Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.32, NO.2, Apr. 2022 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2022.32.2.329

가명정보 생명주기에 따른 개인정보보호 중심 설계 적용 원칙에 관한 연구

김 동 현†‡
한국인터넷진흥원 (책임연구원)

A Study on the Principle of Application of Privacy by Design According to the Life Cycle of Pseudonymization Information

Dong-hyun Kim†‡
KOREA INTERNET & SECURITY AGENCY (General Researcher)

요 약

최근 개인정보가 데이터로 활용되면서 다양한 신산업 등이 발굴되고 있지만 체계적인 관리체계 구축 미흡 등으로 개인정보 유출 및 오남용 사례가 연이어 발생되고 있다. 또한 지난 '20년 8월, 데이터 3법 시행 이후 개인정보를 가명·익명 처리하여 활용하는 서비스가 등장하고 있지만 불충분한 가명처리 및 가명정보 처리에 대한 안전성확보조치, 혐오표현 등의 민감정보의 처리 미흡으로 개인정보 이슈가 발생하고 있다. 이에 본 연구는 개인정보를 안전하게 활용하기 위하여 캐나다의 Ann Cavoukian[1]이 제시한 개인정보보호 중심 설계(Privacy by Design, 이하 PbD) 원칙을 기반으로 가명정보 생명주기에 적용할 수 있는 새로운 PbD원칙을 제안하였다. 또한, 제안한 방법에 대하여 개인정보보호 관련 전문가 30명을 대상으로 설문조사를 통하여 제안 방법의 유의미함을 확인할 수 있었다.

ABSTRACT

Recently, as personal information has been used as data, various new industries have been discovered, but cases of personal information leakage and misuse have occurred one after another due to insufficient systematic management system establishment. In addition, services that use personal information anonymously and anonymously have emerged since the enforcement of the Data 3 Act in August 2020, but personal information issues have arisen due to insufficient alias processing, safety measures for alias information processing, and insufficient hate expression. Therefore, this study proposed a new PbD principle that can be applied to the pseudonym information life cycle based on the Privacy by Design (PbD) principle proposed by Ann Cavoukian [1] of Canada to safely utilize personal information. In addition, the significance of the proposed method was confirmed through a survey of 30 experts related to personal information protection.

Keywords: Personal Information, Pseudonymization Privacy by Design, De-Identification Privacy by Design

I. 서 론

지난 '20년 8월, 이른바 데이터 3법의 개정과 함께 데이터 산업진흥 및 이용촉진에 관한 기본 법률도 '22년 4월 시행을 앞두고 있어 더욱 많은 산업에서 다양한 데이터가 활용될 예정이다. 또한, 디지털 뉴딜 등 우리나라를 비롯하여 전 세계적으로 데이터 경제를 강화하기 위한 다양한 개인정보 활용 정책들이 등장하고 있다[2]. 이 중 이종산업 간 가명정보의 결합 및 자율주행차, 스마트시티 등의 산업은 개인정보가 방대하게 활용되는 경우로 보다 안전한 개인정보에

Received(01. 26. 2022), Accepted(02. 21. 2022)
† 주저자, kdonghyun@kisa.or.kr
‡ 교신저자, kdonghyun@kisa.or.kr(Corresponding author)

대한 관리체계가 요구된다. 개인정보를 안전하게 활용하기 위한 방법으로는 가명처리를 통해 특정 개인을 알아볼 수 없도록 처리하여 활용하는 방법이 있지만 가명정보는 데이터 상황에 따라 재식별 위험이 언제든지 존재하고 있다[3]. 이러한 위험성을 최소화하기 위해서는 개개인이 의식적으로든 무의식적으로든 발생시키는 정보를 누가 얼마나 통제할 수 있는가가 문제점으로 발생하고 있으며 이러한 문제점을 해결하기 위한 방안으로 개인정보보호 중심 설계(Privacy by Design, 이하 PbD)가 화두가 되고 있다. PbD는 온라인 서비스를 포함하는 IT기술의 발전에 따라 서비스 기획 단계에서부터 폐기 단계까지의 전체 생애주기(Life-Cycle)에 걸쳐 이용자의 프라이버시와 데이터를 보호하는 기술 및 정책을 적정하게 적용하는 것으로 개인정보를 취급하는 모든 측면에 있어 개인정보가 적절하게 취급되는 환경을 '사전적(事前的)'으로 만드는 것이라고 정의[4]하고 있으며, '10년 캐나다의 Ann Cavoukian이 이에 따른 7개의 기본원칙을 제시한 바 있다[1].

이러한 PbD 원칙은 일반적인 개인정보에 대한 생명주기를 기반으로 수집, 저장, 이용 또는 제공, 파기의 모든 단계에서 검토가 가능하며 이를 준수하기 위한 다양한 관리적·기술적 보호조치가 법률 또는 지침 등을 통해 제공되고 있다. 하지만 지난 '21년 10월 마련된 '가명정보 처리 가이드라인[5]'에 따른 가명정보에 관한 생명주기의 경우 가명처리 대상을 추출하고 특정 개인을 알아볼 수 없도록 데이터 상황을 고려한 위험도 평가 및 가명처리 후 검토, 사후관리를 수행하는 등의 단계로 구성되어 있어 일반적인 개인정보처럼 PbD를 적용하는 것이 어려우며 이를 위한 별도의 지침도 부재한 상황이다.

이에 본 논문에서는 개인정보에 기반을 둔 PbD 원칙에 대한 현황을 살펴보고, 국제표준에서 제시하고 있는 비식별 조치 생명주기를 기반으로 가명정보 처리 시 적용할 수 있는 PbD원칙을 새롭게 제안하고자 한다. 이를 위해 본 논문의 2장에서는 PbD에 대한 국외현황을 3장에서는 국제표준에 기반을 둔 가명처리의 생명주기별 고려사항을, 4장에서는 고려사항에 따른 새로운 가명처리 PbD원칙을 제안, 검토한 다음 5장을 끝으로 결론을 맺고자 한다.

