연구보고서 2016-001

개인정보 비식별화기술의 쟁점 연구

A Research on De-identification Technique for Personal Identifiable Information

이 현 승 / 송 지 환

2016. 8

SPRi 소프트웨어정책연구소

이 보고서는 2016년도 미래창조과학부 정보통신진흥기금을 지원받아 수행한 연구결과로 보고서 내용은 연구자의 견해이며, 미래창조과학부의 공식입장과 다를 수 있습니다.

목 차

제1장 서 론 ····· 1

제1절 연구배경 ····· 1

제2절 연구범위 ····· 3

제2장 빅데이터의 등장과 활용 ····· 4

제1절 빅데이터란 ····· 4

제2절 빅데이터 관련 기술 ····· 6

제3절 빅데이터의 활용 사례 ····· 8

제4절 빅데이터 산업 전망 ····· 10

제5절 데이터 브로커 산업 전망 ····· 12

제6절 빅데이터 활용과 개인정보 보호와의 관계 ····· 15

제3장 개인정보의 권리화와 보호강화 ····· 16

제1절 프라이버시권과 개인정보 자기결정권의 등장 ····· 16

제2절 국내 개인정보 자기결정권의 발전 ····· 18

제3절 정보통신의 발달과 개인정보 보호의 강화 ····· 22

제4절 빅데이터 활용과 개인정보 비식별화 ····· 25

제4장 개인정보 비식별화기술 ····· 28

제1절 비식별화기술의 개요 ····· 28

제2절 비식별화기술의 불완전성 ····· 35

제3절 비식별화기술의 평가척도 ····· 37

제4절 비식별화 소프트웨어의 현황 ····· 44

제5장 비식별화기술의 쟁점과 동향 ····· 50

제1절 비식별화기술과 개인정보법령 준수여부 ····· 50

제2절 정부의 개인정보 관련 정책동향 ····· 54

제3절 일본의 개인정보 관련 정책동향 ····· 56

제6장 제도개선의 시사점 ····· 59

제1절 개인정보 정의와 범위의 개선 ····· 59

제2절 비식별정보 유통의 관리체계 ····· 60

제3절 법체계와 컨트롤타워의 재정립 ····· 61

표 목 차

<표 1-1> 개인정보의 경제·사회적 가치 유형 ····· 2

<표 2-1> 빅데이터 요소 기술 구성 및 분류 ····· 7

<표 2-2> 세계 빅데이터 시장 규모 ····· 11

<표 2-3> 국내 빅데이터 시장 규모 ····· 11

<표 2-4> 미국 9개 데이터 브로커 ····· 13

<표 2-5> 미국 9개 데이터 브로커의 매출구성 ····· 13

<표 3-1> 각국 개인정보 규제 비교 ····· 20

<표 3-2> 개인정보 자기결정권의 세부내용 ····· 21

<표 3-3> 국내 개인정보 유출사건 일람표 ····· 23

<표 3-4> ISO 27001과 ISMS의 비교 ····· 24

<표 3-5> PIMS와 PIPL 통합 변경사항 ····· 24

<표 3-6> 개인정보보호법과 정보통신망법의 제재규정 ····· 25

<표 4-1> 비식별화기술 17종 위험도 비교 ····· 34

<표 4-2> 매사추세츠주 공무원 병원 진료기록 ····· 36

<표 4-3> 매사추세츠주 케임브리지시 선거인 명부 ····· 36

<표 4-4> 비식별화 소프트웨어 현황 ····· 45

<표 5-1> 일본 개정 개인정보보호법의 주요 정의규정 ····· 57

<표 6-1> 개인정보 관련 법 비교 ····· 62

그 림 목 차

[그림 2-1] 빅데이터의 6V 특성 ····4
[그림 2-2] 빅데이터의 가치창출 과정 ····5
[그림 2-3] 빅데이터 처리 프로세스 ····7
[그림 2-4] 개인정보 통제권에 대한 긴장관계 ····15
[그림 4-1] 데이터 식별가능성 스펙트럼 ····28
[그림 4-2] 무작위화와 일반화의 예 ····30
[그림 4-3] 비식별화 주요 기술 17종 ····31
[그림 4-4] 17종 비식별화기술의 재분류 ····33
[그림 4-5] 비식별 조치 및 사후관리 절차 ····38
[그림 4-6] 비식별화 이후 3-익명성 보장 ····39
[그림 4-7] 비식별화 이후 3-다양성 보장 ····41
[그림 4-8] 2-다양성과 쏠림공격 ····42
[그림 4-9] t-근접성 예시 ····43
[그림 4-10] 질병에 대한 분류 트리 예 ····44
[그림 4-11] ARX Data Anonymization Tool 주요 화면 ····47
[그림 4-12] Cornell Anonymization Toolkit 주요 화면 ····47
[그림 4-13] TIAMAT 주요 화면 ····48
[그림 4-14] SECRETA 주요 화면 ····48
[그림 4-15] μ-ARGUS 주요 화면 ····49
[그림 4-16] PARAT 주요 화면 ····49
[그림 5-1] 약학정보원 관련 사건의 개요도 ····51
[그림 5-2] SK텔레콤 전자처방전 사건 개요도 ····53

요 약 문

그간 ICT기술의 발전으로 수많은 기업과 공공기관들은 대량의 데이터를 축적하여 왔는데, 이러한 빅데이터는 최근 '21세기의 원유'라고 불릴 정도로 경제적·사회적 가치를 인정받게 되었다. 빅데이터 중에서도 개인에 관한 일반적인 정보, 의료정보, 위치정보, 신용정보 등은 개인정보로 통칭되며, 빅데이터의 수집·분석과 관련된 빅데이터산업은 급속히 성장하고 있으며, 개인정보의 수집과 판매에 집중하는 데이터 브로커들도 미국을 비롯해 전세계적으로 활동하고 있다.

국가권력으로부터의 자유를 의미하던 전통적인 프라이버시권은 이러한 상황에 발맞추어 개인정보 자기결정권이라는 개념으로 진화하였고, 세계 각국은 개인정보 자기결정권을 보호하는 한편, 집적된 개인정보를 활용하여 각종 산업을 육성하고 사회적 편익을 증대시키고자 노력하고 있다. 잇따른 대규모 개인정보 유출사고를 겪은 우리나라는 개인정보 보호체계를 강화해 왔으나, 최근에는 개인정보의 활용을 증대시키고자 하는 움직임이 경제 관련 부처들로부터 시작되고 있다. 가까운 일본의 경우에는 개인정보 활용을 촉진하기 위해 개인정보보호법을 대폭 개정해 2017년 전면 시행예정이며, 인권을 중시하는 EU는 개인정보 관련 권리를 구체화하고 개인정보 관리감독을 강화하기 위해 GDPR을 제정, 2018. 5. 25.부터 EU회원국들과 EU 내에서 활동하는 사업자들에게 시행할 예정이다.

우리나라에서는 이미 수집된 개인정보를 제3자에게 제공할 때 정보 주체의 사전동의가 필요하지만 비식별화하여 개인을 알아볼 수 없을 때는 예외이기 때문에 개인정보의 비식별화 기술에 대한 관심이 높아져 가고 있다. 하지만 다른 정보와 결합하여 개인이 식별될 경우에는 법적 규제와 처벌의 대상이 될 수도 있고, 실제로 SK텔레콤의 전자처방전 사업과 약학정보원의 처방정보 유출사건에서 비식별화 및 개인정보 유출 여부가 쟁점으로 떠오른 바 있어, 산업계에서는 보다 명확한 기준을 요구해 왔다.

이에 정부에서 2014년에 발표한 '개인정보 비식별화 기술 활용 안내서'에서는 무작위화 방법과 일반화 방법으로 대표되는 비식별화기술 18종과 함께 비식별 적정성 평가기준으로 k-익명성(k-anonymity), l-다양성(l-diversity), t-근접성(t-closeness)을 제시하였다. 그러나 기술안내서에서는 개인정보의 정의와 범위에, 적정성 평가의 제도운용에 관한 가이드라인이 존재하지 않았다.

본 보고서에서는 지난 6월 30일 발표된 ‘개인정보 비식별조치 가이드라인’에 언급된 17종의 비식별화기술과 3종의 비식별 적정성 평가기준을 자세히 소개하였다. 특히 외국에서 비식별 데이터가 공개된 정보와 결합하여 재식별된 사례들을 통해 비식별화기술의 한계와 적정성 평가기준의 필요성을 보다 쉽게 이해할 수 있도록 하였다. 아울러 현재 사용할 수 있는 비식별화 소프트웨어 현황도 오픈소스 소프트웨어 위주로 자세히 정리하였다.

또한 새 가이드라인에서는 지금까지 개인정보의 범위를 매우 확장했다는 비판을 받던 “다른 정보와 쉽게 결합하여 개인을 식별할 수 있는 정보”(‘결합용이성’)의 요건으로 ① 입수가능성과 ② 재식별의 합리적 가능성을 제시해 개인정보의 범위가 보다 분명해졌지만, 여전히 개인정보 해당 여부에 관한 사법적 판단의 참고자료라는 한계를 가지고 있음을 지적하였다.

그러므로 실제로 비식별화기술과 관련하여 개인정보의 보호와 산업적 활용을 조화시키는 방안으로서, ① 개인정보의 정의에 관한 법개정을 적극 검토해야 하며, ② 비식별정보의 유통에 대한 당국의 관리체계가 필요하고, ③ 법체계 정비와 함께 개인정보보호위원회를 개인정보의 컨트롤타워로 강화할 필요가 있음을 제시하였다.

SUMMARY

In recent years, big data, sometimes called 21st century's oil, is highly evaluated with its economic and social values. With rapid evolution of ICT technology, many companies and organizations have been collecting these valuable, enormous data competitively. As a category of big data, Personally Identifiable Information (PII) consists of general, medical, location, and access to credit information about people. The big data industry including collection and analysis processes are rapidly growing. In addition, data brokers that focus on collecting and selling personal data are running business actively in the US and other countries as well.

The right to Privacy gets evolved from the meaning of freedom against government power to the right to the protection of personal data concerning himself or herself. These days many countries try to protect informational self-determination. At the same time, these countries also try to promote various industries and improve social benefits by using the accumulated personal data. Korea government has been reinforcing the legal protection system for personal data after serveral cases of huge personal data leakage. Recently, economy-related ministries of Korea has announced a few plans to increase the utilization of personal data. In Japan, the Personal Information Protection Law was significantly amended and will be enforced in 2017. EU, traditionally emphasizing human rights, enacted General Data Protection Regulation (GDPR) to concretize the rights related personal data and to strengthen the management and oversight of personal data. In 2018, the GDPR will be enforced to the all EU member states and the foreign companies that want to operate in the member states.

In Korea, the prior consents of data subjects must be required by the law whenever their personal data legally collected are provided to new 3rd parties not listed at the consent time. Since de-identified information converting personal data by de-identification techniques are not personal data anymore, these de-identification techniques becomes in the spotlight as a detour to strict prior consent of offering personal data to 3rd parties. However, a legal risk like compensation for damages or a criminal punishment still exists if

a subject may be identified by combining the provided de-identified information with other information. In the SKT case and IMS health Korea case, main issues are whether encryption is a kind of de-identification or not and whether de-identified data by encryption is personal data or not. Watching these cases, Korean companies have required more explicit and legally effective guidelines to government.

Korean government released the first guideline for personal data de-identification techniques in 2014. It included 18 de-identification techniques and suggested the adequacy evaluation method for guaranteeing the level of de-identification such as k-anonymity, l-diversity, and t-closeness. However, it had some limitations like no clear definition and scope of personal data, and no systematic management for adequacy test.

In this report, we introduce 17 de-identification techniques and 3 adequacy evaluation methods from a brand-new ‘Guideline for personal data de-identification’ published by Korean government on last June. Moreover, this report includes some well-known incident cases of re-identification by combining public data and de-identified data. From the cases, the limitation of de-identification techniques and the need of adequacy evaluation standard are shown. In addition, de-identification applications including many open-source software everyone can currently use are listed.

The latest guideline showed more clear scope of personal information. That is, about combinability, the guideline provided ① possibility of acquisition and ② reasonable possibility of re-identification as the requirement of additional information that the person can be identified by easily combining de-identified information with. The combinability was criticized because it extended the scope of personal data excessively. However, the guidline still has no legal effect to the judiciary.

Therefore, we suggest measures to harmonize personal data protection and industrial use of the personal data or de-identified data with the de-identification techniques as follows: ① the amendment for the clear definition of personal data, ② requiring a management system of the government for trading de-identification information, and ③ strengthening the Privacy Commissioner of Personal Information as a main authority of personal information with maintenance of legal system.

CONTENTS

Chapter 1. Introduction

Chapter 2. The Advent of Big Data and Its Practical Use

Chapter 3. Rights and Protection of Personal Data

Chapter 4. De-identification Techniques for Personal Data

Chapter 5. Issues about De-identification Techniques

Chapter 6. Proposals of Legal System Improvement

제1장 서 론

제1절 연구배경

1. 개인정보와 빅데이터산업

□ IT산업의 발달로 컴퓨터와 SW로 거의 모든 정보처리가 이루어지는 지금, 우리나라는 2000년대 이후 여러 차례의 개인정보 유출사고를 겪으면서 개인정보보호에 관한 논의가 더욱 활발해져 왔음

○ 특히 금융권에서 개인정보 유출사건이 지속적으로 발생해 '한국인의 개인정보는 중국 양쯔강의 노인들도 하나씩 갖고 있다'는 말도 나옴[1)]

○ 개인정보보호의 필요성을 인식한 정부에서는 정보통신망법과 신용정보법 등 개별법을 강화해 오다가 2011년에 개인정보보호에 관한 일반법인 개인정보보호법을 제정·시행하였으며, 유출사고가 발생할 때마다 보호조치 관련 규정 외에도 손해배상, 과징금, 형사처벌 등의 제재규정을 지속적으로 강화해 왔음

○ 주민등록번호, 이름, 주소 등의 개인정보가 헌법 상 기본권인 개인정보 자기결정권의 보호대상이기 때문에 이러한 법규제의 강화는 국민들의 기본권 수호를 위한 측면이 컸음

□ 한편으로는 개인정보를 포함한 빅데이터의 축적 및 활용을 통해 소비자의 편익이 증대될 수 있다며 개인정보의 활용을 촉진하는 규제개선이 필요하다는 목소리도 높아져 가고 있음

○ 인터넷의 확산·보급 등에 힘입어 예전에는 처리·분석할 엄두도 내지 못했을 정도의 개인정보를 포함한 대규모 정보를 축적한 사업자들이 등장했고, 이들은 자신들의 빅데이터를 발달된 IT기술로 분석·활용해 다양한 서비스를 제공해 오고 있음

---

1) 경향신문, "'한국인 개인정보는 중국 노인들도 하나씩…' 시대에 뒤떨어진 금융사의 보안 인식", 2014. 1. 20.자 기사 참조
http://biz.khan.co.kr/khan_art_view.html?artid=201401201647031&code=920100&med=khan

○ 이에 따라 개인정보의 경제·사회적 가치가 새로이 주목받으면서 빅데이터 산업과 데이터 브로커 산업이 발전해 오고 있음

<표 1-1> 개인정보의 경제·사회적 가치 유형

가치 유형	주요 내용
국제 현안 해결	● 재난재해, 실업과 식량 안보와 같은 국제적 위기에 대한 이해와 대응 촉진 ※ 구글 감기예측, UN Global Pulse의 실업, 식량안전 분석 등
효율성 향상	● 모든 산업에서의 조직의 효율성 및 생산성 향상 ● 미국은 개인 데이터 유통을 통해 매년 7천억 달러 또는 전체 지출의 30%의 건강비용 절감 ● 금융서비스에서는 사기 예방을 통한 비용 절감과 온라인 거래와 지불을 촉진하여 효율성을 상당히 향상
예측능력 향상	● 개인화된 맞춤형 상품 및 서비스 개발 촉진 ● 도서추천 서비스, 운전 행태에 따른 개인 맞춤형 보험 상품 개발, 개인 맞춤형 뉴스, 기사 서비스 등 ● 고객의 행태 분석을 통해 개연성 있는 사건에 대비
대중화된 정보접근	● 검색 엔진, 이메일, 뉴스 사이트와 소셜 네트워크 등과 같은 '무료' 서비스 이용 ● 사실상 대부분의 표면상 무료 서비스는 자신과 자신의 행태에 관한 데이터 제공을 통한 타겟 광고의 대가임
개인의 권한 강화	● 소극적인 행위자에서 집단적인 소통방식으로 기업과 대화 ● 자신과 믿음, 선호도 등에 관한 정보 공유를 통해 세상과 연결

출처 : 한국정보화진흥원, 빅데이터 시대의 개인 데이터 보호와 활용, 2013을 보완

□ 우리나라는 개인정보의 수집·활용·제3자 제공에 정보주체의 사전동의와 목적 내 이용을 원칙으로 삼고 있지만, 통계 및 학술연구 등을 위해 비식별화(또는 익명화[2])한 정보[3]는 예외로 하고 있어 개인정보 비식별화기술에 대한 관심이 높아져 가고 있음

2) 일각에서는 비식별화라는 용어와 익명화라는 용어를 엄격하게 구분하여 사용하지만, 본 보고서에서는 동일한 의미로 간주하며, 앞으로는 '비식별화' 라는 용어를 사용하기로 함

3) 개인정보보호법 제18조(개인정보의 목적 외 이용·제공 제한)

② 제1항에도 불구하고 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 **개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공**할 수 있다. 다만, 제5호부터 제9호까지의 경우는 공공기관의 경우로 한정한다.

4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우

○ 정부에서도 개인정보 비식별화기술 활용 안내서, 사례집, 적정성 자율평가 안내서를 배포하는 등, 비식별화기술 보급에 힘써 왔고,

○ 2016년 7월 1일부터 적용된 “개인정보 비식별 조치 가이드라인”(이하 ‘비식별조치 가이드라인’이라 함)를 통해 개인정보 비식별조치 기준과 비식별정보의 활용 범위 등이 좀 더 명확해졌음

□ 그러나 비식별화기술만으로는 개인정보로 이뤄진 빅데이터 활용의 걸림돌이 완전히 제거되었다고 할 수 없으며, 산업계에서는 지속적인 규제완화를 주장하고 있음

○ 여러 법률에 규정된 각종 제재규정이 강력하고, 기소권을 가진 검찰은 어떤 정보의 개인정보 해당여부에 대해 상당히 폭넓게 보고 있음

○ 따라서 개인정보 관련 사업을 기획·집행할 경우 형사소송의 대상이 될 가능성이 높고, 유무죄 여부를 결정짓는 개인정보의 정의 및 범위에 대한 최종적 해석권한은 행정부가 아닌 법원에 있어 사업자로서는 높은 불확실성 때문에 개인정보나 비식별정보의 합법적 이용마저 주저하게 됨

제2절 연구범위

□ 이에 본 보고서에서는 개인정보 보호와 활용 논의의 배경, 비식별화기술의 개요 및 한계, 비식별화기술의 쟁점들을 살펴보고자 함

□ 먼저, 개인정보 활용의 가장 대표적 수혜자인 빅데이터 산업과 데이터 브로커 산업의 현황, 개인정보 보호의 이론적·법적 근거, 비식별화기술의 개요 및 한계를 개관함

□ 다음으로 비식별화 여부가 논란이 된 사건들을 통해 개인정보의 정의 및 범위의 해석 논란이 해결되어야 비식별화기술이 적극적으로 쓰일 수 있음을 살펴봄

□ 마지막으로 개인정보의 보호 및 활용에 관한 쟁점과 개선방향에 대한 시사점을 도출하고자 함

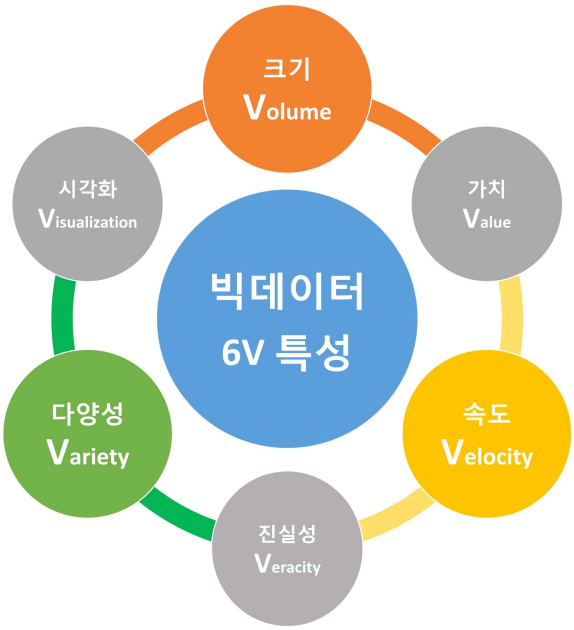
제2장 빅데이터의 등장과 활용

제1절 빅데이터란

□ (빅데이터, big data) 기존 방식으로는 저장 및 관리, 분석이 거의 불가능할 정도로 규모가 크고, 변화의 속도가 빠르며, 다양한 형태를 갖는 데이터를 일컬음

○ 좀 더 넓은 의미로 데이터로부터 가치 있는 정보 추출, 결과 분석, 시각화하는 등의 방법을 포함하기도 함

[그림 2-1] 빅데이터의 6V 특성



○ META Group(現 Gartner)의 더그 레이니(Doug Laney)는 빅데이터의 주된 특성으로 크기(volume), 다양성(variety), 속도(velocity) 등 3V 제시[4)]

- (크기, volume) 실제 데이터의 물리적 크기로서 기업 데이터, 센싱 데이터, SNS 데이터 등 최근 그 규모가 페타바이트(petabyte)[5)]나 엑사바

4) D.Laney, 3D Data Management: Controlling Data Volume, Velocity and Variety", Gartner, 2001. 2. 6.

