발간등록번호
11-1079930-000021-01

개인정보의 비식별화 처리가 개인정보 보호에 미치는 영향에 관한 연구

[최종보고서]

2015. 12. 10

제출자: 책임연구원 고 학 수 (서울대학교 법학전문대학원 교수)
공동연구원 최 경 진 (가천대학교 법과대학 교수)

아래의 연구결과물을 ‘개인정보의 비식별화 처리가 개인정보 보호에 미치는 영향에 관한 연구’에 대한 최종보고서로 제출합니다.

2015. 12. 10.

제출자: 책임연구원 고 학 수 (서울대학교 법학전문대학원 교수)
공동연구원 최 경 진 (가천대학교 법과대학 교수)

개인정보보호위원회 귀중

<목 차>

제1장 서론

제1절 연구의 필요성

오늘날 빅데이터 등 개인정보가 비식별 처리되어 활용되는 사례가 증가하고 있다. 개인정보 처리에는 각종 제약이 붙어 처리가 어려운 반면 비식별 된 개인정보는 제약이 적어 처리가 용이하기 때문이다. 이에 대처하기 위하여 각종 가이드라인이 등장하였고 개인정보를 비식별 하는 방안을 제시하고 있다. 하지만 이러한 가이드라인의 실효성에 대한 연구가 부족하다. 설령 가이드라인을 준수하더라도 개인정보보호 관련법령에 부합하게 되는 것인지 확신을 주지 못하고 있다.

그 결과 실제 공공·민간 영역에서는 가이드라인을 준수하여 개인정보를 비식별처리하고 이를 이용하려는 노력에 소극적인 모습을 보이게 되었다. 그러므로 비식별화가 개인정보 보호에 어떠한 기여를 하는지에 대한 연구가 필요하다. 비식별화 처리가 각종 가이드라인에 따라 잘 이루어지고 있다고 하더라도 이것이 개인정보 보호에 어떠한 영향을 미치는지를 잘 알 수 없는 상황이기 때문이다. 이러한 실효성을 국내외 법·제도 및 사례조사를 통해 분석해볼 필요가 있다.

제2절 연구의 목표 및 범위

우선 개인정보 비식별화 처리가 개인정보 보호에 미치는 영향에 관해 연구한다. 국내외 법·제도, 적용사례 등을 통해 조사·분석한다. 다음으로 국외 법·제도, 적용사례를 면밀히 검토하여 개인정보 비식별화 및 이와 관련된 개인정보 보호 제도를 정리하고, 해외 사례를 유형화 한다. 마지막으로 해외 사례의 유형화를 통해 국내에 적용할 수 있는 방안을 모색한다. 이 때 주민등록번호라는 단일 식별자가 있는 등 외국과는 상이한 국내환경을 감안하여, 실효성이 있는 방안을 모색한다.

본 연구는 크게 ① 개념정의 ② 현황조사 및 분석 ③ 개선방안 제시로 구성된다. 우선 개념정의 부분에서는 개인정보(personal information)가 무엇인지, 그리고 비식별화(de-identification)와 익명화(anonymization)가 구별되는 개념인지 등에 관하여 살핀다. 이를 바탕으로 비식별화 처리의 현황을 조사하고 분석한다. 국내외 공통된 기술적인 논의를 우선 한데모아 정리한 후, 주요 국가별 논의현황을 각각 조사하여 살펴본다.

개선방안을 제시하는 부분에서는 합리적인 비식별화 규제 체제를 모색한다. 앞서 조사한 비식별화의 현황(기술적인 방법론과 국가별 대응방안)에 비추어볼 때 국내 규제체제의 문제점은 무엇이며 외국 규제체제들의 특징이 각각 무엇인지 비교하여 분석한다. 이를 토대로 외국 규제체제가 국내에 바로 적용될 수 있는지 그 가능성을 검토한 후, 국내 상황에 적합한 비식별화 규제가 어떤 것일지 모색한다. 국내 상황에 부합되는 합리적인 비식별화 규제체제를 살필 때에는 기존 연구결과를 바탕으로 기본적인 접근법을 정리한 후 절차적 접근법의 고려사항을 감안하여 국내 상황에 적용하는 절차를 따른다.

제2장 개념정의

제1절 개인정보

개인정보란 개인에 관한 정보로서, 개인을 식별할 수 있는 정보를 의미한다. 개인정보가 어느 범위까지 인정되는지에 대해서는 국가별로 약간의 차이를 보이고 있다. 더욱이 오늘날 급변하는 대내외적인 환경으로 인하여 개인정보의 정의방식과 인정범위는 조금씩 변화하고 있다. 일단 현행 국내법상 개인정보의 정의와 인정범위[1]를 살펴본다. 이후 각 국가별로 어떠한 차이를 보이는지에 대하여 법규정을 중심으로 검토[2]한다.

개인정보란 살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다)를 말한다. (개인정보 보호법 제2조 제1호)[3] 이를 분설하면, ① 살아 있는 개인에 관한 정보[4]로서 ② 특정 개인과 관련이 있는 정보[5]이며 ③ 식별정보[6]이거나 식별가능정보[7]인 경우를 의미한다.

현행 법령을 기초로 개인정보의 범위를 밝힌 대표적인 판례로서 '휴대전화번호 사건'과 'IMEI 사건'이 있다. 휴대전화번호 사건은 휴대전화번호 뒷자리 4자리만으로도 그 사용자가 누구인지를 식별할 수 있는 경우가 있고,

---

1) 인하대학교 산학협력단, 개인정보 범위에 관한 연구, 개인정보보호위원회(2014.10) 등
2) 한국정보화진흥원, 각국의 개인정보 보호법상 개인정보의 정의 및 해석(2014.10) 등
3) 이는 살아 있는 개인을 전제로 하고 있으면서 그 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통해 개인을 알아볼 수 있는 정보이다. 이 때 개인을 알아볼 수 있는 정보이기만 하면 단순히 해당 정보가 개인을 직접적으로 알 수 있는 정보일 경우뿐만 아니라 다른 정보와의 결합을 통해 개인을 알아볼 수 있는 경우 그 정보 또한 개인정보에 해당한다. (안전행정부, 「개인정보 보호법령 및 지침·고시 해설」, 2011.12,)
4) 현행 개인정보보호법상 보호대상이 되는 개인정보는 살아 있는 개인에 관한 정보만을 대상으로 한다. 따라서 권리의무의 주체가 아닌 사자(死者)의 개인정보, 권리의무의 주체라 하더라도 법인정보 및 사물의 위치정보 살아 있는 개인에 관한 정보가 아닌 경우에는 이 법의 보호대상에 해당하지 않는다.
5) 어떠한 정보가 특정한 개인에 대한 사실, 판단, 평가 등 그 개인과 관련이 있는 정보일 때 개인정보가 된다. 그러므로 특정개인과 관련이 없이 나열되거나 기록된 정보는 개인정보에 해당하지 않는다.
6) 식별정보란 성명, 주민등록번호, 여권번호 등 특정한 개인을 식별할 수 있는 정보를 말한다.
7) 어떤 정보가 특정한 개인을 식별할 수 없는 정보라고 하더라도 다른 정보와 '쉽게' 결합하여 개인을 식별할 수 있는 정보를 식별가능정보라고 한다.

특히 그 전화번호 사용자와 일정한 인적 관계를 맺어온 사람이라면 더욱 그러할 가능성이 높다는 취지의 판시를 하였다.[8] IMEI 사건은 IMEI나 USIM 일련번호는 휴대폰 가입신청서 등 가입자정보에 나타난 다른 정보와 어려움 없이 쉽게 결합됨으로서 개인을 특정할 수 있게 되는 이상 이들을 개인정보라 봄이 상당하다는 판시를 하였다.[9] 이 두 판례는 하급심 판례라는 한계를 가지고 있고, 이에 관한 비판적 시각 또한 존재한다. 이 두 판례는 현행법상 개인정보 보호법의 적용대상인 개인정보의 범위가 매우 넓게 해석될 수도 있다는 취지로 흔히 인용된다.

국내법상의 개인정보 정의는 다른 나라의 개인정보 정의규정과 비교할 때 근본적인 차이가 있지는 않다. 독일 연방개인정보보호법상 개인정보 정의규정[10]이나 호주 연방프라이버시법상 개인정보 정의규정[11], 캐나다의 프라이버시 법상의 규정[12] 등은 국내법상의 규정과 대체로 유사하며, 그 해석에 있어서도 개인과 관련된 정보를 광범위하게 개인정보로 포섭하고 있는 태도를 보인다. 이는 1980년 정립된 OECD 8원칙에 기초하여 1995년 등장한 EU Directive[13]의 개인정보 정의규정(개인정보란 식별되거나 식별가능한 자연인(정보주체)과 관련된 모든 정보)이 전 세계 개인정보 보호법제에 널리 수용되어 왔기 때문인 것으로 분석된다.

그런데 ICT환경이 변화함에 따라 정보처리기술이 급격하게 발전하면서 개인정보로 인정되는 범위를 좁히거나 세분화하여 규제강도를 달리하려는 움직임이 나타나고 있다. 영국은 개인정보 보호법상 정의규정[14]을 ICO (Information Commissioner's Office) 지침(2012.12.12)[15]을 통하여 세

8) 대전지방법원 논산지원 2013. 8. 9. 선고 2013고단17 판결
9) 서울중앙지방법원 2011. 2. 23. 선고 2010고단5343 판결
10) 식별된 또는 식별가능한 자연인(정보주체)의 인적·물적 환경에 관한 모든 정보
11) 정보 또는 의견의 진실여부 및 물체에 기록되었는지 여부에 상관없이, 식별된 개인 또는 합리적으로 식별가능한 자에 관한 정보 또는 의견
12) 모든 형태로 기록된 식별가능한 개인에 관한 정보로서, 인종, 국적, 민족, 피부색, 종교, 연령, 개인의 결혼상태, 교육, 의료, 범죄경력 또는 근무경력, 개인식별번호, 상징(기호) 또는 기타 개인에게 부여된 것, 주소, 지문, 개인의 혈액형, 개인적 의견 등
13) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data
14) 해당 정보로부터 또는 해당 정보와 개인정보처리자가 보유하고 있는 다른 정보 또는 처리 과정에서 생성된 정보와 결합하여 개인을 식별하는 정보

분화하였다.[16] 일본은 개인정보 보호법을 전면개정(2015.9.3.)하여 기존 개인정보 보호법(2003)상 모호하고 단일했던 정보유형[17]을 명확히 규정[18]했으며 세분화[19]하며 규제강도를 달리[20]하였다. 이러한 경향은 소위 빅데이터 시대에 개인정보 보호와 이용의 균형점을 찾으려는 시도와 함께 점차 활발해지고 있다.

제2절 비식별화(de-identification)와 익명화(anonymization)

1. 문제점

비식별화(de-identification)와 익명화(anonymization)의 개념 구분을 둘러싸고 논란이 발생하기도 한다. 비식별화와 익명화의 개념을 명확히 구별할 수 있다는 입장에서는 ① 비식별화라는 용어사용이 현재 개인정보의 회색지대를 부정하는 개념으로 사용되고 있어 잘못된 개념[21]이며 ② 미국의 몇몇 특정 법률에 국한되어 사용되는 용어인 비식별화를 법제가 다른 우리나라에서 사용하는 것은 부적절[22]하고 ③ 가이드라인에 비식별화라는 용어를 이용

---

15) ICO, "Determining what is personal data" (2012.12.12.)

16) ① 전자형태의 정보 ② 파일링시스템의 일부를 구성하는 정보 ③ 접근 가능한 기록의 일부를 구성하는 정보(상기 두 가지 유형의 정보를 제외) ④ 공공기관이 보유하고 있는 기록된 정보를 개인정보로 추가하였고, 개인정보인지 여부를 판단할 수 있는 8가지 기준(식별성, 관련성, 명백히 특정인에 관한 정보, 개인과 관련된 정보, 처리의 목적, 이력의 중요성, 개인에 주안점을 둔 정보, 개인에 영향을 미치는 처리)을 마련

17) 이 법률에 있어서 「개인정보」란, 생존하는 개인에 관한 정보로서, 그 정보에 포함된 성명, 생년월일 이외에 기록 등에 의해 특정 개인을 식별할 수 있는 것(다른 정보와 손쉽게 조합하여서 그것에 따른 특정 개인을 식별할 수 있는 것을 포함한다.)을 말한다.

18) (1) 이 법률에서 "개인정보"란 생존하는 개인에 관한 정보이며, 다음 중 하나에 해당하는 것이다. 1) 해당 정보에 포함된 이름, 생년월일 기타 기술 등으로 특정 개인을 식별할 수 있는 것(다른 정보와 쉽게 조회 비교할 수 있으며 그것으로 특정 개인을 식별할 수 있는 것을 포함한다.) 2) 개인식별부호가 포함될 것 (2) 이 법률에서 "개인식별부호"는 다음 중 하나에 해당하는 문자 번호 기호 기타 부호 중, 정령으로 정하도록 한다. 1) 특정 개인의 신체 일부의 특징을 전자계산기용으로 제공하기 위하여 변환한 부호이며, 해당 특정 개인을 식별할 수 있는 것 2) 개인에게 제공되는 역무의 이용 혹은 개인에게 판매되는 상품의 구입에 대해 할당되거나 개인에 발행되는 카드 기타 서류에 기재되거나 전자적 방식에 의해 기록된 부호로서 그 이용자나 구매자 또는 발행을 받는 사람마다 다르게 배정 받아 또는 기재되거나 기록됨으로써 특정 이용자나 구매자 또는 발행을 받는 사람을 식별할 수 있는 것

19) 기존에 개인정보 보호법상 단일(개인정보)했던 정보유형을 ①개인정보 ②배려가 필요한 개인정보(본인의 인종, 신조, 사회적 신분, 병력, 범죄 경력, 범죄로 인한 해를 입은 사실 기타 본인에 대한 부당 차별, 편견 기타 불이익이 생기지 않도록 그 취급에 특히 배려가 필요한 것으로 정령으로 정하는 기술 등이 포함되는 개인정보) ③익명가공정보(특정 개인을 식별할 수 없도록 개인정보를 가공하여 얻는 개인에 관한 정보이며, 해당 개인정보를 복원할 수 없도록 한 것)로 세분화 하였다.

20) 배려가 필요한 개인정보(이른바 기밀정보) 유형을 신설하여 보호를 강화했고, 익명가공정보 유형을 신설하여 적절한 규율 하에 개인정보의 유용성을 확보했다.

21) 이은우, 국회토론회 자료집(2015.8.19), 13p (나)목

22) 이은우, 국회토론회 자료집(2015.8.19), 9p (2)

하는 현황이 용어사용을 통해 법령상의 정의를 수정[23]하려 하는 것이어서 부당하다고 한다. 그러므로 비식별화와 익명화가 혼용되어 쓰이고 있지만 정확히 말하면 양자는 다른 개념이므로 구분하여 사용해야 할 것이라고 한다. 이러한 시각은, 익명화된 정보는 일반적으로 개인정보가 아닌 것으로 평가되는 반면, 비식별화된 정보는 상황에 따라 그 평가가 달라질 수 있다고 구분하여 보는 입장을 흔히 반영한다.[24] 이하 비식별화와 익명화의 개념구분과 관련된 각종 논의를 소개한다. 뒤에서 다시 언급하겠지만, 이러한 방식의 개념구분은 명확하지 않고, 개념구분에 대한 국내외에서의 일반적인 관례가 형성된 것은 아니다.

2. 국내 동향

국내에서는 한동안 비식별화와 익명화를 명확하게 구별하지 않고 혼용하였다. 대표적으로 '정부 3.0 추진 기본계획'(2013.6.19)은 비식별화와 익명화를 병기하여 사용하였다.(ex 개인정보의 익명화/비식별화 방법(예시) : 데이터 마스킹, 가명처리 등) 이러한 경향은 행정안전부의 '공공정보 공유·개방에 따른 개인정보 보호지침'(2013.8.30)이 등장하면서부터 달라지기 시작하였다. 동 지침은 용어의 정의 중 비식별화[25]만 소개하였으며, 익명화라는 용어를 별도로 사용하지 않았다.[26] 그 후로 '방송통신위원회 가이드라인'(2014.12.23) 등 대부분의 정부 자료에서는 주로 '비식별화'라는 용어를 사용하고 있으며, 비식별화된 경우 이를 개인정보가 아니라는 입장을 보인다. 2015년 이후 발의된 법안들에도 모두 '비식별화'라는 용어가 사용되고 있다. 그러므로 오늘날 국내의 정부 자료에서는 주로 '비식별화'라는 용어가 사용된다고 볼 수 있다.

3. 해외 동향

---

23) 이은우, 국회토론회 자료집(2015.8.19), 11p (나)목

24) 엄열, 국회토론회 자료집(2015.8.19), 56p 2. 1)

25) 동 지침 중 "비식별화"란 개인정보의 일부 또는 전부를 삭제하거나 다른 정보로 대체함으로서 다른 정보와 쉽게 결합하여서도 특정 개인을 식별하기 어렵도록 하는 일련의 조치를 말한다, 공공정보 공유·개방에 따른 개인정보 보호지침(2013.8.30), 3p

26) 동 지침 중 27p에 소개된 영국의 익명화 규약에 소개된 내용을 두고 동 지침이 비식별화와 익명화를 구별하여 사용하였다고 보기에는 어렵다고 생각한다. 이는 붙임2에 소개된 자료일 뿐, 동 지침에서 다루고자 했던 주요 사항이 아닌 부분이었기 때문이다.

미국은 HIPAA 등 주요한 입법을 통해 '비식별화'(de-identification)라는 용어를 주로 사용해 왔다. 대표적으로 미국의 표준화 담당기관인 NIST(National Institute of Standards and Technology) 자료는 '비식별화' 개념을 사용하고 있으며, 이를 '데이터 관리인이 데이터세트의 식별정보를 제거하거나 대체하여 데이터 이용자가 정보주체의 식별을 하기 어렵도록 만드는 처리과정'이라고 정의[27]하고 있다. 미국의 정부자료나 공공기관 자료에 비식별화 용어와 익명화 용어가 모두 등장하는 경우는 흔치 않다. 미국이 익명화가 아닌 비식별화를 용어로 선택한 데에 별도의 이유가 있는 것인지는 명확하지 않다. 용어에 대해 소개한 미국 교육부(Department of Education) 자료[28]를 보면 비식별화와 익명화 용어가 모두 나타나는데, 비식별화(De-identification)는 식별성을 제거하는 절차(the process of removing or obscuring any personally identifiable information)로, 익명화(Anonymization)는 비식별을 하는 절차(the process of data de-identification which produces de-identified data)로 구별하고 있어, 구체적으로 어떤 실무적 차이가 있는 것인지 불명확하다.

영국은 ICO의 익명화 규약(Anonymisation : managing data protection risk code of practice(2012))을 발표하며 '익명화'(anonymization)라는 용어를 본격적으로 사용하기 시작하였다. 동 규약 본문에 언급된 용어는 아니지만 익명화 기법의 예시를 하고 있는 Annex 3은 '비식별화'(de-identification) 과정에서 정보주체의 익명성 보호와 데이터의 유용성 유지 간 상충관계(the trade-off between preserving data utility and preserving anonymity)가 발생한다는 점을 언급하고 있다. 결국 영국에서

27) De-identification is a process by which a data custodian alters or removes identifying information from a data set, making it harder for users of the data to determine the identities of the data subjects. Once de-identified, data can be shared with trusted parties that are bound by data use agreements that only allow specific uses. In this case, de-identification makes it easier for trusted parties to comply with privacy requirements. Alternatively, the de-identified data can be distributed with fewer controls to a broader audience. In this case, de-identification is a tool designed to assist privacy-preserving data publishing (PPDP). (Simson L. Garfinkel, De-Identification of Personally Identifiable Information, National Institute of Standards and Technology, 2015, 1p)

28) http://ptac.ed.gov/sites/default/files/data_deidentification_terms.pdf

는 주로 익명화(anonymization)라는 용어를 사용하고 있지만 비식별화(de-identification)라는 용어가 사용되기도 하는 것으로 보인다. 다만, 유럽연합(EU) 차원에서는 익명화라는 용어가 주로 사용된다. 하지만, 익명화라는 용어의 사용에 있어 비식별화와 명확한 개념 구분이 전제되는 것인지는 불분명하다.

일본은 2015년 9월 개인정보 보호법을 개정하면서 익명가공정보(匿名加工情報)라는 개념을 새로 도입하였다. 2년여 기간 동안 법률 개정안을 논의하는 과정에서 발표되어 온 회의 자료를 보면 이는 개인 특정성 감소 데이터(個人特定性低減データ)에 해당하는 정보를 의미하는 것임을 알 수 있다. 이는 개정법 요강을 보면 '특정 개인을 식별할 수 없도록 개인정보를 가공하여 얻는 개인에 관한 정보이며, 해당 개인정보를 복원할 수 없도록 한 것'으로 정의된다.[29] 이는 비식별화 내지 익명화된 개인정보를 익명가공(anonymization)정보라고 정의한 것으로 볼 수 있다. 결국 일본에서는 주로 익명화(anonymization)라는 용어가 사용되고 있지만, 비식별화(de-identification)와 명확한 개념 구분을 하여 이용되는 것으로 보이지는 않는다.

4. 검토 및 소결

이상의 논의에 비추어볼 때 비식별화와 익명화는 명확하게 구별될 수 있는 용어나 개념이라 하기는 어려운 것으로 보인다. 즉, 비식별화와 익명화의 개념을 구분하는 학술적 논의가 별도로 있었던 것도 아니고 일반적인 관례가 형성된 것도 아니다. '비식별화' 라는 용어를 주로 사용하는 미국에서도 '익명화' 라는 용어가 이용되기도 하며, '익명화' 라는 용어를 주로 사용하는 영국과 일본에서도 개념상 이를 '비식별화' 와 명확하게 구별하고 있지는 않고 있는 것으로 보인다. 여기에 비추어 한국의 현실을 살펴보면, '비식별화' 라는 용어를 주로 사용해온 관행에 문제가 있는 것으로 보기는 어렵다.[30] 지금까지 살펴보았듯, 적어도 아직까지는 비식별화와 익명화는 엄

29) 아래 [별첨 3] 참조
30) 그와 동시에, 익명화라는 용어를 사용했더라도 별다른 문제가 있지는 않았을 것이다.

격하게 구별되는 개념이라 보기는 어렵기 때문이다. 개념 구분의 실익이 있을지에 관해서는 향후에 더 많은 연구와 논의가 필요할 것이다.

제3장 비식별화 처리의 현황 및 분석

제1절 비식별화 방법론

1. 비식별화의 기술적 방법론

비식별화란 데이터 내에 개인을 식별할 수 있는 정보가 있는 경우, 이의 일부 또는 전부를 삭제, 또는 일부를 속성 정보로 대체 처리하는 등 다양한 방법을 통해 개인을 식별하기 어렵게 만드는 조치를 일컫는다. 이하 ICO 지침(2012)[31], EU Article 29 WP Opinion(2014)[32], NIST(2015)[33]을 비롯하여 비식별화의 통계적인 방법론이 담겨있는 해외 주요 자료와 관련된 국내 자료[34]를 함께 정리하여 소개한다.

(1) 비식별화 적용대상

비식별화는 데이터 내의 식별 가능한 특징을 제거하거나 변형시키는 과정을 통해, 개인과 여러 정보를 연결시켜 개인의 정보가 드러나지 않게 하거나 하나의 특징 정보를 여러 개인과 연결시켜 개인의 식별을 방지하는 조치라 할 수 있다. 이러한 비식별화는 개인정보, 즉 ‘그 자체로 개인을 식별할 수 있는 정보 및 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 정보’를 그 대상으로 한다.

우선 “그 자체로 개인을 식별할 수 있는 정보”에는 (1)쉽게 개인을 식별할 수 있는 정보(이름, 전화번호, 주소, 생년월일, 사진 등) (2)고유식별정보

31) Information Commissioner’s office, Anonymisation: managing data protection risk code of practice, 2012
32) ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 05/2014 on Anonymisation Techniques, Adopted on 10 April 2014
33) Simson L. Garfinkel, De-Identification of Personally Identifiable Information, National Institute of Standards and Technology, 2015
34) 미래창조과학부&한국정보화진흥원 & K-ICT 빅데이터센터, "빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서 Ver1.0", 2015.6.10. 등

(주민등록번호, 운전면허번호, 의료보험번호, 여권번호 등) (3)생체정보(지문, 홍채, DNA 정보 등) (4)기관, 단체 등의 이용자 계정(등록번호, 계좌번호, 이메일 주소 등) (5)기타 유일 식별번호(군번, 사업자등록번호, 별명과 같은 특성, 식별코드인 아이디나 아이핀 등) 등을 생각할 수 있다.

다음으로 "다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 정보"에는 (1) 개인특성(성별, 생년, 생일, 연령, 국적, 고향, 거주지, 시군구명, 우편번호, 병역여부, 결혼여부, 종교, 취미, 동호회, 흡연여부, 음주여부, 채식여부, 관심사항 등) (2)신체특성(혈액형, 신장, 몸무게, 허리둘레, 혈압, 눈동자 색깔, 신체검사 결과, 장애유형, 장애등급, 병명, 상병코드, 투약코드, 진료내역 등) (3)신용특성(세금 납부액, 신용등급, 기부금, 건강보험료 납부액, 소득분위, 의료급여자 등) (4)경력특징(학교명, 학과명, 학년, 성적, 학력, 직업, 직종 등) (5)전자적 특성(PC사양, 비밀번호, 비밀번호 질문/답변, 쿠키정보, 접소일시, 방문일시, 서비스 이용기록, 위치정보, 접속로그, IP주소, MAC주소 등) (6)가족특성(가족정보, 법정대리인정보 등) (7)위치특성(GPS데이터, RFID 리더 접속기록 등) 등을 포함하는 다양한 준식별자(quasi-identifier)가 포함될 수 있다.

여러 유형의 잠재적 준식별자(quasi-identifier)는 데이터의 유형이나 이용 맥락 등 여러 가지 요소에 따라 실제로 유효한 준식별자로 기능할 수도 있고, 그 반대로 준식별자로서의 역할을 하지 못할 수도 있다. 이는 식별자(identifier)의 경우에도 마찬가지이다. 따라서 특정한 유형의 정보에 대해 식별자 또는 준식별자로 분류하고 항시적으로 식별자 또는 준식별자로 기능할 것으로 전제하여 분석하는 것은 오류의 가능성이 높다. 그 반대로 개별적이고 구체적인 상황에 따라 식별자 및 준식별자는 변화할 수 있다고 생각하는 것이 좀 더 현실을 정확히 반영하는 것이 된다.

(2) 비식별화 적용시기 및 적용기준

비식별화는 빅데이터의 수집·활용의 모든 단계에서 개인정보가 식별되는

경우 또는 이후 정보의 추가 가공 등을 통하여 개인이 식별되는 경우에 적용한다. 가령 개인정보의 수집 및 저장 시, 개인정보가 포함되어 있을 수 있는 데이터의 활용 시, 다른 기관과의 정보 공유 시, 기관내의 서로 다른 부서간의 정보 공유 시 등의 상황에서 비식별화가 필요할 수 있다.

이 때 1) 그 자체로 개인식별이 가능한 정보는 삭제해야 할 것이다. 다만 수집 시에 개인정보에 대한 자체이용, 제3자 제공 등 활용에 대한 이용자 동의를 받았을 경우에는 비식별화 없이 활용할 수 있다. 2) 다른 정보와 결합에 따른 위험을 최소화한다. 이는 보유한 개인정보의 분석을 위한 동의 등이 곤란한 경우 분석목적의 달성을 위해 비식별화 처리가 필요할 수 있음을 의미한다.[35] 유의할 점은 3) 정보가 재식별될 수 있는 리스크를 고려하여 비식별화 과정에서의 관리 및 사후적인 관리를 철저히 하여야 한다는 점이다.

(3) 비식별화 주요기술

비식별화의 주요 기술로는 개인정보 중 주요 식별요소를 다른 값으로 대체하여 개인식별을 곤란하게 하는 방법인 1)가명처리(①휴리스틱 익명화, ② k-익명성, ③암호화, ④교환방법), 데이터의 총합 값을 보임으로써 개별 데이터의 값을 보이지 않도록 하는 2)총계처리(⑤총계처리, ⑥부분집계, ⑦라운딩, ⑧데이터 재배열), 데이터 공유·개방 목적에 따라 데이터세트에 구성된 값 중에 필요없는 값 또는 개인식별에 중요한 값을 삭제하는 3)데이터 값 삭제(⑨속성값 삭제, ⑩속성값 부분 삭제, ⑪데이터 행 삭제, ⑫식별자 제거를 통한 단순 익명화), 데이터의 값을 범주의 값으로 변환하여 명확한 값을 감추는 4)범주화(⑬범주화, ⑭랜덤 올림 방식, ⑮범위 방법, ⑯제어 올림), 공개된 정보 등과 결합하여 개인을 식별하는데 기여할 확률이 높은 주요 개인식별자가 보이지 않도록 처리하여 개인을 식별하지 못하게 하는 5) 데이터 마스킹(⑰임의 잡음 추가, ⑱공백과 대체) 등의 방법이 있다.

35) 여기서 '위험' 또는 '리스크'라는 표현을 사용하는 이유는, 비식별화 또는 익명화 작업의 결과가 시공을 초월하여 완벽하게 재식별의 가능성을 제거하는 경우는 상상하기 어렵기 때문이다. 가령 현시점에서 재식별이 불가능해 보이는 정보에 대해서도, 기술이나 기법의 발전에 따라 10년후에는 재식별이 될 수도 있고, 5명의 연구팀을 투입해서는 재식별이 가능하지 않지만 500명의 연구팀이 투입되면 재식별이 가능한 상황도 생각할 수 있기 때문이다.

1) 가명처리(Pseudonymization)

가명처리란 개인식별이 가능한 데이터에 대하여 직접적으로 식별할 수 없는 다른 값으로 대체하는 기법이다. 주로 성명, 기타 출신학교나 근무처 같은 고유특징에 적용된다. 가명처리의 장점은 그 자체로는 완전 비식별화가 가능하며 데이터의 변형 및 변질의 수준이 적다는 점이다. 반면 단점으로는 일반화된 대체값으로 가명 처리를 하게 되어 성명을 기준으로 한 분석에 있어 어려움이 존재한다는 점이다.

①휴리스틱 익명화(Heuristic Pseudonymization)

식별자에 해당하는 값들에 몇 가지 정해진 규칙을 적용하거나 사람의 판단에 따라 가공하여 개인정보를 숨기는 방법이다. 예를 들어 성명을 홍길동, 임꺽정 등 몇몇 일반화된 이름으로 대체하여 표기하거나 소속 기관명을 화성, 금성 등으로 일반적 명칭을 쓰지 않는 몇몇 대명사로 대체하도록 사전에 규칙을 정하여 수행한다. 이 방법은 식별자의 분포를 고려하거나 수집된 자료를 사전에 분석하지 않고 모든 데이터를 동일한 방법으로 가공하기 때문에 사용자가 쉽게 이해하고 활용할 수 있다는 장점이 있다. 하지만 휴리스틱 익명화 기법을 적용한 이후에는 데이터의 유용성이 떨어지게 되며 활용할 수 있는 대체 변수에 한계가 있다는 단점이 있다. 또한 이 기법은 다른 값으로 대체하는 일정한 규칙이 노출될 수 있다는 취약점이 있다는 점에서, 개인을 쉽게 식별할 수 없도록 규칙 수립에 있어 세심한 고려가 필요하다.

②k-익명성(k-anonymity)

하나의 데이터세트 안에 동일한 속성값을 가지는 데이터를 k개 이상으로 유지하여 데이터를 공개하는 방법이다. 지정된 속성을 보유한 데이터 숫자를 일정 수준(k개) 이상으로 유지함으로써 프라이버시의 누출을 방지하고자 하는 방법이다. 가령 30개의 데이터가 포함된 데이터세트에서 3-anonymity

를 수행하는 경우, 최소한 3개 이상의 데이터들이 같은 속성값으로 대체되어 전체 자료가 10개의 대표 데이터로 표현되도록 하는 기법이다. 이 때 같은 속성값으로 대체하기 위해 범주화(suppression), 일반화(generalization)같은 기법을 사용하게 된다. 이 k-익명성은 기존의 익명화 처리 기술과는 다르게 프라이버시 보호를 구체적으로 염두에 두고 개발된 개념이어서 뒤에서 더 자세하게 설명할 예정이다.[36]

③암호화(Encryption)

정보를 가공할 때 일정 규칙의 알고리즘을 적용하여 암호화함으로써 개인정보를 대체하는 방법이다. 오리지널 데이터의 재사용을 염두에 둔다면 복호가 가능하도록 암호화/복호화 값(key)이 필요할 것이어서, 키(key)에 대한 방안도 요구되는 기법이다. 물론 양방향 암호화에 대해서는 복호화 키(key)와 관련된 기술적 방법론은 물론 이를 누가, 어떻게, 어떤 절차에 따라 관리하는 지에 관한 절차적 측면이 매우 중요해진다. 활용하는 목적에 따라서는 단방향 암호화(hash 등)를 사용할 수도 있다. 이 경우 이론상 개인정보로의 복호화가 원천적으로 불가능하게 된다. 단방향 암호화는 개인정보의 식별성을 완전히 제거하는 조치이며, 양방향 암호화에 비해 더욱 안전하고 효과적인 비식별화 기술에 해당한다.

④교환방법(Swapping)

데이터베이스의 레코드를 미리 정해놓은 변수/항목들의 집합과 연계하여 교환함으로써 가명처리를 하는 비식별화 기술이다. 기존 레코드 값들 간의 교환이 아니라 사전에 정의된 외부 값으로 대체된다는 점에서 총계처리의 세부기법인 데이터 재배열과 구별된다. 이는 민감한 속성을 지닌 그룹에서 내부교환만 이루어질 경우 전체 그룹이 식별되어 버릴 수 있는 위험성이 있을 때, 사전에 정해놓은 외부 값으로 대체하여 민감정보를 비식별화하여 식별될 위험성을 줄이기 위해 사용된다.

---

36) 프라이버시 보호모델을 설명하는 단락에서 더 자세하게 후술한다.

2) 총계처리(Aggregation)

총계처리란 개인정보에 대하여 통계값(전체 혹은 부분)을 적용하여 특정 개인을 판단할 수 없도록 하는 기법이다. 총계처리의 장점은 다양한 통계분석(전체, 부분)용 데이터세트 작성에 편리하다는 점이다. 반면 단점으로는 집계 처리된 데이터를 기준으로 해서는 정밀한 분석을 하기가 어려우며, 집계 수량이 적을 경우 데이터 결합 과정에서 개인정보의 추출 또는 예측이 가능할수도 있다는 점이다.

⑤총계처리(Aggregation)

수집된 정보에 민감한 개인정보가 있을 경우 데이터 집단 또는 부분으로 집계처리를 하여 민감성을 낮추는 비식별화 기법이다. 가령 특정 나이값이 있는 경우 집단의 평균 나이값(대표값)을 구한 후 각 개인정보의 속성값을 구해진 대표값으로 대체한다. 또는, 소득 데이터의 경우, 해당 집단의 소득에 대해 전체 평균을 구한 후 일정규칙에 따라 오차를 가감한 후 각 개인정보의 소득 속성값을 변환한다. 다만 대부분의 데이터 값이 유사한 특정 속성을 지닌 데이터의 경우에는 단체의 대표 속성이 개인의 정보를 그대로 대변할 수 있어서 이 방법의 적용에 한계가 있을 수도 있다.

⑥부분집계(Micro Aggregation)

분석 목적에 따라 부분 그룹만 비식별처리를 하는 기법을 말한다. 이 경우 다른 속성값에 비하여 오차 범위가 큰 항목이나 속성값에 대해서는 통계값인 평균 등을 활용하여 값을 변환하게 된다. 가령 다양한 연령대의 소득 분포에 있어서 40대의 소득 분포 편차가 다른 연령대에 비하여 매우 크거나 특정 소득 구성원을 포함하고 있을 경우, 40대의 소득에 대한 평균값을 구한 후 40대에 해당하는 각 개인정보의 소득 속성값을 해당 평균값으로 대체함으로써 식별이 가능한 소득을 가진 40대의 일부를 비식별 처리하게 된다.

⑦라운딩(Rounding)

집계하여 처리된 값에 대하여 올림, 내림, 반올림 같은 라운딩 기준을 적용하는 최종 집계 처리기법이다. 라운딩은 통상적인 총계처리를 위해 흔히 쓰이는 방식이며, 세세한 정보 보다는 데이터 전체에 대한 통계정보가 필요한 경우에 주로 이용한다. 예를 들어 23, 41, 57, 26, 33 등 세세한 나이의 속성값이 있는 경우 이를 20, 30, 40, 50 등의 각 대표 연령대로 표기하여 집계 처리하는 방식이다. 범주화 중 랜덤 올림 방법과도 방식이 유사하여, 같은 의미로 사용하기도 한다.

⑧데이터 재배열(Rearrangement)

데이터세트 전체로 볼 때 기존 정보의 값은 유지하면서 개인정보와 연관이 되지 않도록 데이터를 재배열하는 비식별 기법이다. 특정 개인에 관한 몇 몇 정보를 타인의 정보와 뒤섞어 전체 정보의 손상없이 개인의 정보가 해당 개인과 연결되지 않도록 하는 방법이다. 가령 여러 가지 개인정보 중에서 나이, 소득 등의 특정 속성을 개인별로 교환하여 재배치하게 되면, 개인의 실제 나이와 소득과는 차이가 나는 비식별 자료를 얻게 된다. 하지만 전체적인 통계분석 등에 있어서는 자료의 손실없이 분석을 할 수 있게 된다.

3) 데이터값 삭제(Data Reduction)

데이터값 삭제란 개인정보 중 식별자로 쉽게 쓰일 수 있는 데이터나 식별 또는 누출의 가능성에 특히 유의해야 하는 데이터의 값을 삭제하여 처리하는 방법을 일컫는다. 이는 주로 이름과 같이 개인을 쉽게 식별할 수 있는 정보, 지문이나 홍채 같은 생체정보, 주민등록번호나 계좌번호 같은 이용자 식별번호·계정 등에 적용된다. 데이터값 삭제의 장점은 민감한 개인식별 정보에 대하여 완전한 삭제를 통해 해당 데이터에 대한 예측, 추론 등이 어렵게 된다는 점이다. 반면 단점으로는 데이터 삭제로 인해 데이터의 유용성이 저

하되고, 그로 인해서 분석의 다양성, 분석결과의 유효성, 분석정보의 신뢰성 등이 저하될 가능성이 높다는 점이다.

⑨속성값 삭제(Reducing Variables)

원시 데이터에서 민감한 속성값 등 개인식별항목을 단순하게 제거하는 방법이다. 가령 주민등록번호, 나이, 성명이 나열되어 있는 원시데이터의 경우 분석 목적에 따라 주민등록번호를 나이로 대체할 수 있다면 주민등록번호를 삭제하고 이를 나이나 생년 데이터로 교체할 수 있을 것이다. 다만 이 경우 유의할 점은 남아있는 정보 그 자체로도 분석의 유효성을 가져야 하며 동시에 개인을 식별할 수 없어야 할 뿐만 아니라 인터넷 등에 공개되어 있는 다른 정보와 결합하였을 때에도 개인을 식별할 수 없어야 한다는 점이다.

⑩속성값 부분 삭제(Reducing Partial Variables)

민감한 속성값에 대하여 전체가 아닌 일부값을 삭제함으로써 대표성을 가진 값으로 보이도록 하는 비식별화 기법이다. 가령 상세 주소의 경우 부분 삭제의 기법을 적용하면 대표지역으로 표현할 수 있다. 가령 ‘서울특별시 종로구 세종대로 123’ 주소를 ‘서울특별시 종로구’로 대치하는 방식이다. 이는 아래에서 보는 범주화와도 유사한 방식이다.

⑪데이터 행 삭제(Reducing Records)

다른 정보와 비교하여 값이나 속성에 있어 뚜렷하게 구별되는 식별정보 전체를 삭제하는 방식이다. 이는 특정하게 민감한 속성값 하나만이 아니라 해당 정보를 가진 개인에 대한 내용 전체를 제거하는 기법이다. 가령 소득이 다른 사람에 비하여 두드러지게 구별되는 값을 지닌 정보가 있는 경우 해당 개인정보 전체를 삭제한다. 이 방법은 통계분석에 있어서도 빈번하게 사용된다. 전체 평균에 비하여 오차범위를 벗어나는 자료를 제거할 때도 사용할 수 있는 기법이기 때문이다.

⑫식별자의 제거를 통한 단순 익명화(Trivial Anonymization)

위에서 나열된 형태들을 제외한 모든 제거 형태의 익명화 기법들이 이 범주에 해당한다.

4) 범주화(Data Suppression)

범주화란 단일 식별 정보를 해당 그룹의 대표값으로 변환(범주화)하거나 구간값으로 변환(범위화)하여 고유 정보 추적 및 그 식별을 방지하는 조치이다. 이는 주로 생년월일 같이 개인을 쉽게 식별할 수 있는 정보, 주민등록번호 같은 고유식별정보, 계좌번호 같은 기관·단체 등의 이용자 계정에 적용된다. 범주화의 장점은 범주나 범위가 흔히 통계형 데이터 형식이어서 다양한 분석 및 가공이 가능하다는 점이다. 반면 이 방법의 한계는 범주나 범위로 표현됨에 따라 정확한 수치 값에 따른 분석, 특정한 분석 결과를 도출하기 어려우며, 데이터 범위 구간이 좁혀질 경우 추적이나 예측이 가능할수도 있다는 점이다.

⑬범주화(Data Suppression)

명확한 값을 숨기기 위해 데이터의 평균 또는 범주 값으로 변환하는 방식이다. 이는 은폐화 방식으로도 불린다. 다만 데이터의 평균이나 범주를 이용하여 전체를 표현할 경우 특정 속성을 지닌 개인으로 구성된 데이터세트의 속성 정보를 함께 공개하는 것은 그 집단에 속한 개인의 정보를 공개하는 것과 마찬가지라는 점에서, 범주화를 통해 비식별화 처리를 함에 있어 추가적인 주의가 필요하다. 가령 에이즈 환자 집단을 공개하면서 특정인물이 그 집단에 속하고 있음을 알 수 있도록 표시한다면, 이는 그 특정인물이 에이즈 환자임을 공개하는 것과 다를 게 없다.

⑭랜덤 올림 방식(Random Rounding)

개인식별 정보에 대한 수치데이터에 대하여 임의의 수를 기준으로 올림(round up) 또는 절사(round down)하는 비식별화 기법이다. 이는 민감성이 높은 데이터에 적용하는 범주화 기법, 총계처리 중 라운딩 기법과 유사한 방법이다. 다만 수치데이터 이외의 데이터에도 확장되어 적용될 수 있다는 점에서 차이가 난다. 가령 식별자가 나이, 우편번호 등과 같은 수치정보로 주어진 경우, 이 식별자 중에서 일의 자리, 십의 자리 등 뒷자리 수를 숨기고 앞자리 수만 나타내는 방법을 이용할 수 있다.

⑮범위 방법(Data Range)

개인식별정보에 대한 수치데이터를 임의의 수를 기준으로 하여 범위로 설정하는 기법이다. 이를 통해 해당 값의 분포, 구간으로 표현하게 된다. 가령 소득의 경우 5300만원은 '5000만원에서 6000만원' 이라고 대체하여 표기할 수 있다. 이는 서비스 이용등급, 처방정보(횟수, 기간 등), 위치정보, 유동인구, 사용자 수 등 범위 정보를 통해서도 유용한 분석이 가능한 상황에서 활용될 수 있다.

⑯제어 올림(Controlled Rounding)

랜덤 올림 방법에서, 어떠한 특정 속성값을 변경시킬 때 행과 열의 합이 일치하지 않는 단점을 해결하기 위해 행과 열이 맞지 않는 것을 제어하여 일치시키는 기법이다. 하지만 이는 컴퓨터 프로그램으로 구현하기 어렵고, 복잡한 통계표에는 적용하기 어려우며 해결할 수 있는 방법이 존재하지 않을 수 있다는 점에서, 아직 현장에서는 잘 사용하지 않는 기법으로 알려져 있다.

5) 데이터 마스킹(Data Masking)

데이터 마스킹이란 개인식별 정보에 대하여 전체 또는 부분적으로 대체값으

로 변환하는 조치이다. 이는 주로 이름이나 전화번호 같이 개인을 쉽게 식별할 수 있는 정보, 계좌번호 같이 기관·단체 등의 이용자 계정 등에 적용된다. 데이터 마스킹의 장점은 완전 식별화가 가능하며 원시 데이터의 구조에 대한 변형이 적다는 점이다. 반면 단점은 과도한 마스킹 조치를 적용할 경우 유용한 정보로서의 활용도가 떨어지며, 반면에 마스킹의 수준이 낮을 경우 특정한 값의 추적이나 예측이 가능하다는 점이다.

⑰임의 잡음 추가(Adding Random Noise)

소득과 같은 민감한 개인식별항목에 대해 임의의 숫자 등의 잡음을 추가하여 식별정보의 노출을 방지하는 비식별 기법이다. 가령 생년월일의 경우 실제 생년월일에 대해 사전에 정의한 6개월의 잡음을 추가한다고 하면, 원래의 생년월일 데이터에 1일부터 최대 6개월의 날짜가 추가되어 기존 자료와 오차를 가질 수 있다. 이 방법의 특징은 지정된 평균과 분산의 범위 내에서 잡음이 추가되므로 원 자료의 유용성을 해치지 않으면서 비식별 기법을 적용할 수 있다는 점에 있다. 하지만 경우에 따라 추적이나 예측을 방지하기 위해 마스킹 값으로 적용된 잡음값은 데이터값과는 무관하다는 점에서 분석이나 유효한 데이터로 활용하기 어려울 수도 있다.

⑱공백(Blank)과 대체(Impute)

빅데이터 자료로부터 선택된 비식별 대상 데이터를 공백과 대체의 방식으로 비식별화하는 기법이다. 일단 대상 데이터를 선택한 후 해당 데이터를 공백으로 바꾼다. 이후 공백으로 바뀐 부분을 대체값을 적용하여 채운다. 이 때 공백 이외에도 특수문자로 처리하는 경우도 많다. 이는 주민번호, 사용자 아이디, 성명, 전화번호, 주소 등을 비식별화할 때 유용하게 이용된다.

(4) 프라이버시 보호 모델

1) k-익명성(k-anonymity)[37]

특정 자료의 익명화가 어느 정도 완료가 되어도 그 자료에서 익명화되지 않은 부분과 이미 공개된 다른 자료들 사이에 어떤 공통점을 찾을 수 있다면 이런 공통점을 통해 익명화된 부분의 개인정보도 식별할 수 있는 가능성이 높아진다. 방대한 정보에 대해서 접근하는 비용이 매우 낮아진 현재 이런 연계성 공격에 의해서 비식별화 방식으로 처리된 정보가 예상하지 못한 다른 공개 자료들과의 결합이란 경로를 통해 비식별화된 개인정보가 드러날 수 있는 위험성이 높아지고 있다. 이런 관점에서 공개된 제3의 자료와의 결합에 의한 재식별 가능성을 줄이는 방법으로 2002년에 L. Sweeney가 'k-익명성' 이란 기법을 제안했다. 이 모델은 프라이버시 보호를 위한 통계적 모델의 기본 모형이라 할 수 있다.

'k-익명성' 은 어떤 데이터의 집합들이 있는 경우, 어떤 레코드의 개별값들과 동일한 레코드를 k-1개 이상의 레코드를 가진 경우에 전체적으로 동일한 레코드가 k개 존재하는 경우를 말한다. 결국 이 'k' 수치는 개별 레코드들을 분류하는 기준이 되는 준식별자들의 속성값들이 같은 레코드들의 개수를 의미한다. 그러므로 k수치가 증가한다는 것은 동일한 준식별자들을 가진 레코드들의 개수가 증가한다는 의미이므로 그만큼 데이터가 익명화가 되는 범주가 커진다. 익명화가 되는 영역이 커지므로 어떤 잠재적인 공격자가 자신이 알고 있는 공개된 개인정보를 가지고 있다고 하더라도 어떤 특정의 연계성을 찾기가 어렵게 되는 구조를 가지게 된다. 그 결과 제 3자에 의한 연계성 공격으로부터 개인정보의 익명성을 방어할 수 있는 확률이 상승한다.

하지만 k 수치가 증가하면 해당 데이터세트의 정보로서의 유용성은 보통 하락하게 된다는 한계가 있다. k-익명성의 개념은 기본적으로 일종의 범주화 개념으로 파악할 수 있다. 그런데 많은 개별적 식별 자료들이 범주화를 통해 특정성이 상실되면 정보를 분석, 활용하려는 입장에서는 그 정보 자체의 활용도가 상대적으로 하락하게 되어 데이터의 활용도가 낮아질 수밖에 없다.

37) 여러 범주의 다양한 익명화 기법들이 있지만 이 글에서는 다른 공개된 데이터들과의 연결에 의한 연결성 공격(linkage attack)을 막는 것에 효과적인 기법들의 논의에 집중한다.

그러므로 프라이버시의 보호와 정보의 활용이란 두 가지의 상반된 가치들 사이에 적절한 균형점을 찾는 것이 필요하고, 이런 관점을 'k-익명성' 맥락으로 재해석하면 최적 수준의 k수치를 어떻게 찾을 것인지가 중요한 논점이 된다.[38] 최적 수준의 k수치란 것은 사회적 최적 수준을 찾는다는 관점에서 보면 기술적 평가인 동시에 규범적 평가이기도 하기 때문에 순수한 통계학적 기법에만 집중하는 기술적 접근법으로는 해결책을 찾는 것이 거의 불가능하다.[39] 결국 사회적으로 최적 수준의 k수치를 찾는 과정은 기술적 관점과 규범적 관점 등을 감안한 다양한 관점을 감안한 논의와 실제 데이터에 기초한 분석과정 등 여러 시행착오를 전제로 하게 될 것이다.

2) l-다양성(l-diversity)

통계학적 관점으로 볼 경우 'k-익명성'은 동질성 공격(homogeneity attack)과 배경지식에 의한 공격에 취약하다는 단점을 가지고 있다. 동질성 공격이란 어떤 데이터의 집합에서 동일성을 가진 정보를 이용하여 공격의 상대방의 민감한 정보를 파악할 수 있는 공격이다. 이런 현상이 가능한 이유는 'k-익명성'에 의해 레코드들이 범주화가 되더라도 범주화의 기초가 되지 않은 민감한 정보들이 모두 같은 값을 가질 수 있기 때문이다. 범주화의 기초가 되는 준식별자들에 대해서는 여러 다양한 값들이 혼재가 되어 있어서 연계성 공격에 의한 식별이 어렵지만 이 준식별자와 연결된 민감한 정보들은 처음부터 'k-익명성'의 범주화의 기초가 아니기 때문에 발생할 수 있는 현상이다.

배경지식에 의한 공격은, 별도의 공개된 정보가 없어도 공격자가 별도로 가지고 있는 배경지식이나 상식을 이용해 식별이 가능해질 수 있는 경우를 의미한다. 배경지식에 의한 공격이 성공할 수 있는 경우, 그 기본적인 원인은 동질성 공격의 경우와 마찬가지로 일부 데이터들의 다양성이 약화되어 외부

38) Sasha Romanosky and Alessandro Acquisti, "Privacy Costs and Personal Data Protection: Economic and Legal Perspectives", 24 Berkeley Technology Law Journal 1061, 1071 (2009).
39) Robert Gellman, "The Deidentification Dilemma: A Legislative and Contractual Proposal", 21 Fordham Intellectual Property Media & Entertainment Law Journal 33, 37 (2010).