II. PbD(개인정보 보호 중심 설계) 현황

2.1 PbD 기본 원칙

PbD는 캐나다의 Ann Cavoukian이 주장한 개념으로 프라이버시 정보를 취급하는 모든 측면에 있어 프라이버시 정보가 적절하게 취급되는 환경을 '사전적(事前的)'으로 만드는 것을 의미한다[4]. 여기서 사전적이란 의미는 프라이버시 정보를 이용하려고 하는 단계에서가 아닌 그 이용이 예상되는 시점에서의 대응을 구체화라는 것을 의미한다. 즉 프라이버시 보호를 내재하고 있는 기술과 시스템, 상품 등의 개발을 통하여 프라이버시 침해를 예방하고 최소화하는 제도로 해석될 수 있다[6].

이러한 PbD원칙은 다음과 같이 7개의 원칙으로 구성되어 있다. 첫째, 'Proactive/Preventative'는 사후대응이 아닌 사전에 대비를 해야 하며 문제점을 개선하는 것이 아닌 예방을 하는 것이다. 둘째, 'By Default'는 Privacy as the Default Setting으로 프라이버시 보호를 위한 기본기능이 내재화 되어 있어야 한다. 셋째, 'Embedded'는 시스템의 설계 시 프라이버시 보호가 이미 전제되어야 한다. 시스템 개발 시 SDL(Secure Development Life-cycle)을 적용하고 표준에 근거한 프레임워크를 설계하고 검토하여야 한다. 넷째, 'Positive Sum'은 프라이버시와 보안이 한쪽을 강화하면 한쪽이 약해지는 반비례 관계로 접근하지 말고 두 가지 모두 윈-윈 할 수 있는 방법을 검토하여야 한다. 다섯째, 'Life-cycle Protection'은 전체 생명주기를 범위로 지속적으로 보호 되어야 하며 End-to-End로 관리가 되어야 한다. 여섯째, 'Visibility/Transparency'는 가시성과 투명성이 보장돼야 한다. 가시성과 투명성은 책임과 신뢰를 확립하는데 필수적이며 특히 책임과 개방성, 법률 준수사항들을 준수하고 투명하게 공개하여야 한다. 일곱째, 'Respect for Users'는 개인의 프라이버시를 존중하고 사용자 중심의 설계와 운영을 하여야 한다.

2.2 PbD 관련 해외 법제도 현황

2.2.1 유럽연합

'18년에 시행된 EU GDPR[7]에서는 제25조(Data protection by design and by default)

를 통해 개인정보의 처리에 관련된 개인의 권리와 자유를 보호하기 위해 적절한 기술 및 관리조치를 요구하고 있으며 이 법을 준수하고 있음을 입증하기 위해 개인정보처리자는 개인정보보호 중심의 디자인

Table 1. The Contents of Article 25 of GDPR

Article		Contents	Summary
1	①	Consider the risks that may arise to individual rights and freedoms according to processing, such as the latest technology, cost, personality and scope, situation, and purpose of personal information processing	Prevention of infringement of rights and freedoms
	②	Controllers must implement privacy principles such as minimizing data processing in an effective way and meet the requirements of the GDPR when determining and processing means	Minimize processing
	③	Appropriate technical and management measures should be implemented to protect the rights of data subjects (designed to include safety measures necessary for processing, such as pseudonym)	Protection of information subject rights and safety measures
2	①	Data protection should be basically applied when setting the amount of personal information collected, processing range, personal information storage period, and accessible period	Protection based on preferences
	②	Data protection should be basically applied when setting the amount of personal information collected, processing range, personal information storage period, and accessible period	Application part and viewpoint
	③	Ensuring that personal information is not viewed by an unspecified number of people without the intervention of the information subject through by default	Restriction of access

및 설정의 원칙을 충족하여야 한다고 규정하고 있다. 구체적인 내용은 Table 1과 같다.

전반적으로 정보주체의 권리와 자기결정권을 강화하고 사업자들을 대상으로 사용자 중심의 안전조치 의무 및 보호의 기본 설정 등을 규정하고 있다. EU는 이러한 규정을 이행하기 위해서는 개인정보 영향평가(DPIA : Data Protection Impact Assessment)등을 통해 대부분이 준수될 수 있으며 유럽 네트워크 정보보호원(ENISA : European Union Agency for Cybersecurity)을 통해 PbD구현을 위한 PET기술에 대한 보고서[8,9,10]를 지속적으로 발간하고 있어 개인정보보호책임자(DPO : Data Protection Officer)의 관리 하에 지속적인 운영과 관리를 통해 보완해 나가는 것이 바람직하다고 제시하고 있다.

2.2.2 미국

미국은 연방거래위원회(FTC : Federal Trade Commission)를 통해 PbD를 언급하고 있다. 위원회에서는 정보보안, 합리적인 수집 제한, 안전한 정보보유, 정확성 등과 같은 프라이버시 보호 원칙을 관행에 도입해야 한다고 제시하고 있으며 이러한 원칙을 이행하기 위해 상품 서비스의 생애주기 전체에 걸쳐 종합적으로 관리해야 한다고 제시하고 있다. 또한, 캘리포니아에서 '20년 시행된 CCPA(Consumer Privacy Act)에서는 소비자가 개인정보 삭제를 요구하거나 정보를 공유하지 말라고 요구할 권리와 개인정보를 최소한으로 수집하고, 소비자가 사후동의를 통해 중단을 요구하는 것이 아니라 소비자가 원할 때만 정보를 수집할 수 있도록 규정하고 있다. 유럽의 GDPR만큼 광범위하지는 않으나 최근 발생하는 다양한 개인정보 이슈와 관련하여 PbD를 통해 개인정보 및 프라이버시를 보호하려는 노력이 강화되고 있다. 미국도 유럽과 마찬가지로 미국표준기술연구소(NIST : National Institute of Standards and Technology)를 통해 PET기술 지침[11]을 마련하고 있다.

