. Through this process, certain issues that have not been resolved or clarified may perhaps need to be addressed, as discussed below.

Failure to address diverse risks surrounding pseudonymised data

Regarding pseudonymisation, the PIPA and the Credit Information Act appear to postulate a particular type of situations under which data contained in an original dataset is transformed through a pseudonymisation process and a new dataset is created. As such, a typical example of 'additional information' (which could be used for reconstruction of the original dataset) would be cryptographic keys or functions that were used in the process of pseudonymisation (or reconstruction of the original data). Interpreted this way, these statutes arguably failed to explicitly address risks involving (i) reconstruction of the original dataset by utilizing an

auxiliary dataset held by a third party, (ii) re-identification of a single individual without reconstructing the whole dataset, perhaps by utilizing background knowledge, or (iii) inadvertent or unintentional re-identification.[34]

The scope of 'scientific research purposes'

Amid the debates on pseudonymisation, particular attention was paid to the meaning and scope of 'scientific research purposes' under the PIPA (art 28-2(1)) (or 'research purposes' under the Credit Information Act (art 32(6)(ix-2)) as a lawful basis for processing pseudonymised data without consent from the data subject. The PIPA provides that scientific research means research which applies scientific methodology and that it includes technological development and demonstration, fundamental research, applied research and privately funded research (art 2(8)). This broad interpretation is modelled after Recital 159 GDPR. It is unclear, however, whether scientific research would include research with commercial motivations. In the case of the Credit Information Act, while its amendments make it clear that the research purposes include industrial research purposes, it is unclear what the industrial research entails and whether it includes commercial research. Amid controversies, interpretations of analogous provisions of GDPR would provide useful references. They would include, for example, the

[34] There is a requirement, however, (i) not to process pseudonymised data for purposes of identification, and (ii) to stop processing and to reclaim or discard data if identifiable data are produced while processing pseudonymised data. Article 28-5 PIPA.

UK ICO Anonymisation Code of Practice[35] and the European Data Protection Supervisor's preliminary opinion in 2020.[36]

The process of pseudonymisation

About the concept and methodology of pseudonymisation, the 2020 Amendments do not make clear whether, in order to satisfy the legal requirements, (i) it is sufficient to remove or replace only direct identifiers in a given dataset, or (ii) quasi-identifiers and other attributes should also be examined and modified as needed in order to prevent singling out of an individual.[37] The former would entail a relatively straightforward and simple process, while the risk of singling out an individual may linger. The opposite would be true with the latter. If the former approach were to be taken, in consideration of the heightened risk, enhanced scrutiny through detailed procedural safeguards may need to be in place.

The role of designated data linking agencies

It is also unclear whether designated agencies would play a relatively limited role as a trusted third party (TTP), mainly keeping cryptographic keys or functions that are used for

[35] The Code of Practice states the UK Data Protection Act makes it clear that research purposes include 'statistical or historical research, but other forms of research, for example market, social, commercial or opinion research' (The UK Information Commissioner's Office (ICO), *Anonymisation: Managing Data Protection Risk Code of* Practice (2012) 45 <ico.org.uk/media/1061/anonymisation-code.pdf> accessed 16 June 2020).

[36] '[N]ot only academic researchers but also … profit-seeking commercial companies can carry out scientific research' (European Data Protection Supervisor (EDPS), *A Preliminary Opinion on Data Protection and Scientific Research* (2020) 11 <edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en> accessed 16 June 2020).

[37] This is inherently about the concept of identification and also about how to distinguish between anonymisation and pseudonymisation. In the context of anonymisation, quasi-identifiers and other attributes would need to be considered. 'Data controllers often assume that removing or replacing one or more attributes is enough to make the dataset anonymous. Many examples have shown that this is not the case; simply altering the ID does not prevent someone from identifying a data subject if quasi-identifiers remain in the dataset, or if the values of other attributes are still capable of identifying an individual' (Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques [2014] (Opinion 05/2014) 21).

pseudonymisation or for reconstruction of the original data,[38] or whether they would assume more active roles. Based on what is contained in the proposed amendments to the enforcement decrees for the 2020 Amendments, the roles of the data agency under the Credit Information Act appear to be different from those under the PIPA. For example, the data agency under the Credit Information Act has the authority to evaluate pseudonymisation processes. On the other hand, the data agency under the PIPA would be equipped with a research space within its premise and facilitate data analysis conducted under its auspices. Depending on specific roles that are envisaged and are finally determined, different risk factors would need to be considered.

Failure to draw a clear line between pseudonymisation and anonymisation

As mentioned, while the PIPA does not explicitly employ the concept of anonymised data, it provides that the law would not apply to the data when identification cannot take place even through combination with other information and with reasonable consideration of time, cost and technology (art 58-2). As a matter of statutory interpretation, this concept arguably overlaps with the definition of pseudonymised data: personal data processed in such a manner that a specific individual becomes not identifiable 'without the use of and combination with additional information for the reconstruction of the original data' (art 2(i)(*da*), (i-2)). That is, if re-identification takes place with certain additional information and only after spending an exorbitant amount of time and efforts, the data at issue would fall under the definitions of anonymised data and, at the same time, of pseudonymised data. This, however, cannot be the case since pseudonymised data is, by definition, personal data.