의 공격에 대해 익명성에 의한 방어가 적절히 이루어지지 못하는 경우이다. ‘k-익명성’ 이 보일 수 있는 다양성의 한계로 인해 이와 같은 문제들에 봉착할 가능성에 대비하기 위해 등장한 기법이 ‘l-다양성’ 기법이다. 이 기법은 결국 범주화를 통한 익명화 과정에서 데이터값들이 충분히 다양한 값들을 가질 수 있도록 확인하고 조정하는 과정을 거치도록 하는 것이다. 이런 과정을 통해 동질성 공격을 원천적으로 봉쇄할 수 있고 나아가 배경지식에 의한 공격이 있어도 상당히 강한 방어력을 유지할 수 있게 된다.

이 기법에서의 ‘l’ 수치 또한 k 수치와 마찬가지로 커질수록 범주화의 영역이 커지는 효과를 가져와서 익명성을 더욱 강력하게 하는 효과를 발생시킨다. 그러므로 최적의 k 수치를 찾는 것과 마찬가지로 최적의 l 수치는 익명성 강화에 의한 개인정보 보호의 강화와 데이터 가치의 손상이라는 상충되는 목표 사이에서 적절한 균형점을 찾는 것이 된다.

3) t-근접성(t-closeness)

이 기법은 ‘l-다양성’ 기법을 적용해도 식별화가 될 수 있는 상황에 대해 그 가능성을 막기 위해 등장한 접근법이다. 특정 유형의 정보에 어떤 편향성이나 일정한 패턴이 발생하는 경우에 이런 특이 사항들을 해석해서 익명화된 데이터에 대한 공격을 하는 것이 가능할수도 있기 때문에 k-익명성과 l-다양성 개념만으로는 한계가 있을 수도 있다. 공개된 민감한 정보의 편향적인 분포를 통해 매우 높은 확률로 비식별화된 개인정보의 재식별이 가능해지고 그 민감 정보가 어떤 패턴이나 공통점을 가지고 있다면 개인을 식별할 수 있는 새로운 정보를 발견할 수 있다.

결국 이 모형은 전체 데이터의 모습과 다른 패턴을 가지고 있는 특정 데이터 구간을 통해 개인들을 식별할 수 있는 문제를 해결하기 위해 등장했다. 그러므로 이 모형은 특정 구간의 편향성이나 패턴의 특이성을 완화하기 위해 전체 데이터의 분포와 특정 구간 데이터의 분포를 일치화하는 방식을 도입한다.

여기서의 t 수치는 0에 가까울수록 전체 데이터의 분포와 특정 데이터 구간의 분포의 유사성이 강해지기 때문에 그 익명성의 방어가 더욱 강해지는 경향을 가지게 된다. 이 경우 전체 데이터의 재배치를 통해 t 수치를 조정하는 것이기 때문에 범주화의 영역을 넓혀서 익명성을 강화하는 k, l 수치와는 다소 성격이 다르다. 그래서 이 경우에는 익명성 강화를 위해 특정 데이터들을 재배치해도 전체 준식별자들의 값 자체에는 변화가 없기 때문에 일반적인 경우에 정보 손실의 문제는 크지 않을 것으로 예상할 수 있다. k-익명성, l-다양성, 및 t-근접성 개념에 대한 간략한 요약 및 비교는 [별첨 1]에 제시되어 있다.

4) 차분프라이버시(Differential privacy)

차분프라이버시 접근법은 k-익명성과 l-다양성의 취약한 부분을 보완하기 위해 C. Dwork가 제안한 모형이다. k-익명성과 l-다양성 모두 해당 프라이버시 모델들을 잘 이해하고 있는 잠재적 공격자에 의해서 민감한 정보가 유출될 수 있다는 한계가 있다. 두 모형 모두 해당 속성값들이 동일한 레코드의 숫자와 민감한 정보의 숫자들의 조정을 통해 잠재적 공격자에 의한 재식별의 위험성에 대해 방어를 하는 구조이다. 그런데 일부 유형의 정보에 대해서는 이런 수치들의 조정을 통해서 재식별의 위험성을 낮추는 것이 어려운 경우들이 있다. 그래서 단순한 숫자의 변화가 아니라 레코드들 자체의 확률적 변형을 통해 식별가능성을 제한하는 접근법이 등장하게 되었는데 그 대표적인 것이 차분프라이버시 모형이다.

차분프라이버시 개념은 암호학(cryptography)에서 출발한다.[40] 이 방식은 어떤 공개된 결과값이 특정인에 대한 정보를 드러내지 못하는 것을 보장하려는 것에 목적이 있다. 즉, (1) 어떤 특정인에 대한 정보가 포함되지 않은 데이터 집합에서 차분적인 알고리즘의 적용을 통해 획득한 결과와 (2) 그

40) Vito D'Orazio, James Honaker, and Gary King, 'Differential Privacy for Social Science Inference'(pdf), 제 5면 (2015)

특정인에 대한 정보가 포함된 데이터 집합에서 얻은 결과가 구별되지 못하게 하는 체계를 구축하는 것이 이 개념의 기본적인 목표이다. 이를 위해서는 정확하게 계산된 양의 노이즈를 통계 기록에 넣어서 개인의 식별성을 없애는 방법을 이용한다.

특히 이 차분프라이버시 접근법은 상호적(interactive)인 데이터베이스 이용의 상황에서 큰 효용성을 가지고 있는 방식이다.[41] 이런 데이터베이스는 흔히 검색(search)이나 쿼리(quiry)가 가능한 유형의 데이터베이스가 된다. 이와 같이 상호적인 관계를 전제로 하는 데이터에 대해서는, 검색이나 쿼리가 반복되는 과정을 통해서 해당 데이터베이스에 포함된 정보가 점차적으로 드러나게 될 것인데 이런 상황에서는 데이터의 잠재적 공격자가 계속적으로 드러나는 정보를 통해 개인의 속성에 대한 예측력을 높일 수 있다. 이런 예측력 증가현상을 방지하기 위해 차분프라이버시 개념을 통해 제안된 방식은 어떤 개인의 존재와 부존재가 데이터베이스의 시스템의 반응에 영향을 미치지 않도록 하는 접근법이다.

좀 더 정확하게는, 어떤 개인이 존재할 때 어떤 데이터 요청에 대해 특정 결과를 낼 확률과 그 개인이 존재하지 않을 때 동일한 결과를 내놓을 확률이 같은 경우에는 차분적 프라이버시의 공개가 발생한다는 말할 수 있다. 이 두 확률이 동일하게 되어 이 두 확률 사이의 상대적 비율이 1이 되는 경우에는 특정 개인에 대한 정보가 공개가 되지 않는 이상적인 상태가 되는 것이라 볼 수 있다. 결국 민감한 정보를 보호하기 위해 차분프라이버시는 체계적으로 무작위 수치를 넣게 되고 이 무작위 수치는 일종이 노이즈의 역할을 하게 된다.[42] 이 노이즈의 삽입을 통해 예컨데 어떤 데이터베이스에 특정인에 관한 정보가 포함되어 있는지 여부에 관계없이 동일한 결과물을 산출하도록 하는 것이다.

---

41) James Bambauer, Krishnamurty Muralidhar, and Rothindra Sarathy, 'Fool's Gold: An Illustrated Critique of Differential Privacy', 16 Vanderbilt Journal of Entertainment & Technology Law 4, 제 9면 (2014)

42) Bambauer, et al., 전게서, 제14면

2. 비식별화의 관리적 방법론

개인정보에 비식별화 기법을 적용하여도 언제나 재식별 위험은 남아있게 된다. 이를 대비하기 위하여 비식별화 수준에 대한 평가, 그에 대한 주기적인 재평가, 필요시 비식별화 정보에 대한 접근통제 등과 같은 “절차적 통제”를 마련하고, 또한 재식별 금지를 위한 규약을 이행하도록 하는 것과 같은 “규범적 통제” 등을 구비하는 것이 반드시 필요하다. 이를 위해 관리적 측면에 대한 고려가 반드시 별도로 이루어져야 한다. 이러한 비식별화 수준에 대한 평가 및 정보에 대한 관리라 함은, 공공 및 민간의 개인정보처리자가 개인정보를 포함하고 있는 데이터에서 개인식별요소를 제거한 후 확인해야 하는 개인정보 비식별화에 대한 적정성 평가, 재식별 위험 관리 등 일련의 안전 조치 방법을 포함하게 된다.[43]

(1) 개인정보 비식별화 적정성 평가

개인정보가 포함된 데이터는 개인정보 식별 요소를 제거하는 등의 기법을 통해 개인을 식별할 수 없는 형태로 데이터를 변경 및 이용할 수 있다. 만일 개인을 더 이상 식별할 수 없도록 데이터를 변경할 수 있다면, 이 데이터는 개인정보보호법상 개인정보에 해당하지 않기 때문에 데이터 개방·공유 및 분석에 제한을 받지 않게 된다. 하지만 데이터에 개인식별요소의 제거 조치가 적절히 수행되지 않았다면 이는 다른 데이터와의 결합을 통해 특정 개인을 식별할 수 있는 위험이 남게 된다. 그러므로 정보주체의 개인정보를 보호하면서 데이터 활용 및 빅데이터 분석을 활성화하기 위해서는 개인정보가 포함된 데이터에 대한 개인정보 식별 요소 제거 조치가 적정하게 이루어 졌는지를 확인하기 위한 체계적이고 객관적인 평가 방안이 요구된다.

개인정보 식별 요소 제거 기법은 간단해 보인다. 하지만 이를 실제 적용하고 그 적용이 적절히 이루어졌는지를 판단하는 것은 매우 어려운 작업이다. 특

43) 2014년 12월에 발간된 한국정보화진흥원의 “개인정보 비식별화에 대한 적정성 자율평가 안내서 v1.0”에는 비식별화에 대한 관리적 접근법에 대해 설명하고 있다. 관리적 방법론에 관한 아래 내용 중 상당 부분은 이를 주로 참조한 것이다. (이 안내서는 법률적 구속력을 지닌 문서도 아니고 이 안내서에서 설명된 관리적 접근법 또한 이 안내서를 발간한 기관의 공식적인 입장은 아니다.)

정 시점에는 비식별화된 것으로 평가된 데이터라도, 공격자의 다양하고 열정적인 시도, 새로운 데이터의 공개, 컴퓨팅 기술의 발전 등에 따라 개인이 재식별될 수 있는 위험이 내포되어 있기 때문이다. 개인정보 비식별화에 대한 적정성 평가는 개인정보가 포함된 데이터에 대한 개인식별요소 제거 조치가 적정하게 수행되었는지를 판단하고 전반적인 절차에 대한 관리가 필요하다는 문제의식에서 출발한다.

적정성 평가는 개인정보 비식별화된 데이터를 인터넷 등에 공개하거나, 제3의 기관 등에 제공하는 경우 및 업무의 필요에 의해 자체 검증하고자 하는 경우 등 그 필요성이 인정될 때 자율적으로 수행하는 것이 가능하다. 또한 업무의 필요에 의해 자체 검증하고자 하는 경우에는 평가위원회를 구성하지 않고 자체적으로 평가를 수행할 수 있다. 평가 대상 데이터에 대한 개인식별요소 제거 조치의 적정성을 평가하기 위한 절차는 주로 ① 평가 기초자료 작성, ② 평가위원회 구성[44], ③ 평가 수행[45], ④ 추가 비식별조치[46], ⑤ 평가 후 관리[47] 등 다섯 단계로 구성된다.

---

44) 국내에서 평가위원회는 평가수행기관의 개인정보보호책임자(CPO)가 지정하는 3인 이상의 평가위원으로 구성하며, 평가위원은 홀수로 지정하여 평가위원의 의견이 1:1로 대립되지 않도록 평가위원회 구성한다. 평가위원의 과반수 이상은 해당 평가 대상 기관 및 평가 대상 데이터의 이용과 관련이 없는 외부의 전문가로 지정한다. 즉, 관련 업무영역의 전문가 1인, 개인정보 비식별화 전문가 1인, 법률 전문가 1인은 필수적으로 포함되도록 평가위원회를 구성한다. 이 때 평가위원장은 외부에서 위촉된 평가위원 중에서 호선으로 선출하며, 평가위원회의 개최, 심사, 종료, 평가결과의 발표 등 평가위원회의 운영과 관련된 전반적인 사항을 관장한다. 또한 평가위원장은 다른 평가위원들과 협의하여 평가위원별 전문성을 고려한 업무 분장을 실시하고 평가업무를 수행할 수 있다. 평가위원회는 착수회의를 포함해서 최소 2회 이상 운영하여야 하며, 추가적인 평가위원회의 개최가 필요할 경우에는 평가위원회의 협의를 거쳐 운영 회수 및 기간 등을 연장할 수 있다.

45) 국내에서 평가위원회는 평가수행기관에서 제공한 '기초자료' 와 '세부 평가 방법' 을 기반으로 평가 대상 데이터에 대한 개인식별요소 제거 기법 및 비식별 수준의 적정성에 대해 평가하고, 최종적으로 '적정' 또는 '부적정' 의견을 제시하며, 필요시 기초자료에 대한 보완을 요청할 수 있다. 평가위원회는 평가 대상 데이터의 특성을 종합적으로 고려하여, '세부 평가 방법' 의 각 평가 단계별 평가지표 및 평가기준에 대해 평가위원회의 협의를 거쳐 조정 및 보완하여 사용할 수 있다. 다만 협의가 이루어지지 않는 경우에는 과반수 이상의 평가위원이 선택한 평가지표 및 기준을 채택해서 사용할 수 있다. 평가위원회가 '적정' 으로 평가한 경우에는 해당 정보만으로는 특정 개인을 알아볼 수 없을 뿐만 아니라, 다른 정보와 결합하여도 특정 개인을 식별 할 수 없는 상태라는 것을 의미한다. 평가위원회가 '부적정' 으로 평가한 경우에는 관련 데이터에 대한 추가적인 개인식별요소 제거 조치가 필요하다는 것을 의미한다. 또한 평가위원회는 추가적인 개인식별요소 제거 조치에 대한 사항 (적용기법 및 비식별 수준)을 구체적으로 제시해야 한다. 그리고 평가위원회는 컴퓨팅 환경의 발전, 평가 대상 데이터의 특성, 연계 가능한 데이터의 공개 가능성 등 다양한 환경변화를 고려하여, 평가 대상 데이터에 대해 일정 시간 경과 후 재평가를 받을 필요가 있는지에 대한 의견을 제시할 수 있다. 이때 만일 재평가가 필요하다고 인정된 경우에는 1년, 3년 등 구체적인 재평가 일정을 제시해야 한다.

46) 국내에서 평가수행기관은 평가결과가 '부적정' 인 경우, 평가위원회의 평가 의견을 고려하여 해당 데이터에 대한 개인식별요소 제거 조치를 추가적으로 수행한다. 그리고 평가수행기관은 평가위원회에서 제시한 개인식별요소 제거 기법 및 비식별 수준을 고려하여 개인정보가 식별되지 않도록 조치하고 관련 내용을 문서로 작성하여 관리해야 한다. 또한 개인식별요소 추가 제거 작업이 완료된 경우에는 관련 '기초자료' 를 보완하고, 평가위원회에 추가적인 개인식별요소 제거 작업이 적정히 수행되었는지에 대한 확인을 요청해야 한다.

(2) 재식별 위험 관리 방안

재식별 위험을 방지하기 위한 관리적 조치로 다음과 같은 관리적 방법을 생각할 수 있다. ① 데이터 제공 및 위탁 계약 시 재식별 금지 관련 조항을 반영하도록 한다. 비식별화된 데이터를 제3의 기관에 제공하거나, 처리를 위탁하는 경우, 개인정보가 재식별될 위험을 관리하기 위한 내용을 계약서에 반드시 반영하도록 하는 것이다.[48] ② 또한 데이터 공개 시 재식별 금지 관련 조항을 게시하도록 한다.[49] ③ 재식별의 가능성을 반복적으로 모니터링 하도록 한다. 비식별화한 데이터를 이용 및 처리하는 기관에서는 개인정보 비식별화에 대한 적정성 평가를 재실시 해야 하는 경우가 발생하는지를 상시 또는 정기적으로 모니터링 및 관리할 필요가 있다. 또한 ④ 개인정보 재식별시의 대응 매뉴얼을 마련하고 시행하도록 한다. 즉, 기관별 데이터 공개, 제공, 위탁 특성을 고려하여 개인정보 재식별시 대응 매뉴얼을 마련하고, 관련 개인정보취급자에 대한 교육을 정기적으로 실시한다.

제2절 비식별화 사례 및 한계

지금까지 비식별화와 관련된 개념정의를 바탕으로 비식별화의 기술적 · 관리적 방법론을 살펴보았다. 비식별화 방법론의 전제는 비식별화 조치가 실제

47) 국내에서 평가위원회 또는 평가수행기관에서 인정한 경우에는 평가 대상 데이터에 대한 사후적인 개인정보 비식별화에 대한 적정성 평가를 재실시 할 수 있다. 개인정보 비식별화에 대한 적정성 평가를 재실시 하는 경우에는 평가위원회를 재구성 할 수 있으며, 평가의 연속성 차원에서 기존 평가 위원 중 최소 1인 이상을 포함하여 구성할 수 있다.

48) 구체적으로 살펴보면 (ㄱ) (데이터 처리시 재식별 금지) 데이터를 제공 또는 위탁 받은 기관에서는 데이터를 비식별화된 상태로 처리하고, 데이터를 이용한 재식별 시도를 금지해야 한다는 것을 명확히 하고 (ㄴ) (재제공 및 위탁시 재식별 금지 명문화) 데이터를 제공 또는 위탁 받은 기관에서 관련 데이터를 제3의 기관에 재제공하거나, 처리를 재위탁하는 경우에도 재식별 시도를 금지한다는 것을 명확히 하여야 하며 (ㄷ) (위험 발견시 통지) 재식별이 이루어지거나, 재식별의 가능성이 높아지는 상황 발생시, 관련 사항에 대해 데이터 이용자가 데이터 제공자에게 통지하는 책임을 명확히 하여야 한다.

49) (ㄱ) 개인정보 비식별화 데이터를 인터넷 등에 공개하는 경우, 관련 데이터를 이용하는 개인 또는 단체가 개인정보를 재식별하기 위한 행위를 금지하는 내용을 이용약관 등의 형태로 게시한다. (ㄴ) (데이터 처리시 재식별 금지) 공개된 데이터를 이용하는 기관에서는 데이터를 비식별화된 상태로 처리하고, 데이터에 대한 재식별 시도를 금지해야 한다는 것을 명확히 한다. (ㄷ) (위험 발견시 통지) 재식별이 이루어지거나, 재식별의 가능성이 높아지는 상황 발생시, 데이터 이용자는 데이터 제공자에게 관련 사항을 통지해야 한다는 점을 명확히 한다.

개인정보 보호에 기여하므로, 이를 통해 개인정보의 보호와 이용 간 조화를 꾀할 수 있다는 점이었다. 아래에서는 실제 사례를 바탕으로 비식별화가 개인정보 보호에 기여하는지, 이를 통해 개인정보 보호와 이용 간 조화를 달성하는 역할을 하는지에 대해 확인해 본다.[50] 나아가 비식별화 조치의 한계라고 불리는 재식별이 문제된 사례도 추가로 검토한다. 이를 통해 개인정보의 비식별화 처리가 개인정보 보호에 미치는 영향에 대해 보다 심층적으로 알아본다.

1. 비식별화 사례 (Heritage Health Prize Case)

(1) 의의

Heritage Health Prize Case는 비식별화된 개인정보를 분석한 실제사례이다. 이는 미국에서 개최된 의료보험 관련 빅데이터 분석 대회인데, 여기에서 참가자들은 주최측이 제공한 데이터를 가지고 빅데이터 분석을 진행하였다. 주최측은 대회 진행에 앞서 비식별화 전문가들에게 대회 참가자에게 제공할 데이터의 비식별화를 요청하였고, 해당 전문가들은 자신이 진행한 비식별의 방법이 어떤 것이었는지, 그리고 이것이 개인정보 보호에 충분한 정도였는지 등에 관하여 스스로 평가하여 보고하는 자료를 준비했다.[51]

일반적으로 건강정보는 민감정보로서 엄격하게 보호되어야 할 대상이다. 하지만 이는 연구와 분석을 통한 이용의 대상으로 의술진보의 원동력이기도 하다. Heritage Health Prize를 대상으로 한 비식별화에 대한 연구는 건강정보에 요구되는 보호의 필요성과 이용의 요청을 적절하게 형량하는 방법을 모색하고, 그 적절성을 평가하고자 하였다. 이 사례는 미국의 상황을 배경으로 한다. 보건의료 분야의 개인정보에 관하여 미국에서는 HIPAA 법이 별도

50) 비식별화의 실제 사례에 관하여 공개가능한 정보를 구하는 것에는 커다란 한계가 있다. 비식별화 작업을 실제로 진행한 사례가 있다고 하더라도 해당 당사자가 그에 관해 공개하여 주목을 받는 것을 원하지 않는 경우가 대부분이기 때문이다. 아래 소개된 미국의 사례는 해당 사례가 학술지에 소개되었을 뿐 아니라, 이 사례를 둘러싼 학술적 논박이 벌어지기도 해서 더욱 흥미로운 사례라 할 수 있다.

51) Khaled El Emam et. al., ' De-identification Methods for Open Health Data: The Case of the Heritage Health Prize', Journal of Medical Internet Research (2012)

로 적용되는데, 이에 관해서는 아래에서 별도로 살펴본다. 여기에서는 해당 사례의 소개를 통해 “개인정보의 비식별화 처리가 개인정보 보호에 어떠한 영향을 미치는지”에 대해 주목한다.

(2) 배경

헤리티지 프로바이더 네트워크(Heritage Provider Network, 이하 “HPN”)는 미국 캘리포니아 주에 본사를 둔 의료보험 및 의료기관 네트워크이다. HPN은 2011년 4월에 의료정보 분석에 대한 대회를 개최하였다. 이 대회의 목적은 현재 및 이전의 청구 데이터를 사용해서 환자의 다음 해 입원 일수를 예측하는 모형을 구축하는 것이었다. 이 대회에서 사용된 핵심 데이터 집합은 113,000명의 환자에 대한 익명화된 HPN 데이터로 구성되어 있다. 해당 데이터는 이 대회의 웹사이트에서 다운로드를 통해 모든 참가자에게 제공되었다. 대회 기간 후 2년이 지난 뒤에 일정한 정확성 기준을 상회하는 모형들 중 가장 정확한 예측 모형을 개발한 팀에게 상금 300만 달러가 수여되는 대회였다.

이와 같은 건강정보 빅데이터 분석 대회에 있어 공개되는 건강정보에서 개인정보가 추출되지 못하도록 하는 장치에 더욱 관심을 기울여야 할 필요가 있었다. 만약 이런 장치를 제대로 구비하지 않아서 개인정보가 유출될 위험성이 높아지게 될 경우 HIPAA의 프라이버시 규칙에 저촉되는 규범적 문제가 발생하게 되기 때문이다. 그래서 건강정보의 공개는 대응하는 안전장치인 비식별화 조치인 HIPAA Privacy Rules에 맞추어 단행 되어야 했다. 그래서 이 대회를 대상으로 비식별화가 개인정보 보호에 기여하는지를 다룬 연구에서는 여러 비식별화 조치들을 적용한 후에 이렇게 처리된 데이터들이 재식별될 위험성을 얼마나 가지고 있는지, 그리고 이는 감내할 만한 것인지를 평가하였다.

(3) 적용된 비식별화 방식

대회 참가자에게 제공된 데이터에는 정보의 형태와 종류 별로 여러 가지의 비식별화 방식들이 적용되었다. 우선적으로 “가명(pseudonyms)처리”를 하였다. 회원(환자) ID, 병원 ID 등은 비가역적(irreversible)인 가명으로 변환되었다. 이런 정보들은 직접적인 식별자가 될 수 있기 때문이다. 이런 ID값들은 의료서비스의 제공 중에 사용되기 때문에 일반적으로 잘 알려지게 될 리스크가 높고, 이런 직접적 식별자 정보들은 금전적 이득 등의 이유로 악용될 위험성이 높다고 판단되었다. 예를 들어 환자ID는 의료서비스를 제공하는 병원을 식별하거나 치료 절차들의 수와 종류를 확인하기 위해 이용될 수 있다. 따라서 가명처리는 우선적으로 고려할 비식별화 방법 중 하나가 된다.

또한 데이터들 중에서 이례적으로 높은 수치를 보이는 양적 정보들의 경우는 이런 특이성으로 인한 식별가능성의 위험을 낮추기 위해 데이터 값들의 한계치를 설정하는 처리를 한다. 이런 방식을 “탑코딩(top-coding)”이라고 한다. 결국 탑코딩이란 일부 데이터가 지닌 극단적인 수치 데이터가 그대로 공개되는 것을 막기 위해 데이터의 수치의 상한치를 설정하는 것을 의미한다. 예를 들어, 건강정보와 관련해서는 이례적으로 긴 입원일수나 눈에 뜨일 정도로 높게 나타나는 치료비 숫자 자체가 다른 정보들과의 결합을 통해 그 데이터 주체의 병명이나 치료종류를 추측하게 할 위험성을 높이게 되기 때문이다.

그 이외에도 매우 위험성이 높은 환자들의 정보는 공개되는 데이터에서 “삭제”한다. 예를 들어, 낙태의 경험이 있는 환자나 희귀한 질병을 보유한 환자들은 자신들의 정보가 공개될 가능성에 특히 민감하게 반응할 수 있으므로 그러한 유형의 정보에 대해서는 삭제처리를 하였다. 또한 일부 정보는 “범주화(suppression)”등의 방식으로 비식별화 처리를 하였다. 병원의 입장에서는 개별 병원 특유의 치료 행태가 있을 수 있으며 이런 병원별 특이한 치료 행태는 그 병원에서 치료를 받는 환자의 개인정보와 연결될 수 있는 위험성을 지니고 있기 때문이다.

(4) 연구의 결과

연구는 해당 사례에서의 개인정보 비식별화 조치가 개인정보 보호에 충분히 기여하는지를 판단하고자 하였다. 여기에서 환자들을 보호하기 위해 필요한 비식별화 형태를 이해하기 위해서는 이 대회 중에 발생할 수 있는 위협들에 대한 판단이 선행되어야 했다. 나아가 이런 판단을 위해 몇 가지를 우선 전제하였다. 첫째, 대회의 모든 참가자들은 공개된 데이터 안에 있는 환자들의 개인정보를 재식별하려는 시도를 하지 않는다는 서약서에 동의해야 한다. 둘째, 잠재적 공격자가 이 HPN 데이터에 특정 환자에 관한 기록이 있는지 여부를 아는 것은 불가능하다. 셋째, 이 데이터에 대한 잠재적 공격자는 이 대회의 직접적인 등록을 통해서나 간접적인 방식으로 데이터의 유출을 통해 HPN 데이터를 습득한다.

이런 기본적인 전제에서 잠재적 공격자의 공격 형태는 크게 세 가지로 분류될 수 있다. 그리고 이러한 재식별의 위험성은 각각의 유형에 따라 다르게 정리되어 개별적으로 평가된다. 각각의 재식별 위험을 지닌 공격들이 위험한지 여부에 대한 평가 기준은 재식별의 위험성 확률 5%를 기준으로 하였다.[52] 재식별의 위험성 확률이 5%와 같거나 낮을 경우는 연구의 배경이 되었던 미국의 비식별화 기준인 HIPAA Privacy Rules에 비추어 볼 때 규범적으로는 재식별의 위험성을 무시해도 좋은 경우에 해당되는 것으로 보았다. 그 결과 비식별화 관점에서 보면 해당 비식별화 기법이 적용된 데이터는 HIPAA 프라이버시 규칙에 위배되지 않는 것으로 판단된다.

첫 번째 유형의 공격은 공격의 대상에 대한 어느 정도의 기본적인 정보를 가지고 있는 이웃(neighborhood)에 의한 공격이다. 이런 공격자는 목표가 되는 환자에 대한 추가적인 정보를 가지고 있는 공격자 유형을 의미한다. 예를 들어, 해당 환자의 이웃, 회사동료, 전 배우자, 친척 등이 이 유형의 잠재적 공격자가 될 수 있다. 두 번째 유형의 공격은 투표자등록부(voter

52) 이 5%의 기준은 HIPAA 프라이버시 규칙 중 세이프하버에 대한 일련의 연구들을 참조한 것이다. 이 세이프하버 입안자들은 전체 모집단(population)에서 데이터의 특이성(unique)이 확인되는 확률이 0.04%이면 재식별의 위험성이 매우 낮다고 판단한다. 이 0.04%를 고려해서 본 연구에서 설정한 수용가능한 재식별의 위험성 확률을 5%로 정한 것이다.

registration list)에서 공개된 데이터와의 결합을 통한 공격이다. 특히 캘리포니아 주에서는 이 투표자등록부를 구입할 수 있기 때문에 관련 데이터의 습득이 매우 용이하다. 이 투표자등록부에는 투표자의 생일과 성별정보가 포함되어 있다. 세 번째 유형의 공격은 입원환자데이터베이스(impatient database)와의 결합을 통한 공격이다. 미국의 48개 주들이 입원환자의 데이터를 수집하고 26개 주에서는 별도의 기관을 통해서 병원에서의 퇴원일자를 공개하고 있다. 이 데이터베이스는 연구 또는 기타 승인된 목적으로 이용한다는 전제하에 구입이 가능하다. 예를 들어, 이 대회의 데이터베이스와 입원환자데이터베이스를 비교해서 구체적인 생년월일과 진단 코드와 절차를 발견할 수도 있다.

이런 각각의 공격들을 고려해서 분석한 결과 아래와 같은 형태가 원형 데이터를 최적으로 일반화 방식을 적용해서 비식별화된 데이터로 산출되었다.

Final generalizations in the dataset[53)]

53) http://www.jmir.org/2012/1/e33/#table5

Quasi-identifier	Generalization
Age	10-year interval; 80+
Sex	No change
Days In Hospital Y2	Days to 2 weeks; >2 weeks in year 2
Days In Hospital Y3	Days to 2 weeks; >2 weeks in year 3
Specialty	Grouped specialty
Place Of Service	Grouped place of service
CPT Code[a]	Grouped CPT code
LOS[b]	Days up to 6 days; (1-2] weeks; (2-4] weeks; (4-8] weeks; (8-12 weeks]; (12-26] weeks; 26+ weeks
DSFC[c]	4 weeks
Diagnosis	Primary condition group

[a] Current Procedural Terminology.

[b] Length of stay in hospital.

위와 같이 비식별화된 데이터에서 이 세 가지 유형의 공격에 대한 각각의 재식별의 위험성 확률을 산정한 결과 첫 번째 유형은 0.84%의 위험성을 보였고, 두 번째 유형은 0.0005%의 확률로 위험성이 예측되었다. 세 번째 유형의 경우 다양한 준식별자들의 조합에 대해서 개별적으로 위험성 확률을 산정했는데 전체적으로 0.1%에서 1.7% 사이에 위치하고 있기 때문에 기준선인 5%에 미치지 못했다.[54)]

이 세 가지 유형들의 위험성 판단에서 확인할 수 있듯이 세 가지 경우 모두 기준선인 5%의 확률에 미치지 못하기 때문에 이 대회에서 공개된 데이터에 적용되는 비식별화 조치는 HIPAA 프라이버시 규칙의 관점에서 보면 수용할 만한 것이 되었다고 판단되었다.

54) 구체적인 산정방법에 관해서는, http://www.jmir.org/2012/1/e33/#table6 (Table 6 및 관련 본문) 참조

(5) 연구에 대한 평가

위의 연구는 비식별화 조치가 적절하게 적용되어 해당 데이터에서 개인정보의 재식별의 위험성 수준이 규범적 관점에서 수용할 정도에 이르렀다는 결론을 내렸다. 이를 통해 공개된 HPN 데이터의 재식별 위험성을 우려할 필요가 없다는 의견을 표명했다. 하지만 이러한 연구분석에 대해 비판적인 시각도 있다.[55] 예를 들어 Narayanan은 기본적으로 일회적인 비식별화 조치의 적용만으로 재식별의 위험성이 제거될 수 있다는 접근법 자체에 대해 회의적인 시각으로 바라본다. 특히 병원 또는 의사와 관련된 정보에 대해서는 접근할 수 있는 방식들이 다양하기 때문에 비식별화 조치가 적용된 데이터라 해도 외부의 보충적인(auxiliary) 별도 정보와의 연결을 통해 병원이나 의사의 신분을 특정하는 것이 어렵지 않을 수 있다는 의견을 제시하였다.

Narayanan은 기본적으로 특정한 환자의 개별적인 정보는 병원이나 의사의 경우에 비해서 접근할 수 있는 별도의 정보가 제한되어 있기 때문에 재식별의 위험성은 낮다는 점에는 동의한다. 하지만 병원 또는 의사의 정보들을 통해 간접적으로 환자의 개인정보를 특정할 수 있는 가능성이 있으므로 비식별화 조치만으로 환자가 특정될 수 있는 위험성을 제거했다고는 할 수 없다고 주장한다. 그래서 Narayanan은 위험성을 원천적으로 방지하는 방식이 더 효과적일 수 있다는 견해를 제시한다. 예를 들어, 대회에서 공개되는 데이터의 범위 자체를 줄이는 방식이 더욱 효과적일 수 있다고 본다. 결국 그는 많은 비용을 투자해서 더욱 정교한 비식별화 기법을 개발하는 것은 비효율적인 접근방법일 뿐이라고 일축하며, 잠재적 공격자의 약점을 공략하는 방식이 더욱 효과적이라고 주장한다. 이러한 공략방식의 일환으로 공개되는 데이터의 범위 자체를 제한하는 것을 생각해 볼 수 있다고 한다. 결국 HPN 데이터 비식별화 사례와 그에 대한 Narayanan의 주장은, 비식별화를 둘러싼 전문가 사이의 방법론적 입장 차이를 반영하는 것으로 볼 수 있다.

---

55) 비판적인 시각을 체계화항 논문의 형태로 제시한 것으로 다음 문헌 참조. Arvind Narayanan, 'An Adversarial Analysis of the Reidentifiability of the Heritage Health Prize Dataset'(pdf)(2011)

2. 비식별화의 한계 및 재식별의 위험성 문제

(1) 재식별의 문제의 의의

비식별화 조치의 한계는 비식별화 이후에 나타날 수 있는 재식별의 위험성이다. 아무리 비식별화가 되었다고 하더라도 재식별의 위험성이 크다면 비식별화를 통해 얻고자 했던 개인정보의 보호를 달성하기 어려운 상황이 초래되기 때문이다. 아래에서는 재식별로 인한 문제가 나타난 사례들을 검토한다. 이를 통해 개인정보의 비식별화 처리가 개인정보 보호에 미치는 영향에 대해 좀 더 심층적으로 알아본다.

(2) 재식별의 가능성이 구체화된 세 가지 사례

1) Governor Weld 사건

L. Sweeney의 2000년 연구에 의하면, 1990년 인구조사 정보를 처리한 결과 미국에 있는 87.1%의 사람들이 그들의 우편번호, 생일(생년 포함), 성별 정보를 통하여 식별이 가능하다고 한다.[56] 또한 53%의 미국 거주자에 대해서는 그들이 사는 도시, 생일, 성별에 의하여 식별이 가능하다고 한다. 이러한 연구는 구체성의 정도가 낮은 정보조차 개인의 신원을 확인하는 데에 이용될 수 있음을 보여주는 것이라 할 수 있다.[57] Sweeney는 이러한 맥락의 연구를 통해 개인에 대한 다양한 항목들의 데이터를 조합하여 개인을 식별하는 것이 가능함을 입증했다. 이 연구에 앞서서 1990년대 중반 Sweeney는 이미 비식별화되어 공개된 데이터로부터 특정 개인 – William Weld 당시 메사추세츠 주지사 – 을 재식별해낼 수 있었다. 재식별의 위험성 맥락에서 이 Weld 주지사 사건을 살펴볼 수 있다.

---

56) Sweeney, L., *Simple Demographics Often Identify People Uniquely*. Carnegie Mellon University, Data Privacy Working Paper 3 (2000)

57) Paul Ohm, 'Broken Promises of Privacy: Responding To The Surprising Failure of Anonymization', 57 UCLA Law Review 1701, 1720 (2010).

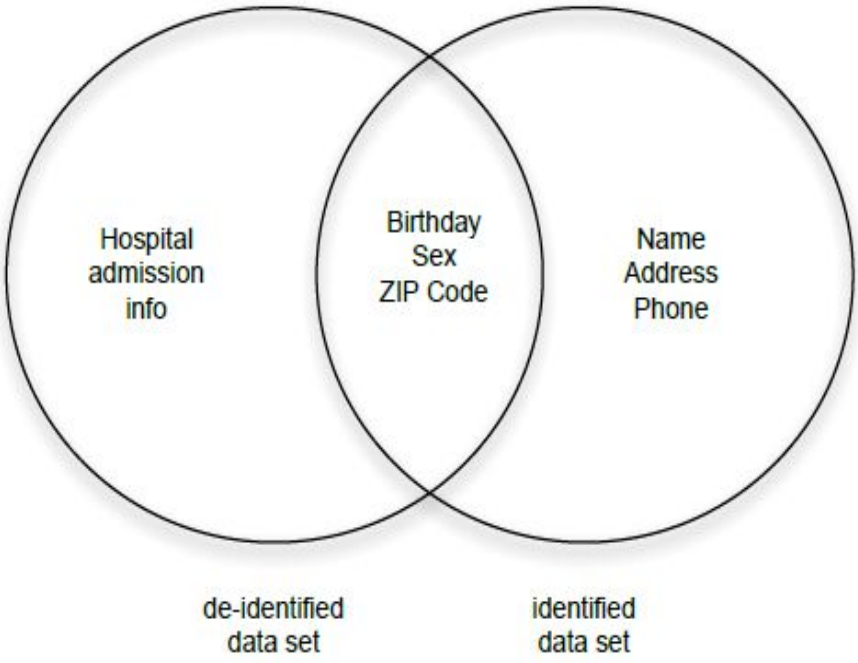
매사추세츠(Massachusettes) 주에서는 주정부 소속 단체보험위원회(Group Insurance Commission (이하 “GIC”))를 통해 매사추세츠 주 공무원들을 위한 건강보험을 제공했다. 1990년대 중반 GIC는 모든 주 공무원들이 병원을 방문한 기록 및 그와 관련된 일부 정보를 일반 연구자들의 연구 목적을 위해 공개하기로 하였다. 실제로 공개된 데이터에는 환자의 병원 방문 기록이 포함되어 있었고, 그와 함께 우편번호, 생일, 성별 등의 정보도 포함되어 있었다. 다만 데이터 공개에 앞서 이름, 주소, 사회보장번호, 그리고 기타 ‘분명한 식별인자’로 여겨지는 정보는 삭제하였다. 주 정부는, 식별자를 삭제하여 비식별화 조치를 취했기 때문에 데이터의 공개를 통해 환자의 개인정보가 유출되지 않을 것이라고 하고 데이터를 공개하였다.

그 당시 매사추세츠 주의 주지사였던 William Weld는 GIC 보유 정보를 공개함에 있어 식별자가 삭제되었기 때문에 정보주체의 개인정보가 충분히 보호될 것임에 관해 공개적으로 자신있게 언급했다. 데이터가 공개되자 당시 대학원생이던 Sweeney는 GIC의 데이터로부터 재식별 가능성을 모색하였다. 그녀는 Weld 주지사의 거주지가 5만4천명의 인구가 거주하는 매사추세츠 주의 Cambridge 시라는 사실과 주지사 거주지의 우편번호 등 주지사가 공인이기 때문에 이미 알려져 있는 정보를 활용하여, Weld 주지사에 관한 정보의 재식별을 시도하였다. Sweeney는 이와 같은 실질적으로 공개된 정보에 더해 20달러를 내고 Cambridge 시의 선거인명부를 구매하였다.[58] 이 선거인명부는 유권자들의 이름, 주소, 우편번호, 생일, 성별을 포함하고 있어서 추가적인 정보를 제공해 주었다. GIC가 공개한 데이터와 이처럼 추가적인 경로를 통해 확보가 가능한 데이터 사이의 비교 및 대조 과정을 통해 Sweeney는 어렵지 않게 Weld 주지사에 관한 정보를 찾아낼 수 있었다. 확보된 데이터에서 오로지 6명의 Cambridge 거주자만 그와 생일이 같았고, 그 중 3명만이 남자였으며, 주지사만이 그 우편번호지에서 거주하고 있었던 것이다.

이런 재식별이 이루어진 방식은 일반적으로 연결(linkage) 공격에 의한 것

58) 미국에서 선거인명부는 판매를 통한 공개가 가능하다

으로 부른다. Sweeney는 GIC 공개 데이터에 포함된 비식별화된 데이터 집합과 선거인명부와 같은 식별성 있는 데이터 집합 사이에 공통적으로 포함된 생일, 성별, 우편번호의 정보를 활용하여 비식별화된 데이터 집합에 포함되어 있는 Weld 주지사를 식별하고 그와 관련된 정보를 확인할 수 있었던 것이다. 이런 연결 공격의 개념은 아래의 그림에서 확인할 수 있다.



출처: Simon L. Garfinkel, 'De-Identification of Personally Identifiable Information', NISTR 8053, 그림 2 , NIST (2015)

2) AOL 사건

2006년에 America Online(이하 "AOL")은 "AOL Research"라 일컫는 새로운 계획을 발표했다. AOL Research는 일반 연구자들에게 편의를 제공하기 위해서 웹사이트에 공개적으로 2천만 개의 탐색 질의를 게시하고 3개월간의 질의들을 요약하여 제공하였다. 이 정보를 대중에 공개하기에 앞서 AOL은 개인정보를 보호하기 위하여 비식별화 처리를 하였다. AOL사용

자명을 가명처리하고 IP주소와 같은 명백한 식별정보를 모두 범주화(suppression)하는 등의 과정을 거쳤다. 가명처리는 아라비아 숫자로 특유의 식별 번호를 부여한 것인데, 이는 공개된 정보의 효용성을 확보하기 위한 것이었다.

검색질의 정보가 공개된 후 블로거들 사이에서는 이런 정보의 공개가 개인의 프라이버시를 침해하는 결과를 가져올 가능성에 대해서 논쟁이 벌어졌다. 이런 논쟁 속에서 어떤 블로그와 그 이후의 뉴스 보도에서 특정 번호의 검색질의 내용들이 좀 더 널리 알려지게 되었다. 예를 들어, 이용자 3505202번이 '우울증과 의료휴가' 에 관해 질문; 7268042번이 '배우자의 부정행위에 대한 두려움' 을 검색; 이용자 17556639번이 '죽은 사람의 사진' 과 '교통사고 사진' 을 검색한 다음 '아내를 죽이는 법' 에 대하여 검색하였다는 등의 내용이 알려지게 되었다.

블로거들 사이의 논쟁은 뉴욕타임즈의 보도를 통해 종식되었다. 뉴욕타임즈 기자인 Michael Barbarno와 Tom Zeller는 'GA Lilburn의 조경사' , '성이 Arnold인 몇 사람' , 'Georgia Gwinnet county의 그림자 호수 구역의 팔린 집' 등의 검색질의를 한 4417749 이용자에 대해 질의 내용 자체로부터 단서를 얻어 해당 이용자의 신원을 파악할 수 있었다. 그들은 이 이용자가, Georgia 주의 Linburn에 거주하는 62세의 Thelma Arnold라는 미망인임을 확인할 수 있었다.[59]

3) Netflix 사건

이러한 AOL의 대실패가 있는지 약 2개월 후, 온라인 영화 대여 서비스 업체인 Netflix가 500만 명의 이용자들이 1999년 12월부터 2005년 12월까지 영화에 어떤 점수를 매겼는지에 대한 1억 개의 기록을 대중에 공개했다.

59) 이 사건의 결과 AOL은 정보를 공개한 연구자 및 책임자를 해고하였다. 기술담당최고책임자인 Maureen Govern은 사임하였다.

각각의 기록에서 Netflix는 평가 대상 영화와 점수(별 1개부터 5개까지), 점수를 매긴 날짜를 공개하였다. AOL과 GIC의 경우처럼 Netflix도 먼저 이용자명과 같은 식별인자를 삭제하고 고유의 이용자 식별번호를 배정함으로써 기록을 익명화하였다. 즉 이용자 1337번이 2003년 3월 3일에 영화 가타카에 4점을 주었고, 2003년 11월 5일에 영화 마이너리티리포트에 5점을 주었다는 정도의 정보를 확인할 수 있는 상태로 하여 데이터를 공개하였다.

이 정보가 공개된 후 Arvind Narayanan과 Vitaly Shmatikov는 '개인이용자에 대하여 조금이라도 아는 공격자라면 [Netflix Prize]의 데이터세트에 있는 개인이용자의 정보를 쉽게 식별이 가능하거나 적어도 개인이용자의 정보가 포함된 소규모의 정보 세트를 식별해 낼 수 있을 것' 이라고 주장하였다. 즉, 데이터베이스에서 사람들을 재식별하는 것은 어렵지 않기 때문에 그들의 영화 취향에 관한 약간의 추가 정보만 가지고도 모든 점수를 매긴 영화를 찾아낼 수 있다는 것이 그들이 주장이었다. 이들은, 만약 개인을 식별화해낼 의도가 있는 공격자가 데이터베이스에 있는 어떤 사람이 잘 알려지지 않은 6개의 영화에 부여한 점수를 알고 있다면, 다른 추가적인 정보가 없이도 이 공격자는 84%의 사람들을 식별할 수 있다고 하였다. 만약 그가 데이터베이스에 있는 어떤 사람이 정확히 언제 6개의 영화에 점수를 매겼는지 안다면, 그 영화가 대중적인 영화인지 여부와 관계없이 그는 당시의 99%의 사람을 식별해 낼 수 있다고 한다. 그리고 점수 매긴 2개의 영화에 관한 정보만으로도, 68%의 이용자들을 재식별할 수 있다고 한다.

이런 주장을 뒷받침하는 사례를 확보하기 위해 Narayanan과 Shmatikov는 Netflix 정보와 이용자들에게 영화 평점을 매길 수 있게 하는 영화 관련 웹사이트인 Internet Movie Database(IMDb)를 통해 확보할 수 있는 정보와 비교하는 방법을 시도했다. Netflix와 달리 IMDb는 Amazon이 이용자가 매긴 책 평점을 공개적으로 게시하는 것처럼 영화 평점을 공개적으로 게시한다.

Narayanan과 Shmatikov는 IMDb 이용자 50명의 영화 평점을 입수했다. 이 작은 표본으로부터 그들은 2명의 이용자들이 거의 통계학적으로 확실하게 Netflix의 데이터베이스에도 포함되어 있음을 파악할 수 있었다. 왜냐하면 두 데이터 집합에 공통적으로 포함된 영화들을 통한 연결을 통해서 Netflix들과 IMDb 사이의 유사한 기록들을 연결함으로써 추론이 가능하기 때문이다. IMDb에 기록된 몇 몇 영화들에 대한 평점 부여를 통해 자신이 보았던 모든 영화들을 의도하지 않게 드러내는 결과가 나타난 것이다. 이는 해당 IMDb 데이터와 Netflix Prize의 데이터가 연결(link)되어 분석될 수 있기 때문에 가능한 것이었다.[60]

(3) 재식별 가능성의 시사점

Governor Weld 사건 등 재식별 맥락에서 반복적으로 언급되는 주요 사건들이 주는 시사점으로 중요한 것은, 첫째로, 비식별화한 데이터에 대한 재식별의 가능성은 상존한다는 것이다. 둘째로, 재식별의 가능성과 별개로, 실제로 이 가능성이 현실화하여 재식별이 이루어지는 경우는 일상적이지 않다는 점이다. 위에서 언급한 세 개의 사건들은 재식별 맥락에서 지속적으로 언급되기는 하지만, 다른 한편으로는 새로운 사건들이 우후죽순처럼 반복적으로 보고되고 있지는 않기 때문이다. 즉 재식별의 개연성에 대해 인정하고 주의할 필요는 있지만, 그렇다고 그 일반성이나 개연성의 정도에 대해서는 과장하여 생각하지 않을 필요가 있다. 셋째로, 데이터의 링크를 가능하게 해주는 준식별자(quasi-identifier)에 대한 관리나 기타 재식별의 가능성을 높여줄 수 있는 사항들에 대해 파악하고 관리하는 것이 중요하다는 점이다. 왜냐하면 위에서 본 주요 사건들에 공통적으로 나타나는 특징은, 재식별을 가능하게 해준 준식별자가 존재한다는 점이기 때문이다.

비식별화 작업을 하는 시점에서 재식별을 불가능하게 하는 것이 설사 가능하다고 가정하더라도, 향후 기술의 발전이나 새로운 식별자(identifier), 준

60) Simon L. Garfinkel, 'De-Identification of Personally Identifiable Information', NISTR 8053, 제 22면 , NIST (2015)

식별자(quasi-identifier)의 존재한다는 점이 발견될 수 있는 가능성이 있다는 것을 고려하면 재식별의 가능성을 영원히 완벽하게 제거하는 것은 불가능하다. 뿐만 아니라 설령 그것이 가능하다고 하더라도, 그렇게 하기 위해서는 데이터의 유용성을 완전히 제거해야 할 가능성이 높기 때문에 그렇게 하는 것이 바람직하지 않을 수도 있다. 그러므로 비식별화의 실질적인 목적은 재식별의 리스크를 완벽하게 제거하는 것이 아니라 이를 최소화하고 또한 지속적으로 관리하는 것에 있다고 보아야 할 것이다.

제4장 비식별화에 관한 국내외 논의 현황

제1절 비식별화에 관한 국내의 논의 현황

1. 개인정보 보호법 제정 이전 – 정보통신망법[61]을 중심으로 한 논의

비식별화에 대한 국내의 논의는 2010년 경 시작되었다.[62] 2010년은 개인정보보호법이 제정되기 이전이어서 개인정보 규제에 정보통신망법 등이 적용되고 있던 당시 법현실이 전제되었고, 비식별개인정보에 대해서는 수집 등에 있어 정보주체의 동의를 받지 않아도 되나, 정보주체가 사후적으로 수집 등의 중단을 요구할 수 있도록 허용하는 옵트아웃(opt-out) 방식의 가이드라인을 제정할 것이 권고되기도 하였다.[63]

2. 개인정보 보호법 제정 이후

(1) 규제완화 담론

2011년 9월 최초의 개인정보 보호에 대한 단일법으로 「개인정보 보호법」

61) 정식명칭은 “정보통신망 이용촉진 및 정보보호 등에 관한 법률”
62) 서울대학교 산학협력단 연구보고서(책임연구자 정상조),「비식별개인정보의 보호 및 활용에 관한 연구」(2010, 방송통신위원회).
63) 같은 보고서

이 제정되었다. 오랜 기간 동안의 심사숙고를 거쳐 도입된 개인정보 보호법은 1980년의 OECD 8원칙을 반영한 개인정보보호법제로서 제정 당시부터 여러 가지의 논의와 논란이 있어왔다. 이를 통해 2005년 소위 지문날인 사건을 통해 헌법재판소가 인정한 개인정보자기결정권이 충실하게 보호될 것이라고 기대되기도 하였지만, 개인정보 보호를 위한 규제가 비현실적이고 과도하여 ICT기술의 발달로 도래한 빅데이터 시대에 역행하는 입법이라는 우려도 있었다.