2.2.3 영국

영국은 정보보호감독국(ICO : Information Commissioner's Office)을 통해 PbD를 언급하고 있다. 구체적으로는 개인정보 및 프라이버시 보호가

어떤 프로젝트의 초기 단계 또는 전체 프로젝트에 대해 PbD가 핵심적인 고려사항임을 강조하고 있으며 이를 구현하기 위한 체크리스트를 Table 2와 같이 제공하고 있다.

그 외 최근 온라인 유해물 규제에 대해서 페이스북, 유튜브와 같은 플랫폼 사업자가 불법적이거나 유해한 콘텐츠 확산을 막는데 '주의 의무(Duty of care)'를 다하지 않으면 규제 당국이 전 세계 매출의 10%까지 벌금을 부과하거나 영국 내 서비스를 차단하는 방안을 마련하고 있다[12]. 특히 어린이에게 안전한 인터넷 환경 이용을 위한 The Children's Code를 도입하여 부모 인증 기능을 Default로 제공하거나 아동정보의 최소 수집 등의 사항을 권고하고 있으며 'The Children's code design guidance'라는 실행규정[13]을 제공하고 있다. 영국은 이러한 PbD 정책을 통해 프라이버시 침해 위험을 최소화 하고 신뢰를 쌓을 수 있으며, 기업에서도 법적 의무를 더 잘 이행하고 위법을 감소할 수 있을 것으로 판단하고 있다.

Table 2. PbD Checklist Provided by ICO

Data protection issues are managed enterprise-wide, including systems, services, products, and business practices
Data is protected by making the core functions of processing systems and services essential components
Predict in advance and take measures to prevent damage before risk and personal information infringement occurs
Data is used only to process personal data necessary for the purpose and to achieve that purpose
It provides the identity and contact information of those responsible for data protection within the organization
All documents are presented in English so that it is easy to understand the work performed with the data of the data subject
Tools are provided to information subjects so that they can check how and policies to use data
It respects information subject data by providing strong privacy preferences and user-friendly options
Guaranteed technology is used to protect data by design
Data protection by design obligations is observed using Personal Information Protection Enhancement Technology (PET)

2.2.4 프랑스

프랑스는 정보보호위원회(CNIL : Commission Nationale de l'Informatique et des Libertés)를 통해 개인정보관리자로 하여금 PbD 원칙을 채택하도록 하는 다양한 정보보호에 관한 지침을 발간하고 있다. '10년 디지털 잊힐 권리(Droit à l'oublie numérique)에 관한 보고서[14]를 통해 PbD원칙에 대한 이행을 권고하였으며, '16년 디지털공화국법을 제정하여 개인의 권리 확장 및 개인정보 자기결정권과 통제권을 강화하였다. 그 외 '18년 인공지능의 윤리 및 데이터 보호에 대한 새로운 선언문[15] 및 2018년 커넥티드 차량의 개인정보 관련 패키지 보고서[16]를 통해 인공지능 시스템에 기본적으로 PbD원칙을 적용하여야 하며, 커넥티드 차량의 경우 차량 이용자들의 데이터에 대한 투명성과 통제권을 확보할 수 있도록 제시하고 있다. 한편 '21년에는 설계 디자인을 통해 아동의 정보와 권리를 강화하는 지침을 마련하기도 하였다[17].

2.2.5 일본

일본은 개인정보 보호 관련 법률에서 PbD에 대한 개념을 담고 있지는 않다. 다만, PbD의 개념은 데이터 주체의 개인정보 보호 권리를 보호하는데 유용하다고 제시를 하고 있다. '20.6월, 일부 개정된 개인정보보호법[18]을 살펴보면 정보주체의 권리 보호를 실효화하고, 국내외 사업자에 대한 위원회의 관리·감독권 강화 및 데이터의 이용과 활용을 촉진하는 내용이 포함되었다. 여기서 PbD의 개념과 유사한 개인의 권리 및 이익을 보호하기 위해 충분한 조치를 해야 한다는 규정과 개인정보의 보호와 이용의 균형을 맞추는 것이 계속해서 중요하다는 것, 그리고 정보주체에게 자신의 개인정보가 어떻게 처리되고 있는 지 파악할 수 있도록 공표 및 통지의 의무를 지정하고 있어 PbD의 원칙을 기반으로 하여 법률이 개정되었다고 분석할 수 있다. 그 외 개인정보처리시스템 개발 시 프라이버시 영향 분석 등을 수행하도록 하고 있어 이러한 개념이 데이터 주체의 개인정보 권리를 보호하는 데 유용한 것으로 이해된다.

2.3 시사점

EU와 미국 정부는 국가마다 다르게 취급하는 프라이버시 보호 기준을 균일하게 맞춰보려는 노력을 공조하고 있다고 정보 정책 리더십 센터장인 보자나 벨라미(Bojana Bellamy)는 이러한 프라이버시 보호 수준을 법과 문화의 측면에서 프라이버시에 대한 균등하고 통일된 이해를 바탕에 두기 위해 PbD가 사용되고 있다고 밝히고 있다[19].

유럽 및 국제단체에서는 프라이버시를 위기관리의 관점에서 바라보고 있으며, 개인정보에 포함된 프라이버시의 잠재위험이 무엇인지, 그런 위험에 대비한 위기관리의 역할이 무엇인지, 어떻게 해야 최고의 결과를 만들 수 있을지, 개인정보를 보호하면서도 사용의 효율성을 높일 수 있는 방법에는 무엇이 있을지에 대한 연구를 지속적으로 진행하고 있다.