이트(exabyte)[6] 이상으로 커지고 있음

- (다양성, variety) 관계형 데이터베이스 등과 같은 정형(structured) 데이터에서부터 HTML, XML, JSON[7] 등과 같은 반정형(semi-structured) 데이터, 이미지, 비디오, SNS, 센싱 데이터 등과 같은 비정형(unstructured) 데이터까지 모든 형태를 가짐
- (속도, velocity) 데이터 수집 및 가공, 분석 등 일련의 과정을 실시간 또는 일정 주기에 맞춰 처리할 수 있는 데이터 처리 능력을 말함

○ 최근 빅데이터의 기존 특성 3V에 진실성(veracity) 혹은 가치(value)를 추가하여 4V, 둘 다 추가하여 5V, 시각화(visualization)까지 추가하여 6V 등으로 확대[8]하기도 함 ([그림 2-1] 참조)

- (진실성, veracity) 의사 결정이나 기업 활동에 활용될 수 있도록 진실하고 정확해야 함
- (가치, value) 비즈니스에 실현될 궁극적 가치에 중점
- (시각화, visualization) 사용자 친화적인 시각적 기능을 통해 빅데이터의 모든 잠재력 활용

[그림 2-2] 빅데이터의 가치창출 과정

□ 빅데이터가 진정한 의미를 갖기 위해서는 큰 규모의 데이터를 수집하는 것

5) 페타바이트(petabyte, PB) : 1,000 테라바이트(terabyte, TB)
6) 엑사바이트(exabyte, EB) : 1,000 페타바이트
7) JSON(JavaScript Object Notation) : 속성-값 쌍으로 이루어진 데이터 객체를 인간이 읽을 수 있는 텍스트 형태로 전달하기 위해 만들어진 개방형 표준 포맷
8) 범지인, 송두한, 최성종, “빅데이터 활용 사례와 시사점”, 농협경제연구소, 2013.08.28

을 넘어 이러한 데이터를 통해 통찰력(insight)을 얻고 실제 기업 활동에 활용하여 가치를 창출할 수 있어야 함 ([그림 2-2] 참조)

- 웹사이트 검색 통계, SNS 데이터를 분석 등을 통해 시장 예측 및 상품 개발에 활용
- 소비자의 방문 및 구매 패턴을 분석하여 마케팅 전략 수립
- 제조 과정에서 발생하는 센싱 데이터를 이용하여 불필요한 작업 제거 및 개선으로 생산성 향상 등

제2절 빅데이터 관련 기술

□ 빅데이터 관련 기술은 수집·공유 단계, 저장·관리 단계, 처리 단계, 분석 단계, 지식 시각화 단계 등 각 처리 프로세스([그림 2-3] 참조)마다 다양한 기술(<표 2-1> 참조) 존재

○ (수집·공유 단계 기술) 내부 조직의 정형화된 데이터와 조직에 필요한 외부의 데이터를 발견 및 수집, 특정 데이터 형식으로 변환하여 정제된 데이터 확보 등의 기술과 서로 다른 시스템 간 데이터를 공유하는 기술

○ (저장·관리 단계 기술) 수집된 대용량의 데이터를 저장하고 데이터 검색, 수정, 삭제 등을 용이하게 제공하는 기술

○ (처리 단계 기술) 엄청난 크기의 데이터를 일괄 혹은 실시간 처리하기 위한 분산 처리 기술

○ (분석 단계 기술) 효과적인 의사결정에 활용하기 위해 빅데이터로부터 의미 있는 정보(지식)을 얻어내는 기술

○ (지식 시각화 단계 기술) 빅데이터에서 얻어낸 가치있는 정보의 의미를 직관적이고 쉽게 알 수 있도록 표현하는 기술

[그림 2-3] 빅데이터 처리 프로세스

출처 : 문혜정, 'Big Data 구축기술과 사례를 중심으로', 2012 재구성

<표 2-1> 빅데이터 요소 기술 구성 및 분류

요소 기술	설명	해당 기술
수집	- 조직내부와 외부의 분산된 여러 데이터 소스로부터 필요로 하는 데이터를 검색하여 수동 또는 자동으로 수집하는 과정 관련된 기술 - 단순 데이터 확보가 아닌 검색/수집/변환을 통해 정제된 데이터를 확보하는 기술	ETL/크롤링 엔진/로그수집기/센싱/RSS, Open API 등
공유	- 서로 다른 시스템간의 데이터 공유	멀티 테넌트 데이터 공유/협업 필터링 등
저장/관리	- 작은 데이터라도 모두 저장하여 실시간으로 저렴하게 데이터를 처리 - 처리된 데이터를 더 빠르고 쉽게 분석하여, 이를 비즈니스 의사 결정에 바로 이용하는 기술	병렬 DBMS/하둡(Hadoop)/NoSQL 등
처리	- 엄청난 양의 데이터의 저장·수집·관리·유통·분석을 처리하는 일련의 기술	실시간 처리/분산 병렬 처리/인-메모리 처리/인-데이터베이스 처리
분석	- 데이터를 효율적으로 정확하게 분석하여 비즈니스 등의 영역에 적용하기 위한 기술로 이미 여러 영역에서 활용해온 기술임	통계 분석/데이터 마이닝/텍스트 마이닝/예측 분석/최적화/평판 분석/소셜 네트워크 분석 등
시각화	- 자료를 시각적으로 묘사하는 학문으로 빅데이터는 기존의 단순 선형적 구조의 방식으로 표현하기 힘들기 때문에 빅데이터 시각화 기술이 필수적임	시간시각화/분포시각화/관계시각화/비교시각화/공간시각화/인포그래픽

출처 : 한국방송통신전파진흥원, 빅데이터(Big Data) 활용단계에 따른 요소기술별 추진동향과 시사점, 방송통신기술 이슈&전망 2013년 제10호, 2013.12.10.

제3절 빅데이터의 활용 사례

□ 구글의 독감 트렌드와 자동 번역 서비스, 검색 통계 등과 월마트의 소비자 패턴분석, 리츠칼튼호텔의 고객맞춤형 서비스, T-Mobile의 고객이탈 사전감지 등은 이미 잘 알려진 대표적인 빅데이터 활용 사례임

○ (구글 독감트렌드) 시간·지역별 검색어에 기반을 둔 독감 예상수치를 구글 독감트렌드 사이트를 통하여 제공

○ (구글 자동번역 서비스) 유럽연합 20여 개 국의 언어로 번역되어 있는 수억 건의 문서를 이용하여 통계적 자동번역[9] 시스템 개발 성공

○ (구글 검색통계) 검색통계를 이용해 오바마 정부의 노후 차량 보상 프로그램의 호응도를 분석하여 정부 예산이 부족할 것을 예측

○ (월마트의 소비자 패턴분석) 월마트의 웹사이트에서 발생하는 거래데이터를 이용한 재고예측 조사시스템을 마련하여 고객선호도 및 수요데이터를 분석하여 점포 운영에 반영

○ (리츠칼튼호텔의 고객맞춤형 서비스) 투숙 고객의 특성을 관찰하여 얻어진 데이터를 종합분석하여 친절한 서비스를 제공

- 기존 고객이 제기한 불만사항을 수집하여 서비스를 개선하고 모든 단계에서 품질개선을 위한 프로세스를 추진함

○ (T-Mobile의 고객이탈 사전감지) 타 통신사로 회선을 옮기는 고객이 보이는 이용패턴을 분석하여 실시간 고객이탈 사전감지 시스템 구축

- 시스템 구축 후 이탈 고객 수가 9만9천 명에서 5만 명으로 감소

□ 최근 사례로는 BoA(Bank of America)의 적합한 상품 제안, 허츠(Hertz)의

9) 통계적 자동번역 : 전문 번역가가 번역한 이중 언어 번역 대응 쌍으로부터 통계적 분석을 통하여 번역 모델의 파라미터를 학습하여 그 모델에 근거를 두고 번역하는 방식

고객 만족도 향상, 오비츠의 사용자 맞춤 검색 결과 제공, UNC헬스케어의 재입원 비용 절감, GE의 지능형 항공 운영 서비스 등이 있음[10]

○ (BoA 사례) 5천만 건, 약 65PB(petabytes)의 고객 데이터를 보유하고 있으며 이를 분석하여 고객에게 적합한 상품 제안

○ (허츠 사례) 글로벌 렌터카 기업인 허츠는 다양한 채널에 퍼져 있는 VOC(고객의 소리, Voice of Customer)를 실시간으로 분석해 고객의 요구 사항에 대해 빠르게 대응할 수 있는 시스템을 운영

- 필라델피아 지점에서 고객에 대한 서비스 지연이 발생하는 가장 큰 요인은 차량 반납에 걸리는 시간 때문이며, 하루 중 구체적으로 어떤 시간에 이런 지연이 발생하는지 파악

- 이러한 정보를 바탕으로 해당 지점의 고객 몰리는 시간대에 직원 수를 늘리고 고객 불만을 원활하게 해결할 수 있는 지점 매니저를 배치하는 등의 대응책 실행

○ (오비츠 사례) 미국의 온라인 여행사이트인 오비츠는 고객의 온라인 활동을 추적하여 특성을 파악하고 이를 통해 고객기호나 지출습관을 예측하여 최적화된 제품/상품을 제시

○ (UNC 헬스케어[11] 사례) 빅데이터를 적극 활용해 선진화된 의료 서비스를 갖추고, 맞춤형 건강관리와 같은 다양한 환자관리 프로그램을 마련

- 유방암과 자궁경부암 부문에서 암진단 건수를 10% 이상 증가시켰으며, 결장암과 같은 다른 암의 진단에 확대 적용

○ (GE 사례) 지능형 시스템 기반의 '지능형 항공 운영(Intelligent Operation)' 실현으로 운영 효율성 극대화

- 예측하지 못한 항공정비로 인해 발생하는 비행지연을 사전에 예방하여 항공기 운영 효율성의 증가, 정시 운영의 증가, 정비 효율의 증가, 유지보수 비용의 감소를 실현

10) 한국정보화진흥원, "2015년 빅데이터 글로벌 사례집", 2015. 5.
11) UNC 헬스케어(University of North Carolina Health Care) : 비영리 통합 의료기관으로 미국 노스캐롤라이나 북부에 설립

제4절 빅데이터 산업 전망

□ 세계 및 국내 빅데이터 시장 규모 모두 크게 성장할 것으로 전망

○ 세계 빅데이터 시장은 2015년 213억 달러 대비 2019년에는 486억 달러로 두 배 이상 성장 예상 (<표 2-2> 참조)

○ 국내 빅데이터 시장 역시 연평균 24.7% 성장하여 2015년 1,527억 원에서 2019년에는 3,583억 원 규모로 예측 (<표 2-3>참조)

□ 기업 간 또는 이종 산업 간의 데이터 융합이 활발할 것으로 전망

○ 금융, 제조, 유통・물류, 의료・건강, 통신・미디어, 안전 등 다양한 산업 간 데이터 융합을 통한 신규 서비스 발굴로 새로운 가치 창조 예상

○ 특히, 인공지능(AI), 기계학습, 딥러닝 등 지능정보기술, O2O(online to offline) 연계, 모바일 데이터 활용이 더욱 활발할 것으로 예측

○ 빅데이터 활용 스마트 서비스 시범사업 공모[12] 등 다양한 정부지원 사업을 통해 빅데이터 기술 조기 확산과 산업 분야별 대형 수요 창출 전망

12) 2016년 빅데이터 활용 스마트서비스 시범사업 공모
http://www.nia.or.kr/BBS/board_view.asp?BoardID=20111202112757833 6&id=16794

<표 2-2> 세계 빅데이터 시장 규모

(단위 : 십억 달러, %)

구분	2014	2015	2016E	2017E	2018E	2019E	CAGR (14-19)
인프라 (비중)	8.9 (51.4)	11.0 (51.6)	13.6 (51.3)	16.5 (50.6)	19.8 (49.7)	23.6 (48.6)	21.7
Compute	3.0	3.7	4.6	5.7	7.0	8.6	23.7
스토리지	4.5	5.6	7.0	8.5	10.2	12.0	22.0
네트워킹	0.9	1.0	1.2	1.4	1.6	1.8	14.9
기타 인프라	0.6	0.7	0.8	0.9	1.1	1.2	17.3
소프트웨어 (비중)	4.0 (23.2)	5.1 (23.7)	6.4 (24.2)	8.1 (25.0)	10.2 (25.6)	12.8 (26.4)	26.2
정보관리 SW	1.7	2.0	2.5	3.2	3.9	4.7	23.3
검색 및 분석 SW	1.9	2.5	3.2	4.1	5.2	6.6	28.0
애플리케이션 SW	0.4	0.5	0.7	0.9	1.1	1.5	28.9
서비스 (비중)	4.4 (25.4)	5.3 (24.7)	6.5 (24.5)	8.0 (24.5)	9.9 (24.7)	12.2 (25.1)	22.7
합계	17.2	21.3	26.5	32.6	39.9	48.6	23.1

출처 : IDC Worldwide Big Data Technology and Service 2015-2019 Forecast(2015. 10), SW산업 주요 통계, 소프트웨어정책연구소, 2016. 4.

<표 2-3> 국내 빅데이터 시장 규모

(단위 : 십억 원, %)

구분	2014	2015	2016E	2017E	2018E	2019E	CAGR (14-19)
인프라 (비중)	58.9 (49.5)	77.0 (50.4)	97.8 (50.2)	122.2 (50.0)	149.2 (50.0)	178.8 (49.9)	24.9
Compute	21.9	28.4	35.9	44.4	54.1	64.1	24.0
스토리지	24.4	32.4	42.0	52.9	64.7	77.9	26.1
네트워킹	7.5	9.6	11.7	14.4	17.7	21.4	23.2
기타 인프라	5.0	6.6	8.3	10.5	12.8	15.4	25.1
소프트웨어 (비중)	27.8 (23.4)	33.6 (22.0)	42.9 (22.0)	53.9 (22.1)	66.0 (22.1)	79.5 (22.2)	23.4
정보관리 SW	16.5	19.1	24.1	30.3	37.0	44.5	21.9
검색 및 분석 SW	9.5	12.3	15.8	19.9	24.4	29.5	25.5
애플리케이션 SW	1.8	2.3	3.0	3.7	4.6	5.5	25.2
서비스 (비중)	32.2 (27.1)	42.1 (27.6)	54.3 (27.8)	68.1 (27.9)	83.5 (28.0)	100.0 (27.9)	25.4
합계	118.8	152.7	195.0	244.2	298.7	358.3	24.7

출처 : IDC Worldwide Big Data Technology and Service 2015-2019 Forecast(2015. 10), SW산업 주요 통계, 소프트웨어정책연구소, 2016. 4.

제5절 데이터 브로커 산업 전망

□ 개인정보를 수집하여 제3자에 공유하거나 유통하는 데이터 브로커 산업도 미국을 중심으로 빠르게 성장하고 있음

○ 마케팅 및 기타 목적을 위해 소비자의 오프라인과 온라인 및 모바일 활동에 대해 매년 엄청난 양의 정보를 수집하고 분석하여 판매하는 기업을 데이터 브로커, 또는 정보 재판매업자(information reseller)라고 하며, 데이터 브로커 간의 데이터유통도 매우 활발함

○ 전세계적으로 약 4천여 개의 데이터 브로커 기업이 활동하고 있으며, 약 2,000억 달러 규모로 추정됨[13)]

□ 미국에는 이미 1990년대부터 데이터 브로커가 존재했으며 빅데이터 시대의 도래로 급성장하고 있음

○ 2013년 10월 기준으로 총사업자 수 약 650개, 2012년 동안 약 1,560억 달러의 추가 매출이익이 발생했고, 고용인원은 약 67.5만명으로 추정됨[14)]

○ 마케팅, 리스크 경감(risk mitigation), 사람찾기 서비스를 주로 제공하며, 미국 공정거래위원회(FTC)가 조사한 9개 주요 사업자(<표 2-4>참조)의 매출은 약 4억 3천만 달러(<표 2-5>참조)[15)]

---

13) "개인 중심의 개인데이터 거래시장의 형성 가능성", ICT Issue Weekly 제514호(2015. 10. 16.), 한국정보화진흥원 참조

14) "온라인 프라이버시에 대한 철학적 배경과 산업적 접근", 정보통신정책연구원, 2013. 64면 참조

15) 액시엄(Acxiom)의 2012년 매출이 11.5억 달러, 인텔리우스(Intelius)의 2008년 매출이 1.28억 달러임을 보면, 미국 FTC의 조사기준이 다르거나 데이터 브로커 업체들이 매출을 축소했을 가능성이 있음
유지연, "미국 데이터 브로커(data broker) 현황", 정보통신정책연구원, 2013. 7. 16. 참조
http://www.kisdi.re.kr/kisdi/fp/kr/publication/selectResearch.do?cmd=fpSelectResearch&sMenuType=2&controlNoSer=2&controlNo=13187&langdiv=1

<표 2-4> 미국 9개 데이터 브로커

회사명	특징
Acxiom	● 마케팅캠페인, 부정사용 탐지를 위한 고객데이터 분석 서비스 제공 ● 전 세계 7억 명의 소비자 정보가 담긴 데이터베이스 보유
Corelogic	● 산업계와 정부에 재무정보와 부동산정보에 기초한 분석서비스 제공 ● 약 8억 건의 부동산 거래정보, 약 1억 건의 담보 데이터베이스 보유
Datalogix	● 거의 모든 미국 소비자의 마케팅 데이터를 제공 ● 2012년 페이스북은 페이스북 이용자의 소셜사이트 상품광고 조회와 오프라인 상점의 구매 관련성 측정 위해 데이터로직스와 협력 발표
eBureau	● 마케터와 재무관련 회사, 온라인유통업체에 수익성이 높은 잠재 고객과 부정 거래 예측 서비스 제공 ● 매달 평균 30억 건이 넘는 새로운 정보 추가 축적
ID Analytics	● 특정인 확인, 부정 거래 확인 서비스 제공 ● 7천억 건의 데이터와 14억 건의 소비자 거래 데이터 보유
Intelius	● 신원 조회와 공문서 정보 제공 ● 20억 건이 넘는 데이터베이스 보유
PeekYou	● 소셜미디어사이트, 홈페이지, 블로그의 콘텐츠를 분석 작성자 확인 서비스 제공
Rapleaf	● 이메일 주소와 함께 이메일 주소 소유자의 연령, 성, 우편번호, 소득, 결혼 여부, 자녀 여부와 취미, 구매 유형 등 정보 제공
Recorded Future	● 소비자와 기업의 과거 이력 데이터 분석을 통해 미래 행동 예측 정보 제공

출처 : 정용찬, “빅데이터 산업과 데이터 브로커”, 정보통신정책연구원, 2015

<표 2-5> 미국 9개 데이터 브로커의 매출구성

2012년 기준, 단위 : 백만달러

상품	마케팅	리스크 경감	사람찾기	합계
매출	196.2	177.8	52.7	426.7
비율(%)	46	42	12	100

출처 : 정용찬, “빅데이터 산업과 데이터 브로커”, 정보통신정책연구원, 2015

□ 미국 FTC는 데이터 브로커 산업의 투명성 부족에 대해 예의주시하고 있으며, 데이터 브로커의 개인정보 거래 등의 활동을 개인이 인지하고 잘못된 데이터를 정정할 수 있도록 행정명령을 내리는 한편, 관련 법률 제정을 권고한 바 있음