(2) 공공데이터법의 제공 및 이용활성화에 관한 법률 제정

개방 · 공유 · 소통 · 협력을 추구하는 정부 3.0 작업이 시작되고 이를 뒷받침하기 위하여 2013년 7월에는 「공공데이터법의 제공 및 이용활성화에 관한 법률」이 통과되었다. 동 법은 공공기관이 보유 · 관리하는 데이터의 제공 및 그 이용 활성화를 촉진하여 국민의 공공데이터 이용권을 보장하고 공공데이터의 민간 활용을 통한 삶의 질 향상과 국민경제 발전에 이바지하기 위한 목적으로 제정되었다. 법률을 구체화하기 위해 후속조치로 작성되어 발표된 「공공정보 개방 · 공유에 따른 개인정보보호지침」은 공공정보 처리 및 분석 시 개인정보를 비식별화하여 보호조치를 하도록 규정하였다. 동시에 비식별화 처리원칙, 비식별화 단계별 조치사항, 비식별화 처리기법 등을 적시하였다. 특히 동 지침은 비식별화를 개인정보의 일부 또는 전부를 삭제하거나 다른 정보로 대체함으로서 다른 정보와 쉽게 결합하여서도 특정 개인을 식별하기 어렵도록 하는 일련의 조치라고 정의한 후, 비식별화 기법으로 가명처리, 총계처리 또는 평균값 대체, 데이터값 삭제, 범주화, 데이터마스킹 기법 등을 소개하였다. 이는 최초로 공식적인 법령과 정부 지침을 통하여 비식별화에 대한 구체적인 방법론을 제시하고 적용하였다는 점에서 의미가 있다.

(3) 방송통신위원회 가이드라인

ICT 기술혁신은 점점 더 가속화되었고 개인정보 보호에 대한 규제가 과도하고 비현실적이어서 빅데이터 산업이 발전할 수 없다는 비판도 늘어났다. 이

에 방송통신위원회는 개인정보를 충실하게 보호하는 동시에 빅데이터 산업을 발전시킬 수 있는 절충점을 모색하기 위하여 2013년 12월 「빅데이터 개인정보보호 가이드라인(안)」을 발표하였다. 이 가이드라인(안)에는, 공개된 개인정보는 정보주체의 동의 없이 이용할 수 있으며, 비식별화를 거친 개인정보는 정보주체의 동의 없이 이용할 수 있다는 내용이 포함되어 있었다.[64] 곧이어 미래창조과학부와 한국정보화진흥원은 2014년 5월 1일에 「빅데이터 활용을 위한 개인정보 비식별화 사례집」[65] 을 발표하였고, 행정자치부와 한국정보화진흥원은 2014년 12월에 「개인정보 비식별화에 대한 적정성 자율평가 안내서」[66]를 발표하였다.

하지만 2014년 7월 30일 개인정보보호위원회는 가이드라인이 개인정보 보

64) 가이드라인(수정안)은 [개인정보의 수집 및 이용] “정보주체 및 정당한 권한이 있는 자에 의해 제한 없이 일반 공중에게 공개”된 개인정보를 ‘공개된 개인정보’(제2조 제1호), “정보통신서비스와 관련하여 이용자가 해당 서비스를 이용하는 과정에서 자동으로 발생되는 인터넷 접속 정보파일, 거래기록 등”의 개인정보를 ‘이용내역정보’(제2조 제2호)로 각각 규정한 후, 정보통신서비스 제공자는 정보주체 및 정당한 권한이 있는 자가 공개대상을 제한하거나 공개 목적을 설정한 경우가 아닌 한 ‘공개된 개인정보’를 정보주체의 동의 없이 수집할 수 있고(제3조), 정보통신서비스와 관련한 계약 체결과 이행을 위하여 필요한 ‘이용내역정보’ 역시 정보주체의 동의 없이 수집하여 조합 또는 분석 처리할 수 있다고 규정(제4조)하고 있다. 또한 정보주체가 거부의사를 표시하지 아니하는 한 정보통신서비스 제공자의 정당한 이익과 상당한 관련이 있고 합리적인 범위를 초과하지 아니하는 경우, 별도로 정보주체의 동의를 얻지 아니하고 정보 조합 · 분석 · 처리시스템을 통해 공개된 개인정보 및 이용내역정보 등을 활용하여 새로운 개인정보를 생성할 수 있다고 규정(제5조 제1항)하고 있다. 또한 [개인정보의 제3자 제공] 가이드라인(수정안)은 제11조 제1항에서 정보통신서비스 제공자는 공개된 개인정보, 이용내역정보, 생성된 개인정보를 이를 제공받는 자, 제공받는 자의 이용목적, 제공하는 항목, 보유 및 이용기간을 정보주체에게 알리고 동의를 받으면 이를 제3자에게 제공할 수 있다고 규정하면서도, 단서에서 ‘공개된 개인정보’는 정보주체의 동의가 없어도 제3자에게 제공할 수 있다고 규정하고 있다. (개인정보보호위원회 2014의결 제16호)

65) 빅데이터 활용단계별 비식별화 조치사항, 비식별화 처리기법, 그리고 각 기법에 대한 상세규칙 등을 제시하고 있으며(표 3) 빅데이터 활용사례별 비식별화 처리 실례를 제공하고 있다. 의료분야 실무 적용사례에서는 성명, 주민등록번호, 연령, 주소, 전화번호, 이메일 주소, 외국인등록번호, 여권번호, 등록번호, 건강보험증번호, 은행계좌번호, 자격/면허번호, 차량번호, 바이오정보(지문, 얼굴, 홍채, 정맥, 음성, 필적 등), 유전자정보, 홈페이지 회원 ID, 사번, 비밀번호, 요양기관기호, 소득, 민감상병, 아이디, 진단명, 약처방날짜, 진단검사날짜, 검사수행날짜 등과 같은 항목에 대한 비식별조치 적용 내용을 제시하고 있다.(정영철, 의료분야 빅데이터 활용을 위한 개인정보 비식별화 규정 현황과 과제, 2015.9.)

66) 개인정보 비식별화에 대한 적정성 평가를 위한 절차와 세부평가방법, 재식별 위험관리방안 등을 주 내용으로 구성되어 있다. 이 중 ①성명(한자, 영문성명 포함), ②주소, ③고유식별정보(주민등록번호, 여권번호, 외국인등록번호, 운전면허번호), ④연월일(생일, 기념일, 사망일, 자격증 취득일등), ⑤전화번호(휴대폰번호, 집전화, 회사전화), ⑥팩스번호, ⑦전자메일, ⑧의료기록번호, ⑨건강보험번호, 복지수급자번호, ⑩계좌번호, 카드번호, ⑪각종 자격증 및 면허번호, ⑫자동차번호,⑬각종 기기의 등록번호 및 일련번호, ⑭IP주소,Mac주소, ⑮홈페이지 URL, ⑯사진(정지, 동영상,CCTV 영상 등), ⑰신체식별정보(지문, 음성, 홍채 등), ⑱기타 유일 식별번호(군번, 사업자등록번호, 식별코드 등) 등 18개 식별자를 예시로 들고 있는바, 이는 HIPAA 프라이버시 규칙에서 제시한 18개 식별자를 고려하여 우리나라 상황에 맞게 수정하여 제시하고 있다. 또한 비식별화를 위한 단계별 조치사항으로는 관계법령 등 검토(제1단계)하여 근거가 없을 경우에는 표 3의 기법을 활용하여 개인식별요소를 삭제(제2단계)하며, 이에 대해 통계 및 수학 등 관련분야 전문가들로 하여금 개인식별 가능성을 검토(제3단계) 한 후 데이터를 활용토록 하고 있다. 이후 재식별화 여부에 대한 지속적인 정기점검/모니터링(제4단계) 단계를 두고 있다. (정영철, 의료분야 빅데이터 활용을 위한 개인정보 비식별화 규정 현황과 과제, 2015.9.)

호법과 정보통신망법의 규정과 입법취지에 부합하지 않는 내용을 일부 포함하고 있음을 지적하여 이를 재검토할 것을 방송통신위원회 위원장에게 권고했고, 방송통신위원회는 이를 수용하여 2014년 12월 23일 행정규칙인 「빅데이터 개인정보보호 가이드라인」을 발표했다. 여기에는 공개된 개인정보를 정보주체의 동의 없이 이용할 수 있다는 내용은 삭제되었으며, 비식별화를 거치면 정보주체의 동의 없이 이용할 수 있다는 내용은 남게 되었다.

3. 개인정보 보호법 등 개정 논의

변화하는 현실에서 더욱 개인정보를 잘 보호하고 동시에 그 이용을 도모하여 혁신을 창출하기 위해서는 개인정보보호 법제를 개선해야 한다는 논의가 본격적으로 시작되었다. 대표적인 논의로는 2015년 2월 5일 강은희 의원 등 21인이 발의한 개인정보 보호법 전부개정법률안, 2015년 3월 5일 부좌현 의원 등 10인이 발의한 개인정보 보호법과 정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안, 2015년 3월 9일 강길부 의원 등 11인이 발의한 정보통신망 이용촉진 및 정보보호 등에 관한 법률 개정안이 있다.

2015.2.5.개인정보 보호법 전부개정법률안 – 강은희의원 등 21인
(1) 통계·연구, 시장조사, 마케팅 등의 목적을 위한 경우에는 개인정보 비식별화 조치를 통해 정보주체의 동의 없이 이를 처리할 수 있도록 하고 개인정보 파기 요건 및 이에 관한 예외 사유를 규정하여, 개인정보 처리 과정에서의 유연성을 부여함(안 제39조 및 제40조). (2) 안전성 확보에 필요한 보호조치를 하지 아니하여 비식별화 처리한 개인정보를 분실·도난·유출·변조 또는 훼손당한 자는 2년 이하의 징역 또는 1천만원 이하의 벌금에 처함(안 제99조 제4항 제7호)

2015.3.5.개인정보 보호법 일부개정법률안 – 부좌현의원 등 10인
(1) 개인정보처리자가 통계작성, 학술연구, 실태조사를 목적으로 개인정보를 처리하거나 이미 공개된 정보를 재가공하는 과정에서 개인정보가 유출되지 아니하도록 개인정보처리자에게 개인정보 비식별화 조치 의무를 부여함(안 제22조의2 제1항 신설). (2) 개인정보처리자가 개인정보를 비식별화하여 처리하는 경우에는 정보주체의

동의를 받지 아니할 수 있도록 함(안 제22조의2제2항 신설).
(3) 개인정보처리자가 개인정보를 비식별화하여 처리하거나 비식별화된 개인정보를 처리하는 때에는 개인정보가 생성되지 않도록 하고, 이 과정에서 안전성 확보에 필요한 기술적·관리적 및 물리적 조치를 하도록 함(안 제22조의2제3항 및 제4항 신설).

2015.3.5.정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안 – 부좌현의원 등 10인
(1) 개인정보를 안전하게 보호하기 위하여 정보통신서비스 제공자의 비식별 조치에 대해서는 방송통신위원회가 인증할 수 있도록 하고, 비식별 개인정보가 다른 정보와 결합하여 특정 개인이 식별되지 않도록 비식별 조치의 기술적·관리적 조치기준을 마련하려는 것임(안제2조제14호·제15호 및 제28조 제1항 제6호 신설, 제47조의3)

2015.3.9.정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안 – 강길부의원 등 11인
(1) “비식별화 방법”을 이용해 빅데이터의 활용을 증진하면서도 개인정보를 안전하게 보호하기 위하여, 정보통신서비스 제공자가 비식별화된 개인정보를 이용하는 과정에서 개인정보가 발생하는 경우에는 이를 파기하거나 다시 비식별화하는 의무를 부과함(안 제24조의3).
(2) 비식별화의 기술적 기준 및 개인정보의 파기 및 추가적인 비식별화에 관하여 필요한 사항은 대통령령으로 정함(안 동조 제3항)
(3) 비식별화 개인정보를 이용하는 과정에서 개인정보가 생성되었음에도 불구하고 이를 지체 없이 파기하거나 비식별화하지 아니한 자에게 3천만 원 이하의 과태료를 부과함 (안 제76조제1항 제2호의4).

이후 2015년 6월 3일 금융위원회는 빅데이터 산업을 활성화하여 핀테크를 통한 혁신을 창출을 도모하기 위하여 신용정보법 시행령을 개정하여 비식별 정보를 개인신용정보에서 제외하겠다고 밝혔다. 곧이어 2015년 6월 10일 미래창조과학부와 한국정보화진흥원은 「빅데이터 활용을 위한 개인정보 비식별화 기술 활용안내서 Ver1.0」[67]을 발표하였다. 하지만 금융위원회는 자체검토 후 신용정보법 자체를 개정하는 방향으로 입장을 선회하였다.

67) 빅데이터 활용을 위해 각 부문별 개인식별 정보항목과 18가지 비식별화 기술 활용방법, 그리고 부문별 활용사례를 안내하고 있다. 이에 비식별화 적용대상 정보를 해당정보 자체로 개인식별 가능한 정보와 다른 정보와 쉽게 결합하여 개인식별 가능한 정보로 나누어 예시를 들고 있으며(표 4), 의료분야 실무 적용사례는 앞의 『빅데이터 활용을 위한 개인정보 비식별화 사례집』 예시와 거의 동일한 내용을 담고 있다. (정영철, 의료분야 빅데이터 활용을 위한 개인정보 비식별화 규정 현황과 과제, 2015.9.)

제2절 해외 논의 현황

1. 미국

(1) 비식별화 처리에 대한 규제 및 논의 동향 개괄

미국에는 개인정보의 보호를 일반적으로 규율하는 연방법은 없다. 개인정보의 보호에 대한 규제는 사안별 또는 분야별로 이루어지는 형태를 가지고 있다. 이와 같이 구체적이고 개별적인 영역에서 각각 별도의 법령이나 판례법을 통해 개인정보를 보호하고 있고, 특히 개인의 신용정보나 의료정보과 같이 민감도가 상대적으로 높은 영역들만을 규율하는 법령들이 연방이나 주(state) 차원에서 제정되고 있다.[68] 예를 들어, 금융 분야의 경우에는 '공정신용평가법(Fair Credit Reporting Act) 및 그램-리치-블라일리법(Gramm-Leach-Bliley Act) 등을 통해 보호를 받고, 의료 분야의 경우에는 '건강정보의 이동과 책임에 관한 법률(Health Information Portability and Accountability Act, 이하 'HIPAA')을 통해 규율하고 있다.

이런 개별적 규제 체제 하에서 개인정보의 비식별화에 대한 규제 또한 분야에 특정되는 개별적인 규제 형태를 가지고 있다. 교육부(Department of Education)가 '비식별화된 학생기록(de-identified student records)'에 대한 관계 법률의 적용가능성을 판단한 경우, HIPAA 프라이버시 규칙에 의해 개인정보의 비식별화에 대해서 규제를 하는 경우 등이 있다. 그래서 개별적인 분야를 규제 대상으로 하는 법령들이기 때문에 각각의 법령들에서 채택한 비식별화에 대한 접근 방식이 다른 분야들에도 일반적으로 통용될 것이라고 결론을 내리는 것에는 무리가 있다. 특히 의료개인정보를 대상으로 만들어진 HIPAA 프라이버시 규칙을 다른 영역에서의 개인정보의 비식별화에 대한 규제에도 적용할 수 있는지에 대해서는 의료개인정보의 특수한 맥락 등을 고려해서 신중하게 접근할 필요가 있다.

---

68) 개인정보보호위원회, '해외 개인정보보호 집행체계 및 개인정보보호 주요 동향조사', 16면 (2012)

미국에서는 일반적으로 보호의 대상이 되는 개인정보를 '개인적으로 식별 가능한 정보(personally identifiable information, 이하 'PII')'로 규정한다. 이 PII 개념을 규정하는 개인정보의 보호에 대한 일반법은 없지만 개인정보의 보호의 규제에 대한 개별 주법이나 분야별 규제들이 대체적으로 PII를 보호대상이 되는 개인정보로 간주하고 있다. PII는 흔히 다음과 같이 정의된다: 어떤 기관이 보유하고 있는 개인에 대한 정보로서 다음과 같은 정보들을 포함한다: (1) 개인의 신분(이름, 사회보장번호, 생년월일, 태어난 장소, 어머니의 성(maiden name), 또는 생체적(biometric) 기록들을 포함)을 구별하거나 추적할 수 있는 용도로 이용될 수 있는 정보; 그리고 (2) 개인과 연결되거나 연결될 수 있는 기타 정보(의료, 교육, 그리고 고용에 대한 정보를 포함).[69] 이와 같은 정의는 NIST의 설명을 따른 것이다. 개별법에 따라서는, PII 개념을 정의하는 방식을 이용하기도 하고 PII에 속하는 정보를 구체적으로 나열하는 방식을 채택하기도 한다.[70]

이와 같은 개념 규정에서 보면 PII의 범위가 광범위하게 정의될 수 있음을 확인할 수 있다. 특히 이 개념 규정의 (2)의 설명 부분에서는 개인과의 연결가능성(linkability)이 있으면 그런 정보도 PII에 해당한다고 정의하고 있다. 사회적 관점에서 보면 개인과의 연결가능성이 있는 정보는 사회의 변화와 기술의 발전에 따라 그 범위가 계속해서 확장되는 경향이 있기 때문에 개인에 대한 정보를 포함한 데이터를 수집하거나 처리하려는 기업의 입장에서는 개인정보의 보호의 규제 대상이 되는 PII가 해당 데이터에 포함될 확률이 매우 높아진다. 이에 따라 기업들은 데이터의 수집과 처리에 있어서 PII로서의 규제를 받게 되어서 데이터의 수집과 처리에 있어 제한과 추가적인 비용들을 부담해야 하는 위치에 있게 된다. 그래서 연결가능성을 사전적으로 막는 방식인 비식별화 처리를 통해 처음부터 처리하려는 데이터에 PII가 포함되지 않게 하려는 접근법이 그 유용성을 가지게 된다.

---

69) NIST Special Publication 800-122.

70) Schwartz, P. and Solove, D., "The PII Problem: Privacy and a New Concept of Personally Identifiable Information", *New York University Law Review*, 2011, 86, pp. 1814-1894 참조.

특히 보호의 대상이 되는 개인정보의 개념을 PII라고 규정하는 미국의 맥락에서 식별화가 될 수 없는 정보는 PII 자체에 해당하지 않는 것이 명백하기 때문에 비식별화 방식의 효용성이 더욱 확연해지는 특징이 있다. 이런 배경에서 비식별화의 방식들에 대한 다양한 연구들이 수행되었고 수학적 모형을 적용해서 통계학적으로 비식별화 문제에 접근하려는 시도들도 나타나게 되었다. 그래서 다양한 통계학적 접근법들이 등장하게 되고 실제로 이런 통계적 적용의 결과물을 비식별화에 대한 규제에 일부 영역에서나마 반영이 된 사례가 등장하기도 했다.[71]

하지만 이런 다양한 통계학적 접근법들의 어느 방식도 비식별화의 가장 큰 쟁점인 재식별의 위험성을 완전하게 해결해주는 것은 아니다. 특히 기술 발전의 속도가 더욱 빨라지면서 과거에는 비식별화되었다고 여겨지는 데이터들이 예상하지 못한 새로운 기술에 의해 재식별되는 위험성이 더욱 커지는 상황에서 어떤 하나의 접근법으로 비식별화 문제를 완벽하게 해결할 수 있다는 인식은 더욱 설득력을 잃고 있다. 동시에 빅데이터 산업의 활성화 정책과 맞물려서 정부기관들이 데이터들을 공개해야 할 부담이 늘어나고 있는 상황에서 개인을 식별할 수 없게 하는 비식별화 조치의 필요성은 더욱 커지고 있다.

이런 분위기에서 2015년 4월에 미국 상무부 산하의 표준화 기구인 NIST에서 비식별화에 대한 최근 20년간의 논의를 정리하는 보고서를 발간하였다.[72] 이 초안을 공개한 후 다양한 개인들 또는 기관들로부터 이 보고서에 대한 의견들을 수렴하였다. 의견들의 수렴과정이 끝난 후 2015년 10월에 이 보고서의 최종안이 발표되었다. 그래서 이 보고서의 내용을 통해 비식별화에 대한 미국에서의 최근 논의의 흐름을 어느 정도 파악할 수 있다.

이 보고서는 최근 20년간의 비식별화의 실제 연구 성과들과 비식별화 기법들을 가치중립적인 관점에서 압축적으로 설명하고 있다.[73] 이 보고서의 전

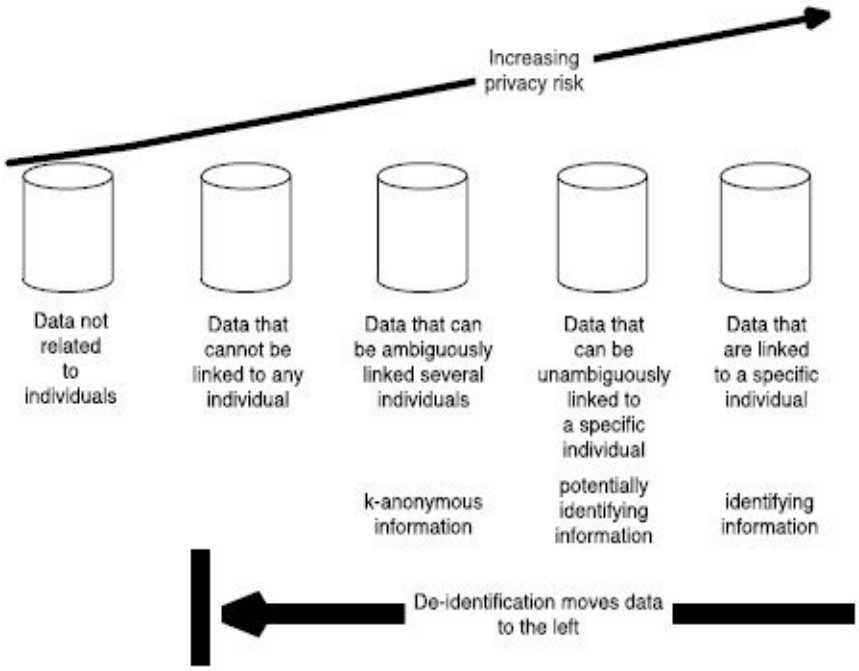
71) 이 구체적인 예는 다음 단락인 'HIPAA 프라이버시 규칙' 부분에서 자세하게 논의한다.
72) Simon L. Garfinkel, "De-Identification of Personally Identifiable Information", NISTR 8053, 3면, NIST (2015)
73) 전게서 60, 제 iii면. 이 보고서의 전체적인 의도를 포괄적으로 설명하고 있는 초록(abstract) 부분에서 이런

체적인 흐름을 살펴보면 기본적으로 모든 데이터 비식별화 기술들은 재식별의 위험성을 가지고 있다는 사실을 기본 전제로 하고 있다. 단지 주어진 맥락(context)에서 어떤 비식별처리 방식이 적용되었는지에 따라 그 위험성의 정도가 차이가 난다고 설명하고 있다. 이런 맥락이라는 것은 다양한 요소들을 포함하지만 기본적으로 해당 데이터가 특정인과 연결되는 정도는 가장 기본적인 요소들 중의 하나라고 할 수 있다. 이런 관점에서 그 데이터가 특정인과 연결되었는지, 특정인과 연결될 잠재적 가능성이 있는지, 특정인이 아니지만 어느 정도의 사람들과 연결될 가능성이 있는지 등에 따라 해당 정보가 식별되어 프라이버시를 침해할 위험성은 달라진다. 특정인과의 연결성을 기준으로 프라이버시 침해로 표현하면 아래의 그림과 같이 표현할 수 있다.

Abstract

De-identification removes identifying information from a dataset so that individual data cannot be linked with specific individuals. De-identification can reduce the privacy risk associated with collecting, processing, archiving, distributing or publishing information. De-identification thus attempts to balance the contradictory goals of using and sharing personal information while protecting privacy. Several U.S laws, regulations and policies specify that data should be de-identified prior to sharing. In recent years researchers have shown that some de-identified data can sometimes be re-identified. Many different kinds of information can be de-identified, including structured information, free format text, multimedia, and medical imagery. This document summarizes roughly two decades of de-identification research, discusses current practices, and presents opportunities for future research.

사실을 밝히고 있다:



출처: Simon L. Garfinkel, "De-Identification of Personally Identifiable Information", 표 1(2015)

하지만 기술 변화를 고려하면 특정 맥락에서 어떤 비식별처리 방식이 더 나은 것인지에 관해 일반화하여 언급할 수는 없다. 다만 외부 정보와의 연결을 통해 식별가능성을 높이는 식별자 또는 준식별자(quasi-identifier)[74]를 데이터베이스에서 존재하지 않도록 하면 재식별의 위험성을 크게 줄일 수 있을 것이다. 그런데 준식별자의 범위 또한 관련 기술의 발전에 따라 확대될 가능성이 높기 때문에 데이터를 관리하는 기업은 준식별자를 찾고 제거하는 기술적 방식뿐만 아니라 여러 가지의 사후적인 보안통제책을 적용하는 관리적 방식도 같이 적용할 필요가 있다. 이런 관리적 방식에는 '데이터 이용에 관한 합의서(data use agreement)'와 같은 계약 형태가 있고 이 계약에 데이터의 이용자가 재식별을 해서는 안 된다는 의무 조항이 포함될 수 있다.

NIST 보고서 초안에 대한 다양한 의견들이 반영되어 확정된 보고서의 최종

74) 준식별자는 간접적인 식별자라는 명칭으로도 이용된다. 그래서 준식별자란 그 자체로는 특정인을 식별할 수 없지만 다른 정보와 연결되어서는 데이터 주체를 식별할 수 있는 정보를 의미한다. NIST 전게서, 제19면 참조

안은 일반적으로 다음과 같은 항목들을 단계적으로 설명하고 있다:[75]

- 비식별화, 재식별화, 그리고 데이터공유모형(data sharing model)의 개념들에 대한 개괄
- 데이터를 비식별화하는 다양한 접근법들(예를 들어, 제거, 마스킹, 또는 이름이나 전화번호와 같은 특정 범주들의 대체(altering))에 대한 설명
- 이미지나 유전자(genomic) 정보와 같은 비도표화(non-tabular) 정보의 비식별화에 대한 도전

이런 기술적인 사항들에 대한 가치중립적인 설명을 통해 비식별화는 완벽하지는 않지만 데이터주체의 프라이버시를 보호할 수 있는 중요한 기술적 통제수단이라는 결론을 내리고 있다. 이 보고서는 이와 같이 비식별화 기술에 대한 일반적인 효용적 가치에 대해 인정을 할 뿐 어떤 특정 비식별화 기술이 다른 기술보다 더 유용하다는 등의 상호비교를 하지는 않는다. 즉 특정 알고리즘이나 어떤 비식별화 기법이 적절하다는 것에 대한 특별한 가치적 논평이나 설명을 하지 않는다.[76] 비식별화 기법들은 데이터의 유용성을 보존하면서 개인을 식별하는 정보를 제거하기 위한 목적을 위해 존재하는 방식들일 뿐이기 때문에 연구의 목적에 따라 판단되는 데이터의 유용성 정도가 달라지는 상황에서 어떤 일의적인 기준에 따라 데이터 기법을 상호 비교하는 것은 무의미하다는 입장이다. 그래서 비식별화는 어떤 단일한 기법이 아니라 개별 연구의 목적이나 상황에 적합한 다양한 접근법들, 도구들, 그리고 알고리즘들의 결합을 통해 달성된다.

이와 같이 개인적으로 식별가능한 정보를 효과적으로 비식별화할 수 있는 것이 데이터의 효용성과 개인정보의 보호를 동시에 만족할 수 있는 핵심적인 요소라는 것을 이 보고서는 인지하고 있다. 하지만 이와 같이 비식별화 방식의 가치를 대체적으로 인정하면서 비식별화에서 항상 문제가 되는 재식

---

75) https://www.huntonprivacyblog.com/2015/10/29/nist-releases-final-report-on-de-identification-of-personal-information/

76) http://www.bna.com/nist-releases-data-n57982063147/

별 가능성의 문제에 대해서도 동시에 주목하고 있다. 그런 면에서 이 보고서의 전체적인 내용은 재식별의 위험성 최소화의 관점에서 비식별화 기법들의 유용성을 확인하는 것이라 볼 수도 있다. 다양한 비식별화 기법들이 등장하고 상황에 따라 어떤 특정한 기법만을 사용하는 것이 아니라 다양한 기법들을 결합해서 사용하는 것은 결국 개별 맥락에 따라 등장하는 재식별의 위험성의 정도와 형태 등이 달라질 수 있기 때문이다.

재식별이 될 경우에 발생하는 구체적인 문제는 크게 두 가지 형태로 나타날 수 있다. 하나는 사적인 정보의 공개와 명성(reputation)의 손실이다.[77] 사적인 정보의 공개는 개인정보가 공개된 당해 개인에게 영향을 미치고 명성의 손실은 일반적으로 비식별화를 이행했던 해당 조직에게 영향을 미친다. 개인의 정보를 보관하고 있는 기관이나 기업이 개인정보의 의도하지 않은 누출로 인해 개인정보관리자로서의 신뢰감을 상실한다는 측면에서 그 기관이나 기업에 대한 손해가 발생하는 것이다.

NIST 보고서는 재식별의 위험성에 대한 인식을 하면서 동시에 재식별의 위험성이 과장되어서는 안 된다는 사실도 강조하고 있다.[78] 또한 재식별의 대상이 공인(public figure)인 경우가 많은 상황에서 이와 같이 공인을 대상으로 재식별의 위험성을 측정한 결과를 가지고 일반인에 그대로 적용하는 것의 문제점을 지적하고 있다. 일반인에 비해서 공인에 대해서는 여러 다른 정보들을 낮은 비용으로 획득하기가 상대적으로 용이하기 때문에 공인을 대상으로 비식별의 위험성을 측정한 수치에 따른 재식별의 위험성을 그대로 일반화할 수가 없다는 것이다.

재식별의 위험성 측정을 위한 기준과 관련해서는 연구자들 사이에서 다양한 의견들이 개진되고 있다. 어떤 연구자들은 재식별의 위험성이란 재식별될 수 있는 비식별화된 기록의 비율을 의미한다고 주장한다.[79] 이런 경우에는 재식별의 위험성은 재식별을 위한 공격의 이행을 통해 직접적으로 측정될 수

77) NIST 전게서, 제 37면
78) NIST 전게서, 제 29면
79) NIST 전게서, 제 38면

있다. 다른 연구자들은 재식별의 위험성을 미래에 재식별될 수 있는 확률이라고 정의한다. 이와 같이 현재의 상황 이외에도 미래의 경우까지 발생될 요소들을 재식별의 위험성 판단에 포함할 경우에 궁극적으로는 위험성을 측정하는 것은 불가능하다. 왜냐하면 위험성은 현재에는 실행할 수 없지만 미래에는 실행될 가능성이 있는 데이터의 유용성을 기준으로 하기 때문이다. 이런 미래적인 요소가 포함되면 어느 정도까지 예측되는 수준을 고려요소들로 정할 수 있을 것인지에 대해 정하는 것은 실질적으로 불가능하다. 그래서 재식별의 위험성을 논의한다고 할 때는 우선적으로 어떤 정의를 전제로 한 논의인지를 확인할 필요가 있다. 즉 이 보고서는 재식별의 위험성은 비식별화의 논의에서 이론적으로 핵심적인 영역임은 분명하지만 이와 관련된 논의들은 아직 시작 단계이기 때문에 좀 더 신중한 접근이 필요하다는 의견을 담고 있다.

이런 논의를 기반으로 이 보고서는 기본적으로 비식별화 기법들이 점차적으로 개인정보를 포함한 데이터의 사용, 공유, 그리고 처리에 있어서 중요해지는 영역이라는 사실을 강조한다.[80] 하지만 약 20년의 역사를 가진 비식별화와 관련된 연구들이 있었음에도 불구하고, 비식별화에 내포된 과학적 이론틀은 아직 정교하게 구축되지 못한 상황이라는 점에도 주목한다. 후술하게 될 HIPAA 프라이버시 규칙에서 규정한 세이프하버 방식에 의한 비식별화 기준과 같이 현재의 관련 기술과 절차들에 체계화된 이론적 근거가 있는 것은 아니다. 즉, 비식별화 절차와 관련해서 일반적으로 받아들여지는 공통적인 기준은 아직 형성되지 않았다. 이런 이론적 근거들에 뒷받침되는 비식별화와 관련된 위험성들을 측정하는 기준들을 찾는 것이 이 분야를 연구하는 전문가들의 과제가 될 것이라고 보고서는 결론을 내리고 있다.[81]

다른 한편, NIST 자료 이외에도 개별 분야의 여러 법령 등을 통해 비식별화에 대해서 규정된 경우들이 존재한다. 우선 교육부(Department of Education)가 “가족과 교육 기록에 대한 프라이버시법(Family and Educational Records Privacy Act)"는 비식별화된 학생기록에는 적용되지

80) NIST 전게서, 제 39면
81) NIST 전게서, 제 39면

않는다고 결정했다. 또한 의료 분야에서는 HIPAA 프라이버시 규칙(HIPAA Privacy Rule)이 비식별화된 의료개인정보의 제한없는 이용을 허용했으며, “경제적 그리고 임상적 보건에 대한 건강정보기술법(Health Information Technology for Economic and Clinical Health Act)” 에서의 보안이나 프라이버시 조건들은 비식별화된 건강정보에는 적용되지 않는다고 명시하고 있다. 그리고 식품매개성(foodborne) 질병에 대한 관리체계에서는 비식별화된 관리데이터에 대해서는 대중의 접근권이 인정되어야 한다고 규정한다. 그리고 연방항공청(Federal Aviation Administration)에 제출된 자발적인 안전보고서에 포함된 데이터가 비식별화가 될 경우에는 대중에 공개가 될 경우 별도의 보호를 받지 못한다.

(2) HIPAA 프라이버시 규칙(HIPAA Privacy Rule)

1) 전체 내용의 개관

HIPAA는 의료 분야의 개인정보 보호에 대해서 규정하는 연방법으로서 HIPAA 프라이버시 규칙은 HIPAA의 내용을 보충하는 행정규칙의 성격을 가지고 있다. HIPAA 프라이버시 규칙과 이 규칙의 내용을 보충적으로 설명해주는 안내서들에 비식별화의 기준들이 나타나있다. 주로 보건부(Health and Human Service, 이하 “HHS”)의 인권국(Office of Human Rights, 이하 “OCR”)이 이런 안내서들을 발행한다.

이 프라이버시 규칙에 대해서는 다른 많은 나라의 규제 기관들이 참고로 하고 있고. 영국의 개인정보보호기관에서 발행한 행동강령에 나타나는 내용과도 상당한 수준의 관련성을 보이고 있다.[82] 의료 분야의 경우에는 환자 개인에 대한 민감한 정보를 포함한 데이터를 학술적 또는 상업적으로 활용할 가치가 높고 그에 따라 의료개인데이터의 유통성이 다른 분야에서 보다 높다. 그래서 다른 분야와는 달리 의료 분야에서는 개인정보의 보호와 데이터 활용의 효용성 사이의 균형점을 비식별화 방식을 통해 해결한다는 것을 규

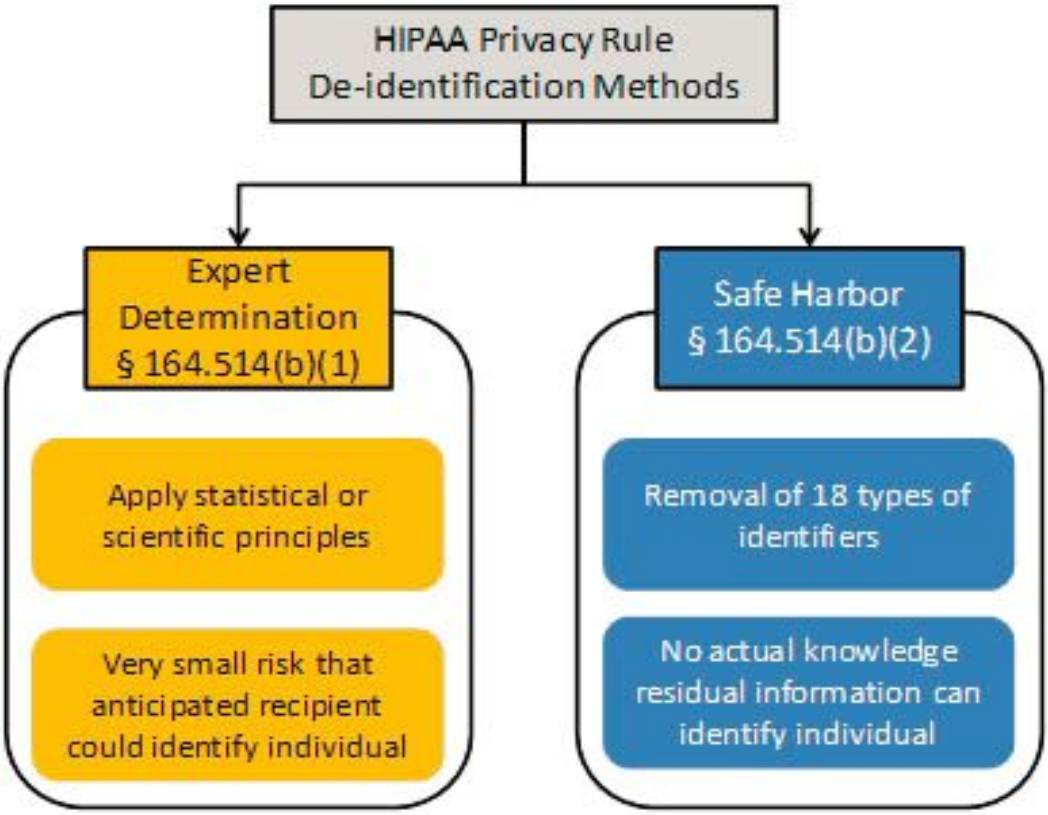
82) Committee on Strategies for Responsible Sharing of Clinical Trial Data et al., 'Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks'(pdf), 208면 (2012)

칙 형태로 명확히 한 것으로 해석할 수 있다.

HIPAA 프라이버시 규칙의 164.514항에 따르면, “개인을 식별하지 않고 개인을 식별할 수 있게 한다는 합리적(reasonable) 근거가 없는 건강정보는 개인적으로 식별가능한 건강정보가 아니다” 라고 규정하고 있다. 규제의 대상이 되는 건강정보에 대한 개념 정의는 위에서 설명한 PII의 접근법을 그대로 따르고 있는 것으로 보인다. HIPAA 프라이버시 규칙의 164.514(b)항에서는 규제의 대상이 되는 주체인 건강정보를 데이터 형식으로 전송하게 되는 의료정보관리기관(covered entity) 또는 협력관계에 있는 사업조직(business associate)이 비식별화 기준을 만족하기 위해서 준수해야하는 구체적인 이행절차들에 대해 규정하고 있다. 특히 HIPAA 프라이버시 규칙은 건강정보가 비식별화된 것으로 인정될 수 있는 두 가지의 접근법에 대해 규정하고 있다.

HIPAA 프라이버시 규칙의 수범자인 의료정보관리기관(covered entity)은 이 두 가지 방식들 중에 원하는 방식을 선택해서 의료개인정보의 비식별화를 할 수 있다. 하나는 절차적 방식에 의한 접근법으로서 관련전문가의 개별적인 판단에 의한 방식이고, 다른 하나는 내용적 방식에 의한 접근법으로서 특정한 개인 식별자들(identifiers) 또는 준식별자들(quasi-identifiers)이 데이터에서 제거되면 비식별화가 된 것으로 간주하는 방식이다. 후자의 방식은 세이프하버(safe harbor) 방식으로도 불린다. 이 두 방식의 얼개는 아래의 <그림 1>에서 확인할 수 있다.

<그림 1> HIPAA 프라이버시 규칙에서의 두 가지 비식별화 접근 방식



출처: HHS 홈페이지
http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html

절차적 방식("Expert Determination" method)에 의한 접근법에 따르면 우선적으로 의료정보관리기관이 대상이 되는 개인데이터가 HIPAA 프라이버시 규칙의 기준에 따라 식별될 수 없는 정보인지를 판단할 전문가를 선임한다. 이 전문가가 되기 위한 특정한 자격 조건이 있는 것은 아닌데, 다만 일반적으로 비식별화 방식에 적용되는 통계학적 그리고 과학적 이론에 대한 지식과 경험을 가지고 있어야 할 것이 요구된다. 이렇게 선임된 전문가는 해당 개인데이터가 식별화될 위험성이 매우 작다(very small)[83]고 판단하면

83) Section 164.514(a) of the HIPAA Privacy Rule. 밑줄과 강조는 추가됨.

(b) Implementation specifications: requirements for de-identification of protected health information. A covered entity may determine that health information is not individually identifiable health information only if:
(1) A person with appropriate knowledge of and experience with generally

이 개인데이터는 식별할 수 없는 데이터로 간주되어 HIPAA의 규제 대상이 되지 않는다. 이 때 전문가는 판단을 위해 적용한 방법들과 판단결과를 문서화해야 한다.

절차적 방식은 개별적 사안별로 전문가를 선임해서 판단해야 하기 때문에 그 만큼 시간과 비용이 소요된다는 단점이 있다. 게다가 의료정보관리기관이 이런 전문가의 판단 결과에 불복해서 계속적으로 다른 전문가들을 선임하는 현상이 생길 경우 그 만큼 사회적 비용도 높아지게 된다는 문제점도 발생한다. 하지만 기술의 발달에 따라 식별가능성을 판단할 수 있는 기준과 범위도 변화한다는 사실을 고려하면 개별적인 사안 별로 그 사안에 적합한 해결책을 찾는 절차적 방식이 적어도 개념적으로는 합리적인 방식이라고 볼 수 있다.

이런 절차적 방식 이외에도 내용적 방식(“Safe Harbor” method)에 의한 접근법도 있다. HIPAA 프라이버시 규칙에 열거된 18가지 유형의 데이터들이 해당 데이터에서 제거가 되고 이 18가지 유형의 데이터가 제거되고 남은 데이터들이 다른 정보와 결합해서 개인을 식별할 가능성이 있다는 사실에 대한 인식을 가지지 않는다면 해당 데이터는 HIPAA의 제한을 받지 않는 상태로 수집과 처리가 가능해진다. 이 18가지 유형의 데이터의 개별적 사항들은 <표 1>에서 확인할 수 있다.[84)]

---

accepted statistical and scientific principles and methods for rendering information not individually identifiable:
(i) Applying such principles and methods, determines that the risk is **very small** that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
(ii) Documents the methods and results of the analysis that justify such determination; or

84) Section 164.514(a) (2)(i) of the HIPAA Privacy Rule. 아래의 표가 <표 1>의 원문이다. <표 1>는 아래의 원문을 요약한 것이다.

(2)(i) The following identifiers of the individual or of relatives, employers, or

<표 1> 세이프하버 방식에서의 18가지 유형의 데이터

household members of the individual, are removed:
(A) Names
(B) All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census:
(1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and
(2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000
(C) All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
(D) Telephone numbers (L) Vehicle identifiers and serial numbers, including license plate numbers
(E) Fax numbers (M) Device identifiers and serial numbers
(F) Email addresses (N) Web Universal Resource Locators (URLs)
(G) Social security numbers (O) Internet Protocol (IP) addresses
(H) Medical record numbers (P) Biometric identifiers, including finger and voice prints
(I) Health plan beneficiary numbers (Q) Full-face photographs and any comparable images
(J) Account numbers (R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section [Paragraph (c) is presented below in the section “Re-identification”]; and
(K) Certificate/license numbers
(ii) The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.

	유형
1	이름
2	주(State) 단위보다 작은 지리적 위치
3	개인과 연관된 날짜들 (생일, 합격일, 해고일, 사망일, 우편번호(최초의 3자리 제외) 등 포함)
4	전화번호
5	팩스번호
6	전자메일주소
7	사회보장번호
8	의료기록번호
9	건강보험플랜(Health Plan)수혜자번호
10	회계번호
11	운전면허번호
12	자동차식별자와 일련번호(자동차번호 포함)
13	장치식별자와 일련번호
14	URL
15	인터넷프로토콜(IP)주소번호
16	생체(Biometric)식별자(지문 포함)
17	전체얼굴사진과 이와 유사한 이미지
18	기타 특이한 식별 번호 또는 코드

위와 같이 제거의 대상이 되는 식별자 또는 비식별자를 선정한 것은 기술적인 방식들에 의한 연구의 결과물들로부터 상당한 영향을 받았다. 이 표에 포함된 개별적 항목들은 각기 별도의 의미를 가지는 것으로, 향후 국내에 유사한 방식의 도입을 고려한다면 각각의 항목이 가지는 의의에 대해 별도의 분석을 할 필요가 있다. 이 중에서, 예컨대 우편번호의 3자리까지는 허용하는 규정의 경우는 위에서 소개한 Sweeney의 Weld 매사추세츠 주지사 식별 연구 등의 영향을 받은 것이다. 이와 같은 연구의 영향을 받아 준식별자로서 5자리 우편번호를 그대로 공개하는 것은 준식별자로서의 위험성이 있음을 인지하게 되어서 3자리까지의 우편번호까지만 식별될 수 없는 정보로 인정하

고 4자리부터는 준식별자에 해당한다고 규정하게 된 것이다.

Safe Harbor를 통한 내용적 접근법은 수범자의 입장에서는 자신이 원형 그대로의 데이터를 어느 범위까지 사용할 수 있고 어느 범위에서는 비식별화 조치를 해야 할 것인지에 관해 비교적 분명하게 파악할 수 있다는 편의성이 있다. 그래서 이런 내용적 접근법은 규제의 실효성 측면에서 볼 때 수범자에게 규제의 준수 비용을 높이지 않아서 전체적으로 낮은 비용으로 규제의 준수를 유도할 수 있다는 장점이 있다. 하지만 이런 내용적 접근법의 한계 또한 존재한다. 정해진 식별자와 준식별자의 범위가 넓을 경우 수범자의 입장에 따라서는 정해진 식별자와 준식별자의 제거 등을 통해서 익명화한 데이터의 가치가 더 이상 데이터로서의 효용성을 상실할 정도로 떨어질 가능성이 존재한다.

그러므로 개별 수범자에 따라서는 일률적으로 정해진 식별자와 준식별자의 규정의 준수가 어려운 경우가 있으므로 비식별화의 모든 경우를 내용적 접근법을 통해서만 해결하는 것에는 한계가 있다. 이런 관점에서 HIPAA 프라이버시 규칙은 수법자가 내용적 방식에 의한 규제의 준수가 어려울 경우를 고려해서 내용적 접근법뿐만 아니라 개별 사안 별로 좀 더 유연성 있는 해결이 가능한 절차적 접근법도 수범자가 사용가능한 규제의 선택사항으로 할 수 있는 대안을 제시한 것으로 볼 수 있다. 내용적 접근법과 절차적 접근법에 대한 자세한 설명을 다음 단락에서 한다.

2) 내용적 접근법

(a) 식별자와 준식별자에 대한 일반적 개괄

위에서 개괄적으로 설명한 것과 같이 HIPAA 프라이버시 규칙에서 규정한 18가지의 식별자 또는 준식별자를 제거하는 방식이 내용적 접근법이다. 그래서 이 내용적 접근법에 대해서 구체적으로 논의하기에 앞서서 HIPAA 프라이버시 규칙의 맥락을 포함한 식별자와 준식별자에 대한 일반적인 파악이

필요하다.

위에서 살펴본 NIST 보고서에 따르면 일반적으로 준식별자와의 관계에 있어서 식별자는 직접적(directly)으로 식별을 가능하게 하는 데이터가 된다. 예를 들어, 이름, 사회보장번호, 이메일 주소가 이에 해당한다. HIPAA 프라이버시 규칙은 이런 식별자의 개념을 더욱 확장해서 18가지의 특정 데이터 유형들도 포함된다고 규정하는 것이다.[85]

이런 식별자는 개인을 외부의 정보의 도움이 없어도 그 자체적으로도 식별할 수 있게 하는 데이터이기 때문에 식별자의 처리의 방식에는 비식별화 기법을 적용한 다른 형태로의 전환뿐만 아니라 제거(removal)의 방식도 있다. 식별자를 처리하는 방식들은 나열하면 다음과 같다:[86]

- 식별자의 제거
- 명백하게(obviously) 일반적(generic)인 데이터나 범주(category) 명칭으로 대체. 예를 들어, 특정 이름을 “인물이름(PERSON NAME)"라는 용어로 대체
- 임의적 표시(symbol)들로 대체. 예를 들어, “*****” 또는 “XXXXX"로 대체
- 무작위(random) 수치로 대체. 만약 동일한 신분(identity)이 두 번 등장할 겨우, 서로 다른 수치로 표시.
- 어떤 체계를 가진 상태에서 가명(pseudonym)으로 대체

이런 식별자들은 일반적으로 아래와 같은 표의 형태로 정리할 수 있다.

85) 이 프라이버시 규칙에 명시적으로 이 18가지 유형의 정보들이 준식별자가 아닌 식별자로 규정되어 있다. 그리고 이 프라이버시 규칙과 이 규칙에 대한 공식적인 OCR의 가이드라인에서도 준식별자 개념에 대한 별도의 언급이나 설명은 없다. 본 보고서에서 식별자와 준식별자에 대해 설명하는 부분은 대체적으로 NIST 보고서의 관점을 반영하는 것이다.

86) NIST 전게서, 제 15면

Direct Identifiers								
Name	Address	Birthday	ZIP	Sex	Weight	Diagnosis	…	…

출처: Simon L. Garfinkel, "De-Identification of Personally Identifiable Information", 표 1(2015)

준식별자는 일반적으로 간접적 식별자를 의미한다. 그래서 그 데이터 자체로는 특정인을 식별할 수는 없지만 다른 정보와 연결되어서는 특정인을 식별할 수 있는 경우에 해당한다.[87] 위에서 설명한 Weld 주지사가 Sweeney에 의해 재식별된 사건에서 재식별에 활용된 생일, 우편번호(ZIP), 그리고 성별이 준식별자들에 해당하고, 식별자를 정리한 위의 표로부터 준식별자를 별도 유형으로 표현한 아래의 표와 같이 새로이 정리할 수 있다.

Direct Identifiers		Quasi-Identifiers						
Name	Address	Birthday	ZIP	Sex	Weight	Diagnosis	…	…

출처: Simon L. Garfinkel, "De-Identification of Personally Identifiable Information", 표 2(2015)

준식별자는 비식별화에 있어서 중요한 도전과제를 던진다. 식별자는 전체 데이터의 집합에서 해당 데이터를 단순하게 삭제하면 되지만, 준식별자는 향후

87) NIST 전게서, 제 19면

의 연구에서는 중요한 가치를 지닐 수도 있는 종류의 정보를 포함하는 경우가 많을 것이기 때문에 삭제를 할 경우에는 데이터의 효용성에 상당한 손실을 가하게 될 수 있다.[88] 왜냐하면 준식별자 자체가 식별자와는 다르게 전적으로 개인의 신원 자체를 인식하기 위해 만들어진 항목이 아니고, 준식별자에는 개인의 신원을 알 수 있도록 하는 부분과 그 이외의 측면에서 효용성을 가지도록 하는 부분이 같이 공존하고 있기 때문이다.

식별자의 대표적인 예라 할 수 있는 이름은 개인의 신원을 다른 사람과 구별하기 위해 개인에게 부여된 특유의 표식이지만, 준식별자의 일반적인 예라 할 수 있는 우편번호와 같은 경우는 그 우편번호 자체는 특정개인을 다른 사람과 구별하기 위한 것이 아니라 특정 지역을 다른 지역과 구분하기 위해 만들어진 항목이다. 그래서 준식별자의 경우에는 삭제의 방식을 일률적으로 적용하기가 어렵게 되어있기 때문에 재식별의 위험성과 데이터의 효용성이란 상반된 가치 사이에서 균형을 추구하는 방식을 모색할 필요성이 있다. 이런 균형적인 접근법 하에서는 일반적으로 다음과 같은 비식별화 방식들이 사용된다.[89]

- 범주화(Suppression): 삭제도 가능하지만 데이터의 효용성이 많이 손상되기 때문에 삭제 대신에 범주화가 더 선호됨
- 일반화(Generalization): 예를 들어, 우편번호의 5자리 전부를 공개하지 않고 특정한 범위에 있다는 정도의 숫자만 공개
- 교환(Swapping): 준식별자들은 지정된 수준의 일반화 내에서 기록들 사이에 교환이 가능
- k-익명성, l-다양성, t-근접성: 바람직한 수준의 프라이버시를 달성하기 위해 처리가 필요한 비식별자들의 양을 산정하는 방식

이와 같이 식별자와 준식별자의 구별을 통해 비식별화의 처리를 하려는 접근법도 최근에 개발되었다. Khaled El Emam과 Bradley은 이와 같이 식별

88) NIST 전게서, 제 20면
89) 이 각각의 개념에 대한 설명은 이 보고서의 앞부분에서 했기 때문에 여기서는 자세한 개념 설명은 생략한다.