한편 미국 연방 정부들 또한 이미 PbD를 어느 정도 적용한 운영을 보이고 있으며, NIST에서는 그 일환으로 프라이버시에 대한 일반 사용자의 이해도를 높이기 위해 보고서나 문건을 만들 때 법적 설명을 줄이고 최대한 간편하고 친절하게 풀어놓으려는 노력을 보이고 있다. 연방거래위원회는 기존의 연방정보처리 표준을 재해석해서 적용하고 있다. 즉, PbD 개념을 포함시키고, 사용하기 쉬운 프라이버시 옵션들을 사용자에게 제공하며, 정보 접근 과정의 투명성을 구축한다는 것이다.

국내에서도 최근 개인정보 유출사고 및 가명정보를 AI로 학습하면서 발생한 오남용 사례가 발생하였다. 정보주체로부터의 목적 외 활용 동의 및 개인식별가능정보, 민감정보 등에 대한 비식별 조치 미흡으로 서비스를 중단하였는데, 개인정보보호위원회에서는 이러한 문제점을 보완하기 위해 '21.6월, '인공지능(AI) 개인정보보호 자율점검표[20]'를 발간한 바 있다. 그 외 PbD에 기반한 스마트도시 구축을 위한 개인정보 6대 원칙[21] 등을 제시함으로써 개인정보 활용에 대한 PbD개념을 필수적으로 탑재할 수 있도록 노력하고 있다.

이처럼 PbD의 개념을 법제화하고 정책을 제시하고 있는 주요국 외에 호주, 싱가포르, 캐나다, 중국 등에서도 법제 개정 시 이러한 개념을 포함하고 있어 향후 PbD에 대한 적용방안 및 기술 연구가 더욱 활발할 것으로 예상된다.

III. 가명처리 생명주기별 고려사항

우리는 본 논문의 2장을 통하여 개인정보 PbD에 대한 원칙과 국내외 현황을 살펴보았다. 본 연구는 개인정보로 취급되는 가명정보를 처리 시 개인정보보호를 중심으로 고려할 사항을 새로운 PbD원칙으로 제시하는데 목적이 있다. 본 장에서는 ITU 국제표준에서 제시하고 있는 비식별정보에 대한 생명주기를 살펴보고, '21년 국내에서 개정된 '가명정보 처리 가이드라인[5]'과의 비교를 통해 가명정보 처리에 대한 생명주기별 고려사항을 정립하고자 한다.

3.1 국제표준 비식별 조치 생명주기

'16년 금융보안원과 KISA는 '개인정보 비식별 조치 가이드라인[22]'을 기반으로 ITU-T X.fdip[1]를 추진하여 '20년 승인되었으며, ISO/IEC 20889 표준[23]에 따라 비식별 조치된 정보의 안전성 보장 및 고려해야할 요구사항을 1.데이터상황과 2.데이터 위험도로 분류하여 기존 절차를 개선한 ITU-T X.rdda[2]를 '19년 추가로 추진하여 '21년 승인된 바 있다.

X.rdda는 Fig.1처럼 비식별정보의 생명주기를 데이터 주체로부터의 데이터 수집, 관리, 사용, 공개 및 활용, 파기로 정의하고 있는데 실무에서의 비식별 정보를 처리하는 절차와 매우 유사하여 바람직하게 구현되었다고 볼 수 있다. 동 표준은 단계별로 비식별 조치 시 고려사항을 Table 3과 같이 제시하고 있다.

즉, Table 3과 같은 고려사항을 검토하여 생명주기 전반에 비식별 조치를 수행할 경우 개인정보 노출 위험을 최소화하고 데이터 공유를 상당히 쉽게 수행할 수 있다고 제시하고 있다. 다만 고려사항별로 구체적으로 어떤 것을 검토해야 하는지에 대한 내용은 제시하고 있지 않으며, 익명정보를 기반으로 제시된 표준이기 때문에 가명정보 처리를 기준으로 볼 때 추가적인 보완사항들이 남아있다는 단점이 있다.

1) ITU-T X.fdip(Framework of de-identification process for telecommunication service providers, 16년 채택, 20년 승인.
2) ITU-T X.rdda(Framework of de-identification process for telecommunication service providers, 19년 채택, 21년 승인.

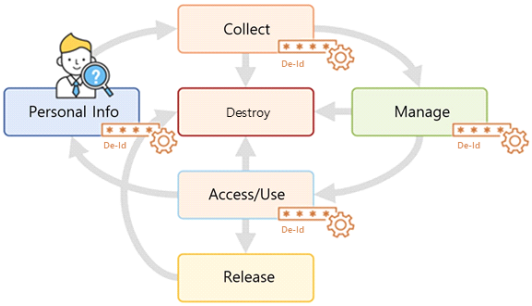

Fig. 1. The De-identification processing point according to the data Life-Cycle

Table 3. ITU-T X.rdda Life-Cycle Considerations

Collect		Delete the minimum items and unnecessary personal information necessary for the purpose
Manage		Data conversion through de-identification measures to achieve data utilization goals
Use	Release	Additional de-identification measures are taken in consideration of the environment using the converted data
	Access	De-identification measures are taken into account the data analyst's environment and other information
Destroy		The destruction of the unidentified information whose purpose has been achieved is completed

3.2 가명정보 처리 가이드라인과 국제표준의 비식별 조치 생명주기 고려사항 비교

'가명정보 처리 가이드라인[5]'은 개인정보를 가명처리 시 고려해야 할 사항을 1단계(사전준비), 2단계(가명처리), 3단계(적정성검토 및 추가가명처리), 4단계(활용 및 사후관리)로 제시하고 있으며, 2단계(가명처리)의 경우 4단계로 세분화하여 위험도 측정에 대해 구체적인 절차를 제공하고 있다. 이와 같은 가명정보 처리 가이드라인과 앞서 살펴본 국제표준에서 제시하고 있는 비식별정보 생명주기별 고려사항을 비교하면 Fig.2와 같이 나타낼 수 있다.