○ 2012. 6. 12. FTC는 스포키오(Spokeo)에게 ‘이력서 이상의 자세한 정보’라고 광고하며 기업 채용담당자에게 개인 프로파일 데이터를 판매한 행위에 대해 개인의 신용에 대한 정확한 정보 작성을 의무화하는 공정신용정보법(FCRA : Fair Credit Reporting Act) 위반으로 80만 달러의 제재금을 부과하

였고[16], 이후 2015년 10월까지 FCRA 위반 혐의로 다른 데이터 브로커에 대해서도 8건을 적발, 추가로 제재금을 부과함[17]

○ 2012. 9. 경 FTC는 9개 사업자에게 개인정보 수집내역과 판매내역을 공개하도록 행정명령을 내렸고, 2013. 6. 경에는 소비자가 데이터 브로커의 데이터 수집 및 활용에 관한 정보를 확인할 수 있도록 하는 'Reclaim your name' 캠페인을 시작함[18]

- 이에 따라, 액시엄(Acxiom)은 'https://aboutthedata.com/' 포털을 개설하여 이용자가 직접 자신의 데이터를 확인·수정할 수 있도록 함

□ 국내의 경우, 개인정보보호법, 정보통신망법 등에서 개인정보 수집 및 제3자 제공을 엄격히 제한하고 있기 때문에 데이터 브로커 산업이 아직 활성화되지는 못하고 있음

○ 한국데이터베이스진흥원이 운영 중인 데이터스토어는 2015년 5월 기준 약 2,200여개의 데이터 상품을 유통하고 있으며 거래건수도 1,140건을 넘었지만, 개인정보를 가공한 DB는 소수에 그침

- SK텔레콤이나 LG CNS도 데이터 가공 및 유통 사업을 진행 중이며, 데이터 컨설팅 업체인 엔코아도 관련 사업을 준비 중임

○ 개인정보의 수집 시 사전동의를 요구하는 법적 제약이 풀리고, 비식별화기준이 명확해진다면 데이터 브로커 산업이 어느 정도 성장할 것으로 전망되며, 직접 개인정보를 수집하지 못하거나 처리할 수 없는 소규모 기업들의 수요가 클 것으로 예상됨

16) 구체적으로는 ① 합법적 목적에만 이용될 것, ② 정확한 정보를 제공할 것, ③ FCRA에 따른 의무를 소비자들에게 고지할 것, ④ 개인에 대한 특정 리포트로 불리한 결과가 발생할 경우 해당 개인에게 알릴 것, 이렇게 4개의 의무위반이 인정되었음. 아래 링크 참조
https://www.ftc.gov/news-events/press-releases/2012/06/spokeo-pay-800000-settle-ftc-charges-company-allegedly-marketed

17) https://www.ftc.gov/news-events/media-resources/consumer-finance/credit-reporting

18) 정보통신정책연구원, "온라인 프라이버시에 대한 철학적 배경과 산업적 접근", 2013. 67면 참조

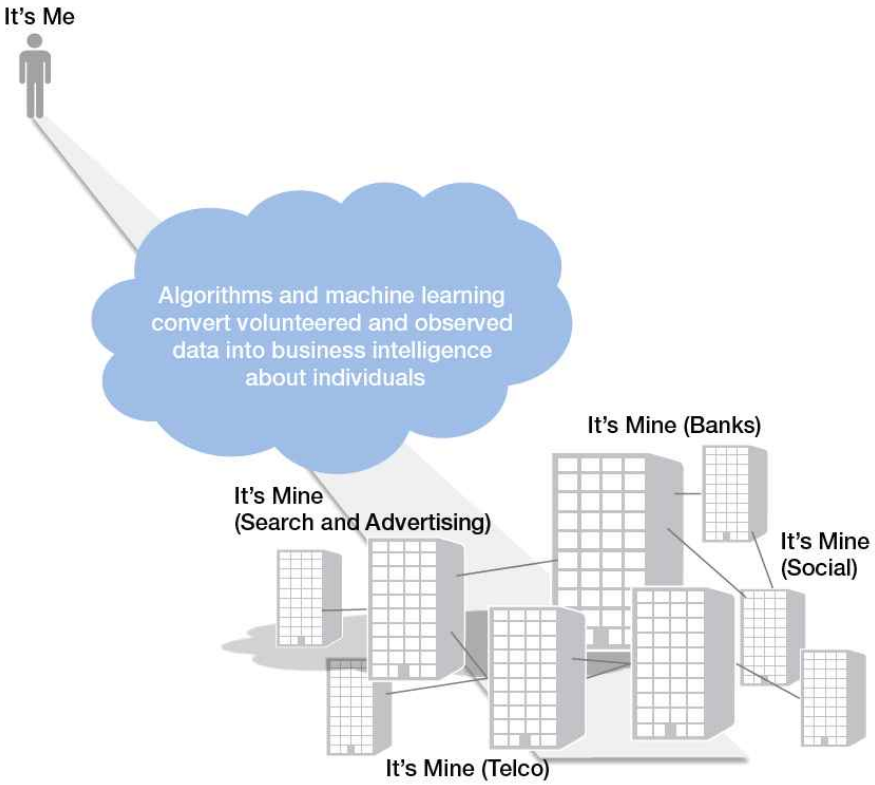
제6절 빅데이터 활용과 개인정보 보호와의 관계

□ 기업이나 기관에서 관심을 가지고 분석·활용하고자 하는 빅데이터는 통상 개인에 관한 정보를 담고 있는 경우가 많음

○ 따라서 빅데이터를 구성하는 개별 개인정보의 소유권자가 누구인지, 그리고 개인정보의 유통으로 해당 정보 주체의 프라이버시가 침해되는지에 관한 논란이 발생함

□ 이에 따라 개인정보에 관한 개인의 권리보호에 대한 논의가 시작되었고, 헌법 상 기본권의 일종인 개인정보 자기결정권의 개념이 확립되게 됨

[그림 2-4] 개인정보 통제권에 대한 긴장관계



출처 : "Rethinking Personal Data : Strengthening Trust", WEF, 2012

제3장 개인정보의 권리화와 보호강화

제1절 프라이버시권과 개인정보 자기결정권의 등장

□ 프라이버시는 시민혁명의 시대에 국가권력으로부터 개인의 사적 영역을 보장하려는 움직임에서 시작되어, 국가 혹은 다른 개인의 불법행위에 의해 침해되는 개인의 권리로 인정되기 시작함

○ 프라이버시의 어원은 '사람의 눈을 피한다'는 뜻의 라틴어 Privatue였으며, 초기에는 '홀로 있을 수 있는 개인의 일반적 권리'라는 소극적 의미를 가졌음

○ 1790년 발표된 프랑스 인권선언문의 사상과 표현의 자유는 서신의 비밀보장을 포함하는 것으로 해석되며, 독일은 1831년 헌법에 편지의 비밀보장을 명시하여 국가권력으로부터 시민의 자유를 보장하고자 함

○ 1890년 미국의 Warren과 Brandeis는 "The Right to Privacy"라는 논문에서 사적 주체에 의한 프라이버시 침해 개념을 도입하고, 프라이버시가 보통법 상 인정되는 권리이기에 불법행위에 기인한 손해배상의 근거가 된다는 것을 주장하면서, 프라이버시권의 개념이 정립되기 시작함

□ 프라이버시는 1960년대 헌법적 기본권으로 격상되었으며, 이후 정보 프라이버시 또는 개인정보 자기결정권으로 확장되었음

○ 1967년 미국 연방대법원의 Katz 판결에서는 헌법제정자는 개인에게 혼자 있을 권리를 부여하였다며 헌법 상 명문의 규정이 없던 프라이버시권을 헌법적 기본권으로 격상하였음

○ 또한 1977년 Whalen 판결에서는 프라이버시가 ① 자신의 중요한 문제를 자율적이고 독자적으로 결정내리는 이익과, ② 사적인 사항의 공개를 원하지 않을 이익을 보호한다고 하여, 개인정보 자기결정권으로 확장하였음

○ 독일에서는 1960년대부터 학설과 판례에서 개인의 자기표현권과 자기결정권

을 인정해 오다 1983년 연방헌법재판소의 인구조사 판결에서 ‘정보자기결정권’을 명시적으로 인정함

□ 1970년부터 세계 각국은 프라이버시 또는 개인정보를 보호하기 위해 법률 제정을 시작함

○ 1970년 독일의 헤센주가 세계 최초로 개인정보보호법을 제정했고, 1974년 미국도 연방기관이 보유하고 있는 개인정보의 보호법규인 프라이버시법을 제정함

○ 1977년 독일 연방정부는 개인정보보호의 기본법인 연방정보보호법을 제정하였으며, 1984년에는 영국이 정보보호법을 제정함

□ 이러한 과정을 거쳐 프라이버시는 공간·결정·정보 프라이버시로 확장되었고, 정보사회인 오늘날 정보 프라이버시에 해당하는 개인정보 자기결정권이 특히 주목받고 있음

○ 오늘날 프라이버시는 ‘사적 영역에 대한 부당한 침해나 공개를 당하지 않고, 개인정보에 대한 외부로부터의 부당한 접근을 방지하며, 자신의 동의하에 자신에 관한 정확한 정보가 유통될 수 있도록 통제할 권리’로 그 정의가 확장됨

- 공간 프라이버시는 개인이 일정한 공간 내에서 자신의 의사에 반하는 부당한 침해로부터 홀로 있을 범위를 말함

- 결정 프라이버시는 개인이 국가 혹은 타인의 간섭 없이 자율적으로 의사결정할 수 있는 자유를 의미함

- 정보 프라이버시는 개인정보가 부당하게 처리되는 것을 통제하기 위한 개인의 요구를 의미함

제2절 국내 개인정보 자기결정권의 발전

□ 국내에서는 프라이버시권에 대한 명문의 규정은 없으나, 제헌헌법과 제정 형법에서도 프라이버시권에 대한 인식은 있었음

○ 제헌헌법 제10조 거주이전의 자유와 제11조 통신의 자유[19], 제정형법 제316조 비밀침해죄는 국가권력으로부터 시민의 자유를 보호하려는 프라이버시권 관련 조항으로 볼 수 있음

○ 1987년 제정된 현행 헌법에서는 제17조 사생활의 자유 외에 제14조 거주이전의 자유, 제15조 직업선택의 자유, 제16조 주거의 자유, 제18조 통신의 자유 조항을 통해 사생활의 자유를 보장하고 있음

□ 이후, 전통적인 프라이버시 외에 자기정보에 관한 통제의 중요성이 널리 받아들여졌고, 대법원과 헌법재판소는 위의 헌법 조항들을 적극 해석하여 개인정보 자기결정권을 도출하고 있음

○ 헌법재판소는 1995년 헌법 제10조와 제17조를 토대로 공판정에 출석한 진술인의 진술녹음결정권을 인정했으며[20], 2005년 지문날인 사건에서 '개인정보 자기결정권'이라는 용어를 처음 사용하였고[21], 같은 해 NEIS 사건에서 개인정보 자기결정권이 헌법 제10조 및 제17조에서 도출된다고 판시하였음[22]

○ 대법원은 1998년 국군보안사 민간인 사찰사건에서 헌법 제10조 (인간의 존엄과 가치)와 제17조를 토대로 고도로 정보화된 현대사회에서 개인이 자신에 대한 정보를 적극 통제할 권리를 가진다고 해석하면서 국군보안사의 민간인 사찰행위를 기본권을 침해한 불법행위로 인정함[23]

---

19) 대한민국헌법[시행 1948.7.17.] [헌법 제1호, 1948.7.17., 제정]
제10조 모든 국민은 법률에 의하지 아니하고는 거주와 이전의 자유를 제한받지 아니하며 주거의 침입 또는 수색을 받지 아니한다.
제11조 모든 국민은 법률에 의하지 아니하고는 통신의 비밀을 침해받지 아니한다.

20) 헌법재판소 1995. 12. 28. 선고 91헌마114 결정 참조

21) 헌법재판소 2005. 5. 26. 선고 99헌마513, 2004헌마190(병합) 결정 참조
다만, 이 결정에서는 개인정보 자기결정권을 헌법 제10조와 제17조 및 헌법원리들에 완전히 포섭시키기 어려운 독자적인 기본권으로 보았으나 이후 NEIS 사건 결정에서 다시 변경됨

22) 헌법재판소 2005. 7. 21. 선고 2003헌마282,425(병합) 결정 참조

- 2006년에도 개인이 자신에 대한 정보를 적극 통제할 권리를 인정하면서, 보험회사 직원이 피해자의 일상생활을 촬영한 것이 피해자의 초상권 및 사생활 자유를 침해한 것으로 판결함[24)]

○ 개인정보 자기결정권은 자신에 관한 정보가 언제 누구에게 어느 범위까지 알려지고 또 이용되도록 할 것인지를 해당 정보주체가 스스로 결정할 수 있는 권리, 즉 정보주체가 개인정보의 공개와 이용에 관하여 스스로 결정할 권리로 정의됨[25)]

□ 정부에서는 이러한 개인정보 자기결정권의 발전에 발맞추어 개인정보 보호 관련 법률을 제정하거나 기존 법률의 내용을 보완하여 왔음

○ 1995. 1. 8. 처음으로 시행된 『공공기관의개인정보보호에관한법률』은 공공기관에서 처리되는 개인정보의 보호를 목적으로 하였고, 민간영역에서의 개인정보 보호는 1999. 7. 1. 처음 시행된 『정보통신망이용촉진등에관한법률』[26)]에서 일부 담당하였음

- 당시의 개인정보에 관한 정의는 현재 시행되는 개인정보보호법 상의 정의와 거의 유사함

○ 이후 개인정보 보호에 관한 일반법인 『개인정보 보호법』이 2011. 9. 30. 시행되었음

- 개인정보보호법은 공공과 민간을 망라하는 개인정보 처리원칙을 규정하여 개인정보 보호의 사각지대를 해소하고 개인정보 유출·오용·남용 등의 침해를 구제하고자 제정되었음[27)]

---

23) 대법원 1998. 7. 24. 선고 96다42789판결 참조
24) 대법원 2006. 10. 13. 선고 2004다16280 판결 참조
25) 헌법재판소 2005. 5. 26. 선고 99헌마513, 2004헌마190(병합) 결정 참조
26) 구 『전산망보급확장과이용촉진에관한법률』을 전부개정하여 제정되었음
27) 개인정보 보호법의 제정이유는 다음과 같다.
정보사회의 고도화와 개인정보의 경제적 가치 증대로 사회 모든 영역에 걸쳐 개인정보의 수집과 이용이 보편화되고 있으나, 국가사회 전반을 규율하는 개인정보 보호원칙과 개인정보 처리기준이 마련되지 못해 개인정보 보호의 사각지대가 발생할 뿐만 아니라, 최근 개인정보의 유출·오용·남용 등 개인정보 침해 사례가 지속적으로 발생함에 따라 국민의 프라이버시 침해는 물론 명의도용, 전화사기 등 정신적·금전적 피해를 초래하고 있는 바, 공공부문과 민간부문을 망라하여 국제 수준에 부합하는 개인정보 처리원칙 등을 규정하고, 개인정보 침해로 인한 국민의 피해 구제를 강화하여 국민의 사생활의 비밀을 보호하며, 개인정보에 대한 권리와 이익을 보장하려는 것임.

□ 개인정보보호법은 개인정보의 수집부터 폐기까지 전 범위에 걸쳐 사전동의(Opt-in)방식을 채택했고, 국내에서 계속 발생한 개인정보유출사고로 인해 제재규정도 엄격해져, 전세계에서도 가장 엄격한 편에 속함

<표 3-1> 각국 개인정보 규제 비교

구분	국가/지침	규제대상 개인정보	비고
정의	한국	생존한 개인의 정보(예: 성명,주민번호,영상등) + 다른 정보와 쉽게 결합해 개인식별가능 정보	·미국 제외, 개인정보 정의에 대해 포괄적 규정 ·일본은'조합'(대조)이란 용어를 사용한 반면, 한국은 '결합' 단어를 사용 : 한국이 개인정보를 가장 넓게 규정한 것으로 볼 수 있음
	일본	생존한 개인의 정보(예: 성명, 생년월일 등) + 다른 정보와 쉽게 조합해 개인식별가능 정보	
	EU지침	식별되거나 식별가능한 자연인 정보(예: 신분증 번호 또는 신체·생리·정신·경제·문화·사회적 요소 정보로 직·간접적 알아볼 수 있는 사람)	
	EU규칙(GDPR)	식별되거나 식별가능한 자연인 정보(예: 이름, 고유식별자, 온라인 아이디, 위치정보, 기타 신체·정신·생리적 특성)	
	미국	개별법상 상이(프라이버시, 금융정보, 소비자정보 의료정보, 통신정보 등)	
동의	우리나라	수집 등 처리에 사전동의 방식	현행법 상 우리나라의 경우만 개인정보의 전체 처리과정에서 사전동의(Opt-In) 방식 채택
	일본	사전동의 및 사후동의 방식	
	EU지침	사전동의 및 사후동의 방식	
	EU규칙(GDPR)	수집 등 처리에 사전동의 방식	
	미국	개별법 마다 상이(제3자 제공 시 사전동의 방식 등)	
제재	우리나라	형벌 및 행정벌 혼재(정보통신망법의 경우 강력한 형사처벌)	·우리나라의 경우 처리과정을 구분해 상이한 제재 부과 ·일본 2014년 법개정으로 완화
	일본	시정권고조치 위반 시 형벌	
	EU지침	없음	
	EU규칙(GDPR)	형벌	
	미국	개별법 상 상이(형벌 또는 행정벌)	

출처 : 한국정보화진흥원, "개인정보보호 법제로 인한 빅데이터 활용 한계사례 조사", 2015 를 보완

□ 개인정보 자기결정권은 익명권, 정보열람청구권 등의 구체적인 권리들을 포함하고 있으며, 국내에서는 개인정보보호법, 정보통신망법, 위치정보법, 신용정보법에 규정되어 있음

○ 2018. 5. 25.부터 시행될 EU의 개인정보보호법(GDPR[28])에서는 이 외에도 정보이동청구권, 정보처리반대권, 프로파일링[29] 관련 정보접근권 및 반대권 등

다양한 권리를 보장하고 있어, 개인정보 자기결정권은 계속 확장되어 가고 있음

〈표 3-2〉 개인정보 자기결정권의 세부내용

권리	권리내용	국내 관련법률
익명권	● 정보주체가 제3자와 온라인 교섭 시 자신의 신원을 밝히지 않고 거래할 수 있는 권리	없음[30]
정보처리 금지청구권	● 기본적인 정보처리 원칙 미충족 시, 개인정보의 수집, 이용, 제공 등 정보처리의 금지를 요구할 수 있는 권리 ● 수집제한의 원칙, 목적구속의 원칙, 시스템공개의 원칙	개인정보보호법 제37조 (정보통신망법 제30조 제3항[31])
정보열람 청구권	● 타인에 의해 처리되는 개인정보의 내용에 관해 정보주체가 열람을 청구할 수 있는 권리	개인정보보호법 제35조 정보통신망법 제30조 제2항 정보통신망법 제30조의2[32]
정보정정 청구권	● 정보주체에 관한 내용이 부정확하거나 불완전할 경우 정정을 요구할 수 있는 권리	개인정보보호법 제36조 정보통신망법 제30조 제2항
정보차단 청구권	● 정보주체가 정보보유기관에 대해, 권한없는 자의 자신의 개인정보에 접근하는 것을 막아줄 것을 요구할 권리 ● 진위여부에 다툼이 있는 동안 해당 정보에의 일반인의 접근을 막을 것을 요구할 수 있는 권리 ● 본인의 의사에 반해 일방적으로 게시된 정보에 대해 일반인의 접근을 차단하는 권리	없음[33]
정보분리 청구권	● 특정 목적을 위해 수집된 개인정보는 원칙적으로 분리된 상태를 유지할 것을 요구할 수 있는 권리	없음 (개인정보보호법 제21조 제3항[34])
정보삭제 청구권	● 정보보유기관이 위법 또는 부당하게 개인정보를 이용할 경우 자기정보의 삭제를 요구할 수 있는 권리	개인정보보호법 제36조 정보통신망법 제44조의2

28) General Data Protection Regulation의 약자임

29) 유럽 개인정보보호법(GDPR)에 따르면 프로파일링은 특정 개인을 분석하거나 행동패턴을 예측하기 위해 개인정보를 활용하는 자동화된 처리기법을 의미하며, 원문은 다음과 같음

‘profiling’ means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;

30) 1994년 12월 6일 제정된 오스트레일리아 프라이버시 헌장 제10조에 익명권이 규정되어 있음.