자와 준식별자의 개념적 구분에 근거해서 11단계의 비식별화 절차를 개발했다.[90] 이 11단계의 절차들은 아래의 표와 같다.[91]

제 1단계	전체 데이터 집합에서 식별자 결정
제 2단계	식별자의 변형(제거 또는 가명으로 대체)
제 3단계	준식별자의 결정을 포함한 위협모형화(threat modeling)의 수행[92]
제 4단계	수용가능한 최소한의 데이터의 효용성 결정
제 5단계	재식별의 위험성 기준의 결정
제 6단계	실험(Sample) 데이터의 도입[93]
제 7단계	실질적인 재식별의 위험성 평가
제 8단계	이 실질적 위험성과 위험성 기준의 비교(제 7단계와 제 4단계의 비교)
제 9단계	새로운 매개수치(parameter)의 설정과 데이터의 변형
제 10단계	결론으로 나온 비식별화 기법 방식에 대한 진단
제 11단계	변형된 데이터를 외부 데이터 집합에 전송

이 11단계 비식별화 절차의 핵심은 초기의 1과 2단계에서 식별자를 우선적으로 구별해서 일률적인 방식으로 변형을 시킨 후 나머지 데이터 중에서 결정된 준식별자를 구별해내어서 4에서 10단계를 통해 데이터의 효용성과의 위험성 사이에 균형점을 찾는 비식별화 방식을 적용한다는 것이다. 즉 준식

90) K. El Emam and B. Malin, "Appendix B: Concepts and Methods for De-identifying Clinical Trial Data," in Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk, Institute of Medicine of the National Academies, The National Academies Press, Washington, DC. 2015

91) NIST 전게서, 제 21면

92) 위협모형화 단계에서 해당 개인정보의 관리주체는 재식별을 할 상대방을 결정하게 되는데 이 경우 이 상대방이 재식별에 사용할 준식별자도 결정하게 된다.

93) 개인정보의 관리주체는 당해 데이터의 재식별 위험성을 평가하기 위해 시험용 데이터를 도입한다.

별자와 식별자 사이에서의 구분을 통해서 각각에 적합한 데이터의 변형 방식을 적용해서 해결하는 것이다. 하지만 식별자와 준식별자의 구분을 명확하게 할 수 있는지 여부가 문제가 되기 때문에 이런 식의 접근법에 대한 비판의 목소리도 존재한다.[94]

(b) HIPAA 프라이버시 규칙의 식별자(또는 준식별자)

① 우편번호(ZIP)

HIPAA 프라이버시 규칙에 따르면 총 5자리의 우편번호들 중 최초의 3자리는 일정한 조건을 만족하면 공개되는 데이터에 포함할 수 있다. 인구조사국(Bureau of the Census)에서 대중에게 공개한 데이터에 따라, (1)최초의 3자리의 동일한 숫자를 가진 모든 우편번호들을 결합에 해당하는 지역이 2만 명보다 많은 인구를 가지거나, (2) 이런 동일한 조건에 해당하는 지역의 인구가 2만 명 이하일 경우에는 그 최초 3자리의 숫자가 000으로 바꿀 경우에는 해당 우편번호는 공개가 가능하다.[95] 즉, 우편번호의 최초의 3자리는 비식별화된 정보로 포함될 수가 있다. 하지만 최초의 3자리가 모두 000으로 표시가 되어야 한다는 조건을 충족할 경우에만 가능하다.

② 일자(dates)

HIPAA 프라이버시 규칙에 따르면 공개가 허용되지 않는 일자들의 요소들에는 일(day), 월(month), 그리고 연도(year)보다 더 특정된 기타 정보가 포함된다.[96] 예를 들어, “2009년 1월 1일”과 같은 수준으로 특정된 일자는 공개될 수 없다. 하지만 이 표현을 “2009”로 비식별화할 경우에는 공개가 가능해진다. 일종의 범주화 또는 일반화 기법이 적용된 예이다.

많은 기록들은 나이를 암시하는 정보를 포함한다. 군복무 기간에 관한 정보

94) NIST 전게서, 제 22면
95) HIPAA Privacy Rule §164.514(b)(i)(B)
96) HIPAA Privacy Rule §164.514(b)(i)(C)

가 그 예가 될 수 있다. 89세가 초과되는 것으로 분명하게(explicitly) 또는 함축적으로(implied) 표현된 나이는 모두 90세 이상으로 기록되어야 한다.[97] 특정 데이터값에 대한 빈도수 자체가 작은 경우에는 특정한 숫자 자체가 특정인을 구별할 수 있는 확률을 높이기 때문에 범주화나 일반화가 적용된 경우라고 할 수 있다. 예를 들어, 만약에 환자가 태어난 연도가 1910년이고 의료보험서비스(healthcare service)의 연도가 2010년이라면, 태어난 연도는 "1920년 또는 그 이전" 으로 기록되어야 한다.[98] 그렇지 않으면 이 데이터를 확보한 주체는 환자의 나이가 약 100세라는 사실을 알게 될 것이기 때문이다.

③ 기타 식별성을 가능하게 하는 숫자, 속성, 코드[99]

이 범주는 HIPAA 프라이버시 규칙에서 명시적으로 나열되지는 않았지만 식별을 가능하게 하는 특징을 지는 기타 요소들에 대한 항목이다. 그러므로 이 프라이버시 규칙의 피규제자는 명시적으로 나열된 식별자의 제거뿐만 아니라 이 범주에 해당하는 기타 식별성을 가능하게 하는 요소들도 제거해야 한다. 이 가이드라인에서는 이 범주에 해당할 수 있는 항목들을 다음과 같이 숫자, 속성(characteristic), 코드로 분류해서 예시하고 있다.[100]

- 식별성이 있는 숫자: 많은 종류의 식별성을 지닌 잠재력을 지닌 숫자들이 있다. 예를 들어, 이 프라이버시 규칙의 서문에서는 의료에서의 임상 실험에서 할당된 번호(clinical trial record number)도 이 범주에 해당한다는 사실을 확인한다.
- 식별성이 있는 코드: 보안책이 갖추어지지 않은 인코딩(encoding) 메커니즘에서 추출된 숫자에 해당하는 토드가 이 범주에 해당한다. 예를 들어, 암호키가 없는 해시함수에서 추출된 코드는 이 범주에 해당할 수

97) HIPAA Privacy Rule §164.514(b)(i)(C)
98) OCR 전게서, 제 25면
99) HIPAA Privacy Rule §164.514(b)(i)(R)

(R) Any other unique identifying number, characteristic, or code, except as permitted by paragraph (c) of this section

100) OCR 전게서, 제 26면

있다.

-식별성이 있는 속성: 개인을 구별해서 식별을 가능하게 하는 모든 요소가 이 범주에 해당한다. 예를 들어, 만약 환자의 직업이 "A 대학의 현 총장" 으로서 기록물에 나열되어 있다면, 이 직업은 이 범주에 해당할 수 있다.

(c) 내용적 절차법에서의 "사실적 인식" 101)

내용적 절차법이 적용되기 위한 두 번째 요건으로서 규정한 사실적 인식은 식별자들이 제거되고 남은 정보가 그 자체만이거나 또는 다른 정보와 결합해서 정보의 주체를 식별하기 위해 이용될 수 있다는 사실에 대한 명확하고 직접적인 인식을 의미한다.[102] 이것은 이 프라이버시 규칙의 피규제자가 남은 정보가 개인의 식별을 위해 이용될 수가 있는지에 대하여 사실적 인식을 가지는 것을 의미한다. 즉, 그 남은 정보가 실질적으로는 비식별화된 정보가 아닌 경우 그러한 사실을 알고 있다는 의미가 된다.[103]

가이드라인에서는 이 사실적 인식의 기준을 충족하지 목하는 가상적 사례의 설명을 통해 이 기준이 어떻게 적용되는지에 대해서 보여주고 있다. 예를 들어, 이 프라이버시 규칙의 피규제자가 환자의 직업이 "A대학의 전 총장" 으로서 기록되어 있다는 사실을 알고 있다고 가정하자.[104] 나이 또는 거주하는 주(state)와 같은 추가정보와 결합한 상태에서 이 정보는 그 환자를 거의 확실하게 식별할 것이다. 이 예에서 피규제자는 단순하게 프라이버시 규칙에 나열된 식별자들을 제거하는 것만 가지고는 비식별성의 기준을 만족하지 못한다. 왜냐하면, 이 경우 식별될 수 있는 위험성의 정도가 매우 명확해서 이 피규제자는 위의 남은 정보가 그 환자를 식별할 수 있을 것이라고 내부적으

101) HIPAA Privacy Rule §164.514(b)(ii)

> (ii) The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.

102) OCR 전게서, 제 27면
103) OCR 전게서, 제 27면
104) OCR 전게서, 제 27면

로는 결론을 내렸을 개연성이 매우 높기 때문이다. 그러므로 이 피규제자가 그 환자의 기록에서 직업을 제거하기 위한 선의(good faith)의 충분한 노력을 하지 않는다면 이 정보는 내용적 접근법의 비식별성의 기준을 만족하지 못하게 된다.

3) 절차적 접근법

위에서 설명한 것과 같이 절차적 접근법은 개인정보의 관리 주체가 외부의 전문가와 해당 정보가 식별될 수 있는 위험성을 평가하는 것과 관련된 계약을 체결해서 위험성을 판단하게 된다. 즉 위험성을 판단하는 주체의 선임에 정부가 직접적으로 관여하지 않고 사적인 영역에서의 전문가 선임 계약을 통해 진행된다. 또한 이 전문가의 자격 조건으로 어떤 정부 기관이나 조직을 특정하지 않고 있다. 그러므로 HIPAA 프라이버시 규칙에서 규정한 절차적 접근법은 사전적으로는 정부가 직접적으로 개입하지 않는 방향을 채택하고 있다. 그러므로 효과적이고 정확한 위험성 판단을 위해서는 선임되는 전문가가 어떤 자격을 지녀야 하는지에 대한 우선적인 고려를 할 필요가 있다.

이 HIPAA 프라이버시 규칙에서는 전문가가 되기 위한 특정한 조건이나 규정을 두고 있지 않다. 단지 동 규칙의 164.514(b)항에서 “비식별화에 대해서 폭 넓게 받아들여지고 있는 통계적 그리고 과학적인 이론과 방식에 대해 적절한 수준의 지식과 경험이 있는 자”가 전문가로서의 판단을 할 수 있다고 규정하고 있다.[105] 이 프라이버시 규칙에는 전문가의 자격에 대해 이런

105) HIPAA Privacy Rule §164.514(b)

> (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:
> (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
> (ii) Documents the methods and results of the analysis that justify such determination

일반적인 내용만을 규정하고 있고, 이 프라이버시 규칙의 가이드라인에서 좀 더 상세하게 설명하고 있다.

그 가이드라인에 따르면 전문가가 되기 위해 어떤 특별한 학위나 자격증이 필요하지 않다고 밝히고 있다.[106] HIPAA 프라이버시 규칙의 규정을 살펴보면 일정 수준의 지식과 관련 경험이 필요한데, 이 관련 경험은 다양한 방식의 교육이나 경험들을 통해 획득할 수 있다고 가이드라인에서 설명하고 있다. 전문가로서의 경험을 획득할 수 있는 방식의 다양성을 인정하면서 그 경험의 분야가 통계, 수학, 또는 다른 과학 영역일 경우에도 상관이 없다고 설명하고 있다. 이와 같이 전문가 선임에 있어서 어떤 구체적인 기준을 고정적으로 규정하지 않고 전문가로 활동할 수 있는 상황을 광범위하게 설정하고 있다. 그래서 개인정보의 관리 주체는 자신들이 보유하고 있는 데이터의 식별성에 대한 위험성 판단에 있어서 자신들이 원하는 전문가를 선임할 수 있다.

하지만 전적으로 사적 자치의 영역으로 둘 경우에는 개인정보의 관리 주체와 해당 전문가 사이에 계약의 형태를 통해 이해관계를 일치시켜서 식별성의 위험성을 왜곡하여 평가할 유인이 발생할 수 있다. 예를 들어, 개인 정보의 관리주체는 자신들의 선호에 맞는 연구 결과물을 발표한 연구자가 일반적인 전문성 기준에 미달해도 이런 전문가를 위험성 평가의 주체로 계약을 체결할 수 있다. 이런 상황이 발생할 수 있기 때문에 사후적인 관점에서는 최소한 정부가 간접적인 수단으로도 통제할 수 있는 장치가 있어야 한다. 그래서 집행의 관점에서 OCR이 사후에 비식별화 방식을 적용한 전문가의 실제 경험과 함께 관련 전문적 또는 학문적 경험들을 살펴볼 수 있다고 한다.[107]

전문가가 위험성을 판단하는 기준은 어떠한지 살펴보자. 위에서 언급한 것과 같이 식별성의 위험성이 ‘매우 작은’ 경우(이하 “이 기준”)라고 판단해야

106) OCR, 'Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule', 제 10면 (2012)

107) OCR 전게서, 제 10면

법에서 규정하는 위험성이 없는 상황이라고 간주할 수 있다. 어떤 상황이 이 기준에 해당하는지에 대해서 가이드라인에 설명이 제시되어 있다. 이 가이드라인에 따르면, 이 기준을 보편적으로 충족시키는 특정한 수치의 형태로 식별의 위험성을 표현할 수는 없다는 사실을 강조한다.[108] 결국 관련된 구체적인 상황에 적합한 방식으로 위험성이 있는지 여부를 판단하는 것이다. 이런 위험성의 정도는 해당 데이터를 식별화하려는 '예상주체' (anticipated recipient)가 가지고 있는 능력에도 의존한다.[109]

식별화하려는 예상주체의 능력은 매우 다양한 요소들에 의존하게 되는데 전문가는 이런 요소들을 전반적으로 고려해서 위험성을 판단해야 한다. 왜냐하면, 특정 환경의 맥락에서 특정 데이터를 위해 결정된 식별의 위험성은 다른 환경에서의 동일한 데이터나 같은 환경에서의 상이한 데이터에서도 동일하게 적용되기가 매우 어렵기 때문이다. 결국 전문가가 적용하게 되는 이 기준은 개인을 식별하려는 예상주체의 능력이란 전제에서 적합한 기준이 되는 것이다. 이와 같이 개인을 식별하려는 잠재적인 주체를 위험성 판단의 핵심적인 요소로 판단하는 것은 후술하게 될 영국에서의 '의도된 공격자(motivated intruder)'기준과 유사한 점이 있다.[110]

이와 같이 모든 프라이버시와 식별성 문제를 보편적으로 해결할 수 있는 단일한 해결책이나 이를 기계적으로 수치화하여 평가할 수 있는 방법은 존재하지 않는다. 실제로는 이런 간단명료한 해결책의 관점이 아니라 다양한 기술과 정책적 장치들의 결합들을 통해 비식별화 문제를 해결하게 된다. 이런

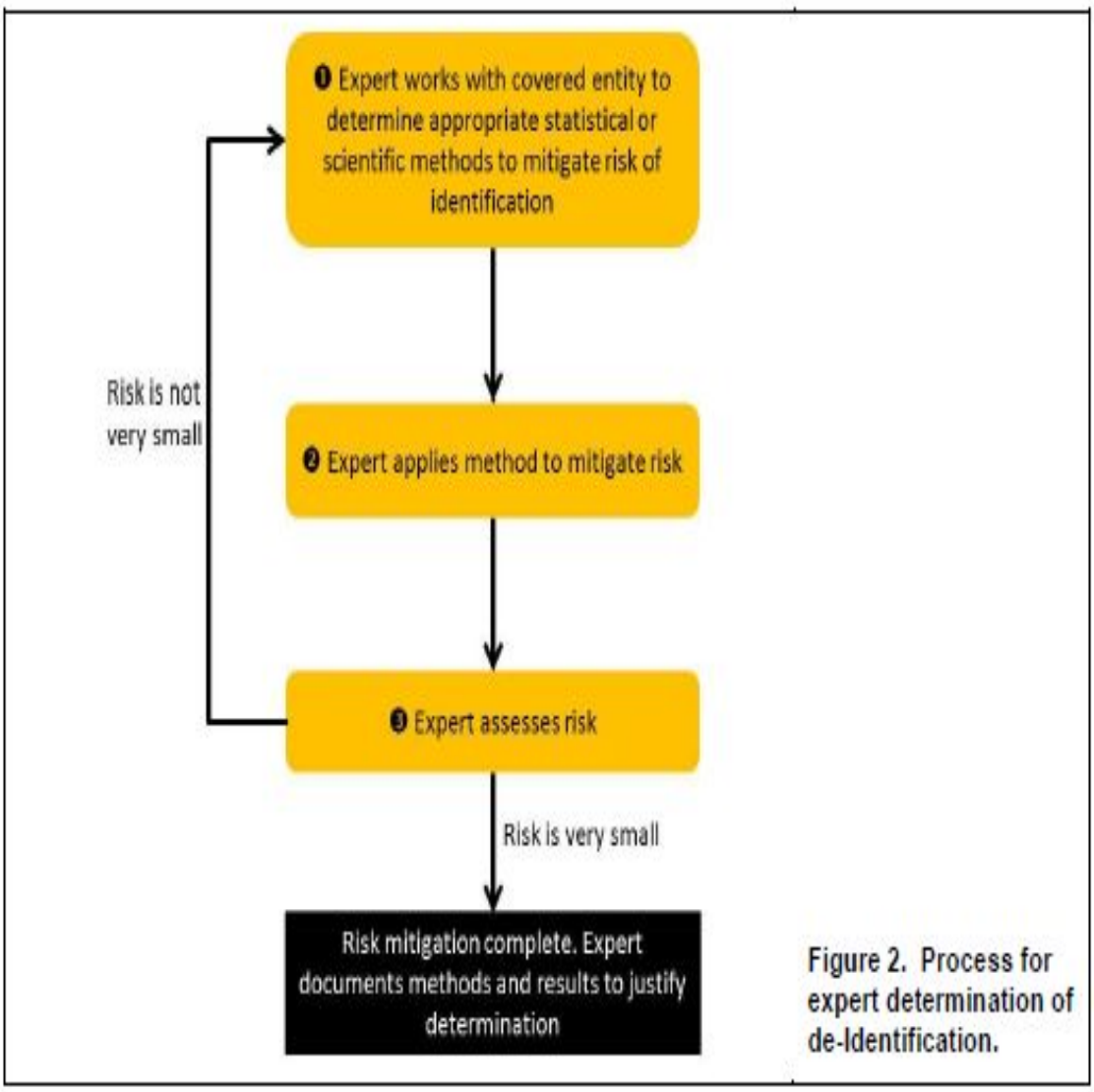
108) OCR 전게서, 제 10면
109) HIPAA Privacy Rule §164.514(b)

> (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:
> (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
> (ii) Documents the methods and results of the analysis that justify such determination

110) 이 부분에 대해서는 영국의 ICO에서 발표한 익명화 기준을 설명하는 단락에서 자세하게 설명한다.

접근법에 따라 OCR도 위험성 판단에 있어서 전문가에게 어떤 특정한 방식이나 기준을 따를 것을 요구하지 않는다. 단지 사후적인 통제적 관점에서 실제 분석에 사용된 방식이나 분석결과를 문서화해서 OCR이 요구할 경우에는 언제든지 제출할 수 있도록 준비되어있을 것을 요구하고 있다.

HIPAA 프라이버시 규칙은 이런 접근법에 따라 절차적 방식에 대해서 어떤 구체적인 설명을 하지 않지만 이 가이드라인에서는 절차적 방식의 일반적인 메커니즘과 전문가가 이 기준의 판단을 할 경우 적용하는 기본원칙들에 대해서 좀 더 구체적으로 설명하고 있다. 우선 절차적 방식의 전체적인 흐름을 아래의 그림에서 확인할 수 있다.



출처: OCR, 'Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule', 그림 2(2014)

위의 표에서 알 수 있듯이, 우선 전문가는 대상이 되는 정보가 해당 데이터를 확보하게 되는 예상주체에 의해 식별될 수 있는 정도에 대해 평가할 것이다. 그 다음으로, 그 전문가는 때때로 해당 개인정보의 관리주체에게 예상되는 식별의 위험성을 축소하기 위해 적용될 수 있는 통계적 그리고 과학적 기법들에 대한 설명을 하게 된다. 이런 설명을 한 후 이 관리주체 기관이나 기업에 소속된 데이터관리자가 그 기법들을 수용할만한 것(deemed

acceptable)으로 인정하는 방식들을 적용하게 된다.[111] 여기서 주의할 점은 전문가가 이 데이터관리자에게 식별성의 위험성을 축소하는 방식들을 설명하는 이유가 이 데이터관리자의 승인이나 허락을 얻으려는 것에 있지는 않다는 사실이다. 관리자가 자신의 판단에 의해서 적용할 방식들을 선택하게 되고 단지 투명성의 관점에서 데이터관리자에게 자신의 방식들을 공개하고 설명을 한 후 데이터관리자의 반응을 단지 고려할 뿐이다. 이 가이드라인에서 이 부분에 대해 명확하게 설명하고 있지는 않지만 위의 표현에서 “deemed acceptable"을 사용함으로써 데이터관리자의 직접적인 승인이 그 목적이 아님을 추측할 수 있다. 마지막으로, 전문가는 해당 정보의 식별성을 평가하여 이 정보가 공개될 경우의 위험성이 이 기준 이하일 것인지 여부를 판단하게 된다. 만약 전문가가 이 기준을 충족하지 못한다고 판단하면 이 기준을 충족할 때까지 위의 과정을 계속적으로 반복하게 될 것이다.

전문가가 위험성 평가를 함에 있어서는 기본적으로 준수해야 할 여러 가지 원칙들이 있을 것인데, 가이드라인에는 적용될 수 있는 원칙들의 예가 나열되어 있다. 나열된 원칙들은 단순한 예시일 뿐이고 여기서 이 원칙들이 적용할 수 있는 원칙들의 전부라고 할 수는 없다. 하지만 가이드라인에 제시된 원칙들은 적용가능한 원칙들을 설정함에 있어 중요한 참고자료로서의 역할을 한다. 이 원칙들은 다음의 표에서 확인할 수 있다.

Table 1. Principles used by experts in the determination of the identifiability of health information.

Principle	Description	Examples
		High: Demographics are highly distinguishing, highly replicable, and are available in public data sources.

출처: OCR, 'Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule', 표 1(2014)

111) OCR 전게서, 제 12면

Replicability	Prioritize health information features into levels of risk according to the chance it will consistently occur in relation to the individual.	*Low:* Results of a patient's blood glucose level test will vary
		High: Demographics of a patient (e.g., birth date) are relatively stable
Data source Availability	Determine which external data sources contain the patients' identifiers and the replicable features in the health information, as well as who is permitted access to the data source.	*Low:* The results of laboratory reports are not often disclosed with identity beyond healthcare environments.
		High: Patient name and demographics are often in public data sources, such as vital records -- birth, death, and marriage registries.
Distinguishability	Determine the extent to which the subject's data can be distinguished in the health information.	*Low:* It has been estimated that the combination of *Year of Birth, Gender,* and *3-Digit ZIP Code* is unique for approximately 0.04% of residents in the United States[9]. This means that very few residents could be identified through this combination of data alone.
		High: It has been estimated that the combination of a patient's *Date of Birth, Gender,* and *5-Digit ZIP Code* is unique for over 50% of residents in the United States[10,11]. This means that over half of U.S. residents could be uniquely described just with these three data elements.
Assess Risk	The greater the replicability, availability, and distinguishability of the health information, the greater the risk for identification.	*Low:* Laboratory values may be very distinguishing, but they are rarely independently replicable and are rarely disclosed in multiple data sources to which many people have access.

위의 표에서 보면 반복성(replicability), 데이터소스에의 접근성(data source availability), 구별성(distinguishability) 요소를 위험성 판단의 원칙들로 나열을 했다. 반복성이란 특정 정보가 시간이 지나도 그 정보의 동일성을 계속 유지할 수 있는지에 대한 항목이다. 데이터소스에의 접근성이란 외부 정보에서도 해당 정보의 식별성을 포함하고 있는지의 여부에 대한 항목이 된다. 마지막으로 구별성이란 특정인의 데이터가 다른 사람의 데이터와 구별이 되는 정도를 의미한다.

이 각각의 항목들을 살펴보면 이 원칙들의 특징이 강하게 나타날수록 해당 데이터의 식별성을 확인할 수 있는 확률이 높아진다는 사실을 추측할 수 있다. 이런 추론에 따라 위 표의 맨 오른쪽 열에서는 각각의 속성에 따라 식별의 위험성이 커지는 경우와 작아지는 경우를 예시로 설명하고 있다. 그래서 이런 요소들을 통해 판단하면 반복성, 데이터소스에의 접근성, 그리고 구별성이 크면 클수록 해당 정보가 식별될 위험성이 더 커진다고 판단할 수 있다. 위에서 강조한 대로 이 가이드라인에서 나열한 위의 3가지 원칙은 하나의 예시들이기 때문에 위험성 판단을 수행할 전문가는 다른 원칙들을 적용해서도 위험성 판단을 할 수 있다.[112)]

이외에도 전문가가 위험성을 판단할 때는 개인에 해당하는 신분을 드러내는 데이터소스와 해당데이터가 연결될(linked) 수 있는 정도를 때때로 고려할 수 있다. 이 연결성은 특정 조건들을 만족하기 위해 필요한 절차이다. 일반적으로 이런 연결성 확인은 아래의 그림에서와 같이 투표자등록부(voter registration records)와 같은 공개된 외부 정보와의 비교를 통해서 이루어지게 된다.

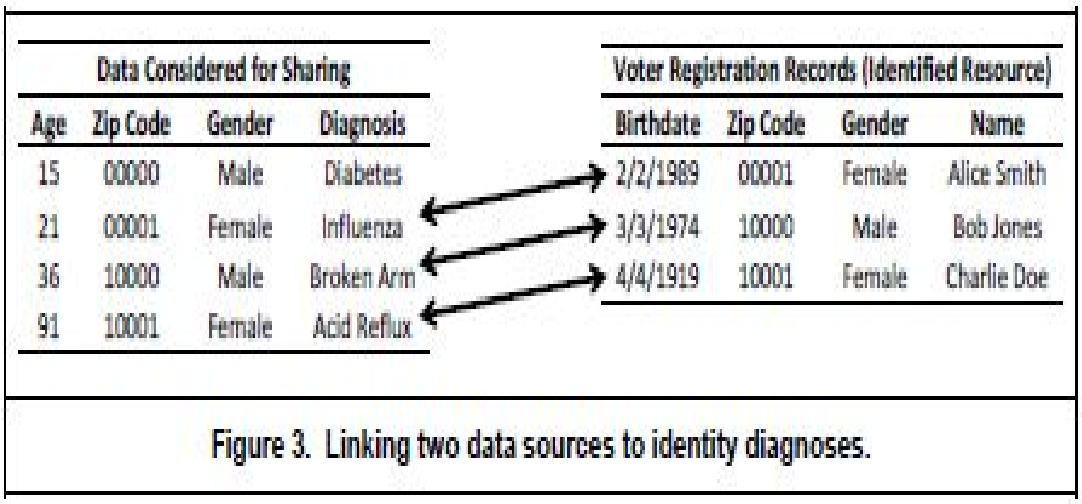


Figure 3. Linking two data sources to identity diagnoses.

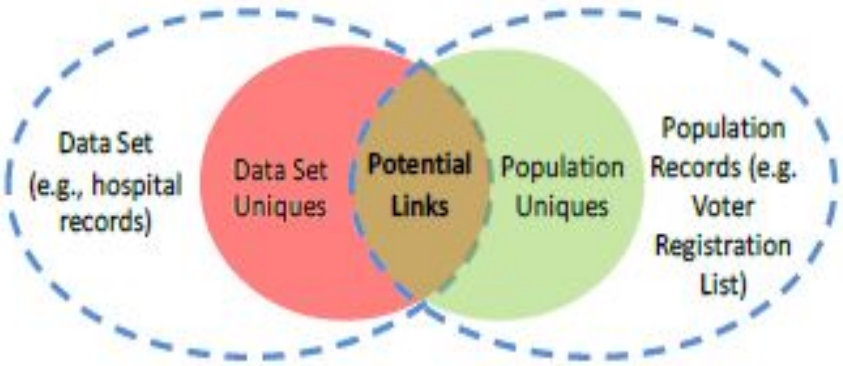
결국 전문가가 위험성 평가를 위해 특정한 방식을 적용할 것이 강제가 되지 않는다. 전문가 자신의 판단으로 해당 데이터와 주변 환경을 고려해서 가장 적합하다고 생각하는 방식들을 사용하게 된다. 전문가가 일반적으로 인정되

112) OCR 전게서, 제 15면

는 통계적 또는 과학적 원칙들을 적용해서 해당 데이터집합(data set)에 있는 특정 기록이 모집단 전체(population) 내부에서 특이성(unique)을 가지는지, 즉 유일한 사람과 연결가능성이 있는지에 대한 확률을 산정한다.

출처: OCR, 'Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule', 그림 3(2014)

만약 데이터집단 전체가 전체집단에 속하지 않을 경우에는 잠재적인 연결성에 대한 확인을 통해 연결가능성이 있는지를 평가하게 된다. 이런 연결가능성의 평가를 통해 다양한 집단들이 서로 다른 영역에 있어도 그 중첩되는 부분 만큼의 위험성에 대한 산정을 하게 되는 것이 위험성 판단의 전체적인 큰 흐름이 된다. 아래 그림을 통해 그 개념을 요약할 수 있다.



출처: OCR, 'Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule', 그림 4(2014)

2. EU

(1) 비식별화 처리에 대한 규제 및 논의 동향 개괄

EU 개인정보보호지침(EU Data Protection Directive, 이하 'EU Directive')의 서문 제26조[113]는 '식별가능한 데이터의 맥락에서의 익명화(anonymous)에 대한 규정을 두고 있다. 이는 다음과 같이 명시적으로 밝힌다. : "(이 지침에 따른) 보호의 원칙은 식별되거나 식별될 수 있는 개인에 적용되어야 한다; 개인의 식별 가능성에 대한 판단은 데이터관리자(controller) 또는 다른 사람이 대상되는 개인을 식별하기 위해 사용할 것이라고 합리적(reasonably)으로 예측되는 모든 수단 들을 고려해야 한다; 이러한 개인정보 보호의 원칙은 익명화된 데이터에는 적용되지 않는데, 이때의 익명화는 더 이상 식별이 되지 않는 방식으로 이루어져야 한다; 행동강령(code of conduct)이 데이터가 익명화되는 방식에 대한 안내서(guidance)의 역할을 할 수 있다."

EU 역내 개인정보를 보호하는 데에 대한 일반 규정이라 할 수 있는 EU Directive에서 익명화되는 데이터는 규제의 대상이 되는 개인정보가 아니라고 명시적으로 표현하고 있다. 하지만 이러한 익명화에 대한 언급은 EU Directive의 서문(preface)[114]에서만 명시되어 있을 뿐, EU Directive의

113)

> Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data
> DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
> of 24 October 1995
> on the protection of individuals with regard to the processing of personal data and on the free movement of such data
>
> THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,
> Having regard to the Treaty establishing the European Community, and in particular Article 100a thereof,
> Having regard to the proposal from the Commission (1),
> Having regard to the opinion of the Economic and Social Committee (2),
> Acting in accordance with the procedure referred to in Article 189b of the Treaty (3),
> … (26) Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible; …

114)

> CHAPTER V CODES OF CONDUCT

본문 조항들에서는 구체적인 설명이 없다. 다만 EU Directive 자체 내에서는 익명화 방식에 대해 좀 더 구체적인 기준들을 규정하지 않고 있을 뿐, 더 자세한 내용들은 EU 회원국들에게 행동강령(code of conduct)이라는 형식으로 위임하고 있기는 하다.

따라서 익명화에 대한 기본적인 원칙은 EU Directive에 명시되어 있기는 하지만, 그 구체적인 내용들은 개별 국가들에 위임되게 되었고, EU 역내에서 개인정보를 수집하거나 처리하려던 사업자들이 EU Directive에만 의존해서 익명화 방식을 적용하기는 매우 어렵게 되었다. 더욱이 만일 개별 국가들이 행동강령을 마련하고 있지 않은 경우에는 신뢰할 수 있는 익명화 방식에 대한 구체적인 기준을 확인할 수 없어 익명화를 통한 데이터의 수집과 처리가 힘들게 되어 난감한 상황에 놓여 있었다. 다행히 2012년 영국을 필두로 익명화에 대한 행동강령이 등장[115]하였고, EU 정보보호 워킹그룹 29에서 의견서(오피니언)[116]을 발표하였다. 이를 계기로 관련분야에 대한 논의가 더욱 활발해지게 되었다.

현재 EU 개인정보 보호법제에 대해 가장 중요한 변화가 될 수 있는 '일반

Article 27

1. The Member States and the Commission shall encourage the drawing up of codes of conduct intended to contribute to the proper implementation of the national provisions adopted by the Member States pursuant to this Directive, taking account of the specific features of the various sectors
.
2. Member States shall make provision for trade associations and other bodies representing other categories of controllers which have drawn up draft national codes or which have the intention of amending or extending existing national codes to be able to submit them to the opinion of the national authority.

Member States shall make provision for this authority to ascertain, among other things, whether the drafts submitted to it are in accordance with the national provisions adopted pursuant to this Directive. If it sees fit, the authority shall seek the views of data subjects or their representatives.

3. Draft Community codes, and amendments or extensions to existing Community codes, may be submitted to the Working Party referred to in Article 29. This Working Party shall determine, among other things, whether the drafts submitted to it are in accordance with the national provisions adopted pursuant to this Directive. If it sees fit, the authority shall seek the views of data subjects or their representatives. The Commission may ensure appropriate publicity for the codes which have been approved by the Working Party.

115) ICO, Anonymisation: managing data protection risk code of practice, 2012

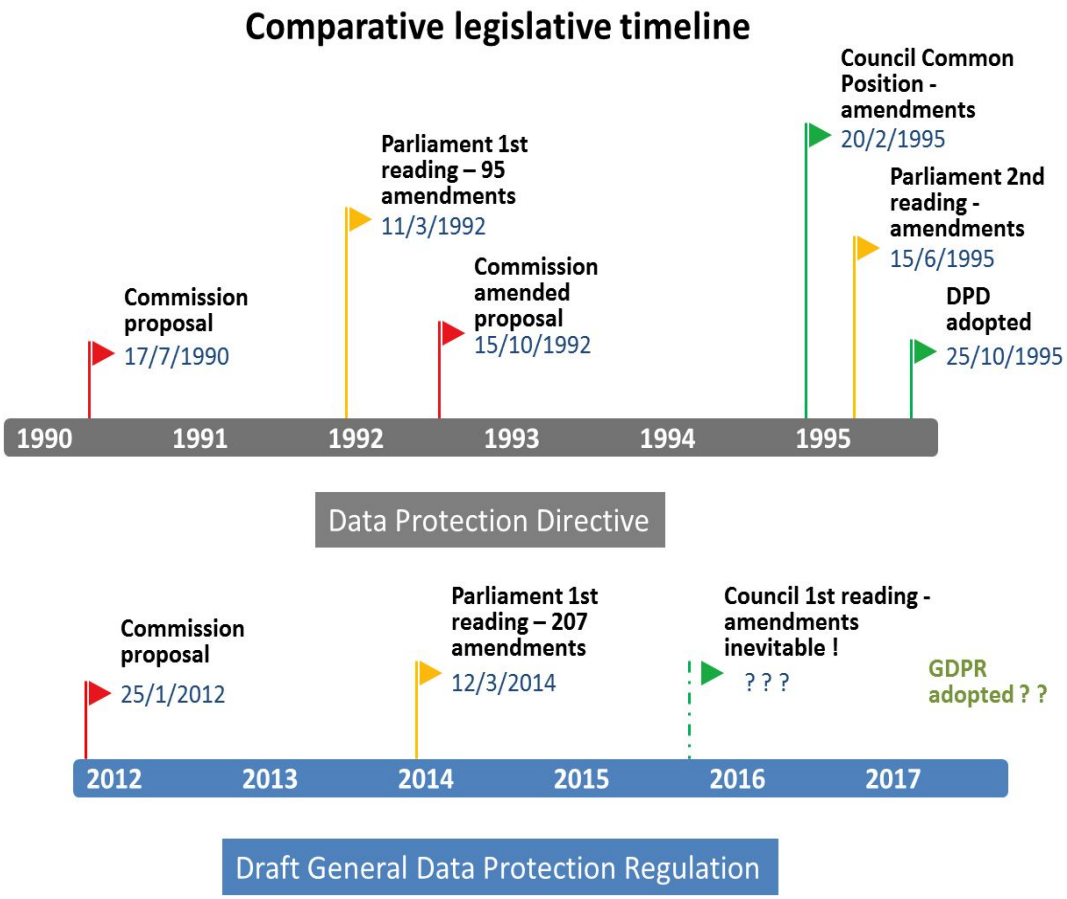
116) ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 05/2014 on Anonymisation Techniques, Adopted on 10 April 2014

데이터보호규칙'(General Data Protection Regulation, GDPR) 도입 논의가 진행되고 있다. 이는 기존 유럽의 개인정보 보호 규범인 EU Directive를 일반데이터보호규칙(GDPR)으로 대체하는 논의라고 정리할 수 있겠다.

2012년 1월 25일 유럽집행위원회가 최초로 제시한 GDPR 안[117]은 유럽뿐만 아니라 전 세계의 개인정보 전문가들 사이에서 엄청난 논의대상이었다. EU의 법체계에 비추어 Directive는 역내 개별 국가에게 지침으로 작용할 뿐 직접적인 강제력은 없는 반면, Regulation은 역내 국가들에게 직접적으로 법적인 구속력을 지녀 강제력을 가지게 되기 때문이었다. 상이한 이해관계를 가지고 있던 개별 국가들과 다국적 기업 그리고 NGO들은 각각 의견을 개진하였다. 공식적으로는 유럽집행위원회(European Commission)가 최초로 제시한 GDPR안에 대해 각료회의(Council of European Union)와 유럽의회(European Parliament)가 각각 수정안을 만들어서 제시한 상태이다. 과거부터 현재까지의 GDPR과 관련된 논의의 진행 사항을 아래의 그림에서 확인할 수 있다.

117)

전반적인 요약 : DPD로부터 GDPR로의 변화 (2015년 12월 현재 여전히 진행 중이며, 확정된 것이 아님)	
Data Protection Directive (DPD)	General Data Protection Regulation (GDPR)
EU 국가들이 가이드로서 활용함 (사실상의 구속력).	모든 EU 구성국들에 적용됨 (법적인 구속력).
EU 에 미침.	전 세계에 미침.
데이터 컨트롤러(Data Controllers)에 적용됨.	데이터 컨트롤러뿐만 아니라, 데이터 처리자(Data Processors), 하청업자(Sub-Processors)에게도 적용됨.
구성국 단위로 비준수에 대해 제재.	제재가 강력함.
각 구성국의 개인정보보호법(DPA)이 승인됨.	하나의 개인정보보호법(DPA)이 승인됨 (or two for >10 Member States).
다양한 유형의 동의(consent)를 인정.	명시적 동의(Explicit consent)만 인정.
이름과 같은 개인정보가 보호됨.	모든 개인정보가 보호됨 (설령 암호화 되었더라도 보호됨).
PII의 제한된 정의.	PII의 확장된 정의.
정보주체에게 사본요청권(Copy request)이 인정됨.	정보주체에게 복사(Copy), 삭제권(deletion), 그리고 데이터 이동권(data portability)이 인정됨.
데이터 프라이버시 영향평가(DPIA)를 할 것이 제안됨.	데이터 프라이버시 영향평가(DPIA)가 요구됨 : 민감정보이거나 대량정보인 경우.
프라이버시 통지(Privacy Notice)가 제안됨 (required with suggestions)	프라이버시 통지(Privacy notice)가 테이블과 특정한 용어를 사용하도록 요청됨
위반사항 통지가 요구되지 않음.	제한된 기한 안에 위반사항 통지.
위반에 대한 패널티가 없음.	위반에 대한 비준수 벌금(fines)이 상당함.
IRON MOUNTAIN, THE IMPORTANCE AND STATUS OF THE GENERAL DATA PROTECTION REGULATION을 참조	



출처: http://creativecommons.org/license.org/by/2.0/uk

이 GDPR의 내용을 2015년 말까지 확정하려는 삼자협상(trialogue)이 유럽 집행위원회, 각료회의, 그리고 유럽의회에서 벌어지고 있다. 아직 GDPR이 확정되지는 않았지만, 각각의 기관들이서 제시한 GDPR안은 기존 EU Directive에 대한 여러 가지 최신의 논의 사항들을 반영하고 있다는 점에서 주의 깊게 살펴볼 필요가 있다. 각각의 GDPR안에 담겨있는 내용을 통해서 개인정보 보호 분야에 있어 최신의 논의동향이 무엇이며 현재 어떻게 논의 중인지를 간접적으로 확인할 수 있기 때문이다. 현재 GDPR을 둘러싸고 첨예하게 논의 중인 쟁점들[118]은 다음의 박스 안에서 확인할 수 있다.

지침(Directive)으로부터 규정(Regulation)까지[119][120]

문제 : 현재 28개의 유럽 구성국들은 1995년 개인정보보호지침(Data Protection Directive)에 근거하여 각 국가 차원에서 법을 제정하고 있다. 독일연방헌법재판소에서 개인정보 자기결정권(informational self-determination)으로 요약된 근본적인 원칙들에 따라, 다양한 법들과 조치들이 EU를 걸쳐 상이한 수준인 개인정보 보호실정에 걸맞게 단행되었다.

해법 : 모두에 대해 동일한 수준의 개인정보 보호를 준수하도록 하는 것이 그 해법이 될 수 있다. 규정(data protection regulation)의 목표는 높은 수준의 개인정보 보호 기준을 달성하는 것이다. 이는 인터넷 시대에 적합하고 잘 어울린다. 단일화된 개인정보 보호 규정은 EU 디지털 싱글 마켓의 일환으로 직접 적용될 수 있으며, 이는 데이터 컨트롤러와 데이터 이용자들이 그들의 권리가 무엇이고 의무가 무엇인지를 훨씬 쉽게 알 수 있도록 도와줄 것이다. 기업들은 더 이상 개인정보 보호 기준이 약한 국가에 그들의 처리를 위한 중앙 데이터 센터를 둘 수 없을 것이다. 더욱이 제안은 EU 개인정보 보호법이 EU시장이 목표가 되고 있는 이상 유럽 내부에서든 외부에서든 유효할 것이라고 예상한다. 더욱 강력한 법집행과 프라이버시 바이 디자인과 같은 개인정보 보호 원칙들은 유럽에서의 개인정보 보호에 대한 시민들과 이용자들의 신뢰를 더욱 강하게 해줄 것이다.

현황 : 유럽연합 집행위원회(EU Commission)는 2012년 1월 그의 법적인 제안을 하였다. 3999 수정안을 이끌어낸 강력한 로비를 한 후에(After intense lobbying which led to 3999 amendments only) 유럽의회(European Parliament)는 거의 만장일치로 2014년 3월에 규정 초안(draft regulation)의 첫 버전을 채택할 수 있었다. 의회에서 EU 구성국들은 오랫동안 답보상태에 있었다. 하지만 2014년 여름 이래로 마침내 그들의 일반적인 접근을 향해 나아갈 수 있었다. 그들은 지금까지 챕터 I, IV, V 그리고 IX에 대하여 합의할 수 있었다. 이는 정책당국(public authorities)을 위한 구체적인 규칙들을 다루었고 다른 특별한 섹터들, 데이터 컨트롤러와 처리자를 위한 의무들을 다루었으며, 나아가 국제적인 데이터 이동에 대해서도 다루었다. 유럽의회(Parliament)와 각료회의(Council) 모두 2014년 여름이 시작하기 이전에 첫 버전에 대한 삼자협상(trilogue negotiations)을 개최하고자 하였다. 그리고 이를 통해 2015년 말까지 입법적인 작업의 결론을 도출하고자 하였다. 이 경우 규정은 모두에게 신법에 대한 적응시간을 주는 유예기간(transition period)에 해당하는 2년 후 모든 EU 구성국들에게 적용될 것이다.

118) EFA, EU General Data Protection Regulation State of play and 10 main issues, 2015 January

주요한 이슈들[121)]

1. 삭제권, 접근권, 정정권(Right to erasure, data access, and correction)

삭제권(Right to erasure), 접근권(data access), 그리고 정정권(correction) : 자신의 개인정보를 지우고자 요청하기를 원하는 자는 누구든지 구글, 페이스북 등과 같은 기업들에 관하여 삭제권을 가져야 한다. 이 경우 데이터 컨트롤러는 이미 데이터를 보내버린 제3자에게 해당 데이터의 삭제 요청을 위하여 필요한 조치를 취하여야 할 것이다. 논란이 되었던 "잊힐 권리"는 의회(Parliament)에 의해 제한되어 왔다. - 오로지 개인정보 보호법을 위반한 채 개인정보를 퍼블리싱한 자들만이 모든 복제본이 삭제되도록 보장할 의무가 있다. 규정은 한편으로는 표현의 자유와 정보의 자유 간 유의미한 균형을 요구한다. 그리고 동시에 개인정보의 보호를 요구한다. 2014년 5월 유럽사법재판소의 구글-스페인 판결에서 언급한 "잊힐 권리"("right to be de-listed")가 언급되는 반면에, 구성국들은 여전히 여기에 대하여 특정한 문구를 더하는 것에 대해 논의를 하고 있다. 더욱이, 보유하고 있던 제공자들은 요청에 따라 빠르게 그리고 비용이 없이 전기적으로 개인정보를 다른 개인에게 넘겨야 한다.

2. '고지된 동의(Informed consent)'

이용자들은 반드시 그들의 데이터에 대하여 무슨 일이 일어나는지, 그리고 그들이 원칙적으로 데이터 처리에 대하여 의식하며 동의할 수 있거나 이를 거절할 수 있어야 한다. 의회(Parliament)가 집행위원회(Commission)에 의해 제안된 "명시적"인 동의를 주장할 때, 의회 버전의 초안인 법은 더욱 많이 막연한 "모호하지 않은" 동의를 예견했다. 이는 사실상 동의를 요청함이 없이 데이터 컨트롤러들(data controllers)에게 값싼 양해를 주게 될 것이다. 웹사이트에 대한 "Do not Track"과 같은 데이터 수집을 거부하는 기술적인 기준들은 EU 레벨에서 자격을 획득할 수 있고 그럼에 따라 일반적인 유효성을 얻을 수 있다. 의회는 데이터 컨트롤러의 " 적법한 이익(legitimate interest)"을 좁게 인정해 왔다. 이런 적법한 이익은 영향 받는 사람들에 의해 합리적으로 기대될 수 있는 것에 대한 동의 없는 데이터 수집과 처리를 가능하게 해 줄 것이다. 다른 한편 구성국들은 심지어 단지 컨트롤러의 "적법한 이익"에 기초한 데이터 처리의 목적을 변경하는 것을 허용해 준다. 이는 집행위원회가 제안해온 것 하에서 개인의 권리를 약화시킬 것이다.

119) Greens/European Free Alliance(EFA), EU General Data Protection Regulation State of play and 10 main issues, 7 January 2015 1p

120) By Biagio Lammoglia, Overview of Roadmap for General Data Protection Regulation, Monday September 21st, 2015

3. 정보에 대한 권리와 투명성(Right to information and transparency)

의회는 EU 집행위원회 보다 정보와 투명성에 대한 더 많은 권리를 요구한다. 만일 제공자가 정보를 공공당국(public authorities) 또는 지성서비스(intelligence services)에 제공했다면 이용자들은 어떻게 그들의 정보가 처리되는지에 대해 이해할 수 있는 정보를 받아야 한다. 데이터 컨트롤러들은 더욱 쉬운 방법으로 이해될 수 있게 무료로 설명해야 한다. 이는 맥락에서 그들이 처리하는 이용자 데이터에 대한 것이다. 이용하는 약관은 이해하기 쉬워야 한다. 의회의 관점에서 표준화된 아이콘들은 프라이버시 정책들에 대한 법적인 언어들이 적혀있는 긴 페이지들을 대체해야 한다.

4. 제3국에 데이터를 이전할 권리(Transfer of data to third countries)

의회는 기업들이 유럽으로부터 직접 제3국들에 데이터를 넘기는 것을 허용해서는 안 된다고 주장했다. 이는 상호적인 법적 조력 협정(mutual legal assistance treaty) 또는 이와 유사한 EU법에 기초한 수단에 의해서만 가능하다. 이러한 유럽 데이터에 대하 외국의 접근에 대한 방어막은 이미 위원회의 제안의 첫 초안에 담겨 있었다. 하지만 이는 미국 정부의 강력한 로비가 있은 후 삭제되었다. 이는 스노든 사건 이후 의회에 의해 되돌려졌다. 구성국들은 국제적인 이전에 대한 챕터의 그들 나름대로의 버전에 이러한 접근법을 이식해 오지 않았다. 하지만 이는 곳 이식될 것으로 보인다.

5. 미래 지향적인 정의(Future-proof definitions)

직/간접적으로 사람에 연결되어 있는 모든 정보는 개인정보 라고 정의된다. 그리고 이는 보호될 필요가 있다. 이는 심지어 빅데이터 시대에서 더욱 중요하다. 이 시기에는 더욱 더 많은 데이터 시트들이 결합되고 분석될 수 있다. 따라서 다른 데이터와 연결되지 않는 가명화된 데이터를 사용하려는 인센티브가 있을 수밖에 없다. 의회는 역시 데이터가 반드시 (심지어 간접적으로) 보호되기 위해서 특정 개인의 정체를 식별할 필요는 없다는 점을 명확히 하였다. - 이는 만일 다수가 소속된 그룹으로부터 개인을 식별(single out)해낼 때 사용될 수 있다.

6. 강력한 제재(Strong sanctions)

불법 데이터 처리 그리고 몇몇 케이스들에서, 기업들은 강력한 제재에 직면하게 된다. 집행위원회(Commission)는 최대 전세계 매출액의 2%가 넘는 제재를 부과받게 된다. 그리고 구성국들은 이것을 고수할 수 있는 자들을 원한다. 의회는 가

능한 제재를 전세계 매출액의 5%까지 올리기를 원했다. 이처럼 강력한 제재는 기업들이 개인정보 보호 위반을 생각함에 있어 낙담하게 한다. 이는 결국 이 기본적인 권리에 대한 이사회의 주의를 보장해준다. 물론, 제재는 항상 부분적이어야 하는 것은 아니다, 그러므로 작은 기업들은 비즈니스에서 밀려날 것을 두려워할 필요가 없으며 개인정보 보호법의 사소한 위반 또는 사고는 또한 두려워할 필요가 없다.

7. 디폴트로서의 프라이버시(Privacy by Design/Privacy by Default)

IT 시스템의 제작자들뿐만 아니라 데이터 처리자들은 그들의 서비스를 데이터 축소(data-minimizing) 과정과 대부분의 데이터 보호-친화적인 사전-셋팅(most data protection-friendly pre-settings)을 하는 방식으로 디자인해야 한다. 목적 제한의 강력한 원칙은 서비스의 제공을 위해 필요한 데이터만이 처리될 수 있다는 의미이다. 의회는 명시적으로 부가적인 데이터 수집 서비스를 제공하는 것에 대한 커플링을 금지해 왔다. 구성국들은 현재 독일의 제안을 고려하고 있다. 이는 서비스 제공자가 그들의 플랫폼들의 이용을 가명화 또는 익명화를 하도록 제안하는 내용이 담겨있다.