Fig.2를 구체적으로 살펴보면 1단계(사전준비)는

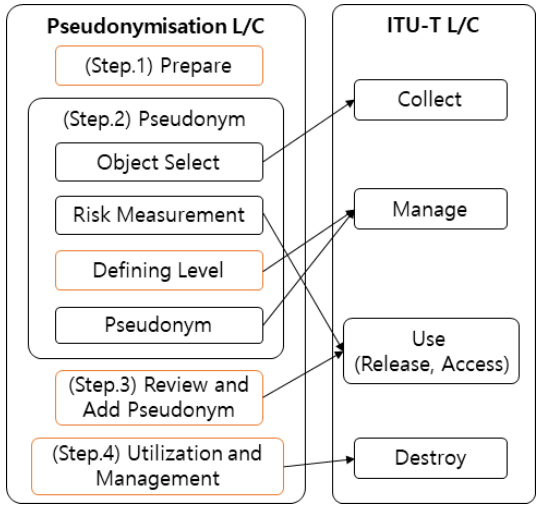

Fig. 2. Comparison of the Guidelines for processing pseudonym information and considerations of the international standard Life-Cycle

가명이 아닌 익명정보 처리를 기반으로 하는 국제표준에서는 매핑할 수 있는 절차가 없다. 가명정보의 경우 법률에서 규정한 목적 내에서만 활용해야하기 때문에 사전준비를 통해 위법사항 등을 검토하지만 익명정보의 경우 별도의 활용 목적 제한이 없기 때문에 비식별 처리 전 사전에 검토하는 절차가 필요 없다. 2단계(가명처리)의 대상선정은 가명·익명처리 목적을 달성하기 위한 최소한의 항목 이용 및 원본에서 데이터 추출 시 불필요한 개인식별정보를 사전에 삭제하는 절차로 국제표준의 '데이터 수집'과 매핑을 할 수 있다. 그리고 위험도 측정은 데이터 사용 시 공개 또는 이용하는 자에 대한 위험도를 측정하는 절차로 국제표준의 '데이터 사용'에 매핑이 된다. 처리수준 정의와 가명처리는 실제로 가명·익명처리를 수행하는 절차로 국제표준의 '데이터 관리'에 매핑을 할 수 있으며, 3단계(적정성검토 및 추가가명처리)는 데이터가 사용되는 환경에 따라 가명처리가 적절히 처리되었는지 검토하는 절차로 추가 비식별 조치를 적용할 수 있는 국제표준의 '데이터 사용'과 매핑할 수 있다. 마지막으로 4단계(활용 및 사후관리)는 국제표준의 '데이터 파기'단계와 매핑을 할 수 있다.

이처럼 국제표준과 가명정보 처리 가이드라인 모두 익명·가명처리 시 제시하는 고려사항이 유사하지만 가명정보의 경우 개인정보 보호법에서 규정하는 개인정보의 일부로써 충분한 가명처리가 적용되지 않을

경우 쉽게 재식별이 가능하기 때문에 활용 목적을 제한하고, 개인정보에 준하는 관리적·기술적 보호조치를 수반하여야 한다. 결론적으로 가명정보 처리는 일반적인 개인정보 및 익명정보 처리와 다른 개념에서의 정보보호 원칙을 적용할 수 있는 방안이 필요하다는 것을 파악할 수 있다.

IV. 가명처리 생명주기 기반의 PbD제안

4.1 가명처리 PbD 제안

앞서 3장을 통해 제시한 가명정보 처리 생명주기에 따라 개인정보 PbD원칙을 적용할 수 있는 절차를 살펴보면 다음과 같다. 1단계(사전준비)는 가명처리 시스템에 대한 보안사항으로 기본설정을 정의할 수 있는 'By Default'원칙과 매칭되며, 2단계(가명처리)는 가명정보에 대한 위험도를 측정하여 사전에 위험을 제거하는 'Proactive/Preventative' 및 가명처리 시 최초 설계를 수행하고 가명처리를 하는 'Lifecycle Protection'원칙과 매칭이 된다. 3단계(적정성검토 및 추가 가명처리)는 검증된 가명처리 기술을 적용해야 하는 'Embedded'원칙 및 가명정보의 안전성과 유용성 검토를 수행하는 'Positive Sum'의 원칙과 매치가 된다. 마지막 4단계(활용 및 사후관리)는 가명정보 처리의 문서화와 가명처리 내역을 공개하여 투명성을 제고하는 'Visibility/Transparency'원칙 및 가명정보의 재식별이 발생할 경우 정보주체에게 대응을 하거나 통지를 하여 추가 사고를 예방할 수 있는 'Respect for User'원칙과 매칭이 될 수 있다. 이러한 사항들을 종합해보면 Fig.3과 같이 나타낼 수 있으며, Fig.3에서 제안하는 원칙을 개인정보 PbD원칙 관점에서 비교해보면 다음과 같다.