출처 : 노동일, 정완, “사이버공간상 프라이버시 개념의 변화와 그에 대한 법적 대응방안”, 경희법학 제45권 제4호, 2010

31) 이용자가 정보통신서비스 제공자 등에 대해 개인정보 수집·이용 등의 동의를 철회할 경우 지체없이 수집된 개인정보를 파기해야 하므로 정보처리금지청구권의 일종으로 파악됨

32) 위 조항은 정보통신서비스제공자에게 부과되는 개인정보 이용내역의 통지의무이나, 정보주체의 정보열람청구권을 실질적으로 보장하기 위한 것으로 보임

33) 인터넷기업들이 자율적으로 시행하는 게시물 임시 차단조치가 유사함

34) 수집한 개인정보가 불필요해지더라도 법령에 따른 보존의무가 있을 경우 다른 개인정보와 분리하여 저장할 것을 규정하고 있음.

개인정보보호법 제21조(개인정보의 파기) ① 개인정보처리자는 보유기간의 경과, 개인정보의 처리 목적 달성 등 그 개인정보가 불필요하게 되었을 때에는 지체 없이 그 개인정보를 파기하여야 한다. 다만, 다른 법령에 따라 보존하여야 하는 경우에는 그러하지 아니하다.

제3절 정보통신의 발달과 개인정보 보호의 강화

□ 정보통신기술의 발전으로 개인에 관련된 정보들의 수집·활용이 용이해지고, 이를 맞춤형 광고 등 기업활동에 활용하고자 하는 욕구가 증대됨

○ 이메일, 게임, 게시판 등의 정보통신서비스는 이용자를 구분하기 위해 자연스럽게 개인정보를 수집·처리하게 되어, 전통적인 공공영역 외에 민간영역에서도 방대한 개인정보가 축적됨

○ 또한 인터넷의 발전으로 이용자들의 프로파일과 행동양태를 수집하거나 검색키워드 같은 컨텍스트(context)까지 고려하여 보다 효과적으로 광고하는 기술이 개발되면서 개인정보의 경제적 중요성이 증대됨

- 인터넷 초창기에는 광고성 이메일은 무작위로 전송되었으나, 2000년대 초에 오버추어(Overture)는 검색키워드를 활용한 맞춤형 광고시스템을 개발·확산시킴

- 오버추어는 야후(Yahoo)에 합병되었고, 이후 구글(Google)의 애드센스(AdSense) 등 다양한 검색광고 서비스가 존재하고, 국내의 네이버와 다음도 검색광고로 수익을 내고 있음

- 아마존(Amazon)과 같은 전자상거래기업, 페이스북(Facebook)같은 사회적 관계망 서비스(Social Networking Service)에서는 이용자의 ID를 기반으로 구매패턴 등의 행동양태를 활용해 맞춤형 광고 및 추천 시스템을 운영하고 있음

□ 이에 따라 개인정보를 불법적으로 수집하거나 한 기업 내에 축적된 개인정보를 불법적으로 입수하여 광고 등에 활용하는 개인정보 침해사례가 갈수록 빈번하게 발생함

---

② 개인정보처리자가 제1항에 따라 개인정보를 파기할 때에는 복구 또는 재생되지 아니하도록 조치하여야 한다.

③ 개인정보처리자가 제1항 단서에 따라 개인정보를 파기하지 아니하고 보존하여야 하는 경우에는 해당 개인정보 또는 개인정보파일을 다른 개인정보와 분리하여서 저장·관리하여야 한다.

<표 3-3> 국내 개인정보 유출사건 일람표

연도	개인정보 유출업체(발생월, 명수/건수)	비고
'06 이전	현대백화점('94, 1200),엔씨소프트('05.5,50만), 국민은행(06.3., 3만여명)	
'08~'10	GS칼텍스('08,1125만), 옥션('08,1800만), 하나로텔레콤('08.4, 600만), 신세계몰('10.3,390만)	하나로텔레콤 배상판결 확정('12.9.)
2011	삼성카드(47만), 현대캐피탈(4월, 175만), 네이트(7월,3500만) 넥슨(11월,1320만), 하나SK카드(5만), 대부업체 등(1900만),한국엡손(8월, 35만), 조폐공사(9월, 92만)	넥슨 과징금 7.71억원, 과태료 1500만원, 시정명령 한국엡손 과징금 3300만, 과태료 900만, 시정명령
2012	SKT·KT(3월, 20만), KT(7월, 870만), EBS(4월, 400만), 웅진코웨이(198만)	KT 1심 배상판결('14.8.22.), 과징금 7.53억원, 시정명령 EBS 과태료 1000만원, 시정명령
2013	한화손보(5만), 메리츠화재(5만), SC은행·시티은행(12월, 13만)	
2014	KB국민카드(1월, 5300만), 롯데카드(1월, 2600만), NH농협카드(1월, 2600만), KT(3월, 1200만), 대한의사협회등(2월, 15만여명), 티켓몬스터 등(3월, 1700만), 재향군인회(3월,1만여명), BBQ(4월, 51만), 천재교육(4월,350만), 스킨푸드(4월, 55만), 토니모리(5월, 50만)	KT 과징금 7000만원, 과태료 1500만원, 시정명령 카드3사 유출사건 1심 배상판결('16.1.22)
2015	약학정보원(진료정보 43억4천만건), 지누스(진료정보 7억2천만건), SKT(처방전 7802만건)	

출처 : SPRi, "4차 산업혁명과 지능정보사회에 대응하는 소프트웨어 중심의 국가혁신전략", 2016

□ 정부는 공공과 민간의 정보보호 및 개인정보 관리체계를 점검하기 위한 국제표준을 도입하고 이를 강화한 별도의 인증제도를 실시해왔음

○ 일반적인 정보보호 인증제도로는 국제표준화기구 ISO의 ISO27001과, 한국인터넷진흥원이 ISO27001을 기반으로 보안요건을 강화한 ISMS (Information Security Management System)이 존재함

○ 개인정보에 특화된 인증제도로는 민간기업이 자율적으로 신청할 수 있는 PIMS(Personal Information Management System), 자영업자를 포함한 중소규모 조직에 대한 PIPL(Personal Information Protection Level)이 각각 존재했으나, 내용이 많이 중복된다는 지적에 따라 2016. 1. 1.부터 PIPL을 폐지하고 PIMS에 통합하여 운영하기로 함

○ 또한 일정 기준 이상의 개인정보파일[35]을 운용하는 공공기관은 정보주체의 개인정보 침해의 위험요인의 분석과 개선사항 도출을 위한 개인정보 영향평

가를 의무적으로 실시해야 하며, 그 외의 개인정보 처리자에게도 권장하고 있음

<표 3-4> ISO 27001과 ISMS의 비교

항목	ISO/IEC 27001[36]	KISA ISMS
제정시기	2005. 10.	BS7799에 토대하여 2002년부터 시행
토대	1995. 2. 영국 BSI, BS7799-1 1998. 2. 영국 BSI, BS7799-2	ISO/IEC 17799:2000, BS7799-2 ISO/IEC 27001
심사항목	11개 분야, 133개 통제항목	18개 분야, 104개 통제항목 (ISO/IEC 27001 항목 모두 포함)
비고		미래창조과학부 관할 2010. 시작되던 G-ISMS를 2014년에 통합 정보통신망법 제47조에 근거 2016. 6. 1. 매출/세입 1,500억원 이상 기업을 의무대상자로 추가 지정

<표 3-5> PIMS와 PIPL 통합 변경사항

항목	통합전		통합후
인증명	PIPL(개인정보보호인증)	PIMS(개인정보관리체계인증)	PIMS(개인정보관리체계인증)
인증대상	모든 개인정보처리자 (정보통신서비스제공자 제외)	정보통신서비스 제공자	모든 개인정보처리자
평가항목	65개	124개	86개
수행기관	NIA(행정자치부)	KISA(방송통신위원회)	KISA (행자부, 방통위)
근거법률	개인정보보호법	정보통신망법	개인정보보호법 정보통신망법
도입시기	2013.11	2011.11	2016.1[37]
인증현황 (2015.10.)	13개 기관	29개 기관	42개 기관

출처 : LG CNS[38]

35) 개인정보 영향평가가 의무화되는 개인정보파일의 기준은 다음과 같다
개인정보보호법 시행령 제35조
1. 구축·운용 또는 변경하려는 개인정보파일로서 5만명 이상의 정보주체에 관한 법 제23조에 따른 민감정보(이하 "민감정보"라 한다) 또는 고유식별정보의 처리가 수반되는 개인정보파일
2. 구축·운용하고 있는 개인정보파일을 해당 공공기관 내부 또는 외부에서 구축·운용하고 있는 다른 개인정보파일과 연계하려는 경우로서 연계 결과 50만명 이상의 정보주체에 관한 개인정보가 포함되는 개인정보파일
3. 구축·운용 또는 변경하려는 개인정보파일로서 100만명 이상의 정보주체에 관한 개인정보파일

36) ISO 27001의 자세한 제정경위에 대해서는 아래 링크 참조
https://www.kab.or.kr/sys_guide/?CodeFlag=0005

37) 새로운 평가항목(86개)는 2017. 1.부터 적용하도록 유예함

38) http://blog.lgcns.com/1074

□ 또한 개인정보 보유기관의 고의 또는 과실로 유출사건이 발생한 경우에 부과하는 과징금, 과태료 같은 행정벌 외에도 벌금 등 형사처벌도 계속 강화되어 옴

〈표 3-6〉 개인정보보호법과 정보통신망법의 제재규정

항목	개인정보보호법	정보통신망법
손해배상	(2015. 7. 24.)[39] 300만원 이하 법정손해배상, 3배 배상제도 도입	(2016. 9. 23.) 3배 배상제도 도입 (2014. 11. 29.) 300만원 이하 법정손해배상 도입
과태료	(2015. 7. 24.) 암호화미조치에 과태료 3천만원 부과	(2016. 6. 2.) ISMS 인증의무자의 미인증 시 과태료 1천만원 -> 3천만원 상향 (2014. 11. 29.) 개인정보 유출 24시간 이내 신고의무 불이행과 거짓소명에 과태료 3천만원 (2012. 8. 18.) 주민번호 수집규정 위반 시 과태료 3천만원 이하. 침해사고 미신고 시 과태료 1천만원 이하.
과징금	(2014. 8. 7.) 주민번호 유출 시 5억원 이하의 과징금	(2014. 11. 29.) 과징금 기준을 매출액 100분의 3으로 상향. 개인정보 유출 시 1억원 이하 과징금 부과가능 조항을 삭제. 위탁자의 관리감독 소홀에도 과징금 부과
형사처벌	(2015. 7. 24.) 제73조 위반행위 시 2천만원 이하 벌금으로 상향	(2014. 11. 29.) 개인정보 미파기자 2년 이하 징역 또는 2천만원 이하 벌금
몰수추징	(2015. 7. 24.) 개인정보보호법위반행위(제70조~제73조)에 관련된 금품, 이익에 대한 몰수, 추징 신설	(2016. 9. 23.) 특정 형사처벌 범죄행위에 관련된 금품, 이익에 대한 몰수, 추징 신설

제4절 빅데이터 활용과 개인정보 비식별화

□ 앞서 서술한 빅데이터의 활용사례들과 맞춤형 광고기술은 대부분 이용자의 개인정보에 근거한 것이어서 개인정보보호법의 규제대상이 됨

○ 월마트의 소비자패턴 분석, 리츠칼튼 호텔의 고객맞춤형 서비스, BoA의 고객 맞춤형 상품제안, 허츠의 VOC 분석사례, 온라인 여행사이트 오비츠의 맞춤형 상품제시, UNC헬스케어 사례가 해당됨

□ 또한 개인정보보호법은 수집제한의 원칙과 목적구속의 원칙에 따라 이미

39) ()안은 시행일을 의미함

구축된 빅데이터를 다른 용도로 활용하는 것과 수집한 기업 외의 다른 기업이 활용하는 것에 대해서도 사용자의 사전동의를 원칙으로 하고 있어 빅데이터의 활용에는 많은 제약이 존재함

○ 개인정보보호법 상 수집한 개인정보는 수집 목적 범위에서 이용할 수 있고(제15조), 정보주체의 동의를 받거나 수집목적 범위 내에서 개인정보를 제공할 수 있으며(제17조), 개인정보를 제공받은 자도 정보주체의 별도 동의나 법률 상 특별한 규정이 없는 한, 목적 외의 용도로 개인정보를 활용하거나 제3자에게 제공해서는 안됨(제19조)

○ 개인정보를 수집목적 외로 이용·제공하려면 개인정보보호법 제18조 제2항[40]에 따른 요건들을 충족해야 하는데, 정보주체의 동의를 얻거나 법률 상 규정이 있거나 공공기관의 업무용도에 한정하는 등 그 요건이 매우 까다로우며, 그 조차도 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때는 허용되지 않음

○ 다만, 제4호에서 통계작성 및 학술연구 등의 목적으로 특정 개인을 알아볼 수 없는 형태로 제공하는 경우에서 허용하고 있음

○ 따라서 개인정보를 보유한 회사가 수집 시 사전동의를 받지 않은 관계 회사나 다른 기업에게 개인정보를 일부라도 제공할 수 있는 합법적 경우는 오직 제4호가 유일하다고 볼 수 있어 개인정보 비식별화기술의 중요성이 높아지

---

40) 개인정보보호법 제18조(개인정보의 목적 외 이용·제공 제한) ② 제1항에도 불구하고 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 **정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외**하고는 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다. 다만, **제5호부터 제9호까지의 경우는 공공기관의 경우로 한정**한다.

1. 정보주체로부터 별도의 동의를 받은 경우
2. 다른 법률에 특별한 규정이 있는 경우
3. 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전 동의를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위하여 필요하다고 인정되는 경우
4. **통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우**
5. 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우로서 보호위원회의 심의·의결을 거친 경우
6. 조약, 그 밖의 국제협정의 이행을 위하여 외국정부 또는 국제기구에 제공하기 위하여 필요한 경우
7. 범죄의 수사와 공소의 제기 및 유지를 위하여 필요한 경우
8. 법원의 재판업무 수행을 위하여 필요한 경우
9. 형(刑) 및 감호, 보호처분의 집행을 위하여 필요한 경우

고 있음

○ 따라서 다음 장에서는 개인정보를 특정 개인을 알아볼 수 없게끔 하는 비식별화 방법과 다른 정보를 결합하여 특정 개인이 식별될 가능성을 평가하는 비식별화 평가방법에 관하여 서술함

제4장 개인정보 비식별화기술

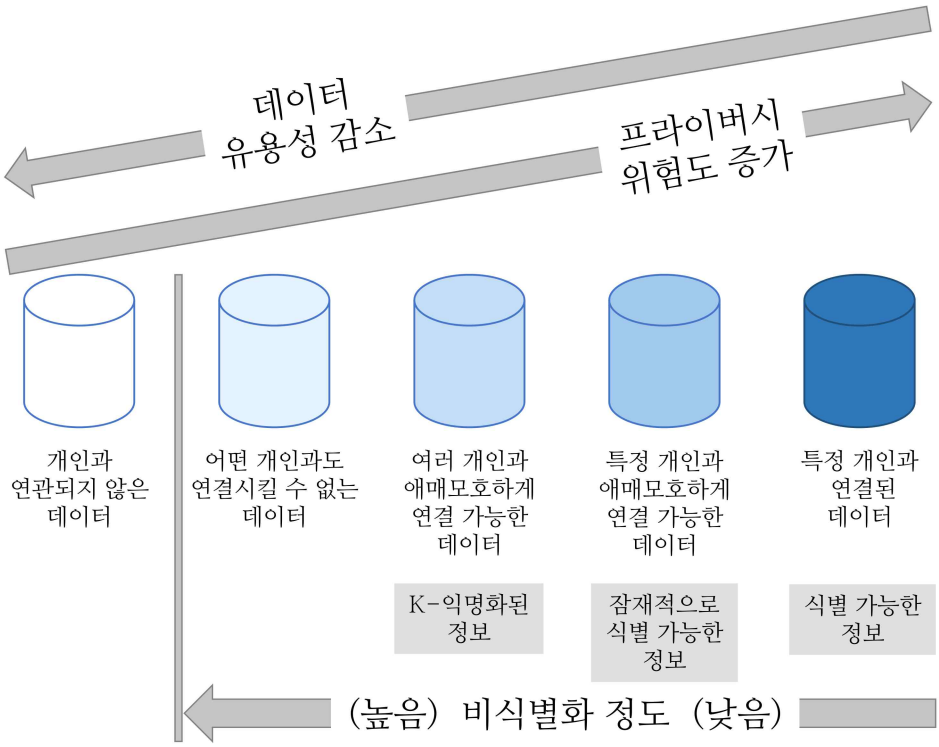
제1절 비식별화기술의 개요

□ (비식별화기술) 수집(collect) 또는 사용(use), 저장(archive), 공유(share)되는 데이터로부터 개인을 식별하지 못하게 조치하는 일련의 방법

○ NIST의 개인정보 비식별화[41] 내부 보고서에 따르면 비식별화의 궁극적인 목표는 [그림 4-1]에서와 같이 데이터를 어떤 개인과도 연결시킬 수 없도록 만드는 것

- 비식별화 정도가 높을수록 데이터의 유용성은 떨어지지만 프라이버시 침해 위험도는 낮아짐

[그림 4-1] 데이터 식별가능성 스펙트럼



41) S. Garfinkel, De-Identification of Personal Information, NISTIR 8053, 2015. 10.

□ 비식별화기술은 무작위화 방법(randomization)과 일반화 방법(generalization)으로 대부분 분류됨[42)]

○ (무작위화 방법, randomization) 데이터의 신뢰성(진실성 또는 정확성, veracity)을 임의로 낮춤으로써 특정 데이터와 개인 간 강한 연결성(strong link)을 제거하는 방식

- 예를 들어, [그림 4-2]에서와 같이 '준식별자'[43)]인 우편번호를 "A", "B", "C" 등과 같은 가명(pseudonymization)으로 대체하거나 또 다른 준식별자인 나이를 ±5 범위의 임의의 값(random value)을 더하여 특정 개인과의 강한 연결성을 제거하는 방식

○ (일반화 방법, generalization) "서울시 종로구 효자동"을 "서울시"로, "2016. 5. 16."을 "2016."과 같이 데이터 값을 보편적인 범위 또는 의미로 변경하여 특정 개인을 식별하지 못하게 하는 방식

- 예를 들어, [그림 4-2]에서와 같이 실제 우편번호를 "*"로 가려 어느 지역인지 정확히 알 수 없게 하거나 나이를 "20대", "30대", "40대" 등으로 범주화하여 개인을 식별하기 힘들게 하는 방식

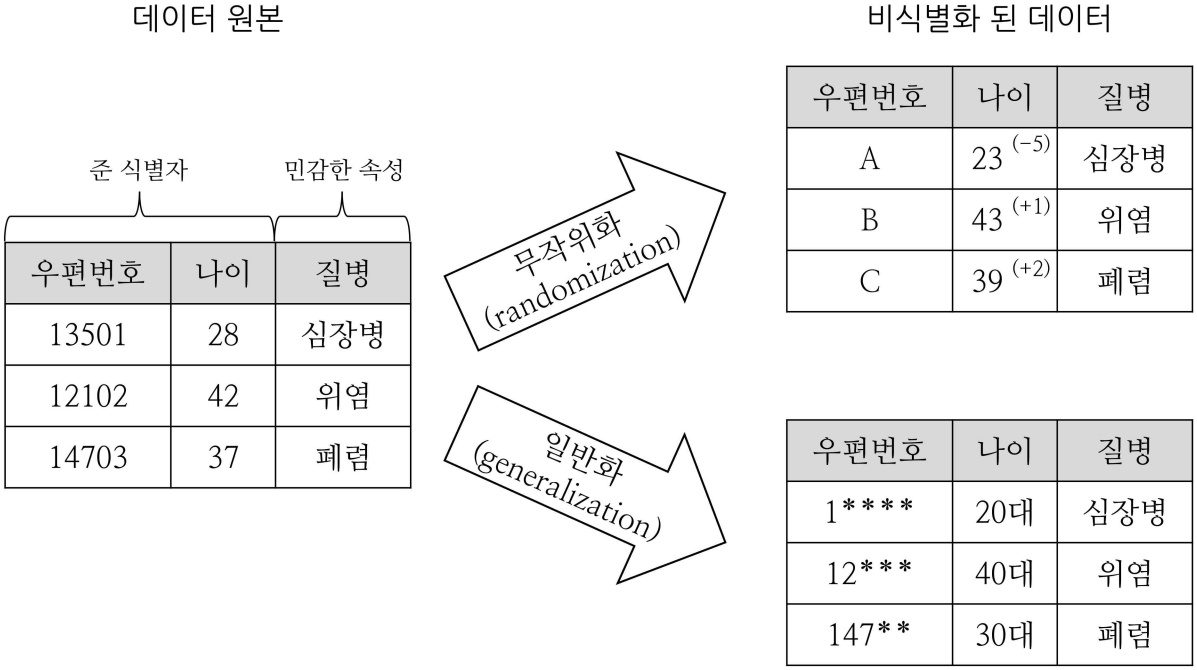
○ 실제 데이터를 비식별화할 때에는 데이터의 특성, 알려진 혹은 앞으로 알려질 데이터 유무 등을 고려하여 무작위화 방법과 일반화 방법을 적절히 조합해서, 비식별화 이후에 재식별화[44)]가 어렵거나 불가능해야 함

42) ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 05/2014 on Anonymisation Techniques, 2014. 4. 10

43) 준식별자 : 직접적으로 대상을 알 수는 없지만 조합을 통해 간접적으로 개인 식별 가능한 생년월일, 성별, 우편번호 등을 말함

44) 재식별화 : 비식별화한 개인정보를 다른 정보 또는 데이터와 비교, 연계, 결합 등을 통해 특정 개인을 알아볼 수 있도록 하는 일련의 조치

[그림 4-2] 무작위화와 일반화의 예



□ ‘비식별조치 가이드라인’은 [그림 4-3]에서와 같이 크게 5가지 범주(세부 기술 17종)의 비식별화기술 소개

○ 세부 기술 17 종에 대한 상세 설명과 예시는 해당 가이드라인 참조

○ 개인정보가 충분히 비식별화가 되었는지에 대한 적정성 평가방법, 전문기관을 통한 기업 간 정보집합물(dataset) 결합 지원, 비식별정보 제공 및 위탁계약 시 준수사항 등 개인정보의 비식별조치 기준과 비식별정보의 활용범위 등을 좀 더 명확히 제시함

[그림 4-3] 비식별화 주요 기술 17종

□ 비식별화기술 17종은 무작위화 그룹과 일반화 그룹으로 분류가능함([그림 4-4] 참조)

○ (무작위화 그룹) 휴리스틱 가명화, 암호화, 교환 방법, 재배열, 임의 잡음 추가 등

- (휴리스틱 가명화, heuristic pseudonymization) 데이터를 정해진 규칙으로 가명처리하여 실제 누구 데이터인지 알 수 없게 하는 기술
- (암호화, encryption) 암호화 알고리즘[45]을 기반으로 개인정보를 암호화하여 숨기는 기술
- (교환 방법, swapping) 민감한 데이터를 사전에 정해진 외부 데이터로 치환하는 기술
- (재배열, rearrangement) 그룹 내 데이터를 임의로 섞어 특정 데이터와 개인 간 연결성을 끊는 기술

45) 암호화/복호화 가능한 양방향 암호화, hash 등과 같은 기술을 이용하여 복호화 불가능한 단방향 암호화 등이 있고, 경우에 따라서는 토큰화(임의로 생성된 값(토큰)을 신용카드번호, 주민등록번호와 같이 민감한 정보 대신 사용하는 기술), 형태보존 암호화(암호화 후 평문의 형태와 암호문의 형태가 동일함을 보장하는 암호화 기술), 일회성 임시 식별자(사용할 때마다 임시 키(key)를 부여 받아 식별자를 암호화하고 사용이 끝나면 임시 키와 암호화된 식별자를 삭제하는 방식) 등도 사용됨

- (임의 잡음 추가 방법, adding random noise) 임의의 노이즈(random noise) 값을 넣어 식별정보 노출을 방지하는 기술
- 모두 무작위화 방법을 이용하여 특정 데이터와 개인 간의 강한 연결성(strong link)을 제거하는 기술임

○ (일반화 그룹) 총계처리 기본방식, 부분총계, 라운딩, 식별자 부분삭제, 감추기, 랜덤 라운딩, 범위 방법, 제어 라운딩, 공백과 대체 등

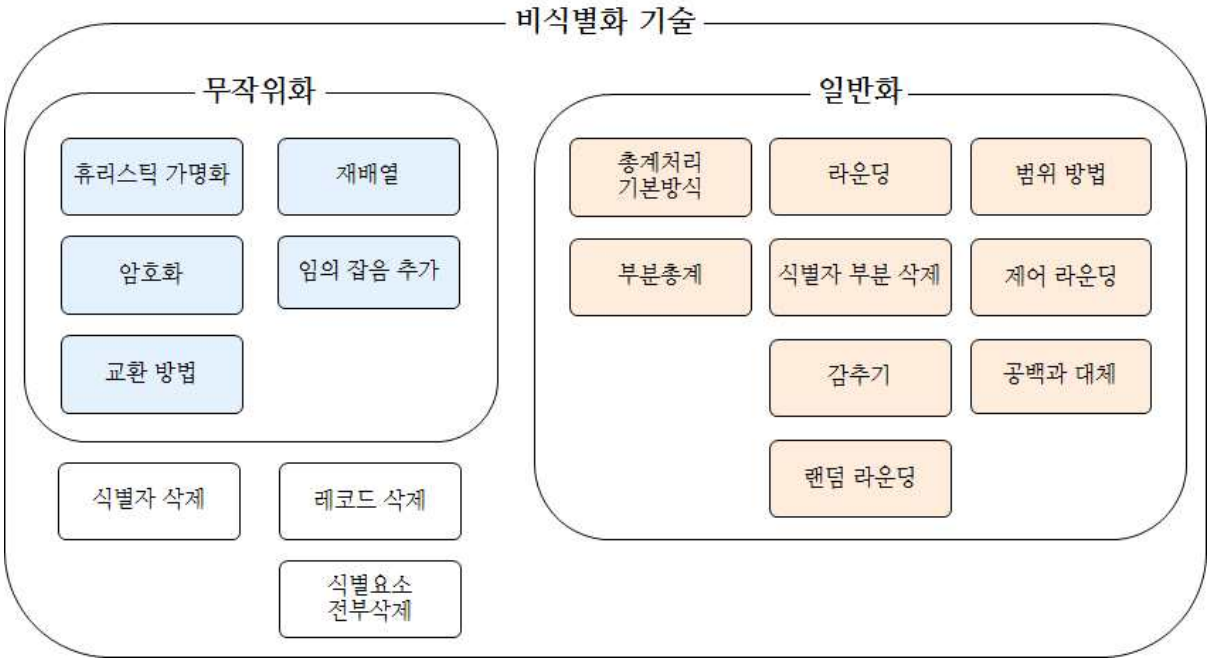
- (총계처리 기본방식, aggregation) 데이터의 총합이나 평균으로 개인의 실제 정보를 숨기는 기술
- (부분총계, micro aggregation) 다른 속성 값에 비해 오차 범위가 크거나 특징적인 경우 해당 속성 값에 대해서만 통계 값을 적용하여 개인을 식별하지 못하게 하는 기술
- (라운딩, rounding) 올림, 내림, 반올림 등의 방법을 사용하여 개인의 실제 정보를 숨기는 기술
- (식별자 부분삭제, reducing partial variables) 속성의 일부 값을 삭제하여 대표성을 가진 값으로 보이게 하는 기술
- (감추기) 데이터의 평균 또는 범주값으로 변환해 일반화하는 기술
- (랜덤 라운딩, random rounding) 임의의 값을 기준으로 해당 값을 올리거나 내려 민감성이 높은 정보를 대표값으로 처리하는 기술
- (범위 방법, data range) 개인 수치데이터를 범위나 구간으로 표현
- (제어 라운딩, controlled rounding) 행과 열의 합이 일치되도록 고려하여 값을 라운딩(rounding)하는 기술
- (공백과 대체, blank and impute) 속성 값 일부를 공백처리하고 특수문자 등으로 채우는 기술 등
- 모두 일반화 방법을 이용하여 데이터 값을 보편적인 범위 또는 의미로 변경하여 특정 개인을 식별하지 못하게 하는 기술임

○ (기타) 식별자 삭제, 레코드 삭제(reducing records), 식별요소 전부삭제 등은

무작위화 그룹도 일반화 그룹에도 포함되지 않음

- 개인을 식별해 낼 수 있는 정보 자체를 삭제하는 방식이기 때문에 무작위화 방법이나 일반화 방법 등에 해당하지 않음

[그림 4-4] 17종 비식별화기술의 재분류



□ 비식별화된 개인정보는 상황에 따라 재식별 가능성이 존재하기 때문에 비식별화기술에 대한 개별화 가능성, 연결 가능성, 추론 가능성 등 3가지 위험성을 다시 검토해야 함

- (개별화 가능성, single out) 데이터셋에서 특정 레코드의 일부 혹은 전체가 한 개인에게 대응될 가능성
- (연결 가능성, linkability) 최소 2개 이상의 레코드들이 동일한 데이터 주체를 가리켜 서로 연결될 가능성 (모든 레코드가 동일한 데이터베이스에 존재할 필요는 없고 특정 개인에 대응될 필요도 없음)
- (추론 가능성, inference) 데이터셋의 속성값들로부터 특정 개인에 해당하는 임의의 속성 값을 상당히 높은 확률로 유추할 가능성

<표 4-1> 비식별화기술 17종 위험도 비교

분류	비식별화기술	개별화 가능성 위험	연결 가능성 위험	추론 가능성 위험
무작위화	휴리스틱 가명화	Yes	Yes	Yes
	암호화	Yes	Yes	May not
	교환 방법	Yes	Yes	May not
	재배열	Yes	May not	May not
	임의 잡음 추가	Yes	May not	May not
일반화	총계처리 기본 방식	일반적으로 No	일반적으로 Yes	일반적으로 Yes
	부분총계			
	라운딩			
	식별자 부분 삭제			
	감추기			
	랜덤 라운딩			
	범위 방법			
	제어 라운딩			
	공백과 대체			
기타	식별자 삭제	해당 없음	해당 없음	해당 없음
	레코드 삭제	해당 없음	해당 없음	해당 없음
	식별요소 전부삭제	해당 없음	해당 없음	해당 없음

출처 : ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 05/2014 on Anonymisation Techniques, 2014. 4. 10, 재구성

□ 비식별화기술 17종에 대한 개별화 가능성, 연결 가능성, 추론 가능성 정도는 각 기술에 따라 다름 (<표 4-1> 참조)

○ 무작위화 방법의 경우 대부분 개별화 가능성에 대한 위험이 있음

- 특히, 가명화 방법은 재식별 위험이 매우 높기 때문에 다른 비식별화기술과 함께 사용해야 함

○ 일반화 방법의 경우 개별화 가능성은 낮지만 연결 가능성과 추론 가능성이 높음

○ 비식별화 시 한 기술만 이용할 경우 재식별화 위험이 있기 때문에 여러 비식별화기술을 조합하여 사용해야 함

제2절 비식별화기술의 불완전성

□ 비식별화된 데이터는 이미 공개된 혹은 앞으로 추가로 공개될 데이터와 결합하여 재식별화될 가능성이 있음

○ 미국의 매사추세츠 주 사례, AOL 사례, Netflix 사례 등이 대표적임

□ 매사추세츠 주 사례(미국, 1997)

○ 미국 매사추세츠 주의 단체보험위원회는 연구 목적으로 공무원 병원 진료기록을 <표 4-2>와 같이 공개

- 식별자인 이름, 주소, 사회보장번호(SSN)은 모두 제거
- 준 식별자인 우편번호(ZIP), 생년월일, 성별 등은 공개

○ 하버드 대학교의 L.Sweeney가 공개된 진료기록에서 개인 재식별 성공

- 매사추세츠주 케임브리지시 선거인 명부(<표 4-3> 참조) 구입 후 비교하여 주지사를 포함해 개인정보 재식별

<표 4-2> 매사추세츠주 공무원 병원 진료기록

Name	SSN	Ethnicity	Date of Birth	Sex	ZIP	Marital Status	Problem	…
		…	…	…	…	…	…	…
		Asian	09/27/64	female	02139	divorced	hypertension	…
		Asian	09/30/64	female	02139	divorced	obesity	…
		Asian	04/18/64	male	02139	married	chest pain	…
		Asian	04/15/64	male	02139	married	obesity	…
		Black	03/13/63	male	02138	married	hypertension	…
		Black	03/18/64	male	02138	married	shortness of breath	…
		Black	09/13/64	female	02141	married	shortness of breath	…
		Black	09/07/64	female	02141	married	obesity	…
		White	05/14/61	male	02138	single	chest pain	…
		White	05/08/61	male	02138	single	obesity	…
		White	09/15/61	female	02142	widow	shortness of breath	…
		…	…	…	…	…	…	…

<표 4-3> 매사추세츠주 케임브리지시 선거인 명부

Name	Address	City	ZIP	DOB	Sex	Party	…
…	…	…	…	…	…	…	…
Sue J. Carlson	1459 Main St.	Cambridge	02142	9/15/61	female	democrat	
…	…	…	…	…	…	…	…

□ America Online (AOL) 사례(미국, 2006)

○ 학술 연구를 위해 65만명의 3개월분 검색로그 2천만건을 공개

- 개인을 식별할 수 있는 ID와 IP 주소는 비식별화함
- 단, 유용성 확보를 위해 ID를 특정번호 식별자(AnonID)로 치환

  * <AnonID, Query, QueryTime, ItemRank, ClickURL> 형태로 공개

○ 뉴욕타임즈 기자 2명이 AnonID “4417749” 사용자 재식별 성공

- 해당 AnonID 사용자는 조지아 주 릴번에 거주하는 62세 미망인 Thelma Arnold 여사였음

  * 질의문에 포함된 “numb fingers”, “60 single men”, “dog that

urinates on everything”, “landscapers in Lilburn, GA”, “homes sold in shadow lake subdivision gwinnett county georgia” 등 다수의 개인 정보를 통해 Thelma Arnold 여사를 식별해 냄

- 검색로그 공개 일주일 만에 데이터 공개중지 및 대중에게 사과

* 관련 연구자와 해당 상사는 해고되고 AOL의 CTO는 사임함

□ Netflix 사례(미국, 2006)

○ 영화 추천 알고리즘의 정확성을 높이기 위한 경연대회 “Netflix Prize”를 위해 50만 이용자의 6년 동안 영화 평점 1억 건을 공개

- 이름 등 식별자 삭제하되, 데이터 처리 내용을 연결하기 위해 독특한 식별자 사용

- 평가점수, 평가일시는 공개

○ 텍사스 대학 연구팀은 온라인 영화전문사이트인 IMDb[46]에 공개된 사용자 리뷰와 Netflix가 공개한 영화평점 데이터를 결합하여 개인 재식별 성공

○ 미국 FTC의 프라이버시 문제 지적으로 제2회 경연은 취소됨

제3절 비식별화기술의 평가척도

□ 비식별화된 데이터라도 재식별화에 대한 잠재된 위험성을 갖고 있기 때문에 비식별화된 데이터의 재식별 난이도에 대한 평가척도 필요

□ 재식별 가능성 여부는 다음과 같은 여러 요소에 영향을 받음

○ 데이터에서 준식별자의 존재 정도 및 비식별화 정도

○ 데이터의 공개 형태

46) IMDb : Internet Movie Database의 약자로 영화뿐만 아니라 TV프로그램, 비디오 게임 등에 대한 정보를 포함한 온라인 데이터베이스

○ 공격자 존재 가능성 및 알려진 혹은 알려질 타 데이터 유무

○ 개인정보 유출 시 침해위험 정도 등

[그림 4-5] 비식별 조치 및 사후관리 절차

출처 : 국무조정실 외, “개인정보 비식별 조치 가이드라인”, 2016. 6. 30.

비식별화 정도가 적절한지를 판단하기 위해서는 위의 재식별 가능성 요소를 고려하여 평가 기준값을 결정하고 계량적인 분석 필요

○ k-익명성(k-anonymity), l-다양성(l-diversity), t-근접성(t-closeness) 등이 비식별화 정도에 대한 계량분석 방법으로 많이 쓰임

□ k-익명성(k-anonymity)

○ 동치류(equivalence class)[47] 내의 레코드 개수가 적어도 k개 이상 존재함을 보장

- [그림 4-6] (가)와 같이 데이터가 주어지고 준식별자로 “우편번호”와 “나이”, 민감한 속성[48]으로 “질병”이라 할 때, [그림 4-6] (나)는 3-익명성이 보장된 데이터임

47) 동치류(equivalence class) : 비식별화 후 동일 값을 갖는 준 식별자 그룹

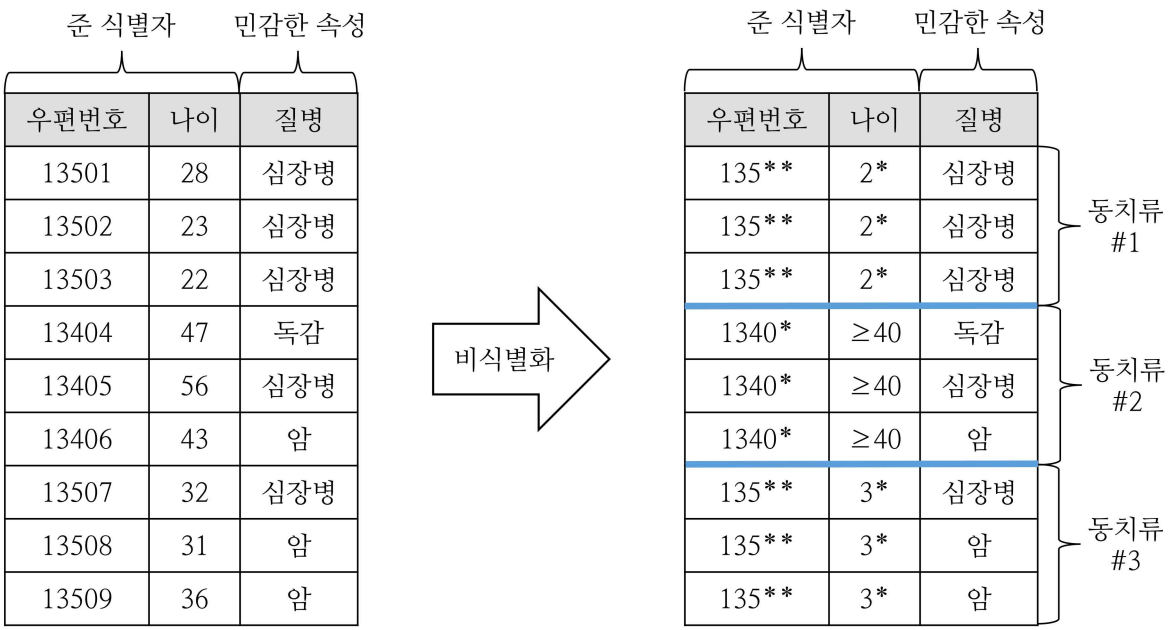
48) 민감한 속성 : 특정한 개인의 사상・신념, 노동조합・정당의 가입・탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 정보(개인정보보호법 제23조)

- [그림 4-6] (나)에서 동치류 #1의 준식별자 “우편번호”는 모두 “135**”, “나이”는 모두 “2*”

  * 즉, 동일 동치류 내에서 각각의 준식별자는 모두 같은 값을 가지며 해당 동치류의 레코드 개수는 3개임

  * 마찬가지로 동치류 #2와 #3 역시 각각의 동치류 내 레코드 개수가 3개이므로 3-익명성 보장

[그림 4-6] 비식별화 이후 3-익명성 보장



○ k의 값이 커질수록 동치류 내의 레코드 개수가 많아지기 때문에 공격자가 개인을 재식별할 가능성이 낮아짐 (반대로 k의 값이 작아질수록 재식별 가능성 높아짐)

○ (취약점 1) 동질성 공격(homogeneity attack)에 취약

- [그림 4-6] (나)의 데이터는 3-익명성을 보장하지만 만약 공격자가 원본

데이터에 포함된 “홍길동”에 대해 우편번호가 “13502”이고 나이가 “23세”라는 추가 데이터를 확보하였다면, “홍길동”이 “심장병”을 앓고 있다고 100% 유추할 수 있음

- 이는 동치류 #1의 민감한 속성인 “질병”의 값이 모두 “심장병”이기 때문에 해당 동치류에 속하는 모든 사람의 “질병”이 의도치 않게 드러나게 됨

○ (취약점 2) 배경지식 공격(background knowledge attack)에도 취약

- 3-익명성을 보장하는 [그림 4-6] (나)의 데이터에서 만약 공격자가 “임꺽정“에 대해 우편번호가 “13508”이고 나이가 “31세”라는 데이터를 추가로 확보하였다면, 이론상 “임꺽정”이 1/3의 확률로 심장병을 앓거나 2/3의 확률로 암환자일 것이라고 유추