8. 요식 행위 감축(Less red tape)

유럽의회에 따르면, 데이터 보호 책임자(Data protection officer, DPO)의 의무적인 임명의 문제는 처리되는 데이터의 양와 그 관련성에 의존해야 한다. 단순하게 기업의 크기에 의존해서는 안 된다. 이런 관점에서 해당 감독 당국에 의한 사전적인 컨설팅은 큰 폭으로 감축될 필요성이 있다. 이는 어떠한 한계점을 넘으면 기업의 데이터 보호 책임자가 의무사항이 되는 것과 마찬가지이다. 각료회의(Council)는 만일 데이터 보호 책임자가 의무적으로 되지 않는다면 구성국에 이를 남겨두어야 할 것이라고 제안해 왔다. 이는 의회(Parliament)에 있어서는 받아들일 수 없는 것이었다. 왜냐하면 이는 또다시 바닥으로의 경쟁을 유발할 것이었기 때문이다. 다른 한편 의회는 DPO가 상근(full-time position)이 아니어도 되며 외부적인 계약자가 될 수도 있다는 점을 명확히 했다.

9. 규칙들의 조화로운 집행(Harmonized enforcement of the rules)

개별 국가들의 개인정보보호의 담당기관들로 이루어진 유럽 개인정보보호 위원회는, 개인정보 보호법의 조화로운 적용을 보장하여야 하고 유럽-차원의 관련성을 지닌 구속력이 있는 결정들을 내릴 수 있어야 한다. 이는 EU 경쟁법과 은행감독에 관하여 이미 어떻게 되었는지와 유사하다. 이러한 방식으로 약한 법집행으로

인한 EU 구성국들의 ‘바닥을 향한 경쟁’의 방식으로는 미래에 적절히 대처하기 어렵다. 의회와 각료회의는 각료회의에게 마지막 말을 주지 않았던 이러한 일반적인 접근법에 동의하였고, 그럼에 따라서 개인정보 보호당국의 독립성은 보장되었다. 하지만 모든 국가들은 여전히 이 “one-stop-shop"의 세부사항이 여전히 논의 중이다. 모든 기관들은 개인정보 보호당국이 더욱 많은 정보들과 더욱 기술적인 전문가와 같은 전문가를 포함한다고 해석된다.

10. 유럽 전체를 위한 하나의 집행기관(One counterpart for all of Europe)

‘one-stop-shop' 접근법은 시민들이 EU 전체를 통틀어 하나의 개인정보 보호당국을 가지게 되는 것을 의미한다. 시민들은 그들 자신의 국가의 개인정보 보호당국에게 EU 어디에서든 벌어진 개인정보 남용을 두고 고소(complaints)할 수 있다. 기업들은 오로지 그들의 주된 설립지(their main establishment)에 있는 당국들에 대처하면 충분할 것이다.

GDPR에 관한 논의 중에서 비식별화와 관련하여 살펴보아야 할 사항은 ‘개인정보 정의’에 대한 부분이다. 개인을 직/간접적으로 식별할 수 있는 정보는 보호되어야 하는데, 그렇지 않은 (비식별화된) 정보는 보호의 대상에서 벗어날 수 있게 되어 이용할 수 있게 되기 때문이다. 특히 익명화와 관련해서 유럽집행위원회의 GDPR안과 EU Directive 사이에는 한 가지 달라진 점이 있다. EU Directive에서는 익명화에 대한 자세한 사항을 행동강령에 위임할 수 있다고 명시해 놓은 반면, GDPR 안에서는 이 부분이 삭제되어 있다는 점이다. 하지만 이 외의 내용들은 동일하다.

EU Directive와 GDPR의 세 가지 안에는 위의 삭제 사항 이외에도 익명화에 대해 규정한 서문의 부분에서 추가된 내용들이 있는데, 이 추가된 부분들에 있어 세 가지 안의 내용은 매우 유사하다. 다만 유럽집행위원회의 GDPR안과는 달리, 각료회의와 유럽의회의 GDPR 안에는 해당 GDPR이 통계 및 연구 목적을 포함하여 식별되거나 식별가능한 자연인(natural person)과는 관련성이 없는 방식으로 익명화가 된 데이터의 처리에는 상관하지 않는다는

121) id, 2p

항목을 추가했다. 기존 EU Directive의 내용과는 큰 차이가 없지만 통계적 분석이나 연구 목적의 경우에는 익명화된 정보의 자유로운 처리가 가능하다는 사실을 명백히 함으로써 이 GDPR의 새로운 체제에서는 익명화 기법의 사용을 좀 더 확대하려는 의도가 있다는 추측을 할 수 있다.

더불어 유럽집행위원회의 GDPR 안 이후 발표된 각료회의와 유럽의회의 GDPR 안에서는 likely to be used to identify 라는 표현이 대두되었다. (*To ascertain whether means are **reasonably likely to be used to identify the individual**, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development.*) 만일 이러한 개연성 내지 가능성(likelihood) 개념이 도입되기 시작한다면, 비식별화 맥락에서 유럽의 법제는 한국의 법제와는 실질적으로 전혀 다른 체계가 되는 것으로 파악할 수도 있으므로, GDPR 논의과정을 보다 관심 있게 지켜볼 필요가 있다. 이에 대한 자세한 사항은 [별첨 2]에서 확인할 수 있다.

(2) 영국 : 행동강령 (Code of Practice)

1) 행동강령 개괄

영국의 정보보호기관(Data Protection Agency)인 Information Commissioner’s Office(이하 ‘ICO’)는 2012년 익명화에 대한 전반적인 내용들을 다룬 안내서를 행동강령의 형태로 출간했다. (ICO, Anonymisation: managing data protection risk code of practice, 2012) 이 행동강령은 상술했던 EU Directive의 서문 26조에 근거하여 마련된 것이다. 서문 26조는 익명화에 대한 구체적인 사항들은 행동강령의 형태로 개별 국가들이 마련할 수 있다고 규정하고 있는바, 동 행동강령은 이러한 서문 26조에 기초하여 마련된 최초의 행동강령이라는 점에서 그 의의가 있다.[122]

122) The Information Commissioner has issued this code under section 51 of the DPA in pursuance

기본적으로 이 행동강령에서 허용되는 익명화의 정도는 식별화 위험성이 영(zero)인 수준이 아니다. 다만 개인정보를 처리하려는 행위주체는 식별의 위험성이 매우 낮은(remote) 수준까지 식별의 위험성을 완화할 수 있어야 한다고 명시한다. 식별의 위험성을 판단하는 기준으로는 합리적 가능성(reasonably likely test) 기준을 채택함으로써 식별의 위험성이 합리적으로 가능할 경우에는 규제의 대상이 되는 개인정보에 해당한다고 한다. 이 기준은 EU Directive가 이미 채택한 기준과 동일하다.[123]

ICO 행동강령은 일종의 안내서이다. 그러므로 이 행동강령 자체가 법적인 강제력을 지니지는 않게 된다. 하지만 이 행동강령에서 규정한 사항들을 준수했다는 사실을 ICO가 확인할 경우 이는 추후 ICO가 개인정보 보호와 관련된 법위반 사실에 대한 조사나 법집행을 할 경우 적극적으로 고려하게 된다. 이런 점에 비추어볼 때, 동 규약은 간접적인 영향력을 지니는 가이드라인의 특성을 지니고 있다고 평가할 수 있겠다. 그 결과 수범자는 동 행동강령에 따라 익명화와 관련된 규정 사항들을 준수한 경우 재식별과 같은 방식을 통해 익명화된 개인정보가 설령 공개될 지라도 이것이 부적절하지 않고 감내할 만한 위험일 수 있다는 확신을 가질 수 있다.[124]

---

of his duty to promote good practice. The DPA says good practice includes, but is not limited to, compliance with the requirements of the DPA. This code was also published with Recital 26 and Article 27 of the European Data Protection Directive (95/46/EC) in mind. These provisions make it clear that the principles of data protection do not apply to anonymised data and open the way for a code of practice on anonymisation. (ICO, Anonymisation: managing data protection risk code of practice, 2012, 10p)

123) The DPA does not require anonymisation to be completely risk free - you must be able to mitigate the risk of identification until it is remote. If the risk of identification is reasonably likely the information should be regarded as personal data – these tests have been confirmed in binding case law from the High Court. Clearly, 100% anonymisation is the most desirable position, and in some cases this is possible, but it is not the test the DPA requires. (ICO, Anonymisation: managing data protection risk code of practice, 2012, 6p)

124) This code gives advice on good practice, but compliance with our recommendations is not mandatory where they go beyond the strict requirements of the DPA. The code itself does not have the force of law, as it is the DPA that places legally enforceable obligations on organizations. Organizations may find alternative ways of meeting the DPA's requirements and of adopting good practice. However, if they do nothing then they risk breaking the law. The ICO cannot take enforcement action over a failure to adopt good practice or to act on the recommendations set out in this code unless this in itself constitutes a breach of the DPA. (ICO, Anonymisation: managing data protection risk code of practice, 2012, 10p)

2) 행동강령의 주요내용

동 행동강령은 개인정보의 이용에 있어 개인정보 보호를 충실히 수행하고 여기에 대해 유발될 수 있는 위험을 관리하는 방식으로 익명화를 소개한다. 우선 동 강령이 어떤 것이며 왜 등장했고 법적인 지위가 어떠한지(Chapter1)를 밝힌다. 그리고 익명화라는 것은 개인정보의 정의와 분리하기 어려운 표리부동의 존재임을 밝힌 후(Chapter2), 개인정보의 익명화가 개인정보 보호에 효과적이라는 점을 설명(Chapter3)한다. 이를 바탕으로 익명화된 정보를 생산하거나 공개할 때에 언제든지 정보주체의 동의가 필요한 것은 아니라는 점(Chapter4)과 공간정보를 개인정보처럼 취급해야 할지를 다룬 후(Chapter5), 개인정보 보호법제에 따라서는 공개해도 되는 정보도 인권법 같은 기타 법령에 의해 공개하지 말아야 하는 때가 있음(Chapter6)을 밝힌다. 끝으로 개인정보 보호를 위해서는 정보의 유형별로 달리 취급될 필요(Chapter7)가 있으며 정보보호를 위한 거버넌스가 중요(Chapter8)하다는 점 및 보호와 이용의 균형을 위한 연구목적상 개인정보 보호법제의 적용예외(Chapter9)에 대해 다룬다.

Chapter1(About this code)의 주요 내용은 5가지로 요약된다. ① 개인정보 보호법은 정보주체가 더 이상 식별되지 않는 방법으로 익명화되었다고 판단되는 데이터에는 적용되지 않는다. ② 개인정보의 익명화는 가능하며 오늘날 현대사회가 직면한 개인정보 이용에 대한 수요를 프라이버시 친화적인 방식으로 충족시켜줄 것이다. ③ 본 규약은 어떤 목적에서든 개인정보를 익명화하고자 하는 모든 조직들에게 도움이 될 것이다. ④ 본 규약은 개인정보의 익명화가 효과적으로 수행되었다는 것을 확신받기를 원하는 사람들에게 반드시 고려해야 하는 이슈들이 무엇인지에 대해 알아낼 수 있도록 도와줄 것이다. ⑤ 본 규약은 개인정보 보호법(Data Protection Act)이 요구하는 법적인 테스트에 초점을 맞추었다.

Chapter2(Anonymisation and personal data)의 주요 내용은 3가지로 요약된다. ① 익명화를 이해하는 것은 개인정보가 무엇인지를 이해하는 것을 의미한다. ② 프라이버시를 보호하기 위해서는 개인정보 그 자체 보다는 익

명화된 데이터를 이용하거나 공개하는 것이 더 좋은 방법이다. ③ 개인정보 보호법을 위반하지 않고 익명화된 데이터를 공개하는 것은 가능하다.

Chapter3(Ensuring anonymisation is effective)의 주요 내용은 3가지로 요약된다. ① 절대적이고 확실하게 재식별 위험을 평가하는 것은 불가능하다. ② 케이스 별 처한 상황에 따라서, 주의 깊게 판단해야 하는 경계선상 케이스들이 많이 있을 것이다. ③ 만일 어떤 주체가 재식별 절차를 통해 개인정보를 만들게 된다면, 그 주체는 데이터 컨트롤러(data controller)로서의 책임을 지게 된다.

Chapter4(Do you need consent to produce or disclose anonymised data?)의 주요 내용은 3가지로 요약된다. ① 동의는 일반적으로 익명화 절차를 적법화하기 위해 필요하지는 않다. ② 심지어 정보주체로부터의 동의를 획득할 수 있을지라도, 여전히 익명화된 데이터를 이용하거나 공개하는 것이 일반적으로 훨씬 안전하다. ③ ICO는 정보주체로부터의 동의를 얻는 것이 매우 어렵고 심지어 불가능할 수도 있다는 것을 이미 알고 있다.

Chapter5(Personal data and spatial information)의 주요 내용은 다음과 같다. 1998년 제정된 개인정보 보호법 아래에서, 우편번호나 GPS 데이터 또는 지도 관련 자료 같은 공간정보를 다루는 것을 다루는 간명한 규칙은 없다. 어떠한 경우에는 이러한 공간정보가 개인정보를 구성할 수 있다. 가령 장소 또는 부동산에 대한 정보가 개인과 연관되어 있는 경우가 있을 수 있다. 물론 그렇지 않은 경우에는 이러한 공간정보는 개인정보가 아니다.

Chapter6(Withholding anonymised data)의 주요 내용은 3가지로 요약된다. ① 개인정보가 아닌 데이터의 경우에, 이를 항상 공개해도 되는 것은 아니다. ② 개인정보 보호법의 개인정보 정의는 데이터가 어느 개인을 식별하지 못하는 상황에까지 연장되어 적용될 수 없다. ③ 유관기관(Public authorities)은 인권법(human rights law)을 준수하는 것을 또한 고려해야 한다.

Chapter7(Different forms of disclosure)의 주요 내용은 4가지로 요약된다. ① 상이한 유형의 익명화된 정보는 상이한 재식별 위험을 제기할 수 있다. ② 간행(publication)은 제한된 접근(limited access)보다 훨씬 위험하다. ③ 제한된 접근은 더욱 풍부한 정보 공개를 가능하게 한다. ④ 제한된 접근은 강력한 거버넌스 약정(robust governance arrangements)을 전제한다.

Chapter8(Governance)의 주요 내용은 3가지로 요약된다. ① 개인정보를 익명화하려는 조직들은 효과적이고 포괄적인 거버넌스 구조가 필요하다. ② ICO는 만일 ICO가 신고(complaint)를 받게 되거나 감사(audit)를 수행해야 할 때 개인정보를 익명화하려는 조직들의 거버넌스에 대해 묻게 될 것이다. ③ 개인정보를 익명화하려는 조직들은 거버넌스 약정들에 대하여 상당한 수준(senior-level oversight)을 확보해야 한다.

Chapter9(The Data Protection Act research exemption)의 주요 내용은 2가지로 요약된다. ① 개인정보 보호법(Data Protection Act)의 연구목적 예외는 연구자들에게 제한적인 하지만 유용한 성격을 지닌다. ② 개인정보를 처리하는 연구자들은 여전히 대부분의 개인정보 보호법뿐만 아니라 그 원칙들을 준수해야 한다.

3) 행동강령 상 재식별 위험성 평가기준

이 행동강령에서 가장 중요하게 다루는 지점은 개인정보를 익명화하는 방식 및 익명화된 데이터를 공개할 경우 감수해야 하는 위험성을 평가하는 방안에 대한 것이라고 보인다. 이런 위험성 평가를 크게 3가지 측면에서 살펴볼 수 있다.

첫째는 관리주체의 측면이다. 강령은 익명화된 데이터를 공개하는 시점과 방식을 두고, 데이터를 공개하는 이유가 무엇인지가 중요하다는 점을 강조한다. 왜냐하면 데이터를 공개하는 이유에 따라 식별화의 원인과 그 결과가 달라질 수 있기 때문이다. 가령 정보의 자유 관점에서 정부가 일반대중에게 정

보를 공개하는 경우 그 정보를 통해 정보주체가 식별될 위험성은 매우 높아지게 된다. 공개된 정보에 접근하고 이를 통해 혜택을 볼 수 있는 사람들이 무작위적이기 때문이다. 반면 연구목적 또는 상업목적으로 정보를 공개하는 경우 그 정보를 통해 정보주체가 식별될 위험성은 상대적으로 낮아지게 된다. 이 경우 정보에 접근하여 혜택을 얻을 수 있는 사람들이 한정될 가능성이 높아지므로 공개된 정보의 유통에 대한 통제력을 쉽게 유지할 수 있기 때문이다.

둘째는 정보유형의 측면이다. 이 강령은 현실적으로 ICO가 식별된 정보의 결과물이 정보주체에 미치는 해악의 규모가 더 큰 사건에 더 많은 관심을 가지게 된다는 점을 은연중 드러낸다. ICO의 규제대상인 개인정보는 생존하는 개인에 대한 정보, 즉 개인정보이다. 그러므로 익명화 측면에서 ICO가 규제를 할지 여부를 고민하는 것은 주로 익명화된 정보가 재식별될 가능성을 판단하는 것이 거의 불가능하거나 애매모호한 경우를 대상으로 할 것이다. 이 때 해당 익명화된 정보의 공개로 인해 공개된 정보주체에게 해악을 미칠 위험성이 있다는 것 자체는 해당 정보가 개인정보에 해당하는지 또는 해당할 수 있는지를 판단할 때 고려되는 요소가 아니다. 하지만 현실적으로 어떤 종류의 정보가 익명화되어 공개되는지 여부는 매우 중요할 수 있다. 어떤 종류의 정보는 다른 종류의 정보보다 잠재적인 공격자에게 더 큰 관심을 받게 되어 집중적인 공격의 대상이 될 수 있고 더 큰 해악의 결과를 가져올 수 있기 때문이다. 그러므로 이러한 정보유형의 측면도 위험성 판단 및 ICO 개입 등을 고려함에 있어 판단기준으로 삼는 것이 합리적이다.

셋째는 잠재적인 공격자의 측면에서 볼 수 있다. 이 강령은 재식별을 의도하는 잠재적인 공격자(intruder)의 측면을 재식별의 위험성을 판단하는 기준에 있어 주요하게 고려한다. 이러한 공격자의 요소를 고려하는 경우 공격자의 공격목적에 따라 재식별의 위험성이 달라진다. 예를 들어, 공격자가 이미 공개된 제3의 개인정보를 가지고 있어 이 개인정보를 통해 익명화된 데이터베이스를 찾는 경우와, 반대로 공격자가 익명화된 데이터베이스를 가지고 공개된 정보에서 대상이 되는 개인정보를 찾으려는 경우를 구별하여 살펴볼 수 있다. 이 두 가지 경우 재식별의 위험성은 각각 다른 확률을 지니게 되며,

이 때 재식별의 위험성을 확실하게 판단하기는 매우 어렵다.

그러므로 재식별의 위험성과 관련된 이런 현실적인 메커니즘을 고려하여 “정보유형에 따른 정보공개의 결과물”과 연결된 “잠재적인 공격자의 의도”를 고려한 “관리적인 측면”의 재식별 위험성 판단기준을 설정할 필요가 있다. 그 결과 동 행동강령은 특정한 익명화 기법의 적용을 통한 문제해결이 아니라 재식별의 위험성을 사후적으로 통제하는 접근법을 사용하여 비식별화의 규제체제를 확립하려는 특징을 보인다. 즉 동 행동강령은 기술적인 접근법이 아닌, 관리적인 접근법을 조금 더 강조하고 있다. 이러한 잠재적인 공격자의 측면을 고려한 판단기준은 “의도된 공격자(motivated intruder)” 기준이라고 불린다. 이 기준은 법률에는 규정되어 있지 않기 때문에 동 행동강령에서 새롭게 규정한 기준이 된다.

“의도된 공격자(motivated intruder)” 기준[125]이 도입되었다는 점은 ICO 행동강령이 가져온 실질적으로 가장 중요한 변화 중 하나라고 할 수 있다. 이는 재식별의 가능성이 무한대로 확장되는 것을 방지해 주는 역할을 하게 되기 때문이다. 동 개념 덕분에 재식별을 시도할 수 있는 모든 가능성이 상정되는 상황은 배제될 수 있게 되었다. 기본적으로 이 기준은 재식별을 시도할 의도가 있는 잠재적인 공격자가 재식별에 성공할 수 있을지에 대한 판단을 포함한다. 여기서의 공격자는 사전적인 인식이 없는 상태에서 특정개인을

125)

Motivated intruder risk: some issues to consider

- What is the risk of jigsaw attack, ie piecing different bits of information together to create a more complete picture of someone? Does the information have the characteristics needed to facilitate data linkage - eg is the same code number used to refer to the same individual in different datasets?
- What other ‘linkable’ information is available publicly or easily?
- What technical measures might be used to achieve re-identification?
- How much weight should be given to individuals’ personal knowledge?
- If a penetration test has been carried out, what reidentification vulnerabilities did it reveal?

Obvious sources of information include

- Libraries
- Local council offices
- Church records
- General Registry Office
- Genealogy websites
- Social media; internet searches
- Local and national press archives
- Anonymised data releases by other organizations, particularly public authorities

(ICO, Anonymisation: managing data protection risk code of practice, 2012, 24p)

식별하기를 원하는 행위자를 의미한다. 이 공격자는 합리적인 수준의 능력을 가지고 인터넷, 도서관, 그리고 모든 공공문서들에 대한 접근성을 가지면서 특정 개인에 대한 추가적인 지식을 가진 사람들에게 질의를 할 수 있는 탐색기술을 가진 인간으로 상정된다. 그러므로 이런 공격자는 어떤 전문가적인 지식이나 해킹 기술을 가질 것을 필수요건으로 하지 않는다. 따라서, 전문기술을 가진 해커는 이 개념에 의해 재식별 가능성의 판단자료에서 배제된다.

이런 전제를 두고 잠재적인 공격자에게 좀 더 매력적인 종류의 데이터 유형을 구분하여, 이런 유형의 데이터를 공격하는 경우 재식별의 위험성이 더 커진다는 전제 하에 개별 사건별로 재식별의 위험성을 평가하게 된다. 더욱 매력적인 종류의 데이터로는 개인에게 수치심을 주는 데이터, 금전적인 이득과 연관된 데이터, 정치적 색채를 가진 데이터 등이 포함된다.

4) 영국 익명화 네트워크(UK Anonymisation Network, UKAN)[126]

영국 익명화 네트워크(The UK Anonymisation Network (UKAN))는 2012년에 익명화에 대한 최선의 관행(best practice)을 구현하는 수단으로 설립되었다. 영국 익명화 네트워크는 개인정보를 취급하고 이를 공유하기를 원하는 누구에게든지 실용적인 조언과 정보를 제공한다. 영국 익명화 네트워크는 ICO에 의해 설립 이후 2년 동안 재정적인 지원을 받았다. 또한 Manchester 대학, Southampton 대학, Open Data Institute (ODI) 및 Office for National Statistics (ONS) 등 네 개의 기관들로 구성된 콘소시움에 의해 공동운영(co-ordinated) 된다. 영국 익명화 네트워크의 목적은 광범위한 이 분야 숙련된 전문가들로부터 익명화에 대한 최선의 관행을 모으고 형성하여 대중의 신뢰를 얻고 프라이버시 위험을 최소화하며 데이터의 가치를 극대화하는 것이다. 영국 익명화 네트워크는 주기적으로 워크숍을 열고 경험을 공유하며 활발히 활동하고 있다.

126) http://ukanon.net/

다음 단락에서는 이 행동강령에서 소개된 비식별화의 구체적인 사례에 대해서 설명할 예정이다. 이 사례들은 ICO가 다양한 기업들을 대상으로 비식별화 사례들의 제출을 요청해서 받은 것이다. 그래서 이 사례들은 관련 업체에서 실제 사용하는 비식별화의 현 상황에서부터 재식별의 위험성에 대해 기업들이 가지는 생각들까지 광범위한 익명화와 관련된 문제점들을 추축할 수 있다는 점에서 논의할 실익이 있다.

5) 케이스 스터디[127]

① 의약정보에의 제한된 접근[128]

임상 분야의 연구에서는 의료분야의 전문가인 임상조사관(investigator)은 오로지 코드화(coded)된 데이터만 그 연구를 지원하는 의약회사에 보고한다. 그래서 어떤 개인정보도 공개되지 않도록 하고 있다. 코드화된 정보를 해독할 수 있는 키는 연구 장소에 있는 임상조사관이 보유하고 있는데 이 조사관은 연구대상의 신분을 공개하지 않을 의무를 부담하고 있다. 해당 연구를 지원하는 회사들은 코드화된 정보를 외국에 있는 제휴회사나, 공동연구 주체들, 그리고 전세계의 보건당국들과 공유할 수 있다. 하지만 어떤 경우에서도 해당정보를 받은 자는 그 정보의 재사용과 재식별에 대한 제한과 비밀엄수의 의무를 가지게 된다. 이런 의무를 부여하는 근거는 계약에 의한 것이든 법령에 의한 것이든지를 불문한다. 이런 안전책들이 마련된 상태에서 연구를 지원한 제약회사가 코드화된 데이터를 제 3자에게 제공할 경우 그 데이터가 재식별될 위험성이 현저하게 낮아진다.

② 핸드폰을 이용한 교통속도 연구[129]

교통속도를 연구하기 위해서 핸드폰에서 생성되는 위치정보가 이용된다. 위

127) 아래 제시된 사례들은 모두 별도의 설명이 없으면 ICO 익명화 행동강령에 포함된 사례들로, 이 자료에 포함된 내용("Annex 2 - Anonymisation case-studies ") 일부를 요약하여 소개한다.

128) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 1: limited access to pharmaceutical data.

129) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 2: using mobile phone data to study road traffic speeds.

치정보는 개인을 식별할 수 있는 잠재적 위험성이 높은 정보이기 때문에 규제의 대상이 되는 개인정보로서 더한 주의가 필요하다. 이런 관점에서 위치정보 또한 익명화된 정보로의 변화를 통한 활용의 필요성이 상당히 크다. 모바일 서비스를 제공하는 통신회사들은 방대한 분량의 가입자 목록들을 보유하고 있는데, 개별 목록의 대표적인 정보 범주는 (1) 핸드폰번호 (2) 근사적 위치, 그리고 (3)시간을 포함한다.

이 항목들 중의 위치정보는 도로 위에서 운행하는 자동차들 안에 있는 핸드폰과 연결된다. 도로의 특정한 두 지점 사이에서 자동차 안에 있는 핸드폰이 이동하는 속도의 산정 결과를 해당 도로에서의 교통 속도로 볼 수 있기 때문에 이 핸드폰의 이동 속도를 분석하면 간접적으로 교통 속도를 확인할 수 있다. 그래서 교통속도의 연구 관점에서 통신회사가 보유하는 개별 가입자의 위치 정보를 확보할 실익이 있는 것이다. 하지만 이런 위치 정보를 연구 목적으로 공개하려면 이 정보에 함께하는 핸드폰번호나 기기번호 등이 같이 공개되어야 할 것이기 때문에 이런 정보를 연구하는 기업은 개인정보보호의 규제를 받을 가능성이 높아지게 된다. 그래서 이와 같은 연구 목적에 따라 위치정보를 필요로 할 경우에도 익명화를 통해 개인정보보호의 규제 대상에서 제외되려는 인센티브가 있다.

결국 이런 연구에서의 데이터 분석을 위해서는 핸드폰번호를 드러내지 않는 익명화 작업이 필요해진다. 어떤 방식을 적용해서 익명화를 하는 것이 좋을지에 대해 해당기관은 언제나 도로교통 속도의 측정이란 연구의 목적을 고려하지 않을 수 없다. 만약에 단순하게 핸드폰 번호를 삭제할 경우에는 이런 연구의 목적에 비추어 데이터의 가치 자체가 상실된다. 핸드폰 번호 자체를 지우게 되면 어떤 핸드폰인지 자체를 구별할 수 없기 때문에 두 지점에 있는 핸드폰이 동일한 핸드폰인지 등에 대한 기본적인 정보 자체를 알 수 없기 때문이다.

그래서 해당 연구기관은 원래의 핸드폰 번호에 대해 일대일로 대응될 수 있는 가짜(dummy) 핸드폰번호로 대체하는 것과 같은 익명화 기법들을 고려하게 된다. 이 가짜 핸드폰번호는 무작위로 부여된 번호이기는 하지만 이런

무작위로 부여된 번호에서 원래의 전화번호를 확인할 수 있는 암호키(cryptographic key)나 참조표(mapping table)를 참조하면 원래의 핸드폰 번호에 대응될 수 있는 접근법인 것이다. 물론 이런 암호키나 참조표의 보관에는 철저한 보안이 필요한 것은 당연하다.

③ 탑승객들의 여행시간 분석[130)]

대중교통업체는 'Go-Card'데이터를 사용하여 특정 연령층의 탑승객들이 다양한 여행들을 위해 소비하는 시간들에 대한 조사를 수행한다. 이 조사에서 익명화된 데이터를 사용해서 관련 데이터들을 분석한다. 여기서의 데이터는 가명처리 또는 해시(hash)처리와 같은 기법들을 사용해서 효과적으로 익명화된 데이터이다. 아래의 <표 A>은 원래의 데이터이고 <표 B>는 해시처리 등이 된 데이터이다.

<표 A>

Go-card no.(탑승객)	생년월일	출발지	도착지	여행시간
WT98765G	01/09/1973	Brooks End	Tree Street	17m 45s
WT45678B	18/09/1933	Brooks End	Tree Street	15m 05s

⇩

<표 B>

ref. no.	나이 구간	출발지	도착지	여행시간
14793X…	35-45	Brooks End	Tree Street	18m
23955P…	75-80	Brooks End	Tree Street	15m

④ 대중에게 공개된 정보와 익명화 위험성[131)]

130) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 3: analysing passengers' journey times.

131) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 4: publicly available information and anonymisation risk.

대중에게 공개된 정보가 특정 데이터의 재식별을 위해 사용되는 경우가 상당히 존재한다. 이런 정보의 대표적인 예가 투표자등록명단(electoral register entry)이다. 이 투표자등록명단은 공개명단의 형태로 외부의 개인이나 기관에게 판매가 될 수 있다. 이 명단에는 투표 자격을 갖춘 대상의 이름과 주소가 포함되어 있다. 예를 들어, 이 명단에 다음과 같은 기록이 포함되어 있다고 가정해보자.

이름: K L Thomas
주소: 1 Sandwich Avenue, Stevenham, SV3 9LK

이렇게 공개된 정보와 부동산 웹사이트에 있는 건물을 대조해서 K L Thomas의 거주지 건물의 시장가치와 같은 재산정보도 파악할 수 있다. 위험성을 영(0)으로 만드는 익명화는 존재하지 않기 때문에 이과 같이 공개된 다른 정보와의 대조를 통해 식별화되는 위험성은 상존한다.

영국의 경우에는 투표자등록명단을 두 가지 형태로 운영하고 있다: 그 중 하나가 '공개명단(open register)'이고, 다른 하나는 '완전명단(full version)'이다. 투표자로 등록을 해야 하는 개인은 자신의 정보를 공개명단의 형태로 할 수도 있고 완전명단의 형태로 할 수도 있다. 공개명단의 경우에는 외부의 어떤 개인이나 기관에 사용목적을 불문하고 판매가 가능하다. 반면에 완전명단의 경우에는 선거나 범죄 방지 차원 등의 특정 목적을 위해서만 공개가 된다. 명단에 포함된 정보는 이 두 명단 모두 동일하다. 이런 맥락에서 보면 이 사례에서는 기본적으로 공개명단을 전제로 한다고 생각할 수 있다.

⑤ 개인정보의 공개에 대한 ICO의 실제 결정 사례 (FS50161581)[132]

2007년에 경찰은 Daisy Land와 Iris Drive에서 2004년에서 2006년 사이

132) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 5: a summary of a freedom of information decision notice relating to the disclosure of personal data.

에 발생한 다수의 주거침입(burglary)사건들에 대한 정보 공개의 요청을 받았다. 이 요청에 대해 경찰은 '정보자유법(Freedom of Information Act(FOIA)'의 제 40조에 근거해 거부했다. 이 조항에 따르면 정보공개가 정보보호의 원칙에 위반될 경우에는 개인정보 공개를 하지 않을 수 있다. 경찰이 정보공개의 요청을 거부하자 ICO에 이 거부행위에 대한 청원(complaint)을 하였고, ICO는 경찰의 이런 거부행위가 FOIA에 위반하는지 여부에 대한 판단을 하게 되었다.

ICO는 해당 통계를 평가한 후 이 통계 자체로부터는 개인을 식별하는 것이 가능하지 않다는 입장을 취하게 되었다. 그런 다음 개인을 식별하게 할 가능성이 있는 기타 외부 요소들과 정보의 맥락에서 그 통계를 살펴보았다. Daisy Lane에는 13 개의 건물들과 Iris Drive에는 83개의 건물들이 있었다. ICO는 해당 통계와 관련성이 있는 지역에 포함된 건물들의 수가 상대적으로 매우 적다는 사실을 확인했다. 이렇게 가능한 지역들의 수가 작다면 그만큼의 오차 범위도 줄어들기 때문에 통계를 통해 특정 지역을 추측할 수 있는 가능성은 높아지는 것은 예상할 수 있다. 하지만 이런 사실에도 불구하고 ICO는 여전히 요청된 정보가 어떤 개인도 식별할 수 있도록 하지는 못한다는 입장을 견지했다. 즉, ICO는 해당 통계가 개인의 식별가능성이 있는 정보가 아니기 때문에 경찰이 정보공개의 요청을 거부할 수 있는 예외조항인 FOIA의 제 40조에 해당하지 않는다는 판단을 했다.

이런 FOIA의 의견에 대해 그 경찰은 그 지역에 대한 지식을 가지고 있는 개인들은 이런 통계에서 확인할 수 있는 정보를 가지고 개인들을 식별할 수 있다고 주장했다. 하지만 이런 주장에 대해, ICO는 요청된 통계정보의 공개가 개인의 식별로 이어질 증거는 없다고 판단을 했다. 결국, 그 해당 통계는 개인정보가 아니기 때문에 40조에 근거한 공개의 예외에 해당하지 않는다는 결론을 내리게 된다.

이런 ICO의 판단에서 주목할 점은 익명화된 정보인지 여부를 판단하는 것은 어떤 일률적인 기준을 통해서 하는 것이 아니라 개별 데이터의 구체적 정황들과 관련 요소들을 통해서 구체적으로 판단한다는 사실이다. 특히 ICO가

강조하는 기준인 의도적 공격자 기준은 일의적으로 적용되는 것이 아니고 개별 사안에 따라 구체화하여 적용된다고 볼 수 있다.

⑥ 정성적(qualitative) 정보의 익명화[133)]

데이터 분석의 일반적인 대상은 정량적인(quantitative) 데이터이지만 이런 양적인 수치가 포함되지 않는 정성적인(qualitative) 데이터도 분석의 대상이 되기도 한다. 정성적인 데이터는 양적인 숫자가 아닌 설명적 또는 묘사적(descriptive) 표현들이 그 대상이 된다. 예를 들어, 키가 170cm인 것은 정량적 데이터이지만 바지가 검은 색인 것은 정성적 데이터가 된다. 정량적 데이터의 분석에 비해 정성적 데이터의 분석은 아직 많이 활성화되지 않았지만 빅데이터 분석 기술이 발전함에 따라 기술적이고 묘사적인 개념들을 분석할 수 있는 기법들이 나타나고 있다. 예를 들어, 텍스쳐 마이닝(texture mining)이 문장들 속에서 사용되는 기술적인 단어들을 분석하는 대표적 방식이 된다.

그러므로 빅데이터 기술의 발전에 따라 정성적 정보의 활용의 정도도 더욱 높아질 것이 예상되기 때문에 정성적 정보의 경우에도 정량적 정보의 경우와 동일하게 공개나 활용을 위해서는 익명화의 필요성을 가지게 될 것이다. 정성적 정보가 어떤 방식으로 익명화가 될 수 있는지를 보여주기 위한 한 가지 예로 한 어린아이와의 인터뷰 내용이 어떻게 익명화가 되는지를 아래의 그림에서 표현한다.

<원문>

Interview recorded : 3pm, 10 October 2011
Interviewee : Julius Smith
DoB : 9 September 2005
School : Green Lanes Primary School

133) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 6: anonymising qualitative data.

I live on Clementine Lane so I walk to school every day. I live in a flat with my parents and my Uncle Jermaine. When I get home from school I watch TV. I don’ t like reading but I like watching Harry Potter films. My favourite subject at school is art. My teacher is Mr Haines and he is very nice. I used to get bullied by Neil and Chris but I told Mr Haines and they stopped.

I play football for Junior Champs, and we are good. I play midfield.

<익명화된 전문>

Interview recorded : October2011
Interviewee ref : 2011/67
School year : Key Stage 1
School local authority area : Lynenham District Council

I live in [LM51 post code] so I walk to school everyday. I live with [family members]. When I get home from school I watch TV. I don’ t like reading but I like watching Harry Potter films. My favourite subject at school is art. My teacher is Mr[teacher’ s name] and he is very nice. I used to get bullied by [other pupils] but I told [the teacher] and they stopped.

I play football for [a local team], and we are good. I play midfield.

위의 예에서 보면 익명화된 전문에서 밑줄 친 부분이 원문과 다른 부분이다. 이 전체적인 인터뷰 내용에서 등장한 사람의 이름이나 특정 지역의 명칭에 대해 삭제를 하거나 특정 우편번호로 대체를 한 것을 알 수 있다. 또한 인터뷰한 사람의 이름, 인터뷰날짜, 생년월일, 학교이름 모두를 특정 숫자나 좀 더 일반화된 표현으로 바꾼 것을 확인할 수 있다. 이와 같이 정성적 정보는 전체 텍스트의 문맥들을 한 번에 확인할 있는 특징이 있다는 사실도 확인할 수 있다.

⑦ 재식별의 위험성 평가에 대한 사전 지식의 중요성[134)]

재식별의 위험성을 평가하는 기준인 의도적 공격자 기준은 잠재적 공격자로 누가 될 수 있는지를 판단하는 것에 있어서 개별 공격자가 가지고 있는 사전적인 지식의 수준을 중요한 요소로 고려한다. 이런 관점에서 실제 사례에서는 이 사전적인 지식이 어떻게 고려되는지를 살펴볼 필요가 있다. 어떤 회사 특정 질병을 유발하는 화학 물질에 근로자가 노출되는 문제에 대해 연구를 수행한다고 가정해보자. 아래의 그림 중 위 쪽에 있는 그림은 기업이 보유하고 있는 개별 근로자의 원래의 정보에 해당하고 아래쪽에 있는 그림은 이 연구를 위해 추출되고 익명화된 정보에 해당한다.

Human Resources summary employee record:

Employee name: F Gradwell
DOB: 01/09/1973
Sex: M
Address: 16 Tree Street, Stevenham, SV8 6QP
Start date: 11/06/1992

Anonymised research database extract:

Age: 39
Sex: M
Postcode: SV8 6QP
Period of service: 20 years 5 months
Contact dermatitis: Positive

위 기업이 위의 근로자에 대한 정보를 기업 외부의 일반 대중들에게 공개할

134) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 7: this shows the importance of third parties' prior knowledge in assessing re-identification risk and illustrates some means of reducing this risk.

가능성은 거의 없다. 즉 이 근로자의 정보는 대중에게 공개된 정보가 아니다. 그런데, 위 근로자의 동료, 친구, 가족들은 이 근로자의 생년월일, 주소, 그리고 입사한 일자(또는 대략의 근사치)를 사전적으로 알고 있을 가능성은 상당히 존재한다. 또한 이 근로자 자신이 자발적으로 이런 자신에 대한 정보를 페이스북과 같은 소셜미디어 웹사이트에 올릴 수도 있다.

이와 같이 다양한 방식으로 이 근로자에 대한 사전적 지식이 형성될 수 있다. 즉 어떤 개인이 해당 정보를 원할 경우, 그들은 사전적 지식과 함께 데이터들을 결합해서 상당한 수준의 가능성으로 이 근로자가 특정 질병을 가지고 있다는 사실을 추론할 수 있다는 사실을 의미한다. 이런 과정을 통해서 근로자에 대한 정보가 외부에 공개가 되지 않아도 주변의 지인들이나 자발적인 행위를 통해 사전적인 지식으로 될 수 있기 때문에 위와 같이 익명화된 연구 목적의 데이터가 아래와 같이 재식별될 수 있다.

Employee name: F Gradwell
DOB: 01/09/1973
Sex: M
Address: 16 Tree Street, Stevenham, SV8 6QP
Start date: 11/06/1992
Contact dermatitis: Positive

이런 관점에서 보면 개인의 식별성을 약화시키는 방향으로 위에서 익명화된 데이터를 일반화, 범주화 등의 방식을 적용해서 익명화 정도를 강화할 필요성이 있다. 아래의 그림과 같이 익명화 정도를 강하게 하면 사전적 지식이 존재한다고 하더라도 데이터들의 결합을 통한 재식별의 위험성은 상당히 감소하게 된다. 물론 익명화를 강화할수록 데이터의 활용 가치는 감소하기 때문에 익명화 과정에서 연구 목적을 충족할 수 있도록 데이터의 가치를 어떻게 확보할 것인지 고민할 필요가 있다.

Age range: 35-45
Sex: M
Location: Stevenham
Period of service: 18 - 22 years
Contact dermatitis: Positive

예를 들어, 이 연구목적이 비율과 같은 특정한 통계 수치 데이터의 산출에 있다면 아래와 같은 데이터만 공개해도 데이터의 익명성과 활용성 모두를 충족하는 경우가 된다.

Stevenham branch: 15% of male employees with 18 - 22 years' service have contracted dermatitis.

⑧ 고객 분석[135)]

이 고객 분석의 사례는 빅데이터의 분석을 분석전문기관이나 업체에 위탁을 할 때 발생하는 개인정보의 보호 문제에 대한 것이다. 빅데이터 분석을 의뢰하는 업체나 위탁을 받는 업체 사이에는 흔히 개인이 식별되는 정보가 포함된 데이터가 이동되기 때문에 그 경우 개인정보보호의 규제 대상이 되고 그에 따라 익명화의 문제도 발생한다. 아래의 예를 통해 이에 대한 구체적인 이해가 가능하다.

한 소매업체가 판매 감소를 겪고 있다. 이 상황을 타개하기 위해서는 판매 증진을 위한 고객의 요구사항들을 이 소매업체는 좀 더 잘 이해할 필요가 있다고 판단했다. 판매 증진을 위한 고민 끝에 이 소매업체는 과거의 특정 시점에서의 거래 데이터에 대한 분석을 통해 고객이 이 소매업체에서 실제

135) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 9: customer analytics.

로 구매하고 있는 것들을 더 잘 이해할 수 있다고 판단을 한다. 하지만 이 소매업체는 데이터 분석을 위한 자체적인 전문 인력이나 기술을 가지고 있기 않기 때문에 외부의 전문 데이터 분석업체에게 데이터 분석을 의뢰하기로 결정한다.

그런데 이와 같이 외부의 업체에게 데이터 분석을 위탁하기 위해서는 결국 데이터 분석을 위해 소매업체가 보유한 원데이터(raw data)가 외부의 업체에게 이전되어야 한다. 그런데 신용카드의 지불에 대한 규제에 따라 이 원데이터를 외부에 이전하는 것이 금지되어 있다. 결국, 이 소매업체는 이런 규제의 제한을 피하기 위해 해당데이터에 포함되어 있는 신용카드데이터를 익명화하기로 결정으로 하고 그 익명화의 방식으로 암호화(encryption)를 적용한다. 이런 암호화 방식이 적용되어 익명화가 된 정보는 더 이상 관련 규제의 대상이 되지 않기 때문에 이 소매업체는 자유롭게 외부 데이터 분석업체에 원데이터를 이전하면서 데이터분석을 의뢰할 수 있다.

이런 분석 작업을 통해 산출된 통계적 정보는 개인을 식별할 수 있는 정보가 아니기 때문에 자유롭게 데이터 분석업체에서 이 소매업체로 이전이 된다. 이런 과정을 통해 이 소매업체는 구매자의 판매 형태를 파악할 수 있게 된다. 이 예는 익명화 기법이 기업의 경제적 행위에 어떤 영향을 미칠 수 있는지도 보여준다.

⑨ 범주화(Suppression) 기법이 데이터에 적용되는 방식[136)]

여론조사에 경우 표현한 질문의 내용에 따라 질문에 답한 형태, 내용 등을 통해 개인의 신분이 드러날 위험성이 있다. 이런 위험성 때문에 여론조사를 위한 설문지를 작성할 경우 설문지를 통해 여론조사에 응한 개인이 식별되지 않도록 주의할 필요성이 있다. 특히 여론조사의 경우에는 질문이나 주어진 문항들을 통해 누가 설문을 했는지를 알 수 없게 하기 위해 이 질문이나 관련 응답 문항들의 구체성을 약화시키는 범주화 방식들을 적용하는 것이

136) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 10: suppression Rules Applied to Data for the Longitudinal Study of Young People in England.

일반적이다.

여론조사에 있어서 어떤 항목을 범주화하고 어떤 항목은 범주화를 하지 않아도 되는지를 여러 요소들을 고려해서 판단하게 되는데 일반적으로 다음과 같은 정보들은 범주화의 대상이 된다.

범주화 대상이 되는 항목	특징	예시
민감성	매우 민감한 항목들은 범주화될 필요가 있다	성관계한 상대방의 수, 성적 지향성
낮은 수치	특정 질문에 대한 전반적인 응답수들이 적어지면 질문 자체가 식별의 가능성을 높이게 된다.	설문에 응한 1000명들 중에 특정 질문에 대한 응답수가 200 이하인 경우
식별이 가능한 기타 경우	민감한 내용을 내포하지는 않지만 낮은 정도로 식별이 가능한 경우가 있다. 해당 항목에서 파생된 항목들에 의해 동일한 정보가 계속 반복되는 경우 등이 해당된다. 이런 반복을 통해 원래의 항목을 통한 식별가능성이 높아질 수 있기 때문이다.	현재의 직업의 수, 주 단위당 근로시간

⑩ 암호적 해시(cryptographic hash) 기법[137]

정보 보안과 관련해서 많이 사용하는 통계학적 기법으로 암호적 해시 함수를 적용한 방식이 있다. 암호적 해시 함수는 원래의 텍스트 데이터에서 생성된 해시 수치만으로는 실질적으로 그 텍스트 데이터를 다시 생성하는 것이 불가능하게 하는 해시 함수를 의미한다. 즉 암호적 해시 함수는 이 해시 수치만으로는 원래의 데이터를 추출할 수 없게 하는 통계적 기법을 의미한다. 이런 특성 때문에 암호적 해시 함수는 원래의 데이터에서 해시 수치는 쉽게 산출할 수 있지만 그 해시 수치로부터 원래의 데이터를 추출하거나, 동일한

137) ICO Code of Conduct, Annex 2 - Anonymisation case-studies, Case study 11: cryptographic hash technique.

해시 수치로부터 상이한 데이터들이 추출될 수는 없는 특징을 가진다.

해시는 이와 같은 일방향적인 특성을 지니고 있기 때문에 비식별화된 정보를 재식별하지 않도록 하는 메커니즘의 설계에 많이 사용되고 있다. 이 암호적 해시 함수의 적용을 통한 익명화 데이터를 생성하는 과정은 다음과 같이 요약하여 예시할 수 있다. 우선 D를 익명화를 원하는 개인정보라고 정하고, K를 데이터관리자(data controller)만 아는 암호키라고 하자. 그런 후에 암호적 해시 함수 H를 선택한다. 함수 H의 선택에 있어서 주의할 점은 사용하는 해시 함수가 충분하게 보안성이 유지되어야 한다는 사실이다. 예를 들어, 이미 외부의 공격에 취약성이 노출된 함수 알고리즘은 더 시상 H의 선택 대상이 되어서는 안 된다. 이과 같이 D, K, H를 선택한 후 H에 의해 처리되는 K와 D의 해시 수치를 산출한다. 이런 H의 적용을 통해 생성된 해시 수치가 익명화된 정보로서 원래의 개인정보를 대체한다. 이런 과정을 통한다면 공개된 해시 수치의 정보만을 알 수 있는 외부자들은 이 해시 수치만으로는 원래의 개인정보를 추출하는 것이 거의 불가능하다. 하지만 해시 수치 이외에 암호키를 보유한 데이터관리자는 개인정보를 추출할 수 있다. 그러므로 이 방식은 보안성 유지를 통한 개인정보의 보호가 가능하면서도 사용을 허가받은 개인은 식별성을 갖춘 데이터를 활용할 수 있게 한다는 특징이 있다.

(3) EU : Article 29 Data Protection Working Party Opinion

1) 주요 내용

위에서 설명한 익명화에 대한 ICO의 행동강령이 2012년에 발행된 이후, 2014년 EU Directive의 제29조 연구반(Working Party)이 익명화에 대한 공식적인 의견서(Opinion)[138]를 발표하였다. 이 연구반이 발표하는 의견서는 개인정보의 보호와 관련한 논의 사항들에 대한 것으로서, EU 내에서는 매우 공신력이 높은 견해로 간주되고 있다. 그래서 익명화 기술에 대한 이 의견서 또한 비록 형식적으로는 법적인 구속력을 가지는 것은 아니지만 EU

138) EU Commission, ARTICLE 29 Data Protection Working Party Opinion 05/2014 on Anonymisation Techniques (2014)

회원국들 각국의 개인정보 보호기구(DPA) 수장이 참여하여 논의한 결과를 바탕으로 한 견해이기 때문에 이 의견서에서 설명하는 내용은 실무적으로는 매우 영향력이 높다.

이 의견서는 개인정보 유출의 위험성을 완화함과 동시에 개인과 사회를 위한 공개된 데이터의 혜택을 누리기 위한 전략으로서 익명화의 잠재적 가치를 인지하고 있다. 하지만 동시에 데이터로서의 효용가치를 유지하면서 동시에 진정 익명화된 데이터베이스를 만들어내는 것은 매우 어렵다는 사실도 인식하고 있다. 이런 현실적인 인식 하에서 동 의견서는 익명화의 기술적인 방식들에 대한 개별적인 설명을 의견서의 첨부문서로 덧붙이는 형태가 아닌 본문의 일부로 포함하여 본격적으로 논의하고 있다. 이러한 기술적인 방식들을 크게 무작위화(randomization)와 범주화(generalization)로 구별한 후 차별적인 프라이버시(differential privacy), 소음추가(noise addition), k-익명성(k-anonymity), l-다양성(l-diversity), t-근접성(t-closeness) 등과 같은 기술적, 통계적 기법들을 소개한다. 개별적 기술의 내용과 함께 그 기술의 장점과 단점, 그리고 공통적인 실수 또는 오류 등도 같이 밝히고 있다.

의견서는 각각의 기술적 방식들의 적용가능성을 세 가지 기준으로 설명한다. 즉, (1) 개인을 구별(signaling out)해낼 가능성, (2) 개인과 관련된 기록물에 연결(linking)될 가능성, 그리고 (3) 개인에 대한 사실을 추론(inference)할 가능성이다. 그 결과 이런 세 가지 위험성에 대한 해결책이 재식별의 위험성에 대한 해결책이 된다. 이와 같이 재식별의 위험성에 대한 접근을 상당히 기술적인 설명을 통해 접근함으로써 동 의견서는 상대적으로 기술적 방식에 의한 접근법을 선호할 수 있다는 인상을 줄 수도 있다.