첫 번째 원칙인 'Proactive/Preventative'는 개인정보 처리 시 사고 발생 후 조치를 하는 사후대응이 아닌 사전에 대비(예방)를 할 수 있도록 하고 있는데 가명정보의 경우 잠재적인 재식별 사고 등을 예측하기가 매우 어렵기 때문에 가명처리에 대한 위험도 측정을 통해 법률에서 정한 목적 범위 내에서 활용이 되는 것인지, 데이터가 불특정 다수한테 공개가 되는 것인지 등의 데이터를 활용하는 환경에 따라 위험도를 측정하는 것이다. 즉 위험도 측정을 통해 가명처리 수준을 사전에 결정하고 처리를 수행함으로써 가명정보 활용에 관한 재식별 등의 사고를 사전에

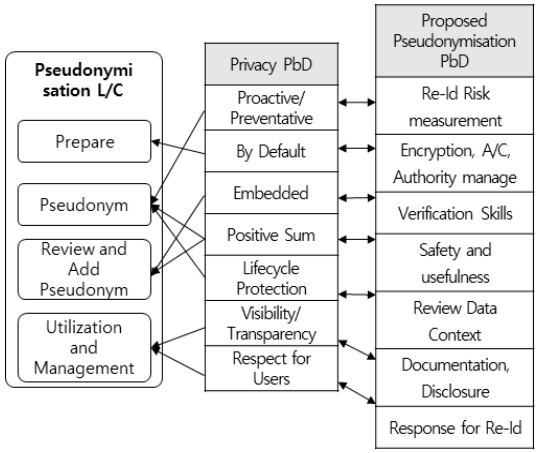

Fig. 3. Relationship between the personal information PbD principle and the proposed pseudonym processing PbD principle

예방할 수 있다. 김동현 등[24]이 제안한 바에 따르면 데이터 상황 기반의 위험도 측정 방법을 활용하는 경우 불충분한 가명처리 및 민감정보 등의 특이치에 대한 검토를 통해 사전에 재식별 가능성을 제거 또는 감소할 수 있다.

두 번째 원칙인 'By Default'의 경우 프라이버시 보호를 위한 기능을 기본적으로 활성화시켜야 한다는 것으로 가명처리 관점에서는 가명처리 대상인 개인정보를 수집 시 암호화 조치를 수행하거나 처리를 수행하는 시스템에 대해서는 기본적으로 접근통제와 접근권한 관리 등의 기술적 보호조치가 기본적으로 적용되어 있어야 한다. 이러한 조치를 적용할 경우 원본 DB에서 가명처리 대상 데이터를 추출하여 보관하는 시스템에 대한 안전성을 보장할 수 있다.

세 번째 원칙인 'Embedded'는 시스템의 설계 시 프라이버시 보호가 이미 전제되어야 한다는 의미로 가명처리 관점에서는 ISO/IEC 20889[23] 및 ENISA[9,10], 국내 가명처리 가이드라인[5] 등에서 제시하고 있는 검증된 가명처리 기술을 사용해야 한다. 기술적으로 검증되지 않은 가명처리의 경우 다양한 연산을 통해 쉽게 재식별이 가능하다.

네 번째 원칙인 'Positive Sum'은 프라이버시와 보안에 대한 상생을 의미하는데 가명처리 관점에서도 개인정보와 유사하게 가명처리 수준이 강화될수록 데이터는 안전해지지만 유용도는 떨어지는 반비례 관계가 있다[11]. 따라서 데이터의 안전성만이 아니라 유용성도 고려하여 안전하고 유용한 데이터를 만들

수 있도록 검토하여야 한다. 이와 같은 절차는 '가명정보 처리 가이드라인[5]'의 2단계(가명처리)단계의 목적과 3단계(적정성검토 및 추가가명처리)에서 제시하고 있는 가명정보의 처리 목적의 달성가능성을 검토하는 절차와 유사하다.

다섯째 원칙인 'Life-Cycle Protection'은 개인정보처리시스템을 설계하는 단계부터 폐기하는 단계까지 End-to-End로 지속적으로 관리가 돼야 된다는 의미로 가명정보 또한 처리하는 과정을 전반적으로 관리할 수 있는 절차가 필요하다. 이를 위해 개인정보보호법 제28조의4에서는 가명정보에 대한 안전조치의무 등을 규정하고 있어 가명정보 활용 시 내부관리계획 또는 가명정보처리 내부지침 등을 마련하도록 하고 있다. 또한 가명정보는 활용되는 목적과 제공받는 대상이 정해져 있기 때문에 사전준비 단계를 통해 내부관리계획 등 수립 시 전체 생명주기에 걸친 검토사항을 사전에 정립하는 것이 가능하다. 가명처리가 된 정보는 단일정보로써 특정 개인을 식별할 수 없는 정보지만 추가정보[3] 등을 통해 재식별 가능성이 존재하므로 사후 모니터링을 주기적으로 실시해주는 절차도 필요하다.

여섯째 원칙인 Visibility/Transparency는 개인정보 처리에 대한 가시성과 투명성을 보장하는 것으로 가명정보 처리 시 사전준비 단계에서 생성한 문서 및 가명처리 수준 결정에 대한 검토의견, 처리 절차별 결과 등을 문서화하고 처리 내용을 홈페이지 등을 통해 공개하는 것으로 대체할 수 있다. 또한 가명정보도 개인정보로 판단하기 때문에 개인정보 보호법 제30조에 따른 가명정보 처리방침 등을 마련하여 공개하여야 한다. 이를 통해 정보주체로부터의 가명처리에 대한 신뢰성을 제고할 수 있다.

일곱째 원칙인 'Respect for Users'는 PbD를 활용한 설계는 정보주체로부터의 개인정보 수집되고 활용되는 동안 정보주체의 자기결정권을 우선적으로 제공하여 사용자 중심이 유지되어야 한다는 의미로 가명정보도 단일 정보만으로는 특정 개인을 알아볼 수 없는 정보지만 추후 재식별이나 가명정보 활용으로 인한 사회적 이슈가 발생하였을 경우 정보주체에게 이를 통지하고 사용을 중지하는 절차가 필요하다. 따라서 가명처리 대상이 되는 개인정보에 대해 정보주체가 가명처리에 대한 처리 중지를 요구할 수 있는 절차와 재식별 사고의 경우 특정 정보주체에게 알려줄 수 있는 기능을 구현해야 할 것이다. 다만, 현재 개인정보 보호법에서는 추가정보를 사용한 재식별을 법률로써 금지하고 있어 제도 개선이 함께 보완되어야 구현이 가능하다.