- 공격자가 배경지식(background knowledge)으로 “임꺽정”에게 심장병이 없다고 알고 있다면, “임꺽정”이 100% 암환자임을 알게 됨

○ 같은 동치류에 속하는 민감한 속성이 얼마나 서로 동일한지를 고려하지 않기 때문에 k-익명성은 동질성 공격과 배경지식 공격에 취약

○ 동질성 공격과 배경지식 공격에 취약한 k-익명성의 한계를 극복하고자 l-다양성 프라이버시 보호 모델 등장

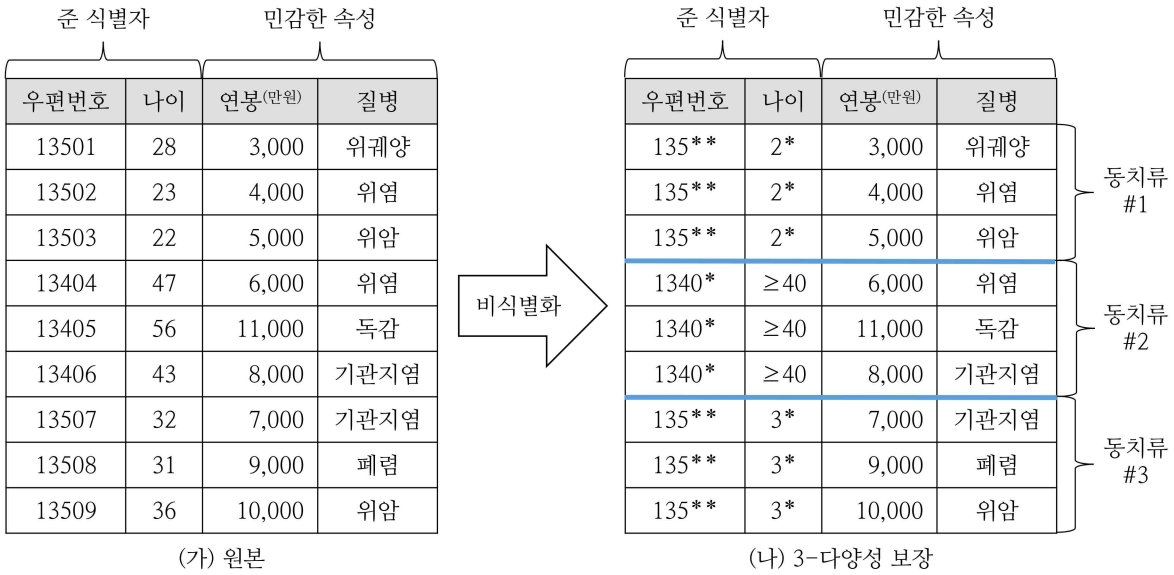
□ l-다양성(l-diversity)

○ 임의의 동치류 내의 서로 다른 민감한 속성 값이 l개 이상 존재함을 보장

○ [그림 4-7] (가)의 데이터와 같이 준식별자로 “우편번호”와 “나이”, 민감한 속성으로 “연봉”과 “질병”이 주어졌을 때, [그림 4-7] (나)의 데이터는 3-다양성이 보장된 데이터임

- 즉, (나)의 동치류 #1, #2, #3은 민감한 속성인 “연봉”과 “질병” 모두 각각 서로 다른 3개의 값을 갖기 때문에 3-다양성 보장

[그림 4-7] 비식별화 이후 3-다양성 보장



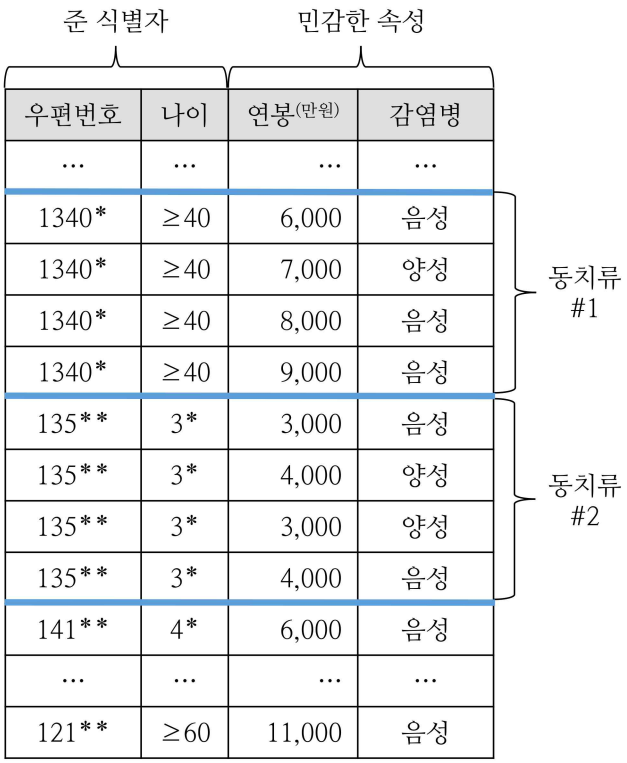
○ l-다양성은 쏠림 공격(skewness attack)에 취약

- [그림 4-8]의 데이터는 2-다양성을 보장하고 민감한 속성인 “감염병”에 대해 10,000개의 레코드 중 1%는 “양성”, 나머지 99%는 “음성”이라고 가정
- 동치류 #1과 #2에 대한 “양성” 레코드의 비율은 각각 25%와 50%임
- “양성” 레코드의 전체 비율인 1%보다 높은 비율로 “양성” 레코드가 동치류 #1과 #2에 쏠림으로써 개인의 민감한 정보가 노출될 가능성이 높아짐

  * 즉, 공격자는 동치류 #2에 속하는 개인의 경우 50%의 확률로 감염병에 걸렸다는 것을 유추할 수 있음

- 이는 각각의 동치류가 전체 분포 또는 비율을 고려하지 않고 l-다양성만을 만족하였기 때문에 발생

[그림 4-8] 2-다양성과 쏠림공격



○ l-다양성은 유사성 공격(similarity attack)에도 취약

- 공격자가 “홍길동”이 비교적 낮은 급여(3,000만원에서 5,000만원 사이)를 받고 있다는 사실을 알고 있다고 가정
- “홍길동”은 41면의 [그림 4-7] (나)의 동치류 #1에 속하기 때문에 공격자는 “홍길동“의 정확한 병명은 알 수 없지만 위장 관련 질환을 앓고 있다는 것을 유추할 수 있음
- 이는 l-다양성이 민감한 정보의 의미유사성(semantical closeness)을 고려하지 않기 때문에 발생

  * 즉, “위궤양”, “위염”, “위암”이 모두 위장과 관련된 질병이라는 의미를 고려하지 않고 서로 다른 속성 값으로만 판단하기 때문에 개인의 민감한 정보가 노출될 수 있음

○ 쏠림 공격과 유사성 공격에 취약한 l-다양성의 한계를 극복하고자 t-근접성 프라이버시 보호 모델이 제안됨

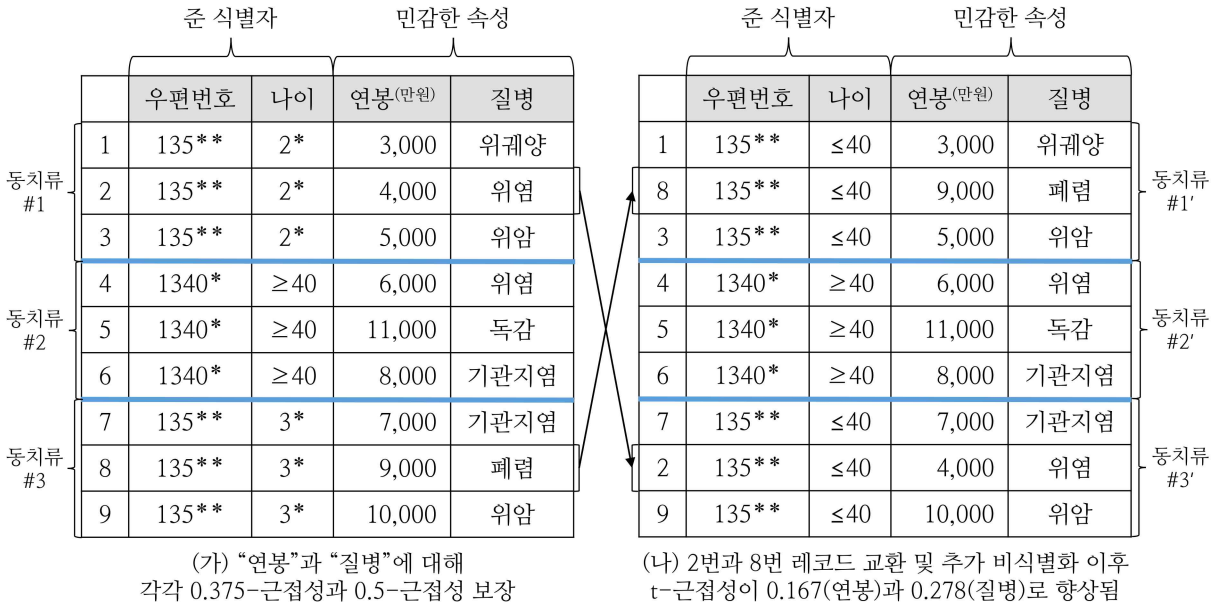
□ t-근접성(t-closeness)

○ 전체 데이터에서 민감한 정보의 분포와 각 동치류에서 민감한 정보 분포의 차이가 t 이하임을 보장. 단, $0 \leq t \leq 1$

- t가 0에 가까울수록 전체 데이터에서의 민감한 정보와 동치류에서 민감한 정보의 분포의 차가 작아짐. 즉, 서로 비슷한 분포를 의미
- [그림 4-9] (가)의 데이터에서 “연봉”의 전체 분포는 [3,000만원, 11,000만원]이고 동치류 #1의 분포는 [3,000만원, 5,000만원]로서 큰 차이가 남. 반면 (나)에서 동치류 #1′의 분포는 [3,000만원, 9,000만원]로서 전체 분포와 좀 더 유사

* “연봉”에 대한 t-근접성이 0.375에서 0.167로 향상됨[49)]

[그림 4-9] t-근접성 예시



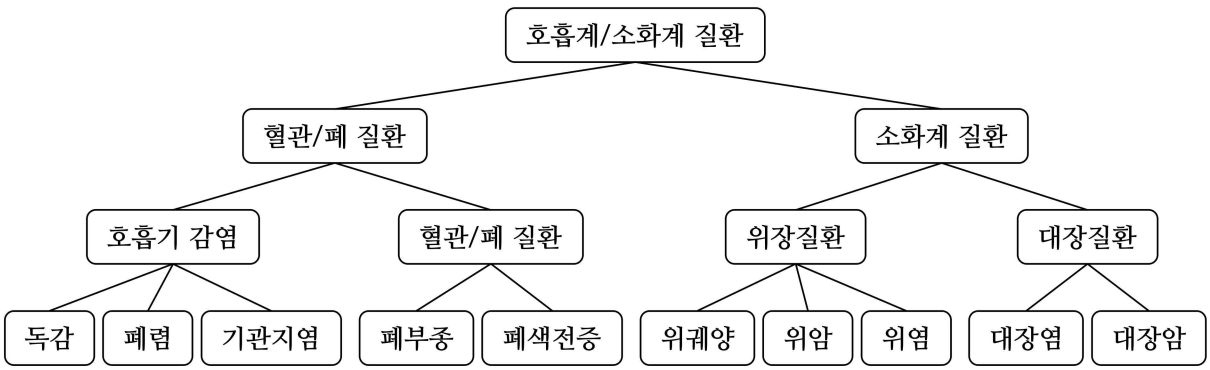
49) N. Le, et al., “t-Closeness: Privacy Beyond k-Anonymity and l-Diversity”, ICDE, vol. 7, 2007. 4.

- [그림 4-9] (가)에서 “질병”의 전체 분포는 호흡기 감염 질환과 위장 질환이지만 동치류 #1은 위장 질환만 포함. 반면 (나)에서 동치류 #1′은 호흡기 감염 질환과 위장 질환을 모두 포함하기 때문에 전체 분포와 좀 더 유사

  * “질병”에 대한 t-근접성이 0.5에서 0.278로 향상

  * 속성 값이 숫자가 아닌 범주 값이기 때문에 [그림 4-10]과 같은 분류 트리(taxonomy tree)를 이용하여 t-근접성 계산

[그림 4-10] 질병에 대한 분류 트리 예



제4절 비식별화 소프트웨어의 현황

□ 학계를 중심으로 개인정보 비식별화 소프트웨어 개발

○ 잘 알려진 프라이버시 모델(k-익명성, l-다양성, t-근접성) 뿐만 아니라 다양한 프라이버시 모델[50]을 기반으로 비식별화 작업을 수행하거나, 재식별화 위험도를 분석하거나, 사용되는 프라이버시 모델에 적당한 파라미터를 추천하는 등 다양한 기능 제공

50) δ-disclosure privacy, δ-presence, (ε, δ)-differential privacy 등

□ 대표적인 비식별화 소프트웨어는 <표 4-4> 참조

<표 4-4> 비식별화 소프트웨어 현황

비식별화 소프트웨어	공개여부
ARX Data Anonymization Tool	공개 소프트웨어
UDT Anonymization Toolbox	
Cornell Anonymization Toolkit	
TIAMAT	
SECRETA	
Open Anonymizer	
ANON tool	
μArgus	
sdcMicro	
PARAT	상업 소프트웨어

○ (ARX Data Anonymization Tool) Java 기반의 비식별화 툴로서 최근 ARX 3.4.1 버전 공개 [그림 4-11],
홈페이지 : http://arx.deidentifier.org/

○ (UDT Anonymization Toolbox) UT Dallas의 Data Security and Privacy Lab에서 개발한 비식별화 툴로 앱과 라이브러리 형태로 공개됨,
홈페이지 : http://cs.utdallas.edu/dspl/cgi-bin/toolbox/

○ (Cornell Anonymization Toolkit) Cornell 대에서 개발하여 공개한 비식별화 소프트웨어로서 다양한 공격자 모델을 대응할 수 있게 한 대화형 디자인이 특징 [그림 4-12],
홈페이지 : https://sourceforge.net/projects/anony-toolkit/

○ (TIAMAT) 데이터를 공개하는 기관이 쉽게 파라미터 값을 선택하여 적절한 k-익명화 등을 수행할 수 있도록 도와주는 공개 소프트웨어 [그림 4-13],
관련 논문 : http://dl.acm.org/citation.cfm?id=1687607

○ (SECRETA) 관계형 데이터베이스 데이터에 대해 비식별화 알고리즘 간 평가

와 비교를 제공하는 공개 소프트웨어 [그림 4-14],
홈페이지 : http://users.uop.gr/~poulis/SECRETA/screenshots.html

○ (Open Anonymizer) k-익명화 개념을 기반으로 데이터 레코드를 일반화하는 비식별화 공개 소프트웨어,
홈페이지 : https://sourceforge.net/projects/openanonymizer/

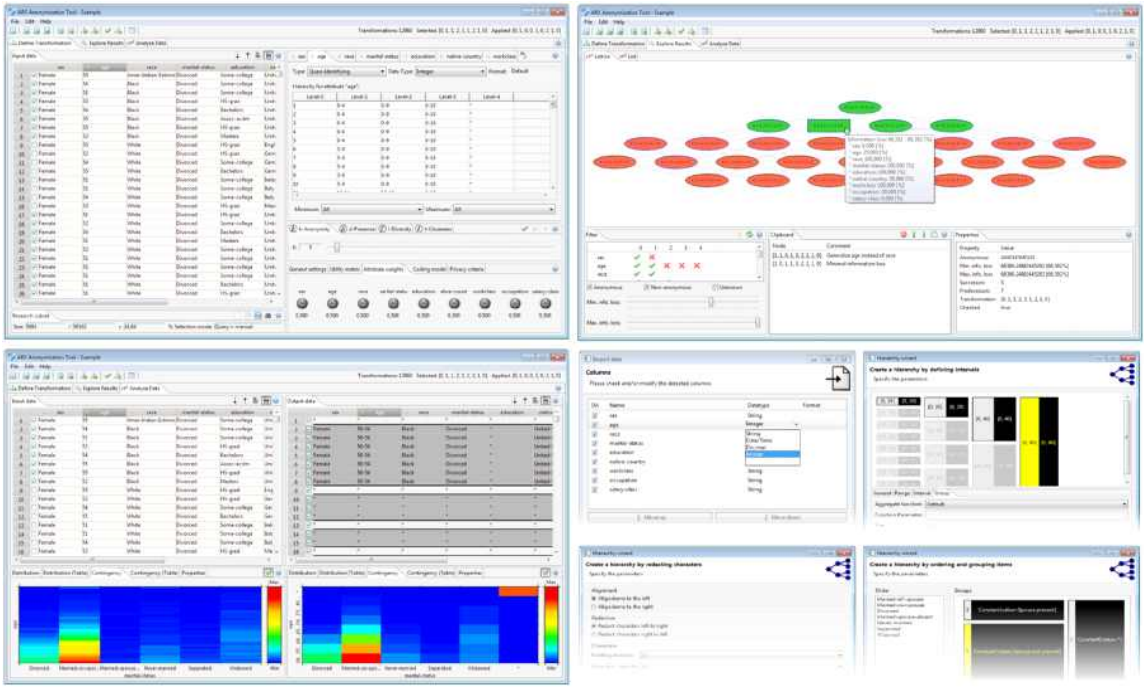
○ (ANON tool) k-익명화 기반 비식별화 공개 소프트웨어,
홈페이지 :
http://www.tmf-ev.de/Themen/Projekte/V08601_AnonTool.aspx

○ (μArgus) 안전한 개인 정보 파일을 생성하기 위해 설계된 비식별화 공개 소프트웨어 [그림 4-15],
홈페이지 : http://neon.vb.cbs.nl/casc/mu.htm

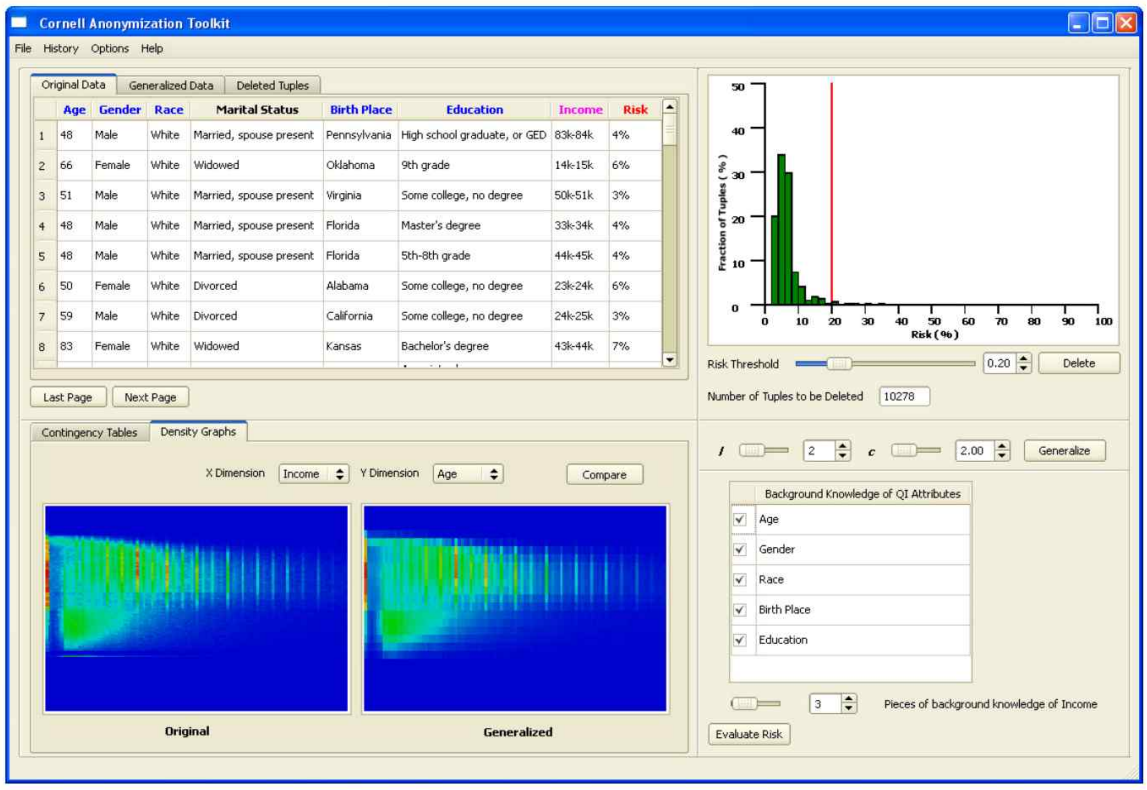
○ (sdcMicro) R 언어 기반의 비식별화 공개 프로그램,
홈페이지 :
https://cran.r-project.org/web/packages/sdcMicro/index.html