하지만 개별적으로 나열된 기술적 방식들의 적용 여부는 구체적인 맥락에 따라서 달라진다는 사실 또한 강조되어 있다. 각각의 기법들이 지닌 장점과 단점이 모두 개별 사건의 맥락에 따라 적용이 달라지기 때문이다. 그 결과 어떤 하나의 기술적 방식이 언제나 우월적인 기법이 되는 것이 아니라, 해당 맥락에 따라 최적의 기법들이 달라진다. 그러므로 동 의견서의 결론에 따르

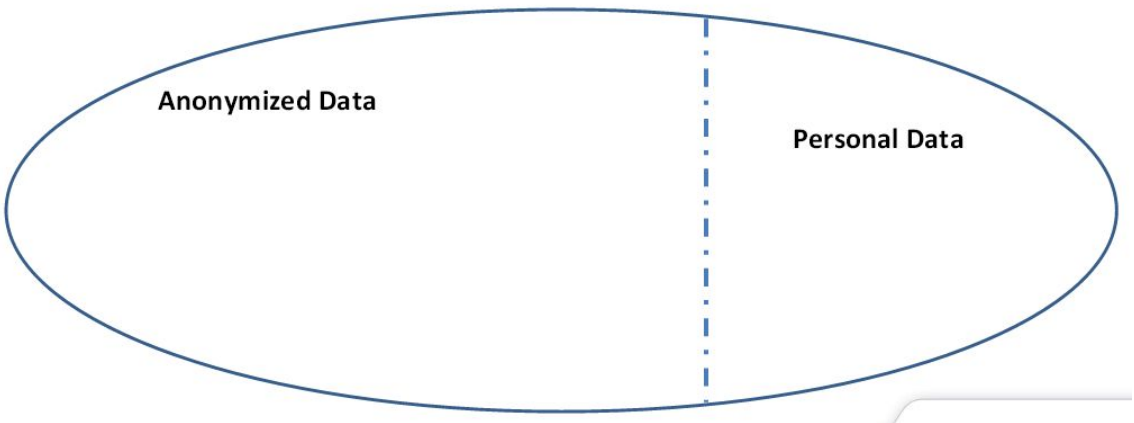
면, 익명화 기법들이 프라이버시의 보호책을 제공함으로써 효과적인 익명화 절차를 창출하는 것에 이용될 수 있게 된다.

그런데 이런 기술적인 접근법이 의도된 효과를 내기 위해서는 익명화 절차의 맥락과 목적이 명확해야 한다. 즉 재식별의 위험성과 관련된, 나아가 비식별화의 규제 체제와 관련된 최적의 해결책은 사안별(case-by-case)로 결정되고, 해당 사안에 따라 위에서 설명한 기법들을 혼합해서 사용하는 것이 최적의 해결책이 될 수 있다. 그러므로 동 의견서는 기술적은 접근법의 가치를 인정하면서도 개별적인 맥락의 중요성에 주목하고 있다. 즉, 사안별 해결책과 절차를 강조하는 관리적 측면 또한 함께 고려되어야 한다는 것이다.

2) 시사점[139]

위에서 본 EU 제29조 작업반 의견서는 익명화의 범위를 한정할 필요가 있다고 주장한다. 개인정보(Personal Data)와 익명정보(Anonymized Data)를 이분법[140]적으로 구분하여 생각하는 경우를 흔히 볼 수 있다. 이분법적으로 파악할 경우 그 사이에 중간영역은 없으며, 익명화된 정보는 개인정보 보호법령의 적용을 받지 않게 된다. 이 때 데이터관리자(data controller)들은 적지 않은 경우에 개인식별정보를 제거하면 익명화가 된다고 생각하는

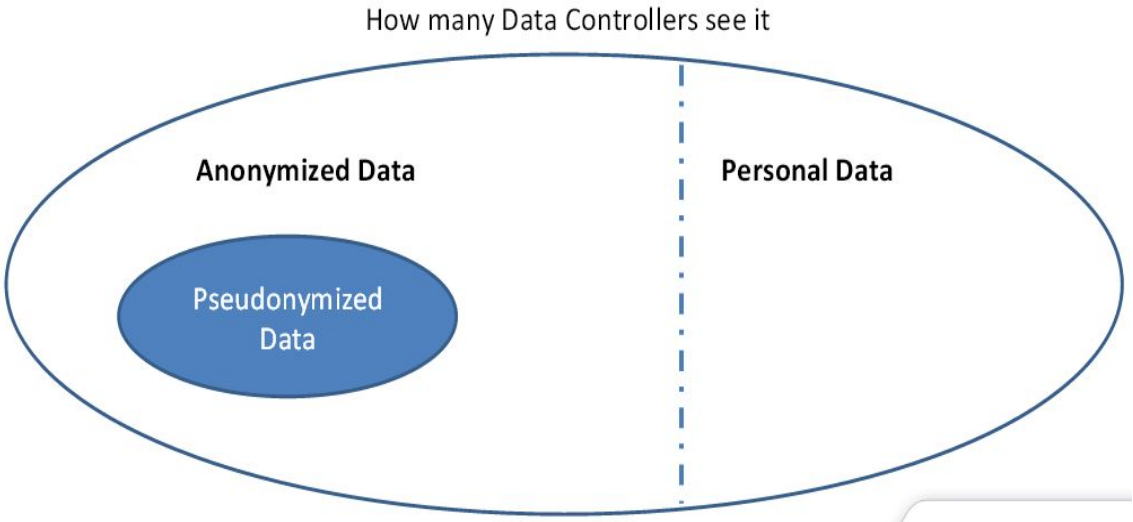
139) CNIL, WP 29 Opinion on anonymization techniques(ppt), 29th May 2015, http://www.phusewiki.org/docs/Paris%20SDE%202015%20Presentations/The%20CNIL's%20Persepctive%20-%20Data%20Anoymisation.pdf



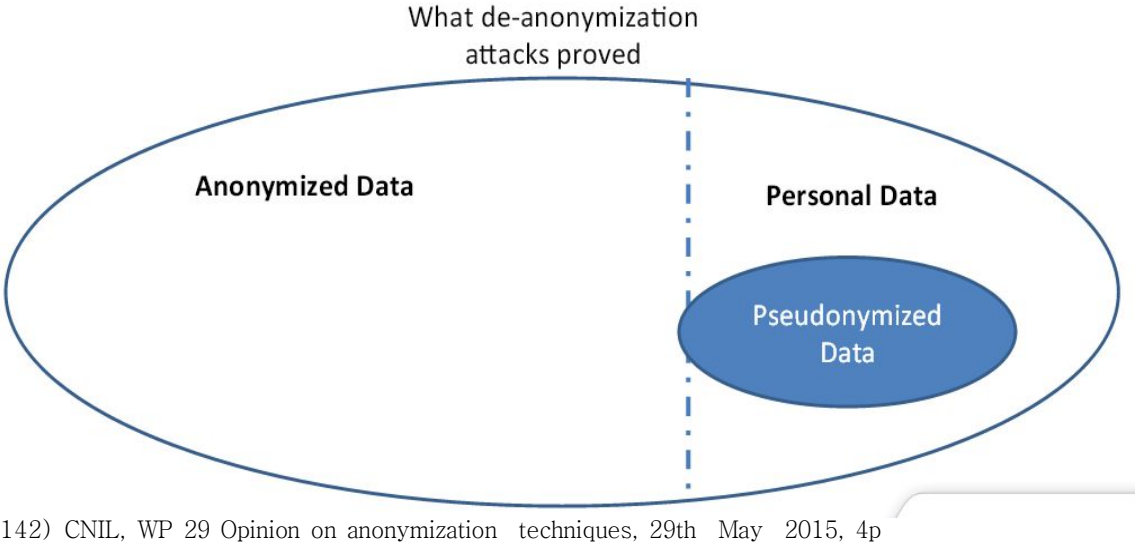
140) CNIL, WP 29 Opinion on anonymization techniques, 29th May 2015, 2p

경향[141]이 있다. 예컨대, 개인정보에서 개인식별정보가 제거되어 가명화된 정보(Pseudonymized Data)도 익명정보의 영역에 속한다고 생각할 가능성이 있다. 하지만 이 의견서는 가명화(Pseudonymized)와 익명화(Anonymized)는 구별되며, 이에 대한 인식에 오류가 있는 경우가 많다고 지적[142]한다.

사실 개인정보와 익명정보 사이에 명확한 구별기준은 존재하지 않는다. 익명화 의견서는 데이터세트가 익명화 되었는지를 체크할 수 있는 방법으로 두 가지 옵션을 제시한다. ① 데이터세트에서 다음 세 가지 속성들이 제거되었는지 확인하는 것이 하나의 방법이다. 이 때, 제거되어야 할 속성은 (1) 식별성(Singling out[143]), (2) 연결가능성(Linkability[144]), (3) 추론가능성



141) CNIL, WP 29 Opinion on anonymization techniques, 29th May 2015, 3p



142) CNIL, WP 29 Opinion on anonymization techniques, 29th May 2015, 4p

(inference[145])이다. ② 또 하나의 방법으로, 재식별 리스크에 대한 분석(analysis of re-identification risk)을 받을 수 있다.

이 때 앞선 기준(①)을 두고 살펴보았을 때 유의해야 할 점은 하나의 익명화 기법만으로는 모든 속성들(식별성, 연결가능성, 추론가능성)을 제거하는 것이 어려운 것이 보통이라는 점이다. 특히, 데이터에서 추론가능성조차 제거한다면 익명화가 빅데이터 분석과 양립가능한 것인지에 관한 근본적인 의문이 들 수 있다. 그럼에도 불구하고 이 경우 추론가능성은 "개인"에 대한 추론가능성이라고 해석하여, 환경(environment)이나 대상(object) 및 소비(consumption) 등에 대한 추론가능성은 제거대상이 아니라고 파악할 수는 있다. 두 번째 기준(②)을 적용한다면 "개인"에 대한 추론가능성을 살린 채 익명화를 수행할 수 있고, 이는 빅데이터와 더욱 용이하게 양립할 수 있다.

3) 비판

이 의견서에 대해서는 비판적인 시각도 존재한다. 가장 일반적으로는, 이 의견서는 단순하게 비식별화 기법들을 나열하고 있을 뿐 관계 기관들이 비식별화 방식을 적용하는 것에 있어서 구체적이고 실질적인 가이드라인을 제시하는 기능을 하지는 못한다는 비판을 받고 있다.[146] 특히 이 의견서에서 명시한 목적은 '현재의 익명화 기법의 효과성과 한계를 문서화하는 것' 이어서 데이터 관리자 또는 데이터 처리자가 데이터의 익명화를 위해 따라야 할 방법들을 제공하지는 않는다.

이와 같이 구체적인 기준들에 대해서 제시하지 못한다는 전체적인 비판은 재식별의 위험성에 대한 규정에도 같이 적용된다. 이 보고서는 수용할 수준의 재식별의 위험성이 무엇인지에 대한 명확한 설명이 결여되어 있다. 이런

143) possibility to isolate some records of an individual in the dataset
144) ability to link, at least, two records concerning the same data subject or a group of data subjects (in the same database or in two different databases)
145) the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes
146) Khaled El Emam and Cecilia Alvarez, 'A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques', International Data Privacy Law, 제 2면 (2014)

부분에 대한 명확한 설명이 없고 서로 모순되는 관계에 있는 모든 종류의 접근법들을 암시하고 있다. 수용가능한 재식별의 리스크는 영(0)이어야 한다는 정도의 내용까지 암시하고 있다. 하지만 이런 영의 위험성은 데이터보호에 대해 규정한 EU 지침(95/46/EC)과 일치하는 것인지 불명확할 뿐 아니라, 실용성 측면에서도 영의 위험성 기준은 한계가 있을 수밖에 없다.

의견서에는 살펴보면 수용할 수 있는 수준의 위험성에 대한 절대적 개념을 영의 위험성 형태로 제시하고 있다.[147] 예를 들어, 익명화 방식의 특징을 비가역적(irreversibly)으로 식별화를 막는 방법인 것으로 설명하고 있다. 또한 익명화의 결과물을 삭제와 같은 수준의 영구적인 상태에 개인정보가 위치할 정도로 되어야 한다고도 규정하고 있다. 이런 표현들은 비식별화된 개인정보는 재식별이 완벽하게 불가능해야 한다는 입장을 반영하는 것으로 해석될 수 있어서, 현실을 외면한 것이라는 취지의 비판에 직면하기도 한다.

3. 일본

(1) 비식별화 처리에 대한 규제 및 논의동향 개괄

일본에서는 개인정보 보호법이 2003년 최초로 제정된 바 있다. 그런데 그 이후로 정보통신기술의 비약적인 발전에 따라 다양한 종류의 방대한 개인정보를 수집·분석하는 것이 가능해졌으며, 이를 통한 신산업·신서비스 창출이 기대되었다. 하지만 사업자는 개인정보로 다루어야 할 대상이 모호하며 이를 다룰 때 수반되는 부담이 크다는 등의 이유로 데이터를 활용하는 유형의 사업을 주저하는 경향이 있었다. 이에 따라, 개인정보 보호법의 개정 필요성에 관한 논의가 대두되었다. 고도 정보통신 네트워크사회 추진 전략본부(IT 종합 전략본부)는 2013년 6월 14일 결정된 「일본 재흥 전략」에 따라 “제도 재검토 방침”을 수립하고 산하에 “개인 데이터에 관한 검토회”를 설치하여 수십여 차례 검토를 거듭하였다. 그 결과 도출된 것이 “개인 데이터의 활용에 대한 제도 개정 대강”[148]이다. 이를 통해 마련된 법

147) El Emam and Alvarez 전게서, 제 2면
148) https://www.kantei.go.jp/jp/singi/it2/pd/index.html

개정안은 약간의 수정을 거친 후 지난 2015년 9월 3일 중의원 본회의에서 가결·성립[149]되었다. 그 결과 2003년 제정된 일본 개인정보 보호법은 12년 만에 개정이 되게 되었다.

이번 개정법은 본격적으로 익명 데이터를 보호하고 이용하기 위한 제도를 도입하였으므로, 개인 데이터를 취급하고 싶어 하는 기업일수록 이번 개정법에 많은 관심을 기울이고 있다. 이하 2015년 9월 3일에 개정된 일본의 개인정보 보호법이 어떻게 만들어졌으며 어떤 내용을 담고 있는지를 익명화를 중심으로 살펴본다. 더불어 개정안에 바탕을 두고 2015년 5월 30일에 제정된 의료영역의 익명화 가이드라인을 소개한다.

(2) 개정 개인정보보호법 (2015)

1) 개인데이터에 관한 검토회

개인 데이터의 활용에 관한 제도 재검토 방침[150][151]에 따라 개인데이터에 관한 검토회가 구성되었다. 「개인 데이터에 관한 검토회」는 2013년 12월에 정리한 「개인 데이터의 활용에 관한 제도 재검토 방침」에 따라 2014년 6월까지 법 개정의 내용을 대강으로 정리할 수 있도록 각 논점에 대해 심층 토론을 실시하였다. 검토회는 ① 빅데이터 시대의 개인 데이터 활용을 위한 검토와 ② 개인정보 보호에 대한 개인의 기대에 부응하기 위한 검토를 추진하였다. 동시에 미국과 유럽을 포함한 외국의 제도 변경과의 일관성을 도모하고자 하였다. 이 중 익명화에 관한 부분은 기술적인 검토가 필요했으므로, 이 부분에 관해서는 별도의 기술 검토 워킹 그룹을 두고 논의를 전개하였다. 이하 기술 검토 워킹 그룹의 논의를 위주로 살펴본다.

2) 기술 검토 워킹 그룹의 논의[152]

149) http://www.cas.go.jp/jp/houan/189.html
150) https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/dec131220-1.pdf
151) https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/gaiyou131220-1.pdf
152) 「고도 정보 통신 네트워크 사회 추진 본부」(IT 종합 전략 본부) 아래 설치된 「개인 데이터에 관한 검

① 논의의 전제

사무국은 “개인정보로 보호되는 개인데이터의 범위”를 ① 현행 개인정보 보호법의 정의를 바탕으로 각 사업자가 정보를 이미 취급하고 있던 현실을 감안하여 기존 정의는 유지하되 ② 개인정보에 해당하는지 판단할 수 있는 가이드라인을 제시하여 명확하게 사전적으로 신속히 대응하도록 하며 ③ 개인정보에 해당하는 지를 판단하는 주체를 명확히 하고, 그 외에 ④ 개인정보에 해당하는지 여부의 판단이 어려운 이른바 회색 지대를 해소하기 위하여 현행법상 개인정보에 해당하지 않는 것이라고 할지라도 특정 개인을 식별할 수 있는 개연성이 높은 것을 (가칭) 준개인정보로 하여 보호되는 대상으로 추가하고자 하였다. 또한 사무국은 “(가칭) 개인 특이성 감소 데이터에 대한 개요”로서 ① 빅데이터의 활용이 부가가치를 창출하고 혁신을 촉진하여 일본 경제성장의 일익을 담당할 것임에 비추어 특히 중요한 개인 데이터의 이용·유통을 도모하고 ② 현행법은 개인정보를 제3자에게 제공할 경우 본인의 동의가 요구되나, 개인 데이터를 가공하여 개인이 특정될 가능성을 줄이는 경우 본인의 동의 없이도 데이터를 제3자에게 제공할 수 있도록 새로운 정보의 유형을 정리하며 ③ 상기 가공된 데이터를 취급하는 자가 제공자 및 수령자로서 일정한 의무[153]를 부담하도록 하고자 하였다.

토회」는 2013년 12월에 “개인 데이터의 활용에 관한 제도 재검토 방침”을 정리했다. 이 재검토 방침에 따라 2014년 6월까지 법 개정의 내용을 대강으로 정리하여 2015년 정기 국회에 법안으로 제출하는 것을 목표로 2014년 3월부터 「개인 데이터에 관한 검토회」는 각 논점에 대하여 심층 토론을 실시하였다. 이 논의 과정인 제7회 검토회(2013년 4월 16일 개최)에서 “개인정보 등 보호되는 개인 데이터의 범위”와 “(가칭) 준개인정보의 새로운 정의” 및 “(가칭) 개인 특이성 감소 데이터의 정의” 등에 대하여 각각 제안이 이루어 졌다. 이를 기술적 관점에서 검토하기 위해 「개인 데이터에 관한 검토회」는 산하 「기술 검토 워킹 그룹」에게 “(가칭) 준개인정보”와 “(가칭) 개인 특이성 감소 데이터”에 대한 기술적인 관점의 검토를 요청했다.

> 5월 하순에 개최될 예정인 「개인 데이터에 관한 검토회」에 제출할 것을 목표로 하여, 다음의 사항에 대하여 「기술 검토 워킹 그룹」에서 기술적인 관점에서 검토하고, 그 결과를 보고하기를 바란다.
>
> 1. 개인정보 등으로 보호되는 개인 데이터의 범위 : (가칭) 준개인정보에 대하여
> 2. (가칭) 개인 특이성 감소 데이터에 대하여

기술 검토 워킹 그룹은 반년 동안 6회에 걸쳐 회의를 하였다. 그 과정에서 두 차례 기술 검토 보고서를 제출하였다. 이하 검토할 보고서는 첫 번째 보고서를 바탕으로 최종 제출된 두 번째 보고서이다. : “(가칭) 준개인정보 및 (가칭) 개인 특이성 감소 데이터 관련 기술적 관점에서의 고찰” : (가칭) 준개인정보의 새로운 정의 등에 대해서, 그리고 (가칭) 개인 특이성 감소 데이터의 정의 등에 대해서 동 검토회에서 기술검토워킹그룹에게 기술적인 관점에서 검토해 달라는 취지의 의뢰를 하였다.

‘기술 검토 워킹 그룹의 논의’ 부분은 그 검토결과를 번역에 가까운 수준으로 정리하여 소개하는 것이다.

이전 기술 워킹 그룹 보고서(2013.12)는 기술의 발전 및 외부 정보의 증가에 따라 다른 정보와 비교하면 프라이버시 침해가 초래될 수 있다는 견지에서, 개인 데이터에 대한 이른바 개인식별의 문제에 대처하기 위하여 “특정”과 “식별”이라는 기준을 도입하였다. 여기에서 “특정”이란 “정보의 주체가 누구인지를 알 수 있는지”를 의미한다. “식별”이란 “정보가 누구한 사람의 정보라는 점을 알 수 있는지(정보가 누구의 정보인지를 알 수 있는지는 논외로 하면서 어느 한 사람의 정보와 다른 사람의 정보를 구별할 수 있는지)”를 의미한다. 이 두 가지 기준에 따라, 개인 데이터를 (1) 식별 특정 정보(그것이 누군가 한 사람의 정보인지를 알 수 있으며, 또한 그 한 사람이 누구인지도 알 수 있는 정보), (2) 식별 비특정 정보(그것이 누군가 한 사람의 정보임을 알 수는 있는데, 그 한 사람이 누구인지 까지는 알 수 없는 정보), (3) 비식별 비특정 정보(그것이 어떤 사람의 정보인지, 나아가 그것이 어느 한 사람의 정보인지 조차 모르는 정보)로 분류하였다.[154]

「개인 데이터에 관한 검토회」(제7회)에서 제안된 (가칭) 준개인정보는 “특정 개인을 식별하지는 않지만 그 취급에 있어 본인에게 권익 침해가 초래될 수 있는 것”이다. 하지만 여기에 대해 본인의 권익 침해는 두 가지가 상정되었다. ① (가칭) 준개인정보에 있어, 어떠한 상황에서 특정 개인이 식별되어 권익이 침해되는 경우와, ② (가칭) 준개인정보에 있어, 특정 개인이 식별되지 않았는데도 권익이 침해되는 경우이다. 이를 개인데이터 3분류(식

153) ⓐ 제공자가 개인정보, (가칭) 준개인정보를 (가칭) 개인 식별성 감소 데이터로 가공하여 이를 제3자에서 제공하고자 하는 경우 가공 방법 등에 관한 정보를 제공한다. (가칭) 준개인정보를 제3자에게 제공하고자 하는 경우 (가칭) 개인 특이성 감소 데이터로 가공해야 한다. ⓑ 수령자가 수령한 (가칭) 개인 특이성 감소 데이터로부터 개인을 특정하는 것을 금지한다. 또한 수령자는 안전 관리 조치를 취하여야 한다. ⓒ 제3자는 영업비밀 등 사업자의 권리와 이익을 침해하지 않는 범위 내에서 제공받은 정보를 공개한다.

154) 「개인 데이터에 관한 검토회」에서 검토 의뢰를 받은 (가칭) 준개인정보에 대하여 제7회 「개인 데이터에 관한 검토회」[자료1-2] 3페이지에서는, “특정 개인을 식별하지는 않으나, 그 취급에 있어 본인의 권익에 침해가 초래될 수 있는 것”(즉 기술검토워킹그룹의 식별 비특정 정보)을 새롭게 유형화한 것 같은 (가칭) 준개인정보는 상기 식별 비특정 정보를 기반으로 한 것이며, 이전 보고서의 연장선상에 놓여있다고 보았다. 따라서 본 보고서에서 사용하는 “특정”이나 “식별”같은 용어는 지난 보고서를 계승하였다. 또한 (가칭) 개인 특이성 감소 데이터는 이전 보고서에서 제안한 제3자에게 제공하는 데에 있어 정보주체의 동의가 필요 없는 데이터는 법 제23조 제1항의 적용제외 정보를 기반으로 하고 있으므로, 여기에 대해서도 이전 보고서를 참고하였다. 다만 이전 보고서에서 검토한 (가칭) 법 제23조 제1항의 적용이 제외되는 정보의 범위는 당시에는 확정되지 않았었다. 이 부분은 이전 보고서에서 향후 「개인 데이터에 관한 검토회」의 검토를 받고 다시 기술적 검토가 필요한 부분으로 지적했었던 지점이었다. 여기에서 논의될 (가칭) 개인 특이성 감소 데이터를 검토할 때에도 (가칭) 법 제23조 제1항의 적용 제외 정보를 기반으로 하고 있다는 점에서 그 범위가 반드시 명확하다고 보기 어렵다는 점에서 같은 전제를 두고 검토하였다.

별 특정 정보, 식별 비특정 정보, 비식별 비특정 정보)와의 관계에서 살피면 ①은 식별 비특정 정보가 특정되어 식별 특정 정보가 된 경우를 말한다. 가령 (가칭) 준개인정보를 이용하는 과정에서 개인에 대한 다른 정보와 매칭이 되고 그 결과 개인이 특정되어(전술한 사무국 안 중 식별 특정 정보에 해당) 어떤 개인에게 권익 침해가 생기는 경우이다. 다른 한편, ②는 식별 비특정 정보를 통한 침해, 즉 (가칭) 준개인정보를 이용하는 과정에서 개인에 대한 다른 정보와 매칭된 결과 개인이 특정되지는 않았지만 개인의 속성이 추정되는 등의 결과로 인해 어떠한 개인의 권익에 침해가 생기는 경우이다.

이처럼 (가칭) 준개인정보와 (가칭) 개인 특정성 감소 데이터에 의한 개인의 권익침해에는 두 가지 경우가 있었다. 본인의 동의 없이 제3자에게 정보를 제공하는 경우, 본인에게 권익 침해가 발생하지 않도록 하려는 안정성의 관점에서는 특정성을 더욱 낮추어야 한다. 하지만 특정성을 낮추게 되면 동시에 식별성도 감소하게 된다. 그런데 식별성조차 없어져 버린 정보는 비식별 비특정 정보가 되어 원래 별도로 보호할 필요가 없는 정보이다. 그러므로 식별 비특정 정보로서 활용하려고 하는 경우에는 안정성을 확보하기 위해 특정성을 낮추어야 하겠지만 유용성을 고려하여 식별성을 어느 정도 남기는 방향으로 검토할 필요가 있다. 이처럼 어느 정도 식별성을 남긴 채 특정성을 저감한 경우, 가공 후의 데이터 그 자체로 인해 문제가 유발될 가능성뿐만 아니라 가공 후의 데이터가 다른 데이터와 대조됨으로 인해 문제가 유발될 가능성에 대해서도 고려할 필요가 있었다.

현재 제3자에게 데이터가 제공되는 경우를 전제하므로, 이러한 조회가능성은 수령자가 어떠한 개인 데이터 등을 제공받은 데이터에 조합할 수 있는지에 따라 좌우된다. 하지만 제공자가 수령자가 어떠한 개인 데이터를 추가로 조합할 수 있는지에 대해 예견하도록 하는 것은 기술적인 견지에서는 매우 어려운 것이다. 그러므로 (가칭) 개인 특정성 감소 데이터에 대해서는 아래와 같은 전제하에 검토하였다. [전제 ①] 사업자가 데이터를 제공하기 위하여 (가칭) 개인 특정성 감소 데이터를 가공하는 경우 등에 있어, 제공받아 조합

되는 데이터가 아닌 제공하려는 (가칭) 개인 특정성 감소 데이터 만의 특정성 저감에 대해서만 다룬다. [전제 ②] 데이터를 제공하려는 사업자는 데이터를 수령하는 사업자가 (가칭) 개인 특정성 감소 데이터에 대해 조합할 수 있는 다른 개인 데이터가 있는지 여부를 예견할 수 없다. 이러한 전제에 기초하여, 만일 데이터를 수령한 사업자가 수령한 (가칭) 개인 특정성 감소 데이터를 다른 정보와 조합하여 개인을 특정하게 된다면, 이는 데이터를 수령한 사업자가 특정화 금지를 위반한 것으로 추정될 수 있는 상황이 된다.

기술검토 워킹그룹은 검토한 결과 다량 또는 다종의 정보가 수집되어 개인이 특정될 우려가 있는 정보를 (가칭) 준개인정보로 새롭게 설정하는 것이 타당하다고 판단하였다. 기존의 법에서는 개인정보는 “특정 개인을 식별할 수 있는 것”이라고 정의되어 있고, 개인정보를 기본적인 보호대상으로 삼고 있었다. 개인정보를 보호대상으로 설정한 취지는 “식별의 용이성”과 “관계의 명확성”에 의해 그 취급으로 인한 본인 권익 침해 가능성이 있기 때문이다. 또한 “개인 데이터의 활용에 관한 제도 재검토 방침”은 제도 재검토의 방향으로 프라이버시 보호를 제시하는데, 개인의 특정성이 사생활 침해의 필요조건이라고 하는 복수의 판례가 공표되었기 때문이다. 이처럼 개인을 특정하게 되면 개인의 권익침해 위험성이 현저하게 높아질 것이므로 기존법은 개인정보취급사업자의 행위를 정하고 이를 적정하게 취급함으로써 개인의 권익침해를 미연에 막고자 했다.

그런데 정보통신기술이 발전하고 있는 상황에서 개인에 관련된 정보의 유통량이 폭발적으로 증가하고 있으며 정보의 수집 및 분석이 용이해져, 어느 시점에서는 개인을 특정할 수 없었던 정보라도 다른 정보와 쉽게 결합시킴으로써 개인을 특정하고 이로 인해 개인의 권익을 침해할 수 있는 개연성이 높아지게 되었다. 그러므로 개인이 특정되지 않는 정보라고 할지라도 개인이 특정될 우려가 있는 정보를 (가칭) 준개인정보라는 이름으로 별도 유형의 정보로 규정할 필요성이 인정되었다.

나아가 개인이 특정될 우려는 대부분의 경우 해당 개인에 대한 다량 또는

다종의 정보를 수집하는 과정에서 커지게 되므로, 이러한 정보수집을 대상으로 다루어야 한다. 이러한 정보수집을 가능하게 하는 식별자는 여러 가지가 있겠지만, 다양한 식별자 가운데에서도 특히 그 성격이나 특성에 비추어 다량 또는 다종의 정보를 수집하게 될 개연성이 높은 것에 한정하여 다룰 필요가 있다. 즉 (가칭) 준개인정보가 되는 것의 식별자가 될 수 있는 것은 모든 식별자가 아니라 상대적으로 다량 또는 다종의 정보를 수집할 수 없는 식별자는 대상에서 제외하고 선정하는 것이 타당하다고 보았다.

또한 비록 개인을 특정하게 되면 개인의 권익침해 위험이 현저히 높아진다는 점은 분명하나, 개인을 특정하지 않는다고 하여 권익침해가 발생하지 않는 것은 아니다. 일본의 법제도는 개인이 특정되지 않고도 권익침해가 생길 가능성을 명시적으로 인정한다. 가령 행정기관이 보유하는 정보공개에 관한 법률인 정보공개법의 제5조 제1호는 “특정 개인을 식별할 수는 없지만 이를 공개함으로써 개인의 권익을 해칠 우려가 있는 것”을 비공개 정보(개시청구의 대상이 되지 않는 정보)로 규정한다. 개인의 인격에 밀접하게 관련된 정보 등이 이에 해당하는 것으로 예시되어 있다.

이처럼 개인을 특정하지 않고도 권익침해가 생길 수 있다. 또한 식별자를 이용하여 다량 또는 다종의 정보를 수집하게 되면 다른 종류의 권리 침해로 이어질 가능성도 충분히 예측할 수 있었다. 그러나 결국 어떤 정보가 “개인의 인격에 밀접”한가 등에 관해서는 기술검토 워킹그룹을 통해 규정하는 것이 적절하지 않다고 판단되어 이는 「개인 데이터에 관한 검토회」의 검토에 맡겼다.

한 사업자 안에 개인정보와 (가칭) 준개인정보가 혼재되어 존재하는 경우 그것들이 일체가 되어 특정 개인에 결부되고 이용·관리되고 있는 것에 관해서는 전체적으로 개인정보로서 취급할 필요가 있다. 특히 여러 항목에 의해 개인이 특정되고 있는 경우 해당 항목에 (가칭) 준개인정보에 해당 항목이 포함되어 있다면 이는 개인정보로서의 취급이 필요하다. 마찬가지로 보통 (가칭) 준개인정보라고 여겨지는 항목일지라도 어떠한 특이상황에서는 그

항목만으로 개인정보가 될 수 있는 경우가 있을 것이다. 가령 성명 및 회사명 등이 명시되어 있는 메일 주소가 이에 해당될 수도 있다.

② (가칭) 준개인정보

(a) 정의에 대한 개념정리

기술검토 워킹그룹은 다양한 식별자 중에서 그 성격이나 특성에 비추어 다량 또는 다종의 정보를 수집하게 되면 개인을 특정하게 될 개연성이 높은 것만을 대상으로 해야 한다고 생각했다. 그렇다면 어떠한 식별자가 "그 성격이나 특성에 비추어 다량 또는 다종의 정보를 수집하게 되면 개인을 특정하게 될 개연성이 높다"고 판단할 수 있는지에 관해 관련된 요소·지표를 밝혀야 한다. 이 점에 대한 검토결과로 제시된 요소·지표는 다음과 같다 : (가) 본인에게 부여되는 것인지, 소유물에 부여되는 것인지[155] (나) 일의성(중첩성)[156] (다) 단사성[157] (라) 공유성[158] (마) 변경가능성[159] (바) 불변성[160] (사) 이용정지가능성[161] (아) 지속성(이용기간)[162] (자) 이용범위

155) 본인에게 부여된 것은 본인 소유물에 부여된 것과 비교하여 본인과의 직접적인 관계가 있어 개인을 특정할 수 있는 개연성이 높다고 볼 수 있다. 또 본인에게 직접 부여된 것은 아니지만 휴대전화 등 항상 소유자가 가지고 있어 타인이 사용하는 경우가 극히 드물다고 생각되는 경우 이러한 휴대전화 등 모바일 통신기기에 부여된 번호 등에 대해서도 본인과 밀접성이 있어 개인을 특정할 개연성이 다른 기기등과 비교하면 높아진다.

156) 남들과 중복되지 않도록 하나의 대상에 하나의 식별자가 부여된 경우, 이를 대상은 다르지만 동일한 식별자가 존재할 수 있는 가능성이 있는 경우와 비교하면 개개의 사람을 찾아낼 수 있는 개연성이 높아진다.

157) 개인의 집합에서 식별자로 사상을 했는데 이것이 일대일 사상관계에 있는 경우, 일대다 혹은 다대일 관계에 있는 것과 비교하면 개인을 특정할 수 있을 개연성이 높다고 할 수 있다. (이는 상기 일의성(중첩성)과 일체적으로 검토한다)

158) 본인이나 소유물 등에 부여된 번호와 기호인 식별자에 대해 해당 식별자를 발행하고 관리하는 사업자에 한하지 않고 발행된 식별자를 복수의 사업자가 독립하여 취득할 수 있는 경우(식별자를 발행하고 관리한 사업자에 의하지 않고도 직접 취득할 수 있는 경우를 포함), 이는 발행하고 관리하는 사업자만 식별자를 이용하는 경우와 비교하면 동일한 식별자를 복수의 사업자가 공유할 가능성이 높아질수록 개인이 특정될 개연성이 높아진다.

159) 부여된 번호 등이 본인에 의사에 의해 변경될 수 있는 경우 이는 변경될 수 없는 경우에 비해 개인의 특정성에 있어 차이가 있다. 또한 본인의 의사에 의해 변경할 수 있는 것이라고 할지라도 쉽게 변경할 수 없는 경우에는 쉽게 변경할 수 있는 경우와 비교하면 개인의 특정성에 차이가 생긴다. 즉 이 경우 개인을 특정할 수 있는 개연성이 높아진다.

160) 본인 의사에 의하지 않고도 시간이나 상태로 인해 동적으로 변하는 것은 외적 요인에 의해 변하지 않는 정적인 것과 비교하면 개인을 특정할 수 있는 개연성이 낮아질 수 있다. 본 지표는 주로 신체적인 특징에 대한 것이라고 생각한다. 정적이고 불변성이 있는 것일수록 변경이 어려운 것이어서 (마) 변경가능성에 포함하여 검토할 수 있다고 생각한다.

161) 일단 식별자가 부여된 후 본인의사 등 어떤 사정으로 해당 식별자를 이용하는 것을 정지할 수 있는 절차가 없는 경우에는 이러한 절차가 있는 경우와 비교하여 개인을 특정할 수 있는 개연성이 더욱 높아진다. 구체적인 예로서 적정한 옵트 아웃 절차가 설치된 경우 본인의 의사에 의해 더욱 쉽게 식별자와 연계된 정보를 떼어낼 수 있다. 이러한 점에 비추어볼 때 본 지표에 대해서도 상기 (마) 변경가능성에 포함하여 검토할 수 있

(데이터규모)[163]. 이를 검토한 결과 (가) 본인에게 부여되는 것인지, 소유물에 부여되는 것인지, (나/다) 일의성(중첩성)/단사성, (라) 공유성, (마/바/사) 변경가능성/불변성/이용정지가능성을 고려할 수 있다고 결론[164]을 내렸다.

사무국에 제안된 (가칭) 준개인정보의 정의는 다음과 같았다.

개인정보에 해당하는 것을 제외하고 생존하는 개인에 관한 정보이며, 다음에 예시하는 것 및 이와 비슷한 것을 포함하는 정보 ① 여권 번호, 면허증 번호, IP주소, 휴대 단말 ID 등의 개인 또는 개인의 정보통신 단말(휴대 전화기, PC 단말기 등) 등에 부번하고 계속해서 공용되는 것 ② 얼굴 인식 데이터, 유전자 정보, 성문 및 지문 등 개인의 생체적 · 신체적 특성에 관한 정보로서 보편성을 가진 것 ③ 이동 이력, 구매 이력 등 특징적인 행동 이력[165]

사무국 안으로 제시된 구체적인 분류 ①②③ 중 ①은 다른 것과 구별하기 위해 식별자로 부여된 성질을 지닌 것들이며, ②는 신체적인 특징 등 남과 다른 계속성을 지닌 것으로서 다량 또는 다종의 정보를 수집하는 식별자의 기능을 할 것으로 인정되었다. ③은 ①②와는 다른 분류가 되므로, 이것이 어떠한 상태로서 다량 또는 다종의 정보를 수집하는 식별자에 상당하는 기능을 할 것인지에 대해 검토할 필요가 있었다. 이 때 여기에서 말하는 이동

다고 생각한다.

162) 계속 같은 번호 등을 사용할 경우 단기간에 이를 삭제하거나 변경할 경우와 비교하면 개인을 특정할 수 있는 개연성이 높아진다고 볼 수 있다. 다만 가장 지속정이 있는 정보가 반드시 개인을 특정 한다고는 할 수 없으며 일시적으로만 사용되지 않는 것이라고 할지라도 개인을 특정할 개연성이 높은 것이 있기 때문에, 이는 결국 다른 성질에 의해서 특정성이 판단되게 된다. 결국 계속성 자체를 지표로 설정하기는 어렵다고 생각한다.

163) 일반적으로 이용범위가 광범위하게 되면 많은 정보가 집적될 것으로 생각되며, 이용범위를 한정하여 데이터를 좁히면 개인을 특정할 개연성이 높아질 경우도 있다. 그러므로 이용범위(데이터규모)를 지표로 설정하기는 그 특성상 어렵다고 생각한다.

164) 기타 전체성, 외관식별성, 외부정보입수가능성, 본인도달가능성과 같은 항목도 있지만 이는 특히 개인을 특정할 개연성을 높이는 성질이나 특성을 가지고 있지는 않다고 파악되므로 구체적인 지표로 삼지는 않았다. 이상을 토대로 9개의 항목에 대해 검토한 결과, 지속성(이용기간) 및 이용범위(데이터규모)는 지표로 기능할지 명확하지 않으므로 제외했다. 일의성(중첩성)과 단사성은 일체로 검토하는 것이 타당하다고 보았다. 변경가능성, 불변성, 이용정지가능성은 거의 동일한 관점에서 검토할 수 있는 것이므로 이들은 하나의 지표로 설정해서 검토해도 무방하다고 생각되었다.

165) 이동 이력과 구매 이력은 사람의 행동 이력이며, 일반적으로 정보가 축적될수록 특정 개인이 식별될 가능성이 높아진다. 또한 타인과 다른 특이한 날짜에 또는 특이한 행동(승강 인원이 적은 역에서 승강, 한 점짜리 쇼핑)을 함으로써 다른 개인과 구별될 가능성이 높은 경우 더더욱 개인을 식별하는 정보로서 기능할 수 있다. 다만 무엇이 특징적인 행동 이력이고 무엇이 그렇지 않은지 일률적인 기준을 고려하기가 어렵다.

이력, 구매 이력 등의 특징적인 행동 이력이란 ①과 같은 식별자를 달리 지니고 있지 않아 이동 이력 또는 구매 이력 등이 특징적인 부분으로 식별자 기능을 하고 있는 경우를 지칭한다.

가. 이동 이력

복수의 위치 정보 이력의 집합인 이동 이력은 정확한 시각과 조합하게 되면, 그리고 해당 정보의 취득 빈도가 높아질수록, 개인의 생활권과 행동 패턴이 나타나면서 개인의 특정으로 이어질 가능성이 높아진다. 또한 이력을 계속적으로 취득하면 반복적인 행동 패턴이 나타나게 되고 그것에 의해서도 개인 특정으로 이어질 가능성이 높아진다. 이동 이력과 구매 이력의 차이점은 복수의 사업자가 동시에 독립적으로 같은 정보를 취득할 수 있는지 여부이다. 예를 들면 휴대 전화에 설치되어 있는 복수의 애플리케이션에서 동시에 독립적으로 시각과 GPS정보를 취득할 수 있는데 이것은 정보를 집적하는 식별자로서 기능할 가능성이 있다. 또한 GPS 정보 등 매우 정밀한 위치 정보는 2지점 사이의 위치정보인 이동이력을 별도로 참고할 필요도 없이 시각과 조합하여 특정 개인을 식별할 가능성이 높은 정보가 될 수 있다. (예를 들면 심야 2시에 정확한 위도·경도 정보는 자택일 가능성이 높다고 여겨진다. 또한 같은 시간대에 동종의 정보 취득이 수차례 반복되면서 같은 위치 정보를 얻을 수 있다면, 그 위도·경도가 거의 확실하게 집으로 추정될 것이다.) 이런 점에서 이동 이력에 대해서는 2지점 이상의 위치 정보인 행동 이력뿐만 아니라 정확한 지점의 위치정보와 시각의 조합만으로도 개인을 특정하는 식별자로서 기능하게 될 수도 있다. 그렇지만 이동 이력 정보의 양·취득 기간· 취득 빈도, 위치 정보와 시각의 정밀도 등이 어떠한 상태일 경우 식별자로 기능할 수 있는 특징적인 상태인지는 일률적으로 정하기 어렵고, 개별 조건 등에 의해 달리 정해질 수밖에 없다.

이동 이력에 관한 정보에 대해, 예를 들어 스마트 폰 상의 애플리케이션을 생각해 볼 수 있다. 이 경우 GPS 위치 정보의 취득에 대해 단말기상 위치정보 취득을 하지 않도록 설정되어 있거나 위치 정보를 취득하는 서비스를 이

용하지 않도록 하는 등 본인의 의사로 쉽게 이용을 정지할 수 있다고 생각할 수 있다. 반대로 취득된 위치 정보의 이용을 정지하는 서비스를 제공자가 본인에게 신청하여 이용 정지하는 기능을 통해 별도로 제공하지 않는 이상 이용 정지를 할 수 없으며, 위치 정보에 관한 서비스를 이용하고 싶다고 생각한 경우 그것을 이용 정지하기에는 쉽지 않을 것이라는 견해도 있다.

나. 구매 이력

구매 이력은 고객 단위로 구입 품명 · 수량 · 날짜 · 상호 등을 기록한 정보이다. 구매 이력에 이름이나 회원 번호 등 개인을 특정할 수 있는 속성 정보가 포함되어 있지 않더라도 지속적인 이력의 취득은 반복되는 행동 패턴 등을 알 수 있게 해주며, 그 결과 개인의 특정으로 이어질 가능성이 있다. 그러므로 장기간 계속적으로 취득된 구매 이력은 신중하게 취급할 필요가 있다. 또한 이력의 취득 기간이 짧다고 하더라도 특징적인 구입이 있는 경우 개인의 특정으로 이어질 수 있다는 점에 유의하여야 한다. 예를 들어 비록 물건을 한 점 구입하였다고 하더라도 구입 수량과 시각, 장소 등이 특이한 경우에는 다른 정보와 대조함으로써 개인의 특정으로 이어질 가능성이 있다. 기타 구매 이력이 개인의 권익에 영향을 준다는 점에서 구매품의 종목에도 유의해야 한다. 의약품 등은 개인의 병력을 포함할 가능성이 있게 되고, 서적 등은 사상과 신념을 지닌다는 점을 암시할 수 있다. 한편 일반 식품과 일용품의 경우 그 구입 품명 · 수량 · 장소 · 시간 등에 특징이 없다면 즉각 개인의 특정에 연결될 가능성은 낮게 된다.

장기간의 계속적인 이력취득 · 특징적인 구입이 있을 경우, 그리고 구입 물품에 관한 유일한 번호를 포함하게 될 경우, 구매자가 공개될 경우 등의 경우에는 구매 이력이 개인을 특정하는 정보로 기능할 수 있으므로 개인의 특정성을 유의하며 다루어야 한다. 하지만 개인의 특정성에 실질적인 영향이 없는 구매 이력도 많다. 그러므로 만일 구매 이력을 (가칭) 준개인정보로 다루게 된다고 할지라도 모든 구매 이력을 (가칭) 준개인정보라고 해석해서는 안 되며, 상술한 개인의 특정에 연결되는 구매 이력에 한정되어야 한다. 하

지만 그 기준을 밝히는 것은 어렵다. 현 시점에는 당장 구매 이력을 (가칭) 준개인정보의 범위에 포함하지는 않고 실태를 바탕으로 한 보다 구체적인 검토를 더하는 것이 필요하다. 나아가 구입 이력에 구매품의 개체를 식별할 수 있는 번호(시리얼 번호 등), RFID 태그를 부착하는 등 상품에 부번을 한 유일한 번호(개체 식별이 가능한 번호) 등이 포함되는 경우 그것들의 번호는 (가칭) 준개인정보 중 ①의 분류에 포함된 정보로서 구입 이력 자체를 (가칭) 준개인정보에 포함하지 않아도 ①의 분류에 포함된 정보에 부수된 정보로서 구입 이력을 (가칭) 준개인정보로 볼 수 있게 된다고 생각된다. 이 밖에 일반 구매 이력에는 구입 상호 등 다양한 정보가 따라붙게 되지만, 구입 상호 등 위치 정보가 될 수 있는 등의 부수되는 정보에 따라서는 개인의 특정으로 이어질 수 있는 가능성도 있으므로 이를 주의하여 다루어야 한다.

이상과 같이 검토한 사항들을 바탕으로 워킹그룹은 (가칭) 준개인정보의 정의에 다음의 취지가 들어가도록 사무국에 제안된 개인정보 개념을 변경할 것을 제안하였다. ① 이는 개인정보가 아니다. ② 이는 생존하는 개인에 관한 정보에 포함된 식별자 또는 식별자에 상당하는 것으로서 (나/다) 일의성(중첩성)/단사성, (라) 공유성, (마/바/사) 변경가능성/불변성/이용정지가능성을 지녀야 한다. 상기 조건을 두루 갖춘 식별자 또는 식별자에 상당하는 것으로서 다음 중 하나에 해당하는 것이다. (1) 개인 또는 개인이 사용하는 통신 단말기기 등에 관한 것, (2) 개인의 신체적인 특성에 관한 것, (3) 상기 2개의 항목 외에 특정 개인의 식별로 이어지는 다량 또는 다종의 정보수집을 가능하게 하는 것.

⑹ 준개인정보의 구체적인 항목

개인정보 취급과 관계된다고 생각되는, 현존하는 식별자 또는 식별자로서의 성질을 갖는다고 생각되는 것들은 다음과 같이 검토되었다.[166]

166) 우선 다음은 개인정보 취급과 관련된다고 생각되는, 현존하는 식별자 또는 식별자로서의 성질을 지닌다고 생각되는 것들을 각 지표에 따라 분류한 것이다. 모든 지표에 부합하는 식별자 등은 특히 개인을 특정할 개연성이 높은 것으로서 (가칭) 준개인정보에 해당한다. 본 검토에서는 생각할 수 있는 식별자 등을 망라하여 나열하고 검토한 것은 아니며, 예를 들어 검토한 것일 뿐이다. 향후 기술의 발전 등에 의해서 상황은 변화될 것이므로, 항상 최선의 상황에 걸맞게 재검토를 할 필요가 있다.

개인 또는 개인이 사용하는 통신 단말기기 등에 대한 것	면허증 번호, 여권 번호, 건강 보험증 번호, 고용 보험 피보험자 번호, 계좌 번호, 신용 카드 번호, 메일 주소, 사용자 ID · 패스워드(복수 사업자로 공용), 유저 ID · 패스워드(한 사업자 내), 차량 번호, 고정 전화 번호, 휴대 전화 번호, FAX번호, Web핑거 프린트(Web브라우저 등 식별할 수 있는 정보나 식별자)정보 통신 단말기 일련 번호(휴대 전화 시리얼 번호 등), MAC주소 정보 통신 단말 ID, IC카드의 고유 ID, 소프트웨어 시리얼 번호, 부동산 구매 이력(부동산 등기 번호)IP주소(V6), cookie
사람의 신체적 특성에 대한 것	얼굴 인식 데이터, 바이오 매트릭스 인증에서 이용되는 신체적 특징 추출 데이터(템플릿), 신체적 특성(글씨, 보행, 성문), 생체 정보(잇자국과 지문, 정맥 패턴, 홍채)DNA, 성별, 피부색, 인종, 가족 구성, 혈액형, 머리 색깔, 혈압, 맥박, 키, 체중
상기 외, 특정 개인식별로 이어지는 다량 또는 다양한 정보수집을 가능하게 하는 것	이동 이력(위치 데이터+날짜), 구매 이력, Web열람 이력

위의 항목을 각 지표에 맞추어 분류한 결과는 다음과 같다.