3) 개인정보의 전부 또는 일부를 대체하는 데 이용된 수단이나 방식(알고리즘 등), 가명정보와의 비교·대조 등을 통해 삭제 또는 대체된 개인정보 부분을 복원할 수 있는 정보(매핑 테이블 정보, 가명처리에 사용된 개인정보 등)

4.2 타당성 검토

본 절에서는 제안하는 가명처리 PbD원칙에 대한 필요성과 기존 개인정보 PbD와의 유사성 및 적합성을 검토해보고자 타당성 조사를 시행하였다. '21.12.27.~'22.1.7.까지 개인정보 PbD원칙을 알고 있는 학계 4명, 산업계 26명 등의 개인정보보호 관련 분야 전문가(평균 11년 경력)를 대상으로 설문조사를 실시하였으며, 리커트 5점 척도를 사용하여 질의에 대한 응답점수가 높을수록 영향도가 높은 것으로 판단하였다. 또한 IBM SPSS 26.0을 이용하여 신뢰도 분석을 실시하고, 분석결과에 대한 Cronbach α값은 각 항목별 결과와 함께 제시하였다. 구체적인 조사결과는 Table 4와 같다.

대부분의 응답자들은 PbD적용에 대해 안전한 개인정보 활용에 도움이 된다고 응답하였으며, 가명정보 처리에 관해서도 이러한 원칙이 제공되면 좋겠다는 응답을 하였다. 그 외 PbD구현을 위한 PET기술 지침도 사례를 기반으로 제공이 된다면 많은 도움이

Table 4. The results of a feasibility study on the PbD principle for processing the proposed pseudonym

Question	Ave Score
Will personal information PbD help you use personal information safely?	4.5
If the PbD principle for pseudonym processing is proposed, would it be helpful for pseudonym processing?	4.2
If PET guidelines according to pseudonym processing PbD are prepared, would it be helpful for pseudonym processing?	4
Are the concepts of the proposed Pseudonym processing PbD principle and the personal information PbD principle properly mapped and appropriate?	Ref, table 5

될 것이라는 응답이 있었다. 타당성 조사 결과에 대한 Cronbach α값은 0.730으로 비교적 높은 수준의 신뢰도를 나타내었다. 다음으로 제안하는 가명처리 PbD와 기존 개인정보 PbD원칙의 유사성 및 적합성은 7개의 설문으로 세분화하여 Table 5와 같이 진행하였다.

제안하는 가명처리 PbD원칙과 개인정보 PbD 원칙과의 유사성 및 적합성에 대한 조사결과는 평균 3.7점, Cronbach α값은 0.811로 높은 수준의 신뢰도를 나타내었다. 유사성 및 적합성이 3점 이하로 도출된 항목의 전문가 의견으로 'By Default'의 경우 개인정보처리시스템 개발 단계에서 구현되는 항목은 맞지만 가명처리는 시스템적인 부분보다는 계약 등의 관리적인 요소로 보완이 가능하기에 기술적 보호조치만이 아닌 관리체계 등에 대한 인증심사 기준을 적용하는 방안이 새롭게 제시되었으며, 'Embedded'의 경우 표준으로 제시되고 있는 기술 외에 다양한 기술들이 활용되기 때문에 검증된 기술을 찾아 적용하는 것이 어려울 것이란 의견이 있었다. 그리고 'Life-Cycle Protection'의 경우 가명정보는 사후 모니터링이 어렵기 때문에 원칙 적용이 힘들 것이라는 의견이 있었다. 그 외 원칙들에 대해서는 유사성 및 적합성이 비교적 준수하다는 의견이 있었으며, 실증을 통해 제안하는 원칙을 검증해야 한다는 의견도 있었다.

Table 5. Similarity and suitability results between the proposed pseudonym processing PbD principle and the personal information PbD principle

No	Privacy PbD	Proposed PbD	Ave Score
1	Proactive/ Preventative	Re-Id Risk measurement	4.5
2	By Default	Encryption, A/C, Authority manage	3
3	Embedded	Verification Skills	2.7
4	Positive Sum	Safety and usefulness	4
5	Life-Cycle Protection	Review Data Context	3.7
6	Visibility/ Transparency	Documentation, Disclosure	4.5
7	Respect for Users	Response for Re-Id	4

Table 6. Supplement to the proposed pseudonym processing PbD principle

No	Privacy PbD	Complemented pseudonymisation PbD principle	Score (before and after)
2	By Default	Certification of the Pseudonym System Compliance with legal technical/administrative requirements	3 → 4.5
3	Embedded	Technology verification through adequacy review	2.7 → 4
5	Life-Cycle Protection	Prepare and Monitoring additional consent for pseudonym	3.7 → 4.4

상기와 같은 1차 조사에서 발견된 문제점들에 대해 Table 6과 같이 최초 제안한 가명처리 PbD 원칙의 내용을 보완하여 2차 조사('22.1.13.~'22.1.21.)를 실시한 결과 유사성에 대한 전체 평균 4.2점, Cronbach α값은 0.732로 준수한 수준의 결과를 도출할 수 있었으며, 새롭게 제시한 원칙에 대해 t-검정을 실시한 결과 'By Default'는 -9.542, 'Embedded'는 -7.779, 'Life-Cycle Protection'은 -4.583으로 유의확률은 모두 0.000이 도출되어 매우 유의미한 결과가 도출되었다.

V. 결 론

'17년 구글의 모회사인 사이드워크랩스는 캐나다 토론토에 스마트시티를 건설하는 프로젝트를 진행하였다. 이 사업은 자율주행, 에너지/교통, 쓰레기 수거, 건물관리 등 사회의 다양한 문제를 첨단기술을 도입하여 해결하고자 하는데 시작하였으나 '22년 현재에도 개인정보보호 정책 등에 반한다는 이유로 전문가들 사이에 논란이 되고 있다. 특히 PbD원칙을 제안한 Ann Cavoukian교수가 프라이버시 위원장으로도 참여하였지만 다양한 데이터가 융복합되어 활용되는 환경에서 발생할 수 있는 재식별 가능성 등의 이슈로 사임을 발표했다. 이처럼 PbD에 대한 원칙은 제시되고 있으나 실제로 PbD를 현실세계에 적용하기란 쉬운 것이 아니다.