○ (PARAT) 비정형화 및 정형화된 데이터 모두에 적용 가능한 상업 소프트웨어 [그림 4-16],
홈페이지 : http://www.privacyanalytics.ca/software/parat/

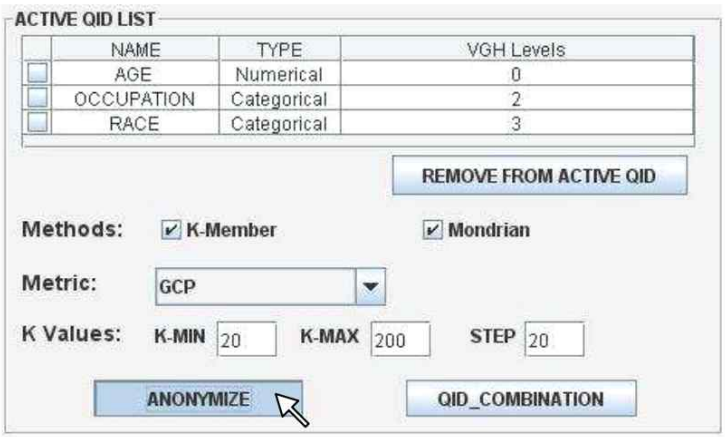
[그림 4-11] ARX Data Anonymization Tool 주요 화면



[그림 4-12] Cornell Anonymization Toolkit 주요 화면

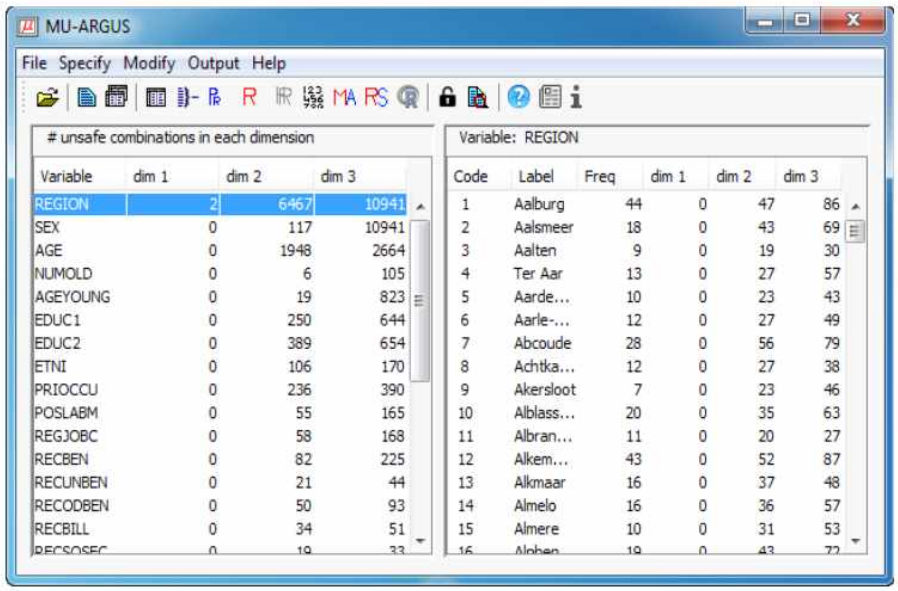


[그림 4-13] TIAMAT 주요 화면

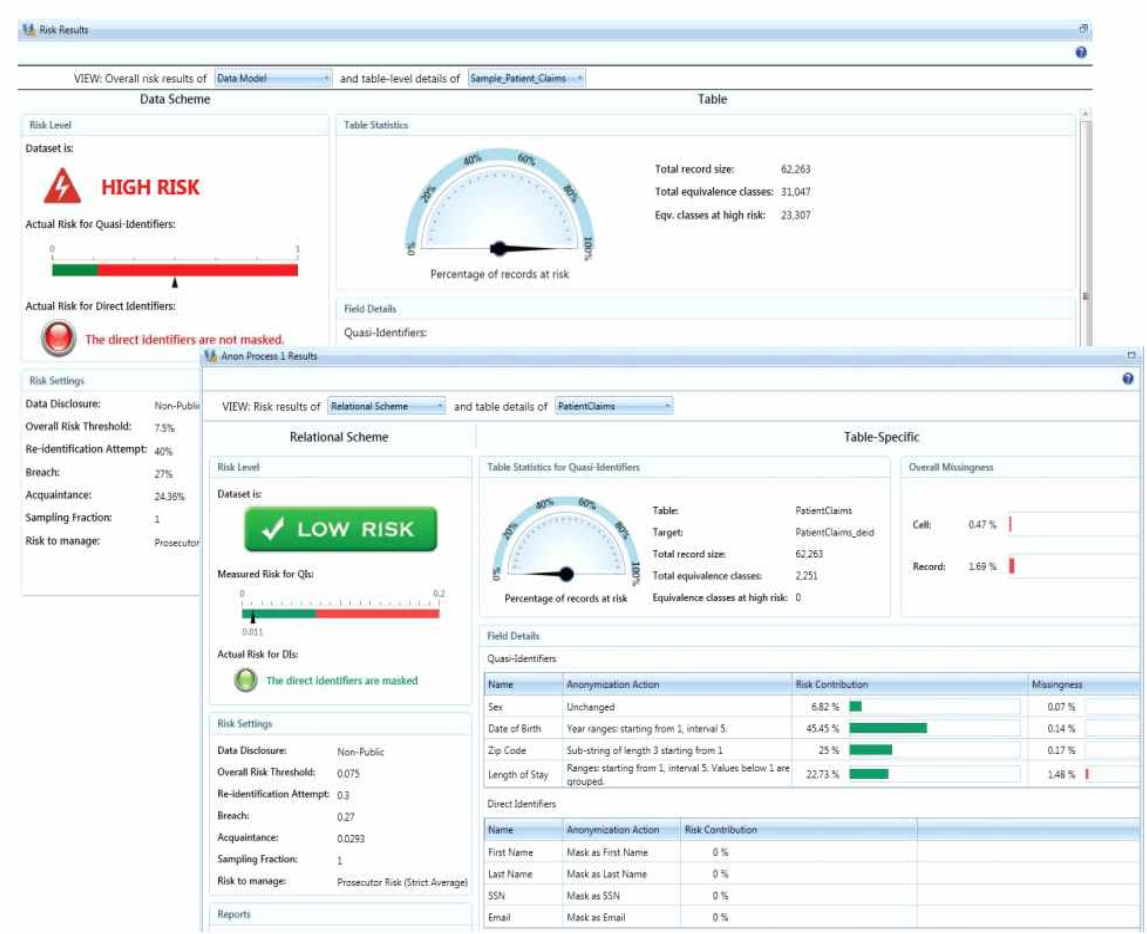


[그림 4-14] SECRETA 주요 화면

[그림 4-15] μ-ARGUS 주요 화면



[그림 4-16] PARAT 주요 화면



제5장 비식별화기술의 쟁점과 동향

제5절 비식별화기술과 개인정보법령 준수여부

□ 개인정보를 비식별화한 비식별정보는 개인정보보호법의 적용대상은 아니지만, 다른 정보와 쉽게 결합하여 식별가능하게 될 경우에는 개인정보로 인정되기 때문에, 비식별화기술의 적용만으로 개인정보보호법 위반 가능성이 없다고 단정할 수 없음

□ 개인정보보호법 관련 사건 중 비식별화기술이 공판과정에서 핵심쟁점이 된 경우[51]는 ① 약학정보원 등의 환자 진료·처방정보 불법 수집·판매 사건과 ② SK텔레콤 등의 전자처방전 사건이 존재함[52]

□ 약학정보원 등의 환자정보 수집판매 사건의 비식별화 관련 쟁점

○ 다국적통계회사 IMS헬스는 2011년 10월부터 2014년 12월경까지 의료프로그램 개발업체 지누스로부터 병원 환자 진료·처방정보(환자들의 성명, 생년월일, 병명, 약물명, 복용량 등) 약 4억 3,019만건을, 재단법인 약학정보원으로부터는 환자 조제정보(환자 주민등록번호, 병명, 약국 조제, 투약내역 등) 약 43억 3,593만건을 제공받음

○ 이 과정에서 주민등록번호가 SHA2-512 일방향 암호화 기법에 의해 13 ~ 15자리의 영문으로 치환되었는데, 한국IMS헬스의 변호인은 A 환자와 B 환자를 구분하는 키(key)값일 뿐 개인을 식별할 수 있는 정보가 아니므로 개인정보가 아니라고 주장함

---

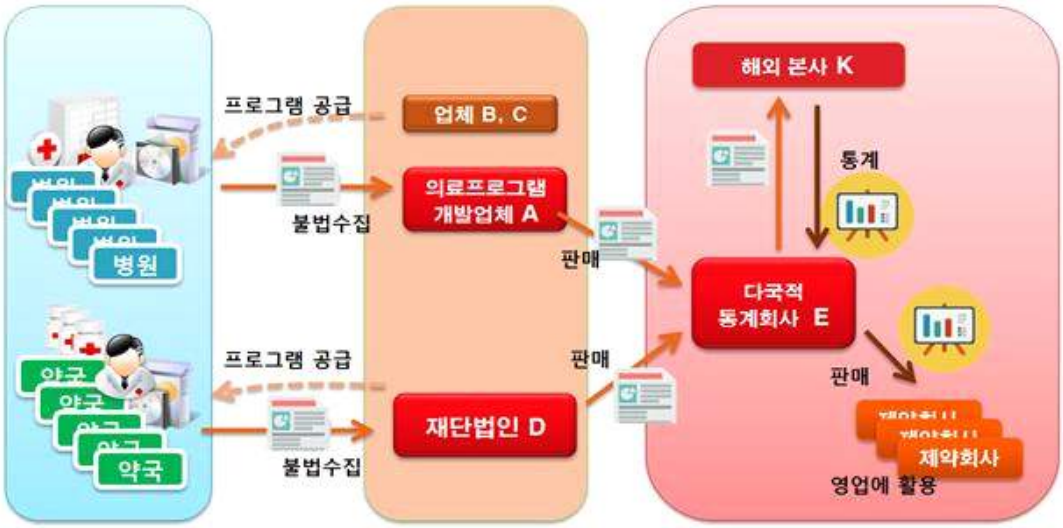
51) 경찰이 도박피의자에게 신고인의 휴대전화번호 마지막 4자리를 알려 준 사건도 도박신고인의 개인정보인 휴대전화번호를 일부만 삭제하여 알려줬기 때문에 데이터 마스킹이라는 비식별화기술과 관련된 사건으로 볼 수 있으나, 이미 많은 연구보고서에서 개인정보의 범위 및 결합용이성과 관련하여 다루어졌으므로 본 보고서에서는 생략함

52) 본 연구에서는 개인정보 비식별화와 관련된 쟁점만을 언급하며, 두 사건의 자세한 내용은 검찰 보도자료와 각종 신문기사를 참고하기 바람
링크 : http://www.spo.go.kr/seoul/notice/notice/notice01.jsp?mode=view&board_no=116&article_no=602083
두 사건에 대해 정리된 기사로는 아래 링크를 참조
http://www.healthfocus.co.kr/news/articleView.html?idxno=59225

○ 검찰은 주민등록번호 등이 일방향 암호화되었더라도 나머지 정보만으로도 다른 정보와 결합하여 식별가능하므로 개인정보로 볼 수 있으며, 약학정보원과 한국IMS헬스 간에 해독값을 공유했으므로 암호화되지 않은 것과 같은 결과라고 주장하며, 암호화는 안전성확보조치의 일환으로 개인정보성과는 무관하다는 입장임

○ 따라서 이 사건의 쟁점은 ① 일방향 암호화된 주민등록번호를 제외한 정보들이 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는 개인정보인 지 여부와 ② 약학정보원이 제공한 해독값으로 주민등록번호를 알아낼 수 있으므로 해독값과 결합하여 개인정보로 볼 수 있는지 여부가 됨

[그림 5-1] 약학정보원 관련 사건의 개요도



출처 : 서울중앙지검 개인정보범죄 정부합동수사단

□ SK텔레콤 전자처방전 사건의 비식별화 쟁점

○ 이동통신회사 SK텔레콤은 2011년 10월부터 2014년 12월경까지 23,060개의 병원에서 약 7,802만 건의 처방전 내역(환자 성명, 생년월일, 병원명, 약품명 등)을 병원 외부의 SK텔레콤 서버로 전송받은 후 가맹점 약국에 건당 50원에 판매하여 환자정보 불법처리 및 전자처방전 정보를 누출한 혐의로 기소

되었음

- 처방전 하단에 SK텔레콤전자처방전이란 문구와 함께 발행번호가 제공되며 약사는 해당 발행번호만 입력하면 손쉽게 전자처방전을 내려 받을 수 있고, 이후 약학정보원의 약국 경영관리시스템(PM2000)에서 해당 전자처방전을 읽어 들여 건강보험청구업무를 수행할 수 있음

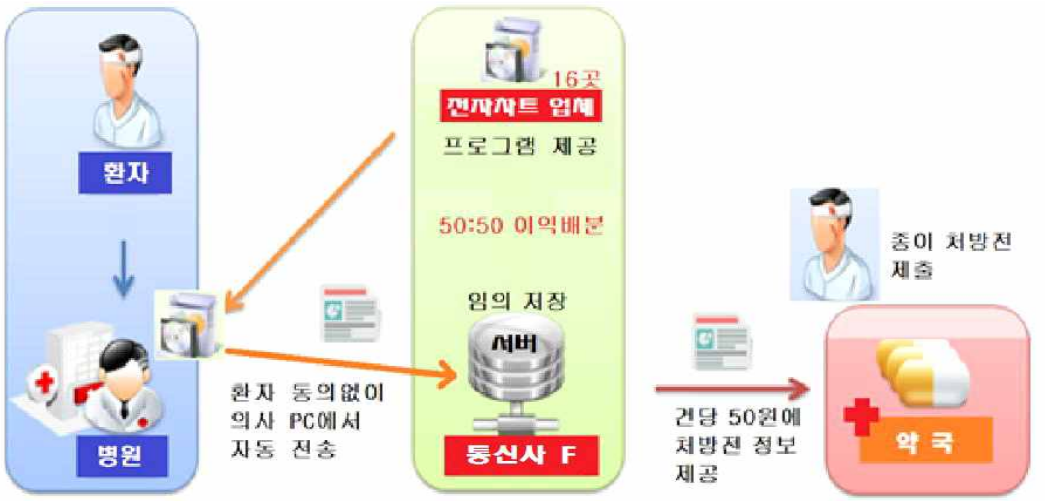
- 처방전 내역이 SK텔레콤 서버로 전송되는 과정에서 병원 측에 설명하거나 환자의 동의를 받은 사실이 없으며[53], 이 전자처방전 사업은 환자, 병원의 업무와 관계없이 오직 약국의 건강보험청구 관련된 자동입력 편의를 위한 것이었음

- 개원내과의사회는 2013년 9월 경 소속 의사회원들에게 전자처방전이 의료법 및 정보통신망 위반에 해당할 수 있으므로 전자처방전 모듈을 삭제하거나 환자로부터 개별동의서를 받아 사용해야 한다는 공문을 보냈고, 이에 천여 곳 이상의 병·의원에서 삭제를 요청함[54]

- 보건복지부도 2014년 8월 20일 경, 의료단체에 보낸 공문에서 의료기기관에서 임의로 전자처방전을 환자가 아닌 자에게 발송했다면 의료법과 개인정보보호법 위반이라는 견해를 밝힌 바 있음[55]

53) 검찰 보도자료에 따른 내용이며, SK텔레콤은 병원 측의 동의를 받았다고 주장함
54) 데일리팜, “의사들, 전자처방전 환자정보 유출 우려…탈퇴 속출”, 2013. 12. 17.자 기사 참조 링크 : http://www.dailypharm.com/News/178849
55) 데일리팜, “전자처방전이 불법?…깜짝 놀란 의원, 당황한 약국”, 2014. 8. 20.자 기사 참조 링크 : http://www.dailypharm.com/News/187636

[그림 5-2] SK텔레콤 전자처방전 사건 개요도



출처 : 서울중앙지검 개인정보범죄 정부합동수사단

○ SK텔레콤 측은 서버에 저장된 처방전은 암호화되고 약사에게 전달된 전자처방전은 종이처방전에 기재된 발행번호 9자리를 입력해야만 복호화되기 때문에[56], 의료법과 개인정보보호법을 위반하지 않았다고 주장함

- 즉, 개인의료정보를 저장한 것이 아니라 암호화되어 비식별화된 정보를 저장 및 전송한 것이므로 개인정보보호법과 의료법을 위반한 것이 아니라는 주장임[57]

○ 따라서 이 사건의 쟁점은 비식별화기술의 일종인 암호화가 적용된 전자처방전이 개인정보인지 여부임

- 구체적으로는 전자처방전 발행번호 9자리와 약국경영관리프로그램(PM2000)과 쉽게 결합하여 개인을 식별할 수 있는 개인정보가 되는지 여부임

□ 위 사건들은 현재 재판이 진행 중이나, 적어도 현재까지는 암호화 등 비식

56) 뉴스타파, “당신의 처방전 정보가 새 나가고 있다”, 2014. 6. 12.자 기사 참조
링크 : http://newstapa.org/12402

57) 덧붙여, SK텔레콤은 병원과의 위수탁 계약 체결과 전자처방전 발행을 담당하는 전자차트 프로그램 설치 시 동의를 받아 의사의 위탁으로 업무를 처리했다고 주장했지만, 검찰은 처방전의 의료정보는 민감정보로 의사가 직접 다뤄야하고 다른 사람에게 위탁할 수 없다고 주장하고 있어, 민감정보의 처리위탁 가능여부 및 절차 준수 여부도 쟁점임

별화기술을 적용하더라도 정보주체의 동의 없이 합법적으로 비식별정보를 제3자에게 제공할 수 있다고 단언할 수 없는 상황임

○ 특히 지금까지 법원의 경향으로는 비식별화기술이 적용된 정보에 대해서도 결합용이성을 지렛대로 삼아 개인정보라고 볼 가능성이 존재함

제6절 정부의 개인정보 관련 정책동향

□ 2016년 5월 18일 열린 제5차 규제개혁장관회의에서는 개인정보 활용 기준을 명확히 하기 위해, 통합 법령 해설서를 2016년 6월까지 발간하고, 하반기에는 개인정보보호법 개정을 추진하기로 하였음

○ 개인정보 범위가 모호하여 신규 사업 진출이 어렵고 비식별화 조치의 적정성 판단이 곤란하다는 인식 하에 통합 해설서를 행자부, 방통위, 금융위가 공동으로 마련하여 '비식별조치 가이드라인'이라는 이름으로 2016년 6월 30일 공개하였음

○ 엄격한 사전동의 규정으로 목적 외 활용 및 제3자 제공이 제한되는 문제점을 해결하기 위해 행자부와 방통위는 개인정보보호법과 정보통신망법의 하반기 개정을 추진 중인데, 사전동의 요건을 완화하거나 포괄적 사전동의를 인정하는 방안을 검토 중일 것으로 예상됨[58)]

□ 금융위에서는 신용정보법 개정을 통해 개인신용정보 보호 강화 및 비식별정보의 활용근거를 마련하기로 하고 7월 중 국회에 개정안을 제출할 예정임[59)]

○ 개인정보와 신용정보를 구분하지 않고, 금융회사가 금융거래 등과 관련하여 처리하는 거래상대방에 관한 정보는 모두 신용정보로 정의하여 신용정보의 범위를 확장하고 개인신용정보보호를 강화함

58) 전자신문, "홍윤식 장관, '빅데이터 활성화 위해 법 개정 추진'", 2016. 5. 25.자 기사 참조 링크 : http://www.etnews.com/20160525000420
59) 금융위, "금융권 개인정보보호체계, 21년만에 확 바뀐다", 2016. 4. 15.자 보도자료 참조

- 그간 대법원에서는 개인정보와 개인신용정보를 엄격히 구분해 '식별정보'는 나머지 신용정보와 결합되는 경우에 한해 '개인신용정보'라고 보아, 학교 졸업앨범의 졸업생의 이름, 주소 전화번호 등[60]이나 인터넷 업체가 보유한 회원들의 성명, 주민등록번호 등[61]이 개인신용정보가 아니라고 판결함