[38] See European Union Agency for Cybersecurity (ENISA), *Pseudonymisation Techniques and Best Practices* (2019) 14 <enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices> accessed 16 June 2020.

Potential conflict with other legislations

The 2020 Amendments may conflict with what is contained in other laws and, as such, harmonisation measures may need to be taken. In particular, in healthcare, de-identification has taken place on a regular basis and there are laws and regulations governing such de-identification process. Specifically, the Bioethics and Safety Act defines 'anonymisation' as (i) deleting personal identification information or (ii) substituting the whole or part of personal identification information with a unique id used within an institution conducting research (art 2(19)). The second part of this definition could possibly be interpreted to mean pseudonymisation under the 2020 Amendments.

Additional notes on technology-based contact tracing in response to the COVID-19 pandemic

In response to the outbreak of the COVID-19 pandemic, various approaches for technology-based contact tracing have taken place around the world. In broad terms, many European countries opted for a decentralized, user-centric approach deploying Bluetooth Low Energy (BLE) to discover and log clients in proximity. This approach can further be divided into a partially centralized approach and a fully decentralized approach. Examples of a partially centralized approach include PEPP-PT (adopted in France and tested-and-discarded in the U.K.) and BlueTrace (adopted in Singapore and Australia). A decentralized approach was adopted in DP3T (Austria) and in Apple-Google Exposure Notification API (Germany, Switzerland, Estonia, and Lithuania). Outside Europe, several states in the U.S. (such as North Dakoda and South Dakoda) and Japan also adopted the Apple-Google Exposure Notification scheme.

Korea, however, alongside Israel, has taken a fully centralized approach. Under Korea's centralized approach, Korea Centers for Disease Control and Prevention (KCDC) would gather and compile data from various sources. Data gathered at KCDC include location data that can be extracted using mobile base station data from mobile carriers and payment card transaction records.

Between a decentralized approach and a centralized approach, in general, the former approach, based on pseudonymised IDs, is more privacy-preserving. On the other hand, the latter approach is more effective in promptly and precisely tracing confirmed cases and those contacted. This approach is also helpful in saving time and resources that epidemiological investigators need to take for interviewing and tracing.[39]

After experiencing the MERS outbreak in 2015, Korea revised the Contagious Disease Prevention and Control Act and inserted a pandemic trigger provision, which authorized embarking a centralized contact tracing system in a pandemic situation.[40] Also, once a pandemic begins, this Act's provisions on data collection would override the general consent requirement under the PIPA (art 15(1)(ii) PIPA). With such a clear statutory basis, Korea could avoid legal controversies surrounding contact tracing.

The contact tracing scheme adopted in Korea served as a crucial enabling factor in the implementation of the country's trace, test and treat strategy.[41] It, however, also revealed what could possibly go wrong when pseudonymised data change hands and are communicated to

[39] Sangchul Park, Gina J. Choi and Haksoo Ko, 'Information Technology–Based Tracing Strategy in Response to COVID-19 in South Korea—Privacy Controversies' (2020) JAMA <doi:10.1001/jama.2020.6602> accessed 16 June 2020.

[40] Ibid.

[41] Ibid.

the public. In particular, pseudonymisation has often been proved insufficient: although names of the confirmed individuals were never revealed, some of them have at times been re-identified based on the age, sex, domicile, and other quasi-identifiers and attributes that are exposed in the public disclosure process.

Prospects

Pseudonymisation and other de-identification methodologies are not novel concepts. Instead, they have been implemented under strict control based on tight professional ethics in highly specialized areas such as the healthcare and pharmaceutical industry. The 2020 Amendments expanded the scope to all other areas. This poses daunting new challenges of harmonizing conflicting concepts and establishing best practices across different research environments and different industrial sectors. There remain a great deal of uncertainties and ambiguities, which may be inevitable in the realm of data. Once the challenges are met, Korea's experience may serve as a useful model for other jurisdictions facing a similar conundrum with the advent of the data-driven society.

[Table A]

Summary of Changes in Regulatory Governance Triggered by the 2020 Amendments

<table>
<tr><th></th><th>Financial data</th><th>Other data collected online</th><th>Other data collected offline</th><th>Location data</th></tr>
<tr><td>Existing laws
(Relevant data protection authority)</td><td>Credit Information Act
(FSC)</td><td>IC Network Act
(KCC)</td><td>PIPA
(PIPC: policy making and coordination /MOIS: enforcement)</td><td rowspan="2">Location Information Act (KCC)
(unaffected by the 2020 Amendments)</td></tr>
<tr><td>The 2020 Amendments
(Relevant data protection authority)</td><td>Credit Information Act
(FSC)</td><td colspan="2">PIPA (PIPC)</td></tr>
</table>