(가) 본인 또는 본인 소유물과의 밀접한 연관성

본인 또는 본인의 소유물과 밀접한 연관성이 있는 것	면허증 번호, 여권 번호, 건강 보험증 번호, 고용 보험 피보험자 번호, 계좌 번호, 신용 카드 번호, 메일 주소, 사용자 ID · 패스워드(복수 사업자로 공용), 유저 ID · 패스워드(한 사업자 내), 차량 번호, 고정 전화 번호, 휴대 전화 번호, FAX번호, Web핑거 프린트(Web브라우저 등 식별할 수 있는 정보나 식별자), 정보 통신 단말기 일련 번호(휴대 전화 시리얼 번호 등), MAC주소 정보 통신 단말 ID, IC카드의 고유 ID, 소프트웨어 시리얼 번호, 부동산 구매 이력(부동산 등기 번호), IP주소(V6), cookie이동 이력(위치 데이터+날짜), 구매 이력, Web열람 이력 얼굴 인식 데이터, 바이오 매트릭스 인증에서 이용되는 신체적 특징 추출 데이터(템플릿), 신체적 특성(글씨, 보행, 성

	문), 생체 정보(잇자국과 지문, 정맥 패턴, 홍채)DNA, 성별, 피부색, 인종, 가족 구성, 혈액형, 머리 색깔, 혈압, 맥박, 키, 체중
상기 이외의 것	

(나/다) 일의성(중첩성)/단사성

일의성/단사성이 있는 것	면허증 번호, 여권 번호, 건강 보험증 번호, 고용 보험 피보험자 번호, 계좌 번호, 신용 카드 번호, 메일 주소, 사용자 ID · 패스워드(복수 사업자로 공용), 유저 ID · 패스워드(한 사업자 내), 차량 번호, 고정 전화 번호, 휴대 전화 번호, FAX번호, Web핑거 프린트(Web브라우저 등 식별할 수 있는 정보나 식별자), 정보 통신 단말기 일련 번호(휴대 전화 시리얼 번호 등), MAC주소 정보 통신 단말 ID, IC카드의 고유 ID, 소프트웨어 시리얼 번호, 부동산 구매 이력(부동산 등기 번호), IP주소(V6), cookie, 얼굴 인식 데이터, 바이오 매트릭스 인증에서 이용되는 신체적 특징 추출 데이터(템플릿), 신체적 특성(글씨, 보행, 성문), 생체 정보(잇자국과 지문, 정맥 패턴, 홍채), DNA, 이동 이력(위치 데이터+날짜), 구매 이력, Web열람 이력
상기 이외의 것	성별, 피부색, 인종, 가족 구성, 혈액형, 머리 색깔, 혈압, 맥박, 키, 체중

(라) 공유성

공유 가능성이 있는 것	면허증 번호, 여권 번호, 건강 보험증 번호, 고용 보험 피보험자 번호, 계좌 번호, 신용 카드 번호, 메일 주소, 사용자 ID · 패스워드(복수 사업자로 공용), 차량 번호, 고정 전화 번호, 휴대 전화 번호, FAX번호, Web핑거 프린트(Web브라우저 등 식별할 수 있는 정보나 식별자), 정보 통신 단말기 일련 번호(휴대 전화 시리얼 번호 등), MAC주소, 정보 통신 단말 ID, IC카드의 고유 ID, 소프트웨어 시리얼 번호, 부동산 구매 이력(부동산 등기 번호), IP주소(V6), cookie, 얼굴 인식 데이터, 바이오 매트릭스 인증에서 이용되는 신체적 특징 추출 데이터(템플릿), 신체적 특성(글씨, 보행, 성문), 생체 정보(잇자국과 지문, 정맥 패턴, 홍채)DNA, 성별, 피부색, 인종, 가족 구성, 혈액형, 머리 색깔, 혈압, 맥박, 키, 몸무게 이동 이력(위치 데이터+날짜)
상기 이외의 것	유저 ID · 패스워드(한 사업자 내), 구매 이력, Web열람 이력

(마/바/사) 변경가능성/불변성/이용정지가능성

변경, 이용 정지 등이 쉽지 않을 것(정적인 것)	면허증 번호, 여권 번호, 건강 보험증 번호, 고용 보험 피보험자 번호, 계좌 번호, 신용 카드 번호, 메일 주소, 차량 번호, 고정 전화 번호, 휴대 전화 번호, FAX번호 정보 통신 단말기 일련 번호(휴대 전화 시리얼 번호 등), MAC주소 정보 통신 단말 ID, IC카드의 고유 ID, 소프트웨어 시리얼 번호, 부동산 구매 이력(부동산 등기 번호)IP주소(V6), 얼굴 인식 데이터, 바이오 매트릭스 인증에서 이용되는 신체적 특징 추출 데이터(템플릿), 신체적 특성(글씨, 보행, 성문), 생체 정보(잇자국과 지문, 정맥 패턴, 홍채), DNA, 성별, 피부색, 인종, 가족 구성, 혈액형, 머리 색깔,[P]이동 이력(위치 데이터+날짜), 구매 이력, Web열람 이력
상기 이외의 것	유저 ID · 패스워드(복수 사업자로 공용), 유저 ID · 패스워드(한 사업자 내), Web핑거 프린트(Web브라우저 등 식별할 수 있는 정보나 식별자), cookie, 혈압, 맥박, 키, 체중

③ (가칭) 개인 특정성 감소 데이터

ⓐ 정의에 대한 개념정리

(가칭) 개인 특정성 감소 데이터의 정의에 대해, 사무국은 다음과 같이 제안했다.

> (가칭) 개인 특정성 감소 데이터는 다음에 제시하는 것을 말한다.
> ① 개인 데이터에 대하여 해당 데이터에 포함된 이름 생년월일 기타 기술 등으로 특정 개인이 식별될 수 있는 것을 삭제하는 등, 정령으로 정하는 방법에 의한 가공을 하여 개인이 특정될 수 있는 가능성을 저감한 것
> ② (가칭) 준개인정보에 대하여 해당 데이터에 포함되는 법의 0조 0항 0호에 해당하는 것을 삭제하는 등, 정령으로 정하는 방법에 의한 가공을 한 것
> ③ ① 또는 ②에 대하여 다른 정보를 더하는 등 가공을 한 것

또한 이 (가칭) 개인 특정성 감소 데이터의 이용 및 유통에 대해서는 아래의 박스 안에 있는 내용과 같이 개인정보 및 프라이버시 보호를 도모할 것을 제안했다.

● 개인정보의 제3자 제공에 대한 본인 동의를 대신하여 개인 데이터를 가공하여 개인이 특정될 수 있는 가능성을 저감한 데이터((가칭) 개인 특정성 감소 데이터))를 본인의 동의 없이 제3자에게 제공할 수 있는 새로운 유형으로 정리한다.
● 별도로 검토되어 있는 제3자 기관은 영업비밀 등 사업자의 권익을 침해하지 않는 범위에서 정보를 공개한다.

이 틀은 동의에 의한 제3자 제공 대신 동의가 필요없는 제3자 제공을 구현하는 틀이며, 제3자 제공의 기본은 여전히 동의를 받는 것이다. 그러므로 본 틀은 동의에 의해 취득된 데이터의 제3자 제공에도 적용되는 것이 아니다.

기술 검토 워킹 그룹은 (가칭) 개인 특정성 감소 데이터의 정의와 개인정보 또는 (가칭) 준개인정보에서 (가칭) 개인 특정성 감소 데이터를 가공할 때 준수해야 할 최소한의 가공 방법에 대해 검토하였다. 우선 출발점은 데이터를 가공하여 유용하게 활용하고자 할 때, 그 유용성과 해당 데이터 내의 개인 특정성 및 식별성 감소는 가치교환관계에 있다는 것이다. 또한 데이터의 종류나 양이 매우 다양하고 많으므로 데이터의 가공 방법은 활용 실태에 입각하여 개별적으로 판단되어야 한다.

(가칭) 개인 특정성 감소 데이터도 마찬가지로 글자 그대로의 완전히 익명화된 데이터로의 가공을 요구하지 않는다. 개인정보 또는 (가칭) 준개인정보에서 (가칭) 개인 특정성 감소 데이터로 가공할 때의 최소한의 가공방법은 쉽게는 (가칭) 준개인정보의 상태가 아닌 데이터로 만드는 것으로 이해할 수 있다. 하지만 (가칭) 준개인정보는 해당하는 항목이 다양하고 하나의 데이터 안에도 (가칭) 준개인정보에 해당할 수 있는 항목이 하나가 있을 뿐만 아니라 복수로 공존하고 있으며 이에 부가된 속성 정보의 양과 질이 많고 다양한 것으로 파악되므로, 이것을 단순화하여 (가칭) 개인 특정성 감소 데이터로 가공하는 방법으로 정리하기는 어렵다.

(가칭) 개인 특정성 감소 데이터에는 다음과 같은 개인 특정 리스크가 존재한다고 생각된다. [개인 특정 리스크 ①] (가칭) 개인 특정성 감소 데이터에

서 직접 개인이 특정되는 경우. [개인 특정 리스크 ②] (가칭) 개인 특정성 감소 데이터와 다른 데이터가 공통 서비스의 유저 ID 등 식별자를 사용하여 기계적으로 개인의 식별자 또는 식별자에 상당하는 것에 의한 복수의 정보를 매칭하고 그 결과 개인이 특정되는 경우. 이는 원래 (가칭) 개인 특정성 감소 데이터 자체가 보유하는 리스크에 해당된다. [개인 특정 리스크 ③] (가칭) 개인 특정성 감소 데이터로부터 수령자의 지식에 의존한 결과 개별적으로 개인이 특정되는 경우. (가령 구매 이력으로 발견한 구매 행동을 SNS 등에서 확인하고 그로부터 개인을 특정하게 되는 경우.) 이는 원래 (가칭) 개인 특정성 감소 데이터 자체가 보유하고 있던 리스크가 아니라 데이터가 제공된 수령자가 (가칭) 개인 특정성 감소 데이터와 조합할 수 있는 개인 데이터 등을 보유하고 있는지 등을 예견할 수 없다는 점에서 발생하는 위험이다.

이러한 개인 특정 리스크에 대처하여 이를 회피하기 위한 가공 방법을 생각해 본다. (1) [개인 특정 리스크 ①을 감소시키는 방법] 개인 데이터에서 (가칭) 개인 특정성 감소 데이터를 가공할 경우 개인을 직접 특정하는 속성을 미리 정한다. (가령 성명, 생년월일, 주소, 소속단체 등) 이것들을 모두 삭제 또는 가공한다.[167] (2) [개인 특정 리스크 ②를 저감하는 방법] 개인정보 또는 (가칭) 준개인정보에서 (가칭) 개인 특정성 감소 데이터를 가공할 경우 식별자 또는 식별자에 상당할 수 있는 것이 다른 사업자와의 관계에서 식별자 또는 식별자에 상당할 수 있는 것에 의하여 복수의 정보 매칭이 되지 않도록 가공한다. 즉 식별자를 삭제 또는 가명화하고 본인과의 관계성을 저감한다. 또한 이력 데이터에 관해서는 이동 이력으로서 기술한다.[168]

---

167) 이 방법을 사용하면 (가령 이름 등에 의한) 개인 특정성을 일정 정도 저감할 수는 있지만, 반대로 미리 정해진 속성 이외의 것을 이용하여 개인 특정이 일어나는 경우는 배제할 수 없다는 점에 유의해야 한다. 본 방법을 채택하는 경우에는 (가칭) 개인 특정성 감소 데이터가 어떤 속성에 해당하는 직접적인 개인 특정성을 저감하는지를 미리 공통으로 정하거나 제3자에게 데이터를 제공하는 사업자가 분명히 하도록 정할 필요가 있다고 생각한다.

상기 사항 외에 미리 정한 속성에서 개인 특정성을 저감하는 방법에는 삭제나 고정적인 가공에 의하지 않고, 지난 번 보고서에 기재했던 다양한 가공 기술을 조합하는 방법에 의해도 된다고 생각한다.

168) 가명화를 통해 본인과 데이터의 밀접성이 감소할 수는 있겠지만 일의성(중첩성)/단사성은 유지된다. 일의성(중첩성)/단사성에 대처하기 위해서는 후술할 개인 특정 리스크 3을 저감하는 방법이 유용하다.

공유성은 가명화를 통해 줄일 수 있다고 생각되지만, 가명의 이용기간 길이와 제공처의 수가 많은지 여부에 따라 개인 특정 리스크가 어떻게 변화할지는 예측하기 어렵다. 그러므로 지속적으로 분석하고 검토할 필요가 있다.

변경가능성/불변성/이용정지가능성은 데이터의 취득 시 결정되는 특성이다. 이는 (가칭) 개인 특정성 감소 데

(3) [개인 특정 리스크 ③을 저감하는 방법] 개인정보 또는 (가칭) 준개인정보로부터 (가칭) 개인 특정성 감소 데이터로 가공할 경우 특징적인 값을 가진 속성 등을 삭제하거나 가공하는 방법으로는 이러한 속성이 조합에 의해 원래의 개인정보와 1대1의 관계를 지니지 못하도록 가공하는 방법인 k-익명화나 일부추출(샘플링)의 방법을 이용하는 것이 유용하다.[169]

또한 이동 이력은 그 특성에 따른 가공이 필요하다. [개인 특정 리스크 ①을 저감하는 방법(이동이력의 경우)] 개인을 직접 특정하는 속성을 미리 정한다. (가령 성명, 생년월일, 주소, 소속 단체 등) 이것을 모두 삭제 또는 가공한다. 만일 위치와 시간의 조합을 통해 주소와 소속 단체의 추정이 가능한 경우, 해당 위치 정보를 삭제 또는 가공한다. [개인 특정 리스크 ②를 저감하는 방법(이동이력의 경우)] 개인정보 또는 (가칭) 준개인정보로부터 (가칭) 개인 특정성 감소 데이터를 가공할 경우 식별자 또는 식별자에 상당할 수 있는 것이 다른 사업자와의 사이에서 식별자 또는 식별자에 상당하는 것에 의해 매칭될 수 있는 성질을 갖지 않도록 가공한다.[170] [개인 특정 리스크 ③을 저감하는 방법(이동이력의 경우)] 개인정보 또는 (가칭) 준개인정보에서 (가칭) 개인 특정성 감소 데이터를 가공할 경우, 특징적인 값을 가진 속성이나 여러 속성이 있는 특징 등의 조합은 삭제하거나 가공한다. 이 경우 모든 속성의 조합이 원래의 개인정보와 1대1의 관계를 지니지 못하도록 가공하는 방법(가령 k-익명화)이나 일부 추출(샘플링 등)이 유용하다.[171]

결국, 개인 특정 리스크 ①과 개인 특정 리스크 ②에 대한 대응 조치는 필수적이라 할 수 있다. 이들은 "(가칭) 개인 특정성 데이터"에 내재된 위험이기 때문이다. 한편 개인 특정 리스크 ③에 대해서는 오로지 데이터 수령자의 행

이터에도 계승된다.

169) 본 가공을 기술적으로 달성하는 것은 난이도가 높다.

170) 위치 정보/시간은 보다 넓은 지역/시간대로 일반화하거나 다른 위치/시간으로 임의 치환, 가명의 짧은 시간으로 갱신, 긴 이력은 삭제 등. 위의 수법을 이용하여 위치 정보를 적절하게 가공하고, 같은 위치 정보(이동 궤적을 포함)가 여럿이 있는 상황을 조성 ※ 본 가공의 결과가 식별자 또는 식별자 상당의 것에 의한 복수 정보의 매칭이 불가능하다는 것을 보증하지는 않는다.

171) 위치 정보/시간 보다 넓은 지역/시간대로의 일반화, 다른 위치/시간으로의 임의 치환. 가명의 짧은 시간으로 갱신, 긴 이력은 삭제 등. 위의 수법을 이용하여 위치 정보를 적절하게 가공하고, 같은 위치 정보(이동 궤적을 포함)가 여럿이 있는 상황을 조성. ※ 본 가공을 기술적으로 달성하는 것은 난이도가 높다.

위에 의해서 생길 것으로 우려되는 위험이므로 데이터 수령자의 행위를 금지하면 적절하게 감당할 수 있는 위험이라고 볼 수도 있다. 또한 식별자의 취급에 있어서는 계속성을 줄이는 것(동일한 식별자를 장시간 사용하지 않게 함), 전체성을 낮추는 것(샘플링과 같은 일부 추출을 이용), 도달 가능성을 배제하는 것(메일 주소 등을 삭제) 등의 가공을 하는 것이 바람직하다. 특히 규모가 큰 데이터에 대한 일부 추출은 효과적인 방법일 수 있다.

지금까지 본 것과 같이 (가칭) 개인 특정성 감소 데이터에 대해서, 모든 데이터에 효과적인 구체적인 가공 방법을 제시하기는 어렵다. 다만 개인정보나 (가칭) 준개인정보의 정의가 특정 개인을 식별하는 혹은 그 개인을 특정할 개연성이 높은 것으로 한 것을 감안하면, (가칭) 개인 특정성 감소 데이터는 그 성질이나 특성에 있어 특정 개인을 식별할 개연성이 낮게 되도록 가공을 하여 특정 개인의 식별이 곤란하게 된 것으로 생각할 수 있다. 이상을 근거로 "(가칭)개인 특정성 감소 데이터" 정의는 다음 사항이 포함되는 방향으로 변경하는 것을 제안한다. ⓐ "개인정보"를 "(가칭)개인 특정성 감소 데이터"로 변경하는 경우에는 특정 개인을 식별하는 정보를 가공하여 개인을 식별할 수 없도록 한다. (다만, 위의 가공을 함에 있어 "개인정보"가 "(가칭) 준개인정보"인 경우에는 더욱 낮은 수준의 가공을 하면 된다.) ⓑ "(가칭) 준개인정보"에서 "(가칭) 개인 특정성 감소 데이터"로 변경하는 경우에는 해당 데이터에 포함되는 식별자(또는 식별자에 상당하는 기능을 가진 것)에 가공을 하여 특정 대상을 식별할 수 없도록 한다. ⓒ "(가칭) 준개인정보"의 정의 변경 방안에서 “ㄷ” 에 해당하는 것을 포함하려는 경우 이력 정보 등은 식별자는 아니지만 식별자에 상당하는 기능을 할 수 있다는 점을 고려할 필요가 있다. ⓓ 상기 2가지 항목을 제외한 (가칭) 개인 특정성 감소 데이터에 가공 등을 실시했을 경우 여전히 계속 (가칭) 개인 특정성 감소 데이터이다. (가공 등으로 개인정보 또는 (가칭) 준개인정보가 된 것은 제외한다)

여기에서 (가칭) 준개인정보와 (가칭) 개인 특정성 감소 데이터를 비교하면 다음과 같다. 양자는 함께 개인을 특정하지 않는 정보이지만, (가칭) 준개인정보에 포함되는 개인을 하나하나 구별하는 정보(예를 들면 식별자)는 외부

의 데이터와 비교함으로써 그 정보에서 특정 개인을 식별할 수 있다. 예를 들어 외부 데이터로서 개인의 이름 등과 메일 주소가 반으로 되어 있는 데이터가 있을 경우 (가칭) 준개인정보에 메일 주소가 포함되어 있다고 하자. 이것을 비교하다 보면 개인의 특정에 이르게 된다. 한편 (가칭) 개인 특정성 감소 데이터에는 개인을 하나하나 구별하는 정보가 가공되어 있기 때문에 (예를 들면 가명화에 의한 생성된 식별자) 그 가공한 제공자만 알 수 있는 정보를 이용하지 않으면 특정 개인을 식별하기가 어려워진다.

(b) 최소한의 구체적인 가공방법[172]

개인정보 또는 (가칭) 준개인정보에서 (가칭)개인 특정성 감소 데이터를 가공하여 제3자에게 제공할 때의 사업자가 취해야 하는 조치에 대해서는 위에 정리한 것과 같다. 데이터의 특성과 가공방법이 다양하기 때문에 (가칭) 개인 특정성 감소 데이터로 가공하기 위한 최소한의 방법을 정의할 수 없으므로, 특정 개인에 대한 식별성을 저감시키는 것 자체와 활용하려는 수요 간 균형을 고려하여 사업자 스스로의 판단과 책임 하에 적절히 가공할 필요가 있다. 그러므로 만일 (가칭) 개인 특정성 감소 데이터가 수령자에 의해 개인이 특정되게 되고 그 개인의 권익 침해가 있게 된다면, 제공된 데이터가 적절하게 가공된 것이었는지가 문제될 수 있으므로, 제공된 (가칭) 개인 특정성 감소 데이터에 대한 가공이 적절했는지에 관해 책임의 명확화 등의 관점에 비추어 증거 보존 등의 대책을 수립하도록 검토할 필요가 있다.

제공된 (가칭) 개인 특정성 감소 데이터가 적정하게 가공되었는지에 대해서는 특정 개인을 식별할 수 없는 상태라는 것이 어떤 것인지에 대한 해석이 제시되어야 한다. 가공 방법 등에 대해 최선의 관행을 제시하고 가이드라인을 책정하며 사전상담을 충실하게 하여, 그 가공방법 등이 유연하게만 정의되지 않고 명확해 지도록 노력해야 한다. 또한 개인정보는 특정성이 있는 데이터에 특별히 가공하지 않더라도 해당 정보주체인 본인의 동의를 얻는 방법을 통해 제3자에게 제공될 수 있다. 본인과 관계 사업자의 상황과 개인 특

172) 개인정보와 (가칭) 준개인정보에서 가공하여 (가칭) 개인 특이성 감소 데이터로 가공할 경우

정 리스크와의 관련성 등을 신중하게 판단한 뒤에, 제3자에게 제공하기 위한 방법을 적절하게 선택해야 한다. 마찬가지로 (가칭) 준개인정보에 대해서도 일률적으로 (가칭) 개인 특정성 감소 데이터로 가공한 경우에만 제3자 제공을 가능하게 할 것은 아니다. 본인의 동의를 얻는다면 제3자에게 제공할 수 있도록 하는 규정을 만드는 것에 대해서도 검토해야 한다. 가공 방법 등에 관한 정보는 제3자 기관에 제출하게 되어 있다.

(c) (가칭) 개인 특이성 감소 데이터를 가공하는 사업자가 장차 설치될 예정인 제3의 기관에게 제출할 사항

(가칭) 개인 특정성 감소 데이터 수령자의 취급으로 해당 데이터에 관한 개인을 특정하는 것의 금지 및 안전 관리 조치를 취하는 것이 제안되고 있다. (가칭) 개인 특정성 감소 데이터에 대하여 개인을 특정하는 것을 금지하는 규정에 관하여 특정 개인을 식별하는 것을 막는 구체적인 방법을 나타내는 것이 효과적이라고 생각되지만, 다양한 활용 사례를 상정한 검토가 필요하다. 동시에 구체적인 방법 등에 관한 최선의 관행을 제시하고, 가이드라인의 책정하는 것 등을 추진하는 것이 중요하다. 한편 (가칭) 개인 특정성 감소 데이터에 관한 안전 관리 조치의 구체적 내용으로서, 개인 데이터에 대해서는 그 유출 · 멸실 또는 훼손의 방지 차원에서 안전 관리가 요구된다. (가칭) 개인 특정성 감소 데이터가 완전히 익명화된 개인을 특정할 수 없는 데이터가 아니라는 특성과 그 취급에 주목하고 멸실 또는 훼손의 방지 같은 관점에서 안전 관리를 해야 한다. 이를 감안하면 ① 개인을 특정하지 않기 위한 적절한 조치를 강구해야 하며, ② 사업자 스스로 또는 용역 업체로부터의 정보 유출이 생기지 않도록 적절한 조치를 강구해야 한다는 점에 있어서 안전 관리 조치가 필요하다는 것을 명확히 해야 할 것으로 생각된다.

나아가 정보 누설 등에 관한 안전 관리 조치의 구체적인 예로서, 데이터의 이용 기한을 정하고 이용 종료 후에는 신속하게 데이터를 파기할 것, 네트워크와 분리된 환경을 구축하여 해당 환경 내에서만 데이터의 분석 등을 실시할 것 등이 꼽힌다. 또한 가공의 정도(특정 개인 식별성의 저감 정도)에 따

라서 필요한 안전 관리 조치는 다를 것이며 그 활용 실태에 따른 적절한 대책을 강구해야 한다. 이 밖에 실제로 정보 유출이 발생된 경우의 절차(주지의 방법, 제3자 기관에 대한 보고절차 등)는 실제로 생길 것으로 예상되는 피해의 정도에 따른 것으로, 가이드라인 등에서 명확화 하는 것이 바람직하다. 이와 함께 수령자가 수령한 (가칭) 개인 특정성 감소 데이터에 개인을 특정하는 정보가 포함되어 있는 것을 발견했을 경우 신속하게 이용을 중지하는 것 이외에, 본건에 대해 제공자 및 제3자 기관에 통보하는 동시에 제공자는 즉시 사태 파악과 개선에 노력해야 한다는 취지의 대응을 하는 것을 검토해야 한다고 생각한다. 이때 수령자에게 일정 기간 동안 해당 데이터를 보존하도록 할 것이 필요할 수도 있다.

(가칭) 개인 특정성 감소 데이터의 제3자 제공에 있어, 제3자 기관은 제공자로부터 가공 방법 등에 관한 정보의 제출받고 그 중 영업 비밀 등 사업자의 권리 이익을 해치지 않는 범위에서 정보를 공개할 것으로 알려졌다. 첫째, (가칭) 개인 특정성 감소 데이터의 제3자 제공이 본인의 동의 없이 제3자에게 제공할 수 있는 새로운 유형으로 정리될 것임을 감안하여, (가칭) 개인 특정성 감소 데이터의 수령자가 누구인지를 파악하는 것은 해당 정보를 파악하기 위한 방법 여하를 불문하고, 필수적이라고 생각된다. 또한 한 사업자가 (가칭) 개인 특정성 감소 데이터(저감 데이터 A)를 제3자 제공할 때 해당 저감 데이터 A에 다른 사업자로부터 받은 (가칭) 개인 특정성 감소 데이터(저감 데이터 B)정보가 포함되는 경우도 생각해 보자. 이 경우, 제3자 기관에 제출되는 정보(또는 제3자 기관이 파악해야 할 정보)에는, 저감 데이터 A의 수령자의 정보 이외, 저감 데이터 B의 제공 업체에 관한 정보도 포함해야 한다고 생각한다. 이는 복잡한 데이터의 유통이 상정되는 가운데, 제3자 기관이 신속 · 정확하게 상황을 파악하는 것이 어떤 사태가 발생할 경우의 권익 침해를 최소화할 수 있도록 하는 것이며, 사업자의 부담을 현저히 증가시키는 것이 아니라는 점에서도 효과적인 수단이기 때문이다.

제3자 기관에서 (가칭) 개인 특정성 감소 데이터에 관계된 정보를 집약하는 의미는 만일 (가칭) 개인 특정성 감소 데이터에서 개인이 특정되고 그것에

의해서 어떠한 권리 침해가 생겼을 경우, 그 원인 조사와 대책을 제3자 기관이 실시하기 위해서 필요한 정보를 미리 보관 유지하는 것, 또 (가칭) 개인 특정성 감소 데이터를 취급하는 사업자가 각각 다른 공개 방법에 의해서 공개 사항을 공표하는 것을 대신하여 예를 들면 제3자 기관이 홈페이지에 일람을 가지고 공표하는 등, 어떠한 개인의 권익 침해는 해당 개인이 알아야 할 것이 많다고 생각되므로 그 때문에 해당 개인이 자신에 관한 데이터가 어떤 사업자에서 활용되고 있는지를 쉽게 확인할 수단을 제공할 수 있기 때문이라고 생각된다.

이 가운데 어떤 사안이 생겼을 경우 조사 등에 필요한 최소한의 정보를 사전에 파악하는 것을 위해 여러 데이터의 종류나 다양한 가공 방법을 고려할 수 있도록 해야 할 것이다. 다만 제3자 기관의 기능과 능력이 아직 불명확한 단계에 있기 때문에 기술적인 관점에서 필수 항목을 정리하지는 않았다. 또한 제3자 기관이 사안을 상세하게 검증하는 경우 제공자로부터 실제로 수령자에게 제공된 (가칭) 개인 특정성 감소 데이터가 없으면 정확하고 상세한 조사가 불가능하다는 점에서 실제 데이터 제공도 고려해볼 수는 있겠지만, 이를 실제로 사업자로부터 제3자 기관이 제공받게 되면 이것을 보관하는 것이 곤란할 수 있으며 실제 데이터까지 제공받아 보관하는 것은 권익침해의 우려정도에서 비추어 현저히 균형을 잃었다고 여겨지므로 실제 데이터의 제공을 요구하는 것은 비현실적이다.

(가칭) 개인 특정성 감소 데이터의 가공에 관하여 사업자로부터 제3자 기관에 제출할 정보 항목으로 상정되는 것으로 다음과 같은 것이 거론된다.

① 제공자(제공원) 정보 ② 제공 업체(수령자) 정보 ③ 제공 기간 및 빈도 ④ 데이터 종별 1(개인 데이터, 준 개인 데이터 및 기밀 정보의 유무) ⑤ 데이터 종별 2(구매 데이터 이동 목록, WEB열람 이력 등) ⑥ 데이터의 규모 ⑦ 가공 방법의 개요(각 항목의 상세함, 저감 정도를 판별 할 정도의 샘플을 포함)

이런 항목들에 대해서는 요구되는 개인정보 및 프라이버시 보호와 사업자의 취급 간 균형을 도모하여 필요 최소한의 것이 되도록 검토해야 한다. 또한 상기 항목에 변경이 생길 경우 사전 또는 소정의 기간 내에 신속하게 제3자 기관에 변경 내용을 제출하도록 해야 한다. 한편, 제3자 기관에서 (가칭) 개인 특정성 감소 데이터에 관해 공개할 사항 중에는 (가칭) 개인 특정성 감소 데이터를 취급하고 있는 사업자 이름도 고려할 필요가 있다고 생각되지만, 그 이외의 공개의 필요성에 대해서는 기술적 차원이 아니라 제도적인 관점에서 판단할 사항이라고 생각한다.

3) 개정 개인정보보호법 내용

2년에 걸친 논의 끝에, 2015년 9월 3일 일본은 개인정보 보호법을 개정하였다. 일본정부 IT종합전략본부가 2014년 6월 9일 발표했던 개인정보 보호법 개정초안은 큰 변화 없이 법률에 반영되었다. 앞에서 살폈듯 일본정부 IT종합전략본부는 개정초안을 작성하기 위해 산하에 개인 데이터에 관한 검토회를 두고 별도의 기술검토워킹그룹을 운영하며 논의를 진행해 왔다. 개인 데이터에 관한 검토회의 논의결과가 개인정보 보호법 개정초안에 달리 반영된 부분은 준개인정보에 대한 부분[173]이다.

당초 검토결과에 따르면 준개인정보라는 새로운 정보유형을 규정하고자 하였다. 이를 통해 준개인정보를 1) 면허증번호, 컴퓨터IP주소 등 지속적으로 사용되는 ID, 2) 음성, 지문, 유전자정보 등 개인의 신체적인 특징, 3) 웹사이트 열람기록, 인터넷 쇼핑몰 구매이력 등의 행동이력이라는 3가지로 크게 분류하고, 이를 취급하는 기업에게 관리를 맡기며 의무를 부과하는 방안이 유력하게 검토되었다. 준개인정보는 업계가 추진 중인 빅데이터 활용에 가장 중요한 부분이라고 고려[174]되어 왔다.

---

173) 닛케이신문, 일본 빅데이터 활용, 적절한 개인정보보호가 과제, 2014. 6. 10
http://www.koba.or.jp/bbs/board.php?bo_table=trade&wr_id=2883&page=15w(2015.10.24.방문)

174) 닛케이 신문에 소개된 자료에 따르면, 카고메(カゴメ, 토마토관련 식음회사)는 아마존재팬과 제휴하여 아마존의 구매이력 데이터를 분석하고 있고, 토마토 음료 개발에 활용하고 있으며, 야후재팬과 아스컬(アスクル,

검토회는 이러한 경우 준개인정보를 개인정보에 준하여 취급하고, 이를 제3자에게 제공할 경우 특정 개인에 관한 정보가 노출되지 않도록 데이터를 일부 가공할 필요가 있다고 보았다. 하지만 산업계는 이에 대해 크게 반발해 왔다. 데이터의 가공이 많아질수록 데이터의 유용성은 현저하게 감소하기 때문이다. 그 결과 2014년 6월 9일에 발표된 개인정보 보호법 개정 초안에는 준개인정보 개념 자체가 삭제되었다. 여기에 속해있던 정보 중 신체적 특징에 대한 정보는 개인정보로, 웹페이지 이력 등은 비개인정보로 분류되게 되었다. 나아가 상기 정보를 이용할 경우 신설될 예정인 개인정보 보호위원회에서 프라이버시 침해 여부를 개별적으로 판단하도록 정리하였다.

이하 2015년 9월 3일에 일본 중의원을 통과한 개정 개인정보 보호법 개정안의 요강[175]을 살펴보면, 다음과 같다. “개인정보 보호에 관한 법률 및 행정절차의 특정 개인을 식별하기 위한 번호의 이용 등에 관한 법률 일부를 개정하는 법률”의 목적은 개인정보의 보호를 도모하면서 개인 데이터의 유용한 활용을 촉진함으로써 새로운 산업·신서비스의 창출과 국민의 안전·안심의 향상을 실현 및 마이 넘버의 이용 사무 확충을 위해서 필요한 개정을 실행하는 것이다.[176] 그 내용은 ① 개인정보의 정의의 명확화,[177] ② 적절한 규율 아래 개인정보 등의 유용성을 확보,[178] ③ 개인정보의 보호를 강화,[179] ④ 개인정보보호위원회 신설 및 권한설정,[180] ⑤ 개인정보 취급의 글로벌화,[181] ⑥ 기타 개정 사항[182]으로 요약할 수 있다. 비식별화에 관련된 주요 개정사항은 아래 표와 같으며, 요강의 주요 내용에 대한 번역 전문

오피스용품 회사)도 공동운영하는 인터넷 쇼핑몰 로하코(ロハコ)의 구매이력 데이터를 분석하여 신상품 개발과 판촉활동에 사용하기 시작했다.

175) http://www.cas.go.jp/jp/houan/150310/siryou2.pdf

176) http://www.cas.go.jp/jp/houan/150310/siryou1.pdf

177) 개인정보의 정의의 명확화, 배려가 필요한 개인정보(이른바 기밀정보)에 관한 규정을 신설

178) 익명가공정보에 관한 가공 방법이나 취급 등의 규정의 정비, 개인정보보호지침의 작성이나 신고 및 공포 등의 규정의 정비

179) 추적성의 확보(제3자 제공에 관한 확인 및 기록의 작성 의무), 부당이득을 꾀할 목적에 의한 개인정보 데이터베이스 등 제공죄 신설

180) 개인정보 보호 위원회를 신설하고 현행의 주무 장관의 권한을 일원화

181) 국경을 초월한 적용과 외국 집행 당국에 대한 정보 제공에 관한 규정 정비, 외국에 있는 제3자 에게의 개인 데이터 제공에 관한 규정 정비

182) 본인 동의를 얻지 않는 제3자 제공(옵트 아웃 규정) 신고 공포 등 엄격화, 이용목적 변경을 가능하게 하는 규정 정비, 취급 개인정보가 5000인 아래 소규모 취급 사업자에 대한 대응

은 [별첨 3]에서 확인할 수 있다.

사항	내용
정의	(1) 이 법률에서 "개인정보"란 생존하는 개인에 관한 정보이며, 다음 중 하나에 해당하는 것이다. 1) 해당 정보에 포함된 이름, 생년월일 기타 기술 등으로 특정 개인을 식별할 수 있는 것(다른 정보와 쉽게 조회 비교할 수 있으며 그것으로 특정 개인을 식별할 수 있는 것을 포함한다.) 2) 개인식별부호가 포함될 것 (2) 이 법률에서 "개인식별부호"는 다음 중 하나에 해당하는 문자 번호 기호 기타 부호 중, 정령으로 정하도록 한다. 1) 특정 개인의 신체 일부의 특징을 전자계산기 용으로 제공하기 위하여 변환한 부호이며, 해당 특정 개인을 식별할 수 있는 것 2) 개인에게 제공되는 역무의 이용 혹은 개인에게 판매되는 상품의 구입에 대해 할당되거나 개인에 발행되는 카드 기타 서류에 기재되거나 전자적 방식에 의해 기록된 부호로서 그 이용자나 구매자 또는 발행을 받는 사람마다 다르게 배정 받아 또는 기재되거나 기록됨으로써 특정 이용자나 구매자 또는 발행을 받는 사람을 식별할 수 있는 것 (3) 이 법률에서 "배려가 필요한 개인정보"라 함은 본인의 인종, 신조, 사회적 신분, 병력, 범죄 경력, 범죄로 인한 해를 입은 사실 기타 본인에 대한 부당 차별, 편견 기타 불이익이 생기지 않도록 그 취급에 특히 배려가 필요한 것으로 정령으로 정하는 기술 등이 포함되는 개인정보를 말한다. (4) " 개인정보 데이터베이스 등 " 의 정의에서 이용방법을 보아 개인의 권익을 해칠 우려가 적은 것으로 정령으로 정하는 것은 제외한다. (5) 개인정보취급사업자의 정의에서 취급하는 개인정보의 양 및 이용 방법으로 보아 개인의 권익을 해칠 우려가 적은 것으로 정령으로 정하는 자를 제외한다는 취지의 규정을 두도록 한다. (6) 이 법률에서 "익명가공정보"란 특정 개인을 식별할 수 없도록 개인정보를 가공하여 얻는 개인에 관한 정보이며, 해당 개인정보를 복원할 수 없도록 한 것을 말한다. (7) 이 법률에서 "익명가공정보 취급사업자"란 특정의 익명가공정보를 전자 계산기를 이용하여 검색할 수 있도록 체계적으로 구성한 것 등을 사업용으로 제공하는 자를 말한다.
익명 가공 정보	(1)익명 가공 정보의 작성 등 개인정보취급사업자는 익명가공정보의 작성 등에 대해서 다음과 같이 한다.

	1) 익명가공정보를 작성할 때는 특정 개인을 식별하는 것 및 그 작성에 이용하는 개인정보를 복원할 수 없도록 하기 위해서 필요한 것으로서 개인정보보호위원회 규칙으로 정하는 기준에 따르고 해당 개인정보를 가공해야 한다. 2) 익명가공정보를 작성했을 때는 가공방법에 관한 정보 등의 누설을 방지하기 위해서 필요한 것으로서 개인정보보호위원회 규칙으로 정하는 기준에 따르고 이들 정보의 안전관리를 위한 조치를 강구하지 않으면 안 된다. 동시에, 개인정보보호위원회 규칙으로 정하는 바에 따라 해당 익명가공정보에 포함된 개인에 관한 정보 항목을 발표해야 한다. 3) 익명가공정보를 작성하고 스스로 해당 익명가공정보를 취급하는데 있어서는 해당 익명가공정보의 작성에 사용된 개인정보에 관련된 본인을 식별하기 위해서 해당 익명가공정보를 다른 정보와 대조해서는 안 된다.
취급사업자등의 의무	(2) 익명가공정보의 제공 익명가공정보취급사업자(익명가공정보를 작성한 개인정보취급사업자를 포함한다. (4)에서도 동일)는 익명가공정보를 제3자에게 제공할 때 개인정보보호위원회 규칙에서 정하는 바에 의하여, 미리, 제3자에게 제공되는 익명가공정보에 포함된 개인에 관한 정보항목 및 그 제공방법에 대해 공표하는 동시에, 해당 제3자에 대해서 해당 제공에 관한 정보가 익명가공정보인 사실을 명시해야 하도록 한다. (3) 식별행위의 금지 익명가공정보취급사업자는 익명가공정보(개인정보를 가공하여 작성한 것을 제외)를 취급하는데 있어서는 해당 익명가공정보의 작성에 사용된 개인정보에 관련된 본인을 식별하기 위해서, 가공 방법에 관한 정보 등을 취득하거나 해당 익명가공정보를 다른 정보와 대조해서는 안 된다. (4) 안전 관리 조치 등 익명가공정보취급사업자는 익명가공정보의 안전관리 때문에 적절한 조치, 익명가공정보의 취급에 관한 민원처리 기타 익명가공정보를 적정히 취급하는 데 필요한 조치를 스스로 강구하고 해당 조치의 내용을 공개하도록 노력해야 한다.
감독	(1) 감독의 주체 및 실시 개인정보취급사업자의 감독을 실시하는 주체를 주무대신으로부터 개인정보보호위원회에 고치고, 익명가공정보취급사업자의 감독을 개인정보보호위원회가 실시하도록 한다. (2) 보고 및 출입 검사 개인정보보호위원회는 일정한 경우에 있어서 개인정보취급사업자 또는

	익명가공정보취급사업자(이하"개인정보취급사업자 등"이라 한다.)에 대한 개인정보 또는 익명가공정보(이하"개인정보 등"이라 함)의 취급에 관하여 필요한 보고 혹은 자료의 제출을 요구하고 또는 그 직원에 해당 개인정보취급사업자 등의 사무소 기타 필요한 장소에 출입시키고 검사하는 등 할 수 있도록 한다. (3) 지도 및 조언 개인정보보호위원회는 일정한 경우에 있어 개인정보취급사업자 등에 대한 개인정보 등 취급에 관해 필요한 지도 및 조언을 할 수 있다. (4) 권한의 위임 개인정보보호위원회는 긴급하고 중점적으로 개인정보 등의 적절한 취급 확보를 도모할 필요가 있음 기타 정령에서 정하는 사정이 있으므로 필요가 있다고 인정될 때는 정령으로 정하는 바에 의해, (2)에 의한 권한을 사업 소관 장관에게 위임할 수 있다. (5) 사업 소관 장관의 청구 사업 소관 장관은 개인정보취급사업자 등에 의한 개인정보 등의 적정한 취급을 확보하기 위해서 필요하다고 인정될 때는 개인정보보호위원회에게 이 법률의 규정에 따르는 적당한 조치를 취하도록 요구할 수 있다.
개인정보보호위원회	(3) 소관 사무 위원회는 (2)의 임무를 달성하기 위하여 다음과 같은 사무를 관장한다. 1) 기본 방침의 책정 및 추진 2) 개인정보 및 익명가공정보의 취급에 관한 감독 및 민원 신청에 대한 필요한 알선 및 그 처리를 실시하는 사업자의 협력 3) 인정 개인정보 보호단체 4) 특정 개인정보 취급에 관한 감시 또는 감독 및 민원 신청에 대한 필요한 알선 및 그 처리를 실시하는 사업자의 협력 5) 특정개인정보 보호 평가 6) 개인정보보호 및 적정하고 효과적 활용에 대한 홍보 및 계발 7) 1)에서까지에 제시하는 사무를 실시하기 위해서 필요한 조사 및 연구 8) 소관 사무에 관한 국제 협력 9) 기타 법률에 의거 위원회에 속하게 된 사무

(3) 비식별 가이드라인 (의료영역, 2015)

사업자단체(사단법인)인 ‘일본사진의료시스템공업회(JIRA)’는 개정법안을

바탕으로 2015년 5월 30일 익명화 기술 가이드를 발표하였다. (의료정보활용의 익명화 기술가이드 Ver1.0) 본 가이드는 의료정보 활용이 의료의 진보를 위해 불가피하다는 점을 밝히며, 의료정보 유용한 활용과 안전한 관리를 양립시키는 대책으로 익명화를 이용할 수 있다고 밝힌다. 그리고 의료정보의 익명화에 관한 사회적 배경과 기술적 내용에 대해 해설하고 있다. 이 비식별화가이드라인의 주요 내용을 아래의 박스 안에 간단하게 요약했다.

-개인정보의 익명화 개념-

현재 일본에는 의료정보 익명화에 관한 가이드라인이 2개 존재한다. 하나는 "의료·개호 관계 사업자의 개인정보의 적절한 취급을 위한 가이드라인"이고, 다른 하나는 "의료·개호 관계 사업자의 개인정보의 적절한 취급을 위한 가이드라인"이다. 여기에는 개인정보를 "생존하는 개인에 관한 정보이며 해당 정보에 포함된 이름, 생년월일 기타 기술 등으로 특정 개인을 식별할 수 있는 것(다른 정보와 쉽게 조회 비교할 수 있으며 그것으로 특정 개인을 식별할 수 있는 것을 포함)"라고 정의한다. 또한 여기에는 개인정보의 익명화를 "해당 개인정보에서 해당 정보에 포함된 이름, 생년월일, 주소 등 개인을 식별하는 정보를 제거함으로써 특정 개인을 식별할 수 없게 하는 것"이라고 설명한다. "의료정보 시스템의 안전관리에 관한 가이드라인"에는 익명정보 취급에 관한 규정이 기재되어 있다.

-일본의 동향-

현재 IT종합전략본부의 개인 데이터에 관한 검토회를 중심으로 개인데이터 활용 규칙의 명확화 및 그 환경정비를 위한 논의가 진행되고 있다. 상기 검토회에서 의료정보에 대해서는 ① 건강정보 개인데이터 이용 및 활용의 문제점(제2차 회의 자료1-3) ② 의료 등 분야에서의 개인 데이터의 활용유형 및 그에 대한 고찰(제2차 기술검토워킹그룹 자료4)를 통해 논의를 했다. 이런 논의를 바탕으로 2013년 12월 20일 IT종합전략본부는 "개인데이터의 유용한 활용을 위한 제도재검토 방침"을 발표했다. 여기에는 익명화된 정보를 본인의 동의 없이 제3자에게 제공하기 위한 법적 정비(개인정보 보호법 개정)을 위한 제언이 담겨있었다.[183)]

-익명화 기술-

기술검토워킹그룹은 익명화 기술에 대해 ① 용어 “익명화”를 개인“식별”과 “특정화”로 구분한다. 다만 ② 어떤 기술적인 수법을 취하고도 일반적인 의미에서의 “완전한 익명화”는 있을 수 없다는 점도 인정한다. 또한 익명화 정도와 관련하여 ③ 익명화에는 여러 수준이 있다는 점을 인정한다. 익명화의 수준과 관련하여 (a) 연결가능 익명데이터(성명 등을 삭제하지만 원래의 정보와 대조함으로써 개인과 연결할 수 있는 것. 가명화라고 부른다) (b) 이른바 익명데이터(성명 등을 삭제하고 원래의 정보와 대응하지 못하도록 한 것. 무명화라고 부른다) (c) 고도의 익명데이터(고도의 익명처리에 의해 특정 개인을 식별하는 것이 곤란하도록 한 것)으로 나누어볼 수 있다. 여기에서 개인정보로 분류되는 정보 범위의 경계는 (b)수준 중에 있다고 할 수 있다. 다만 이를 분명하게 규정하기는 어렵다. (c)은 고도의 익명처리를 의미하며, 대표적으로 "k-익명화"[184]가 있다.

-해외동향-

가이드라인은 추가로 해외사례로서 미국과 영국의 사례를 소개한다. 미국의 경우 HIPAA가 규정한 익명화 규칙(두 가지 방법 중 선택이 가능 : 1) 규정된 18개의 속성을 삭제하거나, 2) 통계 분석 전문가들이 직접 개인이 특정될 위험을 평가하고, 충분히 낮은 것을 판단한 분석의 경과 및 결과를 문서화)에 따른 활용이 많이 실시되고 있다. 영국의 경우 EHR을 활용하기 위한 장치로서 SUS(Secondary Uses Service)가 구축되어 있다. 또 의료 정보의 활용을 위한 가이드라인 정비도 진행되고 있어 의료분야에서의 정보보호와 공개의 양립을 도모하기 위해 NHS가 2013년 2월에 "Anonymization Standard for Publishing Health and Social Care Data"라는 가이드라인을 제정하였다. 그리고 익명화 규칙이 명확하게 규정되어 있다.

제3절 비교분석 및 시사점

183) 제3자 제공에 있어 본인동의 원칙에 대한 예외조건을 추가로 제시하였다. 이는 ① 개인이 특정될 수 있는 가능성을 절감한 개인 데이터의 취급은 ② 이를 제공받는 조직이 제공받은 데이터를 다른 데이터와 대조하여 다시 특정하지 않을 것과 ③ 이를 제공하는 업체에서 제공한 사실의 공표를 요구하는 것을 전제하였다. 이는 이른바 FTC 3원칙을 구현한 것으로서, 2015년 9월 개정법에 그대로 반영되었다.

184) k-익명화에서 k는 특정 개인의 데이터를 k개 미만으로 좁힐 수 있지 않는지를 보이는 익명성 지표이다. k를 크게 하면 개인이 특정될 위험은 줄어들지만 동시에 정보량이 떨어져 실용적으로 사용할 수 없는 데이터가 된다는 문제가 존재한다.

1. 국내 규제체제의 문제점

현재 국내 실정법상 개인정보의 범위[185]는 상당히 넓게 인정될 수 있다. 그렇기 때문에 사실상 개인에 대한 관련성이 조금이라도 있는 정보는 거의 예외 없이 개인정보 개념에 포섭되는 것으로 해석될 수 있고, 그에 따라 개인정보 보호법령의 적용대상이 된다. 개인정보 보호법 제18조[186]와 같은 예외조항이 있기는 하다. 하지만 이러한 예외조항으로 개인정보의 이용이 허용되는 범위가 상대적으로 협소한 것으로 보인다.

---

185) 현행 법령을 기초로 개인정보의 범위를 밝힌 대표적인 판례로서 휴대전화번호 사건(2013고단17)과 IMEI 사건(2010고단5343)

186) 제18조(개인정보의 이용·제공 제한)

① 개인정보처리자는 개인정보를 제15조제1항에 따른 범위를 초과하여 이용하거나 제17조제1항 및 제3항에 따른 범위를 초과하여 제3자에게 제공하여서는 아니 된다.

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 제공할 수 있다. 다만, 제5호부터 제9호까지의 경우는 공공기관의 경우로 한정한다.

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

③ 개인정보처리자는 제2항제1호에 따른 동의를 받을 때에는 다음 각 호의 사항을 정보주체에게 알려야 한다. 다음 각 호의 어느 하나의 사항을 변경하는 경우에도 이를 알리고 동의를 받아야 한다.

1. 개인정보를 제공받는 자
2. 개인정보의 이용 목적(제공 시에는 제공받는 자의 이용 목적을 말한다)
3. 이용 또는 제공하는 개인정보의 항목
4. 개인정보의 보유 및 이용 기간(제공 시에는 제공받는 자의 보유 및 이용 기간을 말한다)
5. 동의를 거부할 권리가 있다는 사실 및 동의 거부에 따른 불이익이 있는 경우에는 그 불이익의 내용

④ 공공기관은 제2항제2호부터 제6호까지, 제8호 및 제9호에 따라 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하는 경우에는 그 이용 또는 제공의 법적 근거, 목적 및 범위 등에 관하여 필요한 사항을 행정안전부령으로 정하는 바에 따라 관보 또는 인터넷 홈페이지 등에 게재하여야 한다.

⑤ 개인정보처리자는 제2항 각 호의 어느 하나의 경우에 해당하여 개인정보를 목적 외의 용도로 제3자에게 제공하는 경우에는 개인정보를 제공받는 자에게 이용 목적, 이용 방법, 그 밖에 필요한 사항에 대하여 제한을 하거나, 개인정보의 안전성 확보를 위하여 필요한 조치를 마련하도록 요청하여야 한다. 이 경우 요청을 받은 자는 개인정보의 안전성 확보를 위하여 필요한 조치를 하여야 한다.

설령 동법 제18조 제2항 제4호에 규정(통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공)되어 있는 것처럼 개인정보를 비식별화(내지 익명화)를 하여 이용하려고 해도 법률에 규정된 비식별화 방법이 없기 때문에 개인정보를 이용하려던 행위주체(정부, 공공기관, 연구기관, 경제주체 등)는 이를 주저하게 된다. 정부 차원에서 제정하고 보급한 가이드라인이 있기는 하지만 이것은 법률이 아니며 어떻게 현행법과 조화롭게 해석되어 실무적으로 유용한 지침을 제공할 수 있을지도 불분명하기 때문이다.

설령 이러한 가이드라인이 현행법과 조화롭게 해석되는 것이 가능하고 이를 철저하게 준수하여 개인정보를 비식별화 한다고 하더라도, 행위주체는 비식별화된 정보를 자유로이 이용할 수 없다. 국내 실정법과 그에 대한 판결례에 비추어볼 때 아무리 철저하게 비식별화(내지 익명화)를 하더라도 재식별이 사후적으로 어떤 식으로든 가능해지면, 해당 정보는 '개인정보'로 회귀하게 되고, 이로 인해 온갖 법적 문제가 발생할 수 있기 때문이다.

결국 국내 실정법과 그에 대한 판결례에 비추어 볼 때, 실무자가 안심할 수 있는 비식별화(내지 익명화)는 사실상 불가능하다. 이와 관련해 빅데이터 분석에 관심을 가질 기업체의 실무진과의 인터뷰를 통해 개인정보에 대한 국내의 규제와 비식별화의 활용에 대한 현장에서의 목소리를 들을 수 있는 기회를 가지게 되었다. 이 인터뷰의 전체적인 의의와 그 내용을 몇몇 단락으로 재구성해서 아래의 박스 안에 요약을 했다.

1. 인터뷰의 의의

비식별화가 개인정보 보호에 어떠한 영향을 주는지를 연구하는 데에 있어, 가장 중요하게 고려해야 하는 것들 중 하나는 현장에서 개인정보를 직접 보호하고 이용하는 실무진의 목소리이다. 2011년 개인정보 보호법이 한국사회에 등장한 이래로 줄곧 개인정보 보호법이 규제 위주이고 이용을 간과하고 있으며, 비식별화

를 통한 개인정보 보호 및 이용에 있어 적극적인 도움을 주지 못하고 있다는 목소리가 현장에서 반복적으로 제기되어 왔다.

하지만 이를 두고 체계적으로 정리한 문헌자료는 직접 접하기 어려웠다. 기업의 입장에서 수요를 명확하게 드러내어 문서화하는 것이 현실적으로 어려웠기 때문이라고 사료된다. 기업이 어떠한 수요를 가지고 있으며 현행 법령이나 제도의 특정 부분이 문제라고 지적하는 것은, (1) 기업의 비즈니스 모델이나 사업계획을 동종업종의 경쟁자에게 드러낼 수 있다는 위험이 있고, (2) 규제기관에게 규제기관의 문제점이 무엇인지에 관해 직/간접적으로 지적하는 모습이 되어 피규제자 입장으로서 조심스러울 수밖에 없을 것이었기 때문이다.