- 이번 법개정으로 금융기관 등이 취급하는 개인식별정보에 대해서도 신용정보법의 적용을 받을 수 있어 신용정보 보호가 강화될 예정임

○ 특히 개인신용정보를 '생존하는 개인에 관한 정보로서 신용정보주체를 식별할 수 있는 정보'로 규정하여 비식별정보는 개인신용정보가 아니라는 점을 명확히 함

- 이에 따라 금융기관 등이 비식별화기술을 활용해 개인정보의 목적 외 활용·제공을 할 수 있는 길이 열림

- 아울러 비식별정보를 제공받은 자의 재식별 금지, 처리 과정에서 개인이 식별되는 경우에는 즉시 삭제의무를 부과하는 등의 개인정보 보호조치를 취함

□ 행자부와 방통위 등 관계부처는 빅데이터 산업 활성화를 위한 TF를 꾸려 개인정보의 범위를 축소하는 개인정보보호법 개정과 비식별화된 개인정보를 전문으로 유통하는 제3기관의 신설을 검토하고 있음[62]

○ '다른 정보와 쉽게 결합하여 알아볼 수 있는 것'이라는 현행법 문구 앞에 '합리적으로 접근가능한 방법을 통해'라는 문구를 삽입해 개인정보의 범위를 축소하는 방식이 검토되고 있음

○ 한편, 비식별화한 개인정보의 전문유통기관 신설에 대해서는 혁신성을 떨어

60) 대법원 2000. 7. 28. 선고 99도6 판결 참조
위 판결은 학원생 모집이라는 상거래를 위해 졸업생 이름, 주소 등 개인정보를 유출해 사용한 피고인에 대해 신용정보법 위반으로 기소하였으나 개인식별정보는 개인신용정보에 해당하지 않는다고 판결한 사건으로 이후 신용정보법 개정으로 개인식별정보도 개인신용정보에 포함되어 처벌이 가능해진 점에 관해서는 아래 링크를 참조
참조링크 : https://www.lawtimes.co.kr/Legal-Info/Legal-Counsel-View?Serial=2630
61) 대법원 2006. 6.15. 선고 2004도1639 판결 참조
62) 매일경제, "개인정보 규제도 푼다...빅데이터 산업 활성화", 2016. 6. 12.자 기사 참조
링크 : http://news.mk.co.kr/newsRead.php?year=2016&no=420288

뜨린다는 반론도 존재함

제7절 일본의 개인정보 관련 정책동향

□ 일본은 빅데이터의 활용과 개인정보보호라는 두가지 목적을 달성하고자 2015년 9월 3일 개인정보보호법[63]을 개정하여 개인식별부호의 정의·범위를 신설하고 개인정보의 정의를 변경하였고, 비식별정보에 대응되는 '익명가공정보'의 정의와 관리체계를 신설하였음[64]

○ 일본도 개인정보의 정의를 명확하게 해야 할 필요성을 인식하고 법률개정에 나선 것으로, 익명가공정보에서 우리나라 일부 기업이 요구하고 있는 일회성 임시식별자[65]와 같은 기술을 '복원불가능한 치환'으로 수용하고 있는 점이 특징임

○ 개정 전에는 개인정보의 취득을 일반적으로 본인의 동의 없이도 할 수 있었으나, 개정법에서는 병력(病歷), 신조(信條) 등을 '배려를 요하는 개인정보'로 별도로 분류하여 사전동의 원칙을 적용해, 민감정보의 취득을 보다 어렵게 하여 개인정보보호를 강화하기도 하였음

○ 민감정보 아닌 개인정보의 제3자 제공의 경우에는 사후배제(opt-out)제도를 채택하고 있었는데, 개정법에서는 개인정보보호위원회의 규칙에 따라 본인에게 통지하거나 용이하게 알 수 있게 조치하면서 위원회에 신고한 경우에는 사전동의 없이도 제3자 제공할 수 있도록 하고 있어 사후배제 제도를 계속 유지하고 있음

---

63) 일본의 개인정보보호법은 민간에 대해서만 적용되고, 내각 등 공공기관들에 대해서는 '행정기관이 보유하는 개인정보보호법', '독립행정법인등이 보유하는 개인정보보호법'이 적용되고 있어, 일반법의 지위를 가지고 있는 우리나라의 개인정보보호법과는 차이가 있음

64) 본 보고서에서는 2015년 9월 9일 최종 개정된 일본 개인정보의 보호에 관한 법률에 관해, 중앙대학교 법학전문대학원 이인호 교수가 개정 전과 비교번역한 것을 활용하였음 (다음 링크 참조)
http://www.pipc.go.kr/cmt/not/ntc/selectBoardArticle.do?nttId=5127&bbsId=BBSMSTR_000000000114

65) 이에 관해서는 다음 문서 및 링크를 참조
"빅데이터 활성화를 위한 법 개정 제언", 김이식, 정보법학회 2015년 6월 정기세미나, 2015. 6. 20.
http://kafil.or.kr/?p=3446&cat=9

<표 5-1> 일본 개정 개인정보보호법의 주요 정의규정

용어	법률상 정의
개인식별부호	● 특정 개인의 신체 일부의 특징을 컴퓨터의 용도에 이용하기 위해 변환된 부호로 당해 특정개인을 식별할 수 있는 것 ● 서비스의 이용이나 상품구입과 관련해 할당되거나, 개인에게 발행된 카드, 서류에 기재되거나 전기적, 자기적으로 기록된 부호로 특정한 자를 식별가능한 것 ● 위의 두 가지 중 정령(政令)[66]으로 정하는 것
개인정보	● 개인식별부호가 포함된 것 ● 성명, 생년월일, 그 밖의 기술(記述) 등에 의해 특정 개인을 식별할 수 있는 것(다른 정보와 용이하게 대조하여 확인하여 특정 개인을 식별할 수 있도록 된 것도 포함)
익명가공정보	● 당해 개인정보에 포함되는 기술(記述) 등의 일부를 삭제하거나 복원할 수 없도록 치환하는 것 ● 개인식별부호를 전부 삭제하거나 복원하지 못하게 치환하는 것 ● 위의 조치들을 취하여 특정 개인을 식별할 수 없도록 개인정보를 가공하여 얻어지는 개인에 관한 정보로서 당해 개인정보를 복원할 수 없도록 한 것

□ 또한 개정법에서는, 개인정보보호위원회를 총리 직속으로 설치하고, 익명가공정보의 작성 및 제3자 제공 관련해 개인정보보호위원회의 규칙을 준수하도록 명확히 규정하고 있음

○ 개인정보취급사업자는 익명가공정보 작성방법에 관한 기준, 가공방법 누설방지를 위한 안전관리조치 기준 등에 대해 개인정보보호위원회의 규칙을 준수해야 하며, 제3자 제공 시에는 포함된 개인에 관한 정보의 항목과 제공방법도 공표해야 하며, 재식별 금지의무도 부담함

○ 익명가공정보취급사업자도 익명가공정보를 제3자에게 제공할 때, 개인 관련 정보 항목과 제공방법을 공표해야 하며, 재식별 금지의무, 안전관리조치 등을 취해야 함

○ 개인정보보호위원회는 사업자들의 법위반 시, 일차적으로는 법을 준수하기 위한 권고와 명령을 발할 수 있고, 이를 지키지 않을 경우에는 6월 이하의 징역 또는 30만 엔 이하의 벌금에 처하도록 규정하고 있음

□ 이처럼 일본은 익명가공의 대략적인 방법과 함께 익명가공정보의 정의와 개인정보가 아님을 명확히 하고, 개인정보보호위원회의 규칙으로 개인정보취급사업자와 익명가공정보취급사업자의 준수기준과 의무를 명확히 함으로

66) 내각이 제정하는 명령으로 우리나라의 대통령령과 유사함

써 개인정보의 보호와 활용을 조화시키고자 함

○ 개정 개인정보보호법의 목적은 “개인정보의 적정하고 효과적인 활용이 새로운 산업의 창출 및 활력있는 경제사회와 풍요로운 국민생활의 실현에 이바지하는 것이라는 점 외에 개인정보의 유용성을 고려하면서 개인의 권리이익을 보호하는 것” 임

제5장 제도개선의 시사점

제1절 개인정보 정의와 범위의 개선

□ 빅데이터산업의 발전과 함께 개인정보도 사회·경제적 가치를 지니게 됨에 따라, 개인정보 자기결정권의 본질을 침해하지 않는 범위에서 개인정보의 활용이 가능하도록 제도를 개선하자는 주장은 일리 있음

□ 개인정보와 개인정보가 아닌 비식별정보를 좀 더 명확히 구분하려면 개인정보의 정의 및 범위를 명확히 해야 하며, 이를 위해서는 개인정보의 법률상 정의의 변경을 우선적으로 검토해야 함

○ 현재의 개인정보의 법적 정의가 유럽 등 다른 국가들과 거의 동일하고 보호가 강화되는 것이 전 세계적인 추세이므로 개정필요성이 없다는 반론이 가능함

○ 하지만, 국내에서는 '쉽게 결합하여 개인이 식별가능한 정보'의 범위에 대한 논란이 형사소송의 형태로 다수 계속되고 있고 행자부, 미래부, 방통위 등의 가이드라인 형태의 유권해석은 사법적 판단의 참고자료라는 한계를 지니므로, 법개정을 통해 그 범위를 명확히 하는 것이 근본적인 해결책임

○ 또한 개인식별부호, 개인정보, 익명가공정보의 기준을 명확히 하고자 법을 개정하였고, 보다 구체적인 내용은 정령에 위임한 일본의 사례를 참고할 필요 있음

□ 한편, '비식별조치 가이드라인'에서 개인정보의 정의 중 '쉽게 결합하여'의 요건으로 ① 입수가능성과 ②결합가능성을 구체적으로 제시한 것은 바람직함

○ 입수가능성이란, 두 종 이상의 정보를 결합해 개인정보를 생성할 때에 합법적으로 입수가능한 정보로 제한한다는 의미로, 해킹 등 불법적으로 취득한 정보들은 입수가능한 정보에 포함되지 않음

○ 결합가능성은 현재의 기술수준이나 비용 측면에서 합리적인 수준이어야 한다는 의미로, 일반적인 사업자가 구매하기 어려운 정도의 고가의 컴퓨터가 필요한 경우 등에는 결합가능성이 인정되지 않는다는 의미임

○ 이러한 요건들은 사법부의 개인정보 여부 판단에 참고자료로 가치가 있다고 보이며, 향후 법개정 시에도 적극적으로 반영될 필요 있음

□ 다만, 법개정을 통하더라도 개인정보 또는 빅데이터 관련 신사업에 대한 법적 위험도의 불확실성을 줄이는 것에 그치며, 사법적 판단의 재량여부를 완전히 없앨 수는 없음

제2절 비식별정보 유통의 관리체계

□ '비식별조치 가이드라인'은 비식별조치를 했더라도 불특정 다수에게 공개될 경우 재식별가능성이 있다고 보아 비식별정보의 공개는 원칙적으로 금지하고 있음

□ '비식별조치 가이드라인'은 ① 외부 평가단을 통한 비식별 조치의 적정성 평가, ② 전문기관을 통한 기업 간 결합DB 작업 지원의 관리체계를 제시하되, ③ 각 산업별 비식별조치의 구체적 내용(k-익명성 수치 등)은 전문기관에 위임하고 있음

○ 한국인터넷진흥원에 개인정보 비식별 지원센터를 설치하고, 각 부처별로 전문기관을 지정·운영하며 전문기관 실무협의회를 운영할 예정임

□ 하지만 실제 개인정보의 비식별조치 후 제3자 제공 현황 또는 기업 간 결합DB 작업 현황에 대한 관리체계는 미흡한 상황임

□ 외부 평가단의 구성은 개인정보 보유기관의 자율에 맡겨져 있고, 개인정보 비식별조치 DB가 유통되면서 기관 내부에서 이뤄지는 융합분석에 대해서는

재식별 금지의무 위반 시 제재규정만 존재함

□ 따라서 개인정보보호법 제18조 제2항 제4호에 근거한 비식별정보의 유통현황을 조사・감독・관리할 체계가 필요함

○ 일본에서 개인정보보호위원회를 새로이 강화하고, 최근 통과된 EU의 GDPR에서 독립적인 개인정보보호 감독기관에게 실질적인 조사권과 교정권을 부여하는 점을 참고할 필요 있음

○ 개인정보보호법 개정을 통해 개인정보보호위원회에 법적 권한을 부여하고, 한국인터넷진흥원을 통해 관리・감독하는 체계가 합리적임

제3절 법체계와 컨트롤타워의 재정립

□ 개인정보와 관련하여, 일반법인 개인정보보호법 외에, 정보통신사업자에게는 정보통신망법이, 위치정보관련 사업자에게는 위치정보법이, 금융기관과 신용정보회사에 적용되는 신용정보법이 각각 존재함

○ 각 법들의 내용은 상당부분 중복되지만, 사용하는 단어(유출, 누출, 누설 등)들이 조금씩 다르고 소관부처가 모두 달라서 개인정보보호 체계의 통일성을 기하기가 어려움

○ 또한 대통령 직속의 개인정보보호위원회도 개인정보보호법 제8조 제1항 제2호, 제4호에 따라 개인정보에 관한 법령의 해석・운용, 정책・제도・법령의 개선에 관한 심의・의결기관이나, 각 법들의 해석의 통일성 및 개선을 추진하기에는 어려움

<표 6-1> 개인정보 관련 법 비교

주요항목	개인정보보호법	정보통신망법	위치정보법	신용정보법
소관부처	행정자치부	방송통신위원회	방송통신위원회	금융위원회
정의	2조 제1호	2조 제6호	2조 제1~2호 (위치정보 관련)	2조 제1호 (신용정보 관련)
주체의 권리	4, 35~38조	30조	24조	35~39조
수집·이용·제공	15~16조, 17, 19조	22~24조,24조의2,30조의2,49조의2,50조	15조,18~20조,29조	15~16조,23~24조,32~33조
목적 외 이용·제공	18조	24조의2	21조	22조의3
보존, 파기	21조	29조	8조,11조,23~24조	20조의2,21조
동의방법	22조	26조의2	26조 (8세이하아동 등)	32조
민감정보	23조			16조
고유정보	24조			34조
주민번호	24조의2	23조의2		34조
업무위탁	26조	25조,50조의3		17조
정보이전	27조	26조		
안전조치	29조	28조	16조	19조
보호책임자	31조	27조		20조
유출통지	34조	27조의3		39조의2
과징금	34조의2	64조의3	14조	42조의2
손해배상금	39조, 39조의2	32조,32조의2	27조	43~43조의2
형사처벌	70~74조, 74조의2	71~73조,75조의2	39~41조	50조
과태료	75조	76조	43조	52조

□ 또한 영리적 목적으로 개인정보의 불법적 수집 및 제공을 영위하는 사이버 흥신소가 성행하는데, 이들의 경우 개인정보 관련 법들을 중복 위반하는 경우가 많아 사실상 별도의 법률을 유지하는 의미가 없음

○ 서울지방경찰청 사이버안전과는 2016년 7월 4일, SK텔레콤의 위치정보서버에 불법적으로 접근해 휴대폰 위치정보를 빼돌리는 등의 사이버흥신소 관련 사건 수사결과를 발표하였는데, 사이버흥신소 관계자들은 위 법률들을 모두 위반한 것으로 보임[67)]

- 적용법조의 구체적 내용은 정보통신망법 제72조 제1항(정보통신망 불법침입), 위치정보법 제40조(불법위치정보 수집), 신용정보법 제50조 제3항(특정인 소재 불법탐지), 개인정보보호법 제71조(개인정보 불법제공)으로 추정됨[68)]

□ 일본의 개정 개인정보보호법에서는 비식별화기준에 대해 사업자들로 하여금 개인정보보호위원회의 규칙을 준수할 것을 명시하고 있고, 위원회의 권고 또는 명령을 위반하지 않는 한 형사처벌도 받지 않는 것을 참조할 필요 있음

○ 이는 강한 처벌보다는 개인정보에 토대한 비식별정보의 제3자 제공 및 활용에서의 투명성 및 법규의 실효성을 더욱 중시하는 선택으로 이해됨

○ 개인정보보호위원회의 규칙을 준수할 경우 민사 상 손해배상의무까지 면책되는 지 여부는 불분명하나 국내와는 달리 형사적으로 강력한 제재를 하지 않는 방향을 선택한 것은 분명함

○ 비식별화 관련 평가기준이 소관부처의 가이드라인인 국내와는 달리, 개인정보보호위원회의 규칙이기 때문에 사법부에 대한 실효성을 확보한 점도 참고해야 함

---

67) "위치정보 해킹....배우자 뒷조사 등 사이버 흥신소 일당 검거", 서울지방경찰청, 2016. 7. 4.자 보도자료 참조
http://www.netan.go.kr/board/boardView.do?board_id=incident&id=5664&page=1&mid=040502

68) 여러 죄를 동시에 저지른 경합범의 경우, 가장 중한 죄에 정한 형의 장기의 2분의 1까지 가중처벌되므로, 개인정보보호법 제71조 위반의 5년 이하의 징역을 기준으로 최대 7.5년의 징역형이 선고될 수 있음

□ 따라서 국내에서도 일반법과 개별법의 법체계를 정립하여 법령해석의 통일성과 법집행의 명확성을 제고할 필요있음

○ 중복되는 내용들은 개인정보보호법으로 단일화하고 각 개별법에는 대상 산업의 특칙을 유지하는 형태로 법체계를 재정립하고, 개인정보보호위원회를 개인정보 관련 컨트롤타워로 격상해 독립성과 권한을 강화해야 함

○ 이로써 사업자들이 복잡하고 중복되는 규제로 인해 겪는 혼란과 비용을 절감하면서 국민들의 개인정보 자기결정권을 보다 실질적으로 보호할 수 있을 것으로 사료됨

참 고 문 헌

국내 문헌

채승병·안신현·전상인 (2012), 『빅데이터:산업 지각변동의 진원』, 삼성경제연구소.

범지인·송두한·최성종 (2013), 『빅데이터 활용 사례와 시사점』, 농협경제연구소.

한국방송통신전파진흥원 (2013), 『빅데이터(Big Data) 활용단계에 따른 요소기술별 추진동향과 시사점』, 방송통신기술 이슈&전망 2013년 제10호.

한국정보화진흥원 (2013), 『빅데이터 시대의 개인 데이터 보호와 활용』.

정보통신정책연구원 (2013), 『온라인 프라이버시에 대한 철학적 배경과 산업적 접근』.

유지연 (2013), 『미국 데이터 브로커(data broker) 현황』, 정보통신정책연구원.

방송통신위원회 (2014), 『빅데이터 개인정보보호 가이드라인』.

한국정보화진흥원 (2014), 『빅데이터 활용을 위한 개인정보 비식별화 사례집』.

한국정보화진흥원 (2014), 『개인정보 비식별화에 대한 적정성 자율평가 안내서』.

한국정보화진흥원 (2015), 『빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서 Ver 1.0』.

김수연 (2015), 『빅데이터 산업 활성화를 위한 개인정보 보호규제 개선 검토』, 한국경제연구원.

한국정보화진흥원 (2015), 『2015년 빅데이터 글로벌 사례집』.

한국정보화진흥원 (2015), 『개인정보보호 법제로 인한 빅데이터 활용 한계사례 조사·분석』.

정용찬 (2015), 『빅데이터 산업과 데이터 브로커』, 정보통신정책연구원.

소프트웨어정책연구소 (2016), 『SW산업 주요 통계』.

국무조정실 등 (2016), 『개인정보 비식별 조치 가이드라인』.

해외 문헌

L.Sweeney (2002), "Achieving k-anonymity privacy protection using generalization and suppression", Int. J. Uncertain. Fuzz., 10(6):571-588.

A.Machanavajjhala (2006), "l-diversity: Privacy beyond k-anonymity", ICDE, p24.

N.Le, et al. (2007), "t -Closeness : Privacy Beyond k-Anonymity and l-Diversity", ICDE, Vol.7.

C.Dai, et al. (2009), "TIAMAT: a Tool for Interactive Analysis of Microdata Anonymization Techniques", VLDB.

ICO (Information Commissioner's Office, 영국) (2012), "Anonymisation: managing data protection risk code of practice".

EU (2014), "Opinion 05/2014 on Anonymization Techniques", EU Article 29 Data Protection Working Party.

S.Garfinkel (2015), "De-Identification of Personal Information", NISTIR 8053.

주 의

1. 이 보고서는 소프트웨어정책연구소에서 수행한 연구보고서입니다.
2. 이 보고서의 내용을 발표할 때에는 반드시 소프트웨어정책연구소에서 수행한 연구결과임을 밝혀야 합니다.