최근 국내외에서는 이러한 사전적 예방을 위한 PbD개념을 법률에서 규정하고, 인공지능(AI) 및 스마트시티 구축 시 고려하여야 할 원칙 등을 제시하고 있어 개인정보 분야에서의 PbD적용은 매우 긍정적으로 작용하고 있다. 그러나 빅데이터 활용에 중심이 되는 가명정보에 대해서는 가이드라인에서 제시하는 절차 외에 사전에 안전성을 담보하기 위한 별도의 연구는 현재까지 제시되고 있지 않다. 가명정보는 개인정보로 취급되는 정보지만 개인정보와는 생명주기가 다르며 적용해야 하는 기술 및 고려사항도 다르게 판단하여야 한다.

우리는 본 연구를 통해 가명정보 처리에 대한 생명주기를 정립하고, 이러한 생명주기에 근거한 가명처리 PbD 원칙을 새롭게 제안하였다. 제안한 원칙은 기존 개인정보 PbD원칙에서 제시하고 있는 틀을 유지하고자 하였다. 향후 본 논문에서 제안한 사전적 원칙이 가명처리 시 고려해야할 사항으로 제시가 된다면 가명정보 활용 시 다양한 산업에서 보다 체계적으로 가명처리를 수행하고 활용과정에서도 정보주체로 하여금 활용에 대한 신뢰성 제고에 기여할 수 있을 것으로 사료된다.

이번에 제안한 가명처리에 대한 PbD원칙의 대한 타당성 및 적합성을 제고하기 위해 기업을 대상으로 실증을 수행하는 후속연구를 수행할 예정이다.

References

〔1〕 Ann Cavoukian, “Privacy by Design by Regulation”, International Conference of Data Protection and Privacy Commissioners, 2010

〔2〕 Soo-jeong Kim, “Big Data, Data Ownership and Data Economy”, The Korean Journal of Civil Law, 96, pp. 3-40, Sep. 2021

〔3〕 Mackey, E and Elliot, M. J, “Understanding the Data Environment,” XRDS: Crossroads, 20(1), pp. 37-39, 2016

〔4〕 Hye-seon Choi, “New Trend of Personal Information Protection -Concept on Privacy by Design-”, Ilkam Law Review, 24, pp. 305-340, 2013

〔5〕 Personal Information Protection Commission, “Guidelines for processing pseudonym information”, Dec. 2021

〔6〕 Na-roo Kim, “A Study on the Introduction and Application of Privacy by design”, SungKyunKwan Law Review, 29(4), pp. 1-30, Dec. 2017

〔7〕 General Data Protection Regulation, Regulation (2016) 2016/679 of the European Parliament and of the Council, Regulation (EU), 2016

〔8〕 ENISA, ″Privacy and Data Protection by Design – from policy to engineering″, Dec. 2014

〔9〕 ENISA, ″Pseudonymisation techniques and best practices″, Nov. 2019

〔10〕 ENISA, ″DATA PSEUDONYMISATION: ADVANCED TECHNIQUES & USE CASES″, Jan. 2021

〔11〕 Simson L. Garfindel, “NIST IR8053 De-Identification of Personal Information”, NIST, Oct. 2015

〔12〕 Sung-ho Jin, “UK Announces Policy Directions for Online Hazardous Products Distribution”, KISDI Policy trends, Jan. 2021

〔13〕 Information Commissioner′s Office, ″Age appropriate design : a code of practice for online services″, Sep. 2020

〔14〕 Rolf H.Weber, ″The Right to Be Forgotton″, Journal of Intellectual Property, Information Technology and E-Commerce Law, 2010

〔15〕 CNIL, “HOW CAN HUMANS KEEP THE UPPER HAND? The ethical matters raised by algorithms and artificial intelligence”, Dec. 2017

〔16〕 CNIL, “CONNECTED VEHICLES AND PERSONAL DATA”, COMPLIANCE PACKEGE, Oct. 2017

〔17〕 Pedro hartung, “The children’s rights-by-design standard for data use by tech companies”, UNICEF Global Insight Issue brief, 5, Nov. 2020

[18] Japan, "Personal Information Protection Act", Jun. 2020
[19] Bojana Bellamy, "Global Data Privacy Law and Practice-Looking Around the Corners", Centre for Information Policy Leadership, Dec. 2014
[20] Personal Information Protection Commission, "AI Personal Information Protection Autonomous Checklist", Jun. 2020
[21] Personal Information Protection Commission, "Six principles of personal information for the establishment of a smart city", Dec. 2020
[22] Office for Government Policy Coordination at el, "Guidelines for De-identification of Personal Data", Jun. 2016
[23] ISO/IEC 20889. "Information technology -Security techniques-Privacy enhancing data de-identification techniques", Nov. 2018
[24] Dong-hyun Kim and Soon-seok Kim, "A New Scheme for Risk Assessment Based on Data Context for De-Identification of Personal Information", Journal of The Korea Institute of Information Security and Cryptology, 30(4), pp. 719-734, Jun. 2020

〈저자소개〉

김 동 현 (Dong-hyun Kim) 종신회원
2013년 2월: 동국대학교 정보보호학 석사
2022년 2월: 중앙대학교 융합보안학 박사 졸업
2010년 10월~현재: 한국인터넷진흥원 데이터활용지원팀 책임연구원
〈관심분야〉 개인정보보호, 가명·익명처리, 데이터 위험관리