본 연구에서는 빅데이터 분석에 관심이 많은 기업에서 비식별화 업무를 담당하는 임원급 실무진을 직접 방문하고 면담하였다. 이를 통해 (1) 현장에서 느끼는 비식별화 현실이 어떠한지, (2) 비식별화에 있어 문제점이라고 생각되는 점이 무엇인지, 그리고 (3) 어떠한 방향으로 관련 법령이나 제도가 변화되기를 바라는지 등을 위주로 정리해 보았다. 면담에 응했던 실무진들의 요청으로 구체적인 기업명이나 사업 분야는 공개하지 않는다.

2. 현장에서 느끼는 비식별화 현실

(1) 우리는 빅데이터를 수집하고 분석하며 활용하고 싶다.

기업이나 공공기관의 입장에서 실무자들은 빅데이터를 수집하고 분석하며 활용하고 싶다. 사업을 하거나 정책을 수립하고 추진하는 과정에서 실무자들은 크게 두 가지 사항들에 직면하게 된다. 하나는 1) 어떻게 하면 주어진 상황에서 가장 최선의 비즈니스 모델을 수립하거나 최선의 경영전략을 수립하여 추진해 나갈지에 대한 "경영판단 또는 정책판단"을 하는 것이다. 두 번째는 2) 어떻게 하면 현재 추진하고 있는 비즈니스나 정책을 소비자나 국민에게 잘 "마케팅을 하거나 홍보"를 하여 최초로 수립했던 비즈니스 모델이나 정책의 목적을 달성하느냐는 것이다.

이러한 1) "경영판단 또는 정책판단"과 2) "마케팅을 하거나 홍보"를 하기 위해서는, "모호하고 산발적"인 데이터를 "구체화하고 구조화"하여야 한다. ICT기술의 발전으로 급속히 성장한 정보화 사회에서 각종 전자기기가 생산해낸 데이터는 다양(Variety)하고 무궁(Volume)하며 신속(Velocity)하게 쌓여가고 있다. 만일 이를 효율적으로 수집하고 분석하여 활용(Value)할 수 있다면, 이는 소위

빅데이터로서 민간과 모호하고 산발적인 데이터를 저렴한 비용으로 효율적으로 구체화하고 구조화함으로써, 민간과 공공에 엄청난 편익을 가져다 줄 수 있다.

(2) 하지만 우리는 빅데이터를 수집하고 분석하며 활용할 수 없다.

정부에서는 빅데이터 산업을 키우기 위해 각종 도움을 주려고 한다. 하지만 그럼에도 불구하고 우리는 빅데이터를 제대로 수집하고 분석하며 활용할 수가 없다. 빅데이터 산업의 본질에 비추어볼 때, 산업 활성화에 걸림돌이 되는 것들이 여전히 방치되어 있기 때문이다. 빅데이터 산업이 활성화되기 위해서는 빅데이터 산업을 활성화할 수 있는 인적 물적 기반이 필요하며, 이를 뒷받침 하는 제도적인 지지가 요구된다. 현재 한국의 현실에 비추어볼 때 빅데이터 산업 활성화에 걸림돌이 되는 것들은 주로 법제상의 이슈들이라고 생각한다. 한국의 각 기업이나 공공기관은 이미 충분한 전문가를 보유하고 있으며 빅데이터 분석 수단을 지니고 있기 때문이다.

빅데이터 산업을 활성화하기 위해서는 빅데이터의 본질에 주목해야 한다. 빅데이터 산업이 활성화되기 위해서는 우선 1) 데이터가 "충분히 확보"되어야 한다. 다음으로 2) 확보된 데이터들로부터 "결합 시너지"를 얻을 수 있어야 한다. 마지막으로 3) 데이터를 이러 저리 돌려보다가 의도치 않게 강력한 상관관계나 인과관계를 얻을 수 있는 "탐험적 성격"이 존중되어야 한다. 이것이 뒷받침 되어야만 데이터로부터 패턴을 발견하여 정보를 추출하고 이를 기초로 통찰(wisdom)을 획득하여 "경영판단 또는 정책판단"과 "마케팅을 하거나 홍보"에 활용할 수 있다. 하지만 현행 법제도의 현실에 비추어 보면, 위의 세 가지는 모두 시도해 보기조차 어렵다. (여기에 대해서는 후술)

(3) 유일한 대안은 "비식별화" 라고 판단된다.

– 실무자는 "비식별화 가능성"에 관심이 많다–

현재 실무자들의 상당수는 당면한 수요에 부합하여 빅데이터 산업의 성과를 보여주기 위한 유일한 대안이 "비식별화"라고 생각하고 있다. 가령 전단에서 언급했던 "데이터 확보" 측면만 두고 살펴볼 것이다. 빅데이터 산업은 결국 데이터의 충분한 확보가 관건이다. 그런데 대부분의 산업에서는 데이터의 자급자족이 불가능하거나 사실상 어렵다. 아무리 거대한 제조업체(삼성, 엘지, 현대 등)나 통신회사(KT, SKT, LGU+ 등)라고 해도 필요한 데이터는 항상 부족하다. 물론 일부 유통업체(백화점, 편의점 등)나 포털(Naver, Kakao), 공공기관의 경우는

유용한 데이터가 자급자족할 수 있을 정도로 충분할 가능성도 있겠으나, 이는 실무자들의 입장에서는 오히려 예외적인 경우라고 보인다.

-공공데이터의 개방만으로는 한계가 있을 수 있다-

결국 수요만큼의 데이터를 확보하지 못한 대다수의 기업이나 공공기관은 데이터의 공개나 거래에 의존하는 수밖에 없다. 그런데 자사 또는 자기관리 이외의 소스로부터 데이터를 공급받기 위해서는 법령상 예외가 있는 경우 등을 제외하고는 "정보주체의 사전동의"를 받아야 한다. 하지만 현실적으로 정보주체의 사전동의를 획득해 놓은 데이터는 거의 찾기 어렵다. 결국 "비식별화를 통해 개인정보가 아닌 것으로 간주되는 데이터를 거래하여 획득하는 방법" 밖에는 없다. 물론 공개된 데이터를 오픈 데이터 소스(개방된 공공데이터 등)에서 찾아볼 수는 있다. 하지만 이는 기업이나 공공기관의 입장에서 1) 필요로 하는 정보가 아닌 경우가 허다하고(가령 통계화된 정보라면, 해당 업무에 대한 맞춤형 통계가 아니라면 별로 의미가 없다. 모든 업무에 맞게 통계화를 할 수는 없기 때문이다.) 2) 실제로 필요한 정보는 내놓지 않는 경우가 빈번하기 때문이다.

3. 문제점이라고 생각되는 점

(1) 오해가 만연해 있다.

-실무자도 개인정보 보호와 프라이버시 존중을 중요하게 생각한다-

빅데이터를 활용하고자 하는 기업이나 공공기관의 실무자 입장에서 답답한 점은 빅데이터 산업에 대한 오해나 불신이 사회에 지나치게 만연해 있다는 점이다. 빅데이터 산업에 종사하는 실무자라고 하면 흔히들 개인정보의 보호는 백안시하며 개인정보의 이용에만 혈안이 되어 있는 사람들이라고 색안경을 끼고 대하는 경우가 많았다. 하지만 기업이나 공공기관이 개인정보 보호를 간과하고 있다거나 프라이버시의 가치를 폄훼하고 있다고 보는 것은 매우 잘못된 인식이다. 오히려 빅데이터 산업에 관심을 가지는 기업이나 공공기관일 수록 개인정보 보호와 프라이버시 존중을 중요하게 고려한다. 사업이나 기관의 평판이 달려 있어 자칫하면 기업이나 기관이 망하거나 입지가 대폭 축소되는 문제가 생길 수도 있다는 우려 때문이다.

-실무자에게는 개인 "특정"정보가 아니라 "식별"정보가 필요하다-

이러한 오해가 생기고 확대재생산하게 된 원인 중의 하나로 1:1 마케팅(소위 타겟마케팅(target-marketing))에 대한 신화가 한 역할을 했다고 생각한다. 지난 10여 년 동안 적지 않게 발생해 온 개인정보 유출 문제로 인하여 고유식별정보가 유통되고 이로 인해 보이스피싱과 같은 부당한 설득에 대한 불안감 이라는 사회문제가 발생하면서 기업의 1:1 마케팅이나 공공기관의 1:1 정책홍보 조차 부정적으로 바라보게 된 것이 아닐까 여겨진다. 하지만 사실 현재 한국의 대다수 기업이나 공공기관은 1:1 마케팅에 별로 관심이 없다. 배너광고나 쿠폰 등을 이용한 마케팅이나 정책홍보가 실제 수익창출이나 효과증대로 연결된 비율이 극히 미미한 것으로 판단하고 있기 때문이다. 오히려 이러한 시도는 대다수의 개인들에게 거부감만을 유발하고 있는 것으로 판명되었다. 결국 현재 빅데이터 산업에 관심 있는 기업이나 공공기관에게 유용한 빅데이터 분석의 활용은 개인화된 일대일 마케팅이 아니라 어느 정도의 통계화된 정보이다. 가령 기업의 경우, 어느 지역 안의 전반적인 소비자들의 행태를 분석하기 위한 개인 데이터들의 획득과 가공이 필요할 뿐이다. (물론 1:1 마케팅이나 홍보의 규모가 장차 커질 수 있다는 가능성은 배제하지 않는다. 다만 정책 입안자들이 이러한 현실을 감안해 주시기를 희망한다.)

(2) 법령이 모호해서 구체적인 가이드를 주지 못한다.

-개인정보 정의가 너무 모호하다 -

우리 법 상 개인정보의 정의는 살아 있는 개인에 관한 정보로서 개인을 알아볼 수 있는 정보이며, 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다. 그런데 다른 정보와 결합하여 특정 개인을 식별할 수 있는 정보가 모두 다 개인정보 보호법제가 적용되는 개인정보에 해당한다고 보게 되면, 사실상 개인정보 보호법의 적용을 받지 않게 되는 개인에 관한 정보는 찾아보기 어렵다. 오로지 비식별화를 통해 특정 개인을 식별할 수 없도록 만든 정보만이 개인정보 보호법의 적용을 벗어날 수 있게 된다. 현행 개인정보 보호법제가 적용되면 엄격한 사전동의 원칙이 적용되어 사실상 빅데이터를 운용할 수 없으므로 기업 및 공공기관은 비식별화에 관심을 기울이고 있다.

하지만 현행 법제에서는 기업들이 비식별화 방식을 자신있게 적용해서 빅데이터의 활용도를 높이기가 어렵다. 현재 개인정보보호법 상에서의 '개인정보'의 개념 규정을 보면 다른 정보와 '쉽게 결합해서' 알아볼 수 있는 정보도 개인정보에 해당한다고 너무 불확실하게 규정되어 있어서 모든 잠재적인 재식별가능성이 있는 정보도 개인정보에 해당한다고 해석할 여지가 충분하다. 그렇다면 기

업들은 자신들이 예상할 수 없는 형태로 미래에 자신도 생각하지 못한 제 3자에 의해 재식별될 가능성이 있기만 하다면 계속적으로 개인정보의 규제 대상이 되는 위험성을 가지게 된다. 이런 상황에서 어떤 기업도 비식별화 방식을 통해서 개인정보를 활용할 인센티브를 가지지 않게 되고, 이것이 현재 국내의 상황이다. 결국 데이터를 획득하고 다른 정보와 결합하여 결합 시너지를 확보하기가 어려울 뿐만 아니라 엄격한 사전동의 원칙의 적용으로 말미암아 빅데이터의 탐험적인 성격도 살리기 힘들다.

-현행 사전동의 접근법의 효과성에 의문이 있다 -

개인정보 보호법은 법령으로 규정해 놓은 몇 가지 예외를 제외하고는 모든 개인정보의 수집, 분석 및 활용에 있어 정보주체의 사전동의를 받도록 요구한다. 그런데 오늘날 빅데이터를 운용하고자 하면 적게는 수천, 많게는 수백만 개의 데이터를 처리해야 한다. 설령 이미 사전동의를 획득하여 보유하고 있던 데이터를 비식별화하여 마이닝(mining)을 하려 하여도 이것이 개인정보의 수집목적에 부합하는 활용에 해당하는지가 모호하게 될 뿐만 아니라, 만일 활용에 해당한다고 하면 데이터 개수만큼의 정보주체로부터 개별/구체적인 사전동의를 받아야 한다.

비용도 만만치 않을 뿐만 아니라, 정보주체는 대부분 동의를 주저하는 경향이 있어 새로운 서비스 개발로 인한 편익도 누리기 어려워진다. 그러므로 사전동의는 비현실적이다. 설령 사전동의를 받는다고 해도 이것이 정보주체의 개인정보 자기결정권 보호를 위해 효과적인지는 의문이다. 이는 비전문가에게 전적으로 사용결정권을 맡기게 되고, 전문가가 위험방지를 위해 개입할 여지를 없애게 한다는 측면에 비추어볼 때, 기업에게 면책을 주는 방편일 뿐 소비자를 보호하기 위해 유익하다고 보기만은 어렵기 때문이다.

4. 기대하는 변화방향

-개인정보 개념의 명확화가 필요하다 -

이런 상황에서 기업들은 정부가 적극적으로 나서서 이와 같은 불확실성을 해소시켜주기를 바라고 있다. 특히 빅데이터 분석 가치의 잠재성이 커지고 있는 상황에서 기업들은 정부가 비식별화와 관련된 구체적인 기준들을 제시해서 기업들이 관련 규제의 불확실성으로 행동을 자신있게 하지 못하는 현상황을 타개해 줄 것을 바라고 있다. 실제 한 기업에서는 자체적으로 습득한 개인정보를 비식별화하려는 행위 자체가 혹시 개인정보의 '활용'에 해당할 수 있지 않을까하는 고민을 가졌다. 만약 이에 해당하면 비식별화 행위 자체도 사전동의의 대상이 될

수 있다. 이런 문제점을 해결하기 위해 기업 내부의 법률팀에게 실제 자문을 구한 적이 있었다. 이와 같이 관련 기준들의 불확실성이 기업 입장에서는 상당한 걸림돌로 작용하고 있는 것이다.

이러한 불확실성의 해소 방향으로 개인정보 보호법상 개인정보의 정의가 좀 더 명확하게 규정되거나 이런 개념을 구체화하는 신뢰할 수 있는 기준들이 제시되기를 희망한다. 기업이나 공공기관이 운용하는 데이터는 기상데이터 등 극히 일부를 제외하고는 대부분 개인에 관한 데이터이다. 그런데 개인에 관한 데이터는 아무리 사소한 것일지라도 개인정보 보호법제상 사전동의 원칙 및 관련규제가 엄격하게 관철된다. 아무리 비식별화를 하여 보호와 이용의 균형을 꾀하더라도 개인정보의 정의가 모호하고 비식별화 방식 및 절차에 대해 확립된 기준이 없어 기업 및 공공기관은 늘상 재식별의 위협을 받게 되고 빅데이터의 운용에 대해 주저할 수밖에 없다. 그러므로 개인정보의 인정범위를 합리적으로 조정하는 방식으로 개인정보의 정의가 실질적으로 명확해지기를 희망한다.

-현재의 사전동의 원칙에 대한 검토가 필요하다-

나아가 사전동의 원칙에 대해서도 재검토를 할 필요가 있다. 개인정보 활용기술은 날로 복잡해지고 있는데 비전문가에게 전적으로 사용결정권을 부여하는 사전동의 원칙이 최선의 방편인지는 실무적 관점에서 의문을 제기할 수밖에 없다. 이는 현실적 관리 능력이 없는 주체에게 '네 것이니 네가 결정하라' 고 한 격이 아닌지 묻고 싶다. 나아가 이는 전문가가 위험방지를 위해 개입할 여지를 없앤 것이어서, 결국 기업이나 공공기관의 면책에 활용될 뿐인 것은 아닌지 재고할 필요가 있다. 전국민이 모두 나서서 고민하지 말고 차라리 이를 (1) 전문기관의 감시에 맡기는 것이 더 좋지 않을까 사료된다. 이는 이미 식품의약품안전처 같은 전문기관이 좋은 선례를 보이고 있는 방식이다. 또한 정보주체가 (2) 선택해야 할 항목을 인지할 수 있도록, 그리하여 무엇을 사용하고 있는지 중요한 것은 무엇인지 나중에라도 바꿀 수 있도록 하는 방식을 도입하는 것을 제안하고 싶다. 이는 이미 영양성분 표시 등의 사례에서 살펴볼 수 있다. 나아가 사후동의의 가능성을 열어주어 (3) 사안발생시 개별동의를 받을 수 있도록 하는 것이 어떨까. 이는 빅데이터의 탐험적 성격과 기업의 변화 노력에 현실적으로 부합하는 방식일 것으로 보인다.

2. 외국 규제체제들의 특징 및 비교

지금까지 연구결과를 바탕으로 볼 때, 해외에서 비식별화를 하여 정보를 이용하는 경우에 크게 네 가지 방법이 사용되는 것으로 생각해 볼 수 있다. 첫 번째는 미국 HIPAA ‘expert determination rule’ 상의 비식별화 방식, 즉 기본적으로 전문가의 경험과 식견에 의존하여 비식별화 수준을 판단하는 방식이다. 두 번째는 미국 HIPAA ‘safe harbor’ 상의 비식별화 방식으로, 식별자 및 준식별자로 사용될 수 있는 정보 유형을 법령에 특정하여, 이를 삭제하고 나면 비식별화된 것으로 의제하는 방식이다. 세 번째는 일본 개정 개인정보 보호법상 비식별화 방식으로, ‘익명가공정보’ 등 새로운 개인정보 개념을 도입하여 정보의 활용가능성을 명시적으로 인정하고, 그와 동시에 제3의 기관인 개인정보 보호위원회를 설립하여 이를 감시하도록 법제도를 설계하는 것이다. 마지막 네 번째는 영국에서와 같이 추상적인 ‘motivated intruder’ 개념을 도입하여, 재식별 가능성에 대해 합리적인 수준의 일정한 기준을 정하고 그에 기초하여 재식별 가능성을 평가하는 방식이다.

그런데, 어떤 방법이건 비식별화를 통해서 재식별의 가능성을 완벽하게 차단하는 것을 전제로 하지 않음에 주목할 필요가 있다. 즉, 비식별화 또는 익명화를 통해서 재식별의 가능성을 완벽하게 차단하는 경우만을 염두에 둔다면, 비식별화 또는 익명화에 관해 별도로 논의할 실익 자체가 없어진다. 왜냐하면, 재식별의 가능성이 완벽하게 차단된 정보는 더 이상 개인정보라고 부를 수도 없고, 그러므로 원칙적으로 개인정보 보호법령의 규제를 받지도 않을 것이기 때문이다.

비식별화의 첫 번째 방법으로 미국 HIPAA expert determination rule 상의 비식별화 방식을 생각해 보자. 이 방식에 따르면 전문가의 경험과 식견에 의존하는 한편, 완벽한 비식별화가 아니라 재식별의 가능성을 약간 남겨둔 상태에서 비식별을 인정하게 된다. 위에서 상세하게 살펴본 것과 같이, 이러한 절차적 방식에 의한 접근법에 따르면 우선적으로 의료정보관리기관은 대상이 되는 개인데이터가 HIPAA 프라이버시 규칙의 기준에 따라 식별가능한 정보인지 여부를 판단할 전문가를 선임한다. 선임된 전문가가 해당 개인데이

터가 식별화될 위험성이 매우 작다(very small)고 판단하면 이 개인데이터는 식별할 수 없는 데이터로 간주되어 HIPAA Privacy Rule의 규제 대상이 되지 않는다. 이 방법은 달리 생각하면, 아주 작은 수준의 재식별 리스크에 대해서는 현실적으로 용인하여 데이터의 활용을 허용하는 것으로 생각할 수 있다. 아주 작은 수준의 재식별 리스크란, 리스크가 영(0) 수준인 완벽한 비식별을 의미하지 않는다. 매우 작은 수준일지라도 재식별 리스크가 있는 상황에서는, 재식별이 흔치는 않지만 드물게는 발생할 가능성을 염두에 두게 되는 것이다.

두 번째로, 미국 HIPAA safe harbor 상의 비식별화 방식을 생각해 보자. 이 방식에 따르면 식별자 및 준식별자를 법령에 특정하여, 이를 삭제하고 나면 비식별화된 것으로 의제하게 된다. 이러한 내용적 방식에 의한 접근법에 따르면 HIPAA 프라이버시 규칙에 열거된 18가지 유형의 데이터들이 해당 데이터에서 제거가 되고, 그 결과 남은 데이터들이 다른 정보와 결합해서 개인을 식별할 가능성이 있다는 사실에 대한 인식을 가지지 않는다면 해당 데이터는 HIPAA의 제한을 받지 않는 상태로 수집과 처리가 가능해진다.

이 방법은 실무적인 처리가 간단하고, 처리에 대한 제3자에 의한 검증도 수월하다는 장점이 있다. 다른 한편, 이 방법의 근본적인 한계는 이 방법에 의할 경우에 재식별 가능성이 얼마나 낮아질 것인지에 대해 확신할 수 없다는 것이다. 재식별 가능성은 여러 가지 요인의 영향을 받아 변화할 수 있는데, 이 방법은 그러한 개별 사안의 특징과 무관하게 18개의 식별자 및 준식별자를 특정하여 일괄적으로 삭제할 것으로 요구하므로, 경우에 따라서는 삭제된 데이터로부터 재식별이 가능해질 수 있는 상황도 존재할 수 있다. 결국, 이 방법도 주요 식별자 및 재식별자의 삭제를 통해 재식별 리스크를 현저하게 낮추기는 했지만 이를 완벽하게 제거한 것이라고는 할 수 없는 방법이 된다.

세 번째 방법은 일본 개정 개인정보 보호법상의 비식별화 방식이다. 이 방식은 익명가공정보 등 새로운 개인정보 개념을 도입하여 정보 활용이 가능성을 열어두는 한편, 제3의 기관인 개인정보 보호위원회를 설립하여 이를 감시

하도록 법제도를 설계한 것이다. 특히 새로 개념이 도입된 익명가공정보의 경우 제3자 제공시 정보주체의 동의를 받지 않아도 되도록 하였다. 그런데, 이러한 익명가공정보를 작성할 때는 특정 개인을 식별하는 것 및 그 작성에 이용하는 개인정보를 복원할 수 없도록 하기 위해서 필요한 것으로서 개인정보보호위원회 규칙이 정하는 기준에 따라 해당 개인정보를 가공해야 한다. 동시에, 개인정보보호위원회 규칙으로 정하는 바에 따라 해당 익명가공정보에 포함된 개인에 관한 정보 항목을 발표해야 하는 등, 익명가공정보 취급사업자 등은 사실명시의무, 식별행위금지의무, 안전관리조치의무를 지게 되었다. 그에 대응하여 개인정보 보호위원회가 설립되게 되어 보고 및 출입 검사, 지도 및 조언 등을 하며 익명가공정보 등에 대한 보호와 관리를 하게 되었다. 이 방법은, 일정 수준 이상으로 비식별화된 정보에 대해 활용가능성을 열어두는 한편, 그 과정에서의 절차적, 규범적 통제를 강화한 것으로 해석할 수 있다. 익명가공정보는 완벽한 비식별화 조치를 전제로 하는 정보라고 할 수는 없다. 다만, 식별행위를 금지하는 의무를 부과하고 안전관리를 위한 의무를 부과하는 등의 규범적, 절차적 과정을 통해 실제로 재식별 상황이 발생할 가능성을 통제하는 것이다.

네 번째는 영국에서처럼 ‘motivated intruder’ 또는 그와 유사한 개념을 도입하여 판단하는 방식이다. 이 방식에 따르면 잠재적 공격자에 의한 비식별 정보의 재식별 가능성에 대해 합리적인 수준의 일정한 기준을 정하여, 그에 기초하여 재식별 가능성을 평가하게 된다. 이 방법을 이용할 경우, 기술적인 접근법이 아니라 관리적인 접근법을 좀 더 강조하게 된다. 여기서 의도된 공격자(motivated intruder)는 해당 데이터에 대해 이미 잘 알고 있는 전문가나 해커 등의 전문기술자는 포함하지 않고, 그와 다르게 합리적 수준의 능력을 가지고 인터넷, 도서관, 공개된 공공문서 등에 대한 접근성을 가지면서 특정개인에 대한 추가 지식을 가진 사람들에 질의할 수 있는 탐색기술을 가진 자를 상정하는 것이다. 이와 같은 의도된 공격자를 전제로 하여 재식별 가능성을 평가하게 되므로 당연히 해당 데이터에 대한 전문가나 기술전문가 등에 의해 재식별이 일어날 가능성에 대해서는 평가하지 않는다. 이는 달리 말하면, 전문가 등에 의해 재식별이 발생할 가능성을 열어두고 있

는 방법이라 평가할 수 있다는 것이다. 그에 따라 이 방법은 재식별의 가능성이 무한대로 확장되는 것을 방지해 주는 역할을 하게 된다.

제5장 합리적인 비식별화 규제 체제의 형태

제1절 외국 규제체제의 국내 적용 가능성 검토

현행 국내법을 전제로 할 때, 앞서 소개한 ① 미국 HIPAA expert determination rule 상의 비식별화 방식, ② 미국 HIPAA safe harbor의 비식별화 방식, ③ 일본 개정 개인정보 보호법상 비식별화 방식, ④ 영국 motivated intruder 개념을 도입하여 평가하는 방식은 어떤 방법이든 허용되기 어렵다. 만일 이런 외국의 규제체제를 국내에 적용하려 한다면 개인정보 보호법 등 관련법령의 개정이 필요할 것으로 보인다.

이 중 첫 번째 방법인 HIPAA expert determination rule 은 명시적으로 재식별의 가능성이 매우 낮은 수준을 평가의 기준으로 제시하고 있고, 따라서 암묵적으로는 재식별의 가능성이 매우 낮지만 그럼에도 불구하고 사후적으로 재식별이 가능할수도 있는 상황의 존재를 인정하는 방법이 된다. 이처럼 재식별 가능성을 인정하는 것을 전제로 비식별화를 하는 것은 국내의 현행 개인정보 보호법령 체계상 허용되는 것으로 보기는 어렵다.

두 번째 방법인 HIPAA safe harbor rule의 경우도 암묵적으로 재식별의 가능성을 전제로 하는 것이라 볼 수 있다. 이 방법에 따르면 명시적으로 나열된 식별자 및 준식별자를 제거한 후 정보를 활용할 수 있게 되는데, 해당 식별자 및 준식별자가 삭제된 후에는 재식별이 불가능할 것이라 보장할 수 있는 장치가 현실적으로는 마련되어 있지 않기 때문이다. 오히려 그 반대로 정해진 식별자 및 준식별자를 제거한 후 추가적인 연구 과정에서 재식별이 가능하게 될 가능성은 상존하는 것으로 보아야 할 것이다. 따라서 HIPAA expert determination rule의 경우와 마찬가지로 재식별 가능성을 인정하고 있는 방법이라 볼 수 있고, 그런 만큼 현행 국내 법령상 인정되기 쉽지 않은 방법이 될 것이다. 다만, 국내 데이터 환경에 기초한 별도의 연구와 논의를 통해 HIPAA safe harbor rule에 상응하는 국내 기준을 마련하는 시도는

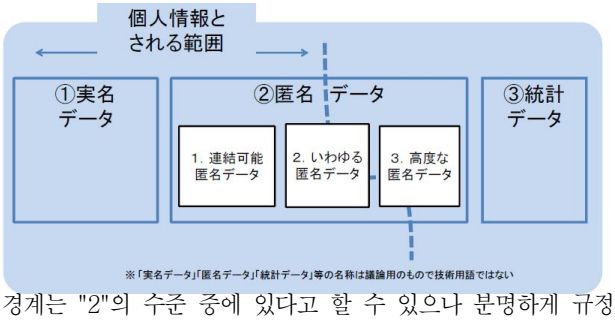
해볼 가치가 있을 것으로 보인다. 이를 통해, 국내 실정법과 조화를 이룰 수 있는 방식으로 식별자 및 비식별자를 특정하는 것이 가능할 것인지 좀 더 구체적으로 모색해볼 필요는 있을 것이다.

세 번째로, 2015년 개정 일본 개인정보보호법상 비식별화 규제체제에 대해 생각해 보자. 동 개정법은 (1) 개인정보와 (2) 개인정보 아닌 정보 사이의 이분법적이고 경직된 접근방식에서 탈피하여, 그 중간 영역에 속하는 정보의 유형을 새롭게 도입한 것이 커다란 특징이다. 이는 식별가능성을 정도(degree)의 문제로 파악하고, 개인정보의 개념 자체에 대해 좀 더 유연하게 생각할 필요가 있다고 주장하는 Schwartz & Solove (2011)의 주장과도 궤를 같이 하는 것이다.[187]

개인정보의 제3자 제공에 있어서 본인동의 원칙에 대한 예외조건을 추가한 동 개정법은 개인정보를 크게 3가지로 나누는 것을 전제[188]하고 있다고 보

187) Paul M. Schwartz&Daniel J. Solove, The PII Problem: Privacy and a New Concept of Personally Identifiable Information, New York University Law Review, Vol. 86, 제 1814면 (2011)

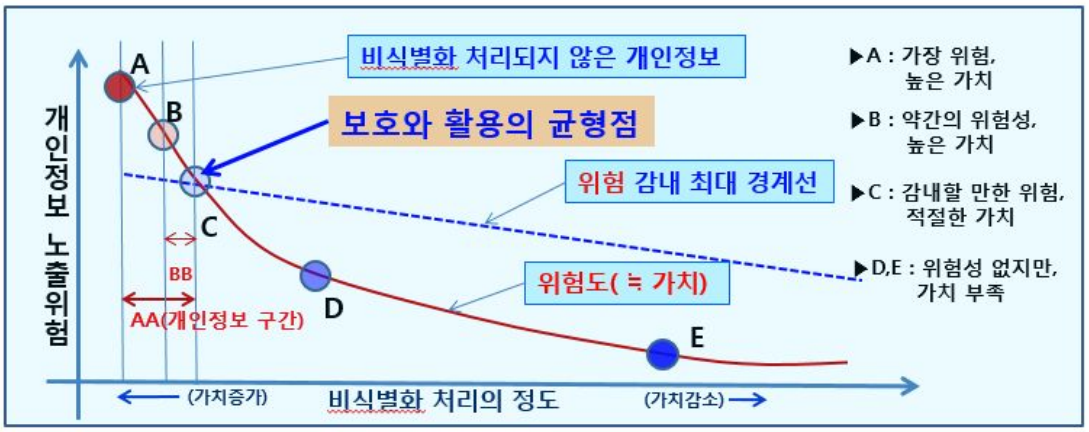
188) 일본사진의료시스템공업회(JIRA), 의료정보활용의 익명화 기술가이드 Ver1.0, 2015. : 입법의 전제가 된 "개인 데이터에 관한 검토회" 기술 검토 워킹 그룹 자료 중 ① 익명화 기술의 현상에 대해서(타카하시 구성원): 제1회 기술 검토 워킹 그룹 자료 2-3, ② 개인 식별할 수 없는 익명 데이터는 작성할 수 있나(타카하시 구성원): 제2회 기술 검토 워킹 그룹 자료 1, ③ 기술 검토 워킹 그룹 보고서: 제5회 자료 2-1을 참고해 보면, 워킹 그룹은 ① 용어"익명화"를 개인"식별"과 "특정화"로 구분하고 있다. ② 어떤 기술적 수법을 취하고도 일반적 의미에서의 "완전한 익명화"는 있을 수 없다는 점을 인정한다. ③ 익명화 정도에 대해 익명화에는 여러 수준이 있음을 인정한다. 여기에서 익명화수준에 대한 내용은 다음 "1""2""3"과 그림 4.1에서 소개한다.



개인정보로 분류된 범위의 경계는 "2"의 수준 중에 있다고 할 수 있으나 분명하게 규정하기는 어렵다. "3"은 고도의 익명화 처리를 의미하며, 그 대표 예로서 "k-익명화"가 있다. 여기서 k는 특정 개인의 데이터를 k개 미만으로 좁힐 수 있지 않는지를 보이는 익명성 지표이며, k를 크게 하면 개인이 특정될 위험은 줄어들지만 동시에 정보량이 떨어져 실용적으로 사용할 수 없는 데이터가 된다는 문제가 존재한다.

1. 연결 가능 익명 데이터 : 성명 등을 삭제하지만 원래의 정보와 대조함으로써 개인과 연결 가능한 것 (가명화라고 부른다)
2. 이른바 익명 데이터 : 성명 등을 삭제하고 원래의 정보와 대응 하지 못하도록 한 것 (무명화라고 부른다)
3. 고도의 익명 데이터 : 고도의 익명 처리에 의한 특정 개인 식별이 곤란하도록 한 것

인다. 이는 국내에 적용해 보면, 개인에 관련된 정보의 유형을 비식별화 수준에 따라 3가지('개인정보,' '익명정보,' '비식별정보')로 나누고, 사용자별 비식별처리의 요구수준을 달리 설정하자는 논의[189]와도 일맥상통하는 것으로 생각할 수 있다. 이를 도식화해 보면 다음 그림과 같다.



[출처] '통계자료의 비밀보호를 위한 익명화 방법들', 통계연구 p153, 통계청, 2004.10.

이 논의에 따르면 사용자별 비식별처리의 요구수준이 달라지며, 그에 따라 정보의 유형을 세분화하여 달리 규율할 필요성[190]이 제기된다. 이 경우 개인정보 여러 항목 중에서 어떠한 항목을 대상으로 비식별화할 것인가(식별

189) 박원환, 빅데이터 분석을 위한 개인정보의 비식별화 방법, 2015.9.14. 발표자료 ; 박원환 & 황조연, 통계자료의 비밀보호를 위한 익명화 방법들, 통계연구 제9권 제2호, 2004, 146-172p

190)

1. 업무 사용자 ※[그림]에서 A	▷(목적) 정보주체에게 서비스를 하기 위해 개인정보를 수집하여 활용 ▷(근거) 관련 법적 근거(보호법 제15조 등)를 준수하여 수집·이용 ▷(범위) 해당 개인정보처리자 외 사용 제한 ▷(정보) 비식별화 하지 않은 **원래의 정보(raw data)**를 주로 활용 ▷(보호) 높은 보호 수준 요구(관련 법령 등)
2. 심층 분석자 ※[그림]에서 B **(Anonymous Data)**	▷(목적) 높은 수준의 정보분석, 그 결과를 이용하는 **연구/통계** 등 분야 ▷(근거) **보호법 제18조제2항제4호**에 따라 제공받아 분석 이용이 가능 ▷(범위) 개인정보처리자가 특정인에게만 제공하고, 그 제공받은 자는 목적 내에서만 활용하도록 하는 경우 ▷(정보) '**특정 개인을 알아 볼 수 없는 형태로 개인정보를 제공**하는 경우' (사용자가 특정 개인만 알아볼 수 없도록 낮은 수준의 비식별화 처리) ※1.이 정보를 **익명화 정보**라고 일부 학자 주장(이인호 교수, 이창범 박사 등) 2. 미국의 국립표준기술원의 기술적 정의: **비식별화와 익명화를 구분**하여 정의 ▷(보호) 높은 보호수준 요구(개인정보이므로 관련 법 준수)
3. 일반 분석자 ※[그림]의 C/D/E **(De-identification Data)**	▷(목적) 일반적인 빅 데이터 분석을 위해 활용하며, 의사결정 등 목적 ▷(근거) 법적 근거 불요 **(개인정보 → 일반 정보)이므로 규율대상 아님 (쟁점)** ▷(범위) 빅 데이터 분석 등 활용을 희망하는 모든 자는 활용 가능 ▷(정보) **다른 정보와 결합 또는 재식별화가 불가능**하도록 비식별화 한 정보 ※ 정보의 가치 손상을 최소화하는 비식별화 처리 방법이 필요(과제) ▷(보호) 해당 없음

[주] '공공정보 개방·공유에 따른 개인정보 보호지침', 행정자치부 개인정보보호과, 2013.9.

자), 해당 항목에 대해 어떠한 방법으로 비식별화 처리 할 것인가(방법 연구), 각 방법으로 어느 수준까지 할 것인가([그림]의 B, C, D 등 사용자 특성에 따라 등급화), B 수준(회색)일 경우 계약서 등에 포함할 내용은 무엇으로 할 것인가 등에 대한 논의가 필요하며, 각각의 사항은 기존의 개인정보 보호 관련법령 만으로는 대처하기 어렵게 된다. 그러므로 만일 일본의 개정법과 유사한 비식별화 규제체제를 도입하고자 할 경우에도 현행 국내법의 해석으로는 어려움에 봉착하게 될 것이다.

네 번째, '의도된 공격자' (motivated intruder)의 개념에 대해 생각해 보자. 위에서 살펴본 것과 같이, 의도된 공격자의 개념을 도입하여 비식별화 수준에 대해 평가하는 것은 대다수의 잠재적 공격자를 암묵적인 전제로 하는 것이다. 달리 말하면, 재식별을 시도하는 모든 가능한 공격자를 전제로 하여 평가하는 것이 아니고, 특히 전문기술자 등에 의한 공격 가능성은 실질적으로 평가 과정에서 고려하지 않는 것이다. 이러한 평가 기준은 현행의 국내법령상 인정되기 쉽지 않을 것이다.

제2절 국내 상황에 적합한 비식별화 규제 및 법제 개선의 모색

1. 기본적인 방향

비식별화[191] 관련 법제도의 기본적인 방향은 (1) 재식별 리스크를 최소화하는 한편, (2) 정보의 유용한 활용은 적절히 허용되고 장려될 수 있는 환경을 조성하는 것에 있을 것이다. 이 중 재식별 리스크를 최소화하는 방법으로는 넓게는 비식별화의 절차적인 측면과 비식별화의 실체적인 측면을 구분하여 생각할 수 있다. 절차적인 측면에서는, 정보에 대한 적절한 접근통제(access control)의 수준 및 방법을 파악하고 결정하는 것, 정보 유형이나 이용 맥락

191) 비식별화와 관련하여 현행법상 합법적인 처리의 가능성을 열어주기 위하여 비식별화가 필요한지 아니면 익명화가 필요한지에 대한 논란이 있다. 이는 비식별화와 익명화의 개념에 관하여 서로 다른 시각을 가지기 때문이다. 그런데 이러한 비식별화와 익명화의 구분은 실질적인 측면에서는 큰 의미를 가지지 아니한다고 할 수 있다. 중요한 것은 비식별화든 익명화든 현행 법률 하에서 허용되는 것이 어느 정도의 식별성이 제거 또는 감소된 것이어야 하는 것인가 하는 점이다. 이러한 실질적인 관점으로부터 비식별화 논의를 시작하여야 할 것이다. 이에 따라 이하 및 이 보고서에서는 비식별화라는 용어로 통일하여 실질적인 측면에서 합법적인 처리가 가능한 개인정보의 범위 및 방안을 논의하고자 한다.

등을 고려하여 비식별화를 위해서 적용이 필요한 적절한 통계학적 데이터 마이닝 기법을 판단하는 것, 그리고 데이터에 대한 실제 관리에 있어 통제나 모니터링 등을 하는 것 등을 포함하여 비식별화 실행에 관한 관리적·절차적 측면이 중요하다. 그리고 비식별화의 실체적인 측면에서는, 실제로 비식별화 작업을 수행하고 비식별화 수준에 관해 평가하고 검증하는 실무적·실체적 측면을 고려하여야 할 것이다.

이와 같은 두 측면에 대한 고려가 중요한 이유는, 흔히 비식별화에 대한 논의와 그에 대한 고려가 지금까지 비식별화 '기법'에 지나치게 치중하여 이루어져온 면이 있기 때문이다. 또한 우수한 기법을 이용하여 일단 비식별화를 하고 나면, 해당 정보에 대해서는 실질적으로 개인정보 보호법제 상 문제가 될 우려가 없이 자유롭게 이용할 수 있는 것이 아닌가 하는 오해를 어렵지 않게 볼 수 있기 때문이기도 하다. 관리적이고 절차적인 측면을 별도로 고려함으로써, 데이터에 대한 비식별화가 일회성 작업을 통해 실질적으로 '면죄부'를 부여받을 수 있는 것은 아니고, 반복적이고 지속적인 관리 및 절차적 타당성을 확보하는 것이 비식별화 기법의 적용 못지않게 중요할 수 있다는 점을 강조하기 위함이다.

물론 개념적으로는 절차적인 비식별화와 실체적인 비식별화를 구분하여 생각할 수 있지만 현실에서 이 두 측면이 명확히 구분되지는 않을 수도 있다. 절차적인 비식별화 맥락에서는 주로 법이나 제도적인 측면이 강조될 것이다. 다른 한편으로는, 실체적인 비식별화 맥락에서는 방법론적 기법이 강조가 될 것이다. 이러한 비식별화 작업은 정부기관이나 독립적인 제3의 기관 등을 통해 관리 또는 검증, 인증이 될 수 있을 것이다. 데이터의 유형이나 이용의 맥락, 비식별화의 정도 등을 고려하여, 매우 제한적인 접근만 허용되는 경우도 있을 수 있다. 그 반대로 매우 광범위한 접근이 허용되는 경우도 있을 수 있다. 그에 대한 전반적인 판단 및 관리가 바로 절차적인 비식별화의 핵심적인 요소가 될 것이다.

접근통제와 관련하여, 예를 들어, 가장 광범위한 접근이 허용되는 경우로서,

통계청의 일부 인구통계학적 데이터를 생각해볼 수 있다. 이러한 데이터는 'release and forget' 형태로 일반에 공개해도 무난한 데이터이다. 그 반대로, 예를 들어, 일부 민감정보에 관해서는 설령 이것을 비식별화한 이후라고 할지라도, 제한적인 접근만 허용되어야 할 것이다. 가장 극단적인 접근제한은, 클라우드를 통한 저장은 배제한 채 데이터 보관 장소를 특정해 놓고, 특정한 물리적 장소와 시간을 정해 데이터에 대한 접근을 허용하는 것이다. 더 나아가 그 경우에도, 데이터에 대한 복사는 허용되지 않을 수 있고, 데이터를 대상으로 한 쿼리 숫자나 내용에 제한을 둘 수도 있을 것이다.

관리적인 측면의 구체적인 사항을 정함에 있어서는 데이터의 유형이나 이용맥락, 데이터의 규모, 비식별화 방식 등 매우 다양한 요소가 고려되어야 할 것이다. 그렇기 때문에 비식별화 방식 및 데이터에 대한 접근, 관리 등을 몇 가지로 간단하게 유형화할 수는 없을 것이다.

또한, 데이터에 대한 관리의 맥락에서, 계약이나 서약을 통한 통제가 포함될 수 있을 것이다. 계약·서약에는 기본적으로 '적어도 현시점에서는 재식별이 가능하지 않은 것으로 알고 있다' 그리고 '향후 재식별을 위한 시도를 하지 않겠다' 는 취지의 내용 등이 포함되게 될 것이다. 이에 관해서는 가령 HIPAA safe harbor rule의 내용을 참조할 수 있고, (조금 맥락이 다르기는 하지만) EU의 Standard Contractual Clauses 및 Binding Corporate Rules의 내용으로부터도 시사점을 얻을 수 있을 것으로 보인다. 또한 강력한 통제가 필요한 유형의 정보에 대해서는 비식별화한 정보에 대한 제3자 제공과 관련하여 정보주체의 동의 없이는 이를 하지 않겠다는 내용을 포함시킬 수도 있을 것이다.

2. 개인정보의 개념 정의

현행법상 비식별화를 어렵게 만드는 가장 큰 법령상의 원인은 개인정보의 개념 정의가 너무 광범위하여 사실상 거의 모든 '개인에 관한 정보'가 식별가능성을 근거로 개인정보에 포함될 수 있다는 점에 있는 것으로 보인다.

이로 인해서 재식별의 가능성이 완벽하게 제거되지 않으면 비식별화를 통해 정보를 활용하는 것이 매우 어렵다. 물론 위에서 본 것과 같이, 비식별화의 과정을 통해서 재식별의 가능성을 완벽하게 차단하는 것은 매우 어렵고 이에 관해 100% 보장을 할 수도 없다. 또한, 재식별 가능성을 확실히 차단하기 위해서는 재식별을 위해 이용될 수 있는 가능성이 약간이라도 있는 유형의 정보를 대부분 제외하여 비식별 작업을 해야 할 텐데, 이를 통해 결국 남게 되는 정보는 유용성 또한 남아있지 않는 무용지물이 될 가능성이 높다.

물론 그렇다고 하여 현행법의 규정과 취지를 넘어서 식별가능한 개인정보의 범위를 과도하게 제한하여 개인정보보호법의 적용을 배제하는 것도 바람직하지 않다. 그러나 입법적으로 개선하지는 못한다고 하더라도 최소한 현행법상의 개인정보의 개념정의에 대한 합리적인 해석을 제시하지 않는다면 개인정보의 범위를 둘러싸고 계속 논란이 이어질 것이고 이러한 모호함은 법규의 실효성을 떨어뜨리는 악영향을 줄 수도 있다. 법적인 예측가능성이 확보되어야 법규 준수의 가능성이나 실효성도 증가시킬 수 있을 것이다. 이러한 측면에서 1차적으로는 개인정보의 개념에 대하여 해석상 가이드라인을 제시하는 방안을 검토할 수 있다. 2차적으로는 개인정보의 개념 정의에 대한 다양한 의견 수렴과정을 거쳐 입법적 개선을 꾀할 필요가 있다.

해석론적 접근방식의 경우에는 현행 개인정보보호법상 개인정보보호법에 대한 해석권한이 있는 개인정보보호위원회가 개인정보의 개념에 관한 가이드라인을 제시하는 방안을 추진할 필요가 있다.[192] 구체적으로 미국의 HIPAA와 같이 일정한 범위의 식별자 및 준식별자만을 개인정보로 한다는 해석은 과도한 해석권한의 행사로 해석될 수 있겠지만, 개인정보의 개념을 해석하는 데 있어서 일정한 기준을 제시하는 것은 가능할 것이다. 즉, 현행 개인정보의 개념을 보면 ‘살아 있는 개인에 관한 정보’, ‘성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보’를 개인정보의 기본 요건으로 설정하면서, “해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것”을 포함하고 있는데,

192) 개인정보보호법 제8조 제1항 제4호에서는 개인정보보호위원회가 “개인정보 보호에 관한 법령의 해석·운용에 관한 사항”의 심의·의결 권한을 가진다고 규정하고 있다.

‘성명, 주민등록번호 및 영상 등’에 모든 정보가 포함되는 것이 아니라 적어도 성명, 주민등록번호 및 영상과 동등하거나 유사한 정도의 식별가능 속성이 있는 정보를 통하여 식별되는 경우를 말하는 것으로 해석하거나, ‘해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것’을 해석에 있어서도 개인정보처리자가 알았거나 알 수 있었을 범위의 다른 정보와의 결합의 용이성을 판단기준으로 삼도록 하여 합리적 기대가능성 요건을 해석에 의하여 부가함으로써 현재의 기술수준이나 상황에 적합한 수준으로 제한하는 해석의 시도가 필요할 것으로 보인다.[193] 비교법적으로는 EU나 일본 등에서 개인정보를 규정하는 경우에도 합리적 방법 등을 전제로 하여 식별가능성을 파악하고 있음에 주목할 필요가 있다.[194]

이러한 해석론적 접근방식과 함께 보다 적극적으로 입법적 개선을 하는 방안도 고려할 수 있다. 그러나 개인정보보호의 사각지대를 만들 수 있는 방향으로의 개인정보 개념의 축소나 전체적인 제재를 줄이자는 시도는 모두 현실적으로 쉽지 않아 보이기 때문에[195] 개인정보를 이원화하는 방향으로 입법적 개선을 시도하는 것도 검토할 필요가 있어 보인다. 예를 들면, 미국의 HIPAA와 같이 일정한 유형의 개인정보를 특정하여 이에 대하여는 현재와 같은 높은 수준의 규제를 유지하는 것이다. 즉, 고유식별정보나 이름, 휴대폰 번호와 같이 직접 식별 가능성이 높은 정보나 민감정보와 같이 개인의 기본적 권리에 대한 침해 가능성이 높은 정보를 구체적으로 열거하여, 제한 열거된 개인정보에 대하여는 현재와 같은 규제체계를 유지하는 반면, 그 외의 나

193) 이인호,「개인정보 보호법상의 ‘개인정보’ 개념에 대한 해석론-익명화한 처방전 정보를 중심으로」,『정보법학』, 제19권 제1호, 2015, 82면에서는 “「개인정보 보호법」 제2조 제1호 괄호 부분의 “다른 정보와 쉽게 결합하여”란 ‘정보처리자가 통상적인 업무과정에서 쉽게 입수할 수 있는 다른 정보와 결합하여’의 의미로 해석되어야 마땅할 것이다.

194) EU의 경우에는 “합리적으로 사용될 것 같은 방법 (means reasonably likely to be used)”을 전제로 하여 식별가능성을 판단한다. EU Directive 및 GDPR 규정 참조. 일본에서의 해석론은 ‘다른 사업자에게 조회(照會)[물어서 알아봄]해야 하는 경우’나 ‘통상적인 업무과정에서 입수하는 것이 곤란한 경우’는 제외되어야 한다”는 입장일뿐더러, 위에서 본 것과 같이 2015년 개정법에는 아예 별도 유형의 개인정보 카테고리를 설정하였다.

195) 문재완, “개인정보의 개념에 관한 연구”, 「공법연구」, 제42집 제3호(2014.2.), 69-74면에 의하면, “인격주체성을 나타내지 않는 정보도 다른 정보와 결합하여 정보주체의 인격주체성을 침해할 가능성이 있으므로 이를 우려하여 개인정보의 범위를 확대하는 것은 필요한 일”이라고 하면서, “개인정보의 정의가 어느 정도 불완전하고, 포괄적이고, 불명확한 것은 현대 정보사회에서 불가피한 일”이라고 한 것도 비슷한 취지로 이해된다.

머지 개인에 관한 정보는 모두 사후 규제로 변경하는 것이다.

이와 관련하여 이러한 개인정보 개념의 차별화를 위한 입법론에 대하여는 명확한 구별 기준의 설정이 곤란하거나 애매한 영역이 존재할 가능성이 충분히 예상되기 때문에 실현 가능성이 희박하다고 비판하는 견해도 있다.[196] 이러한 우려에도 충분히 공감이 가지만, 현재의 개인정보 개념도 이미 그 외연적 측면에서 개인정보에 해당하는가 그렇지 않은가에 대하여 모호함이 있어서 수범주체에게 혼란을 주고 있고, 개인정보처리자에 대한 범위를 보다 명확화 하더라도 결국 수범주체가 어떠한 경우에 처벌되는지 그렇지 않은지에 대해서는 결국 개인정보의 개념에 따라 정해지기 때문에 개인정보의 개념과 범위에 대한 명확화 작업은 의미가 있을 것으로 생각한다. 현실적으로 차별화 기준의 모호함을 회피하기 위해서 위에서 설명한 것처럼 강력한 규제의 대상은 법정화 하는 것이 바람직할 것으로 보인다. 그리고 그 이외의 개인정보는 원칙적으로 수집·이용할 수 있고, 사전동의도 가능하지만 사후 거부도 가능한 방식으로 규율하면서 행위규제로서 심각한 정보주체에 대한 침해가 있는 경우에 제재를 가하는 이원화된 규율방식을 취하는 것이 가장 실현가능한 대안이 될 수 있다.[197][198] 이러한 이원화된 방식의 또 다른 수정방식으로 현재의 폭넓은 개인정보 개념을 유지하면서도 일본의 개정 개인정보보호법처럼 비식별화를 합법적으로 보장하기 위하여 일본의 '익명가공정보'와 같은 비식별화 처리된 정보의 개념을 신설함으로써 예외적으로 허용되는 비식별화 처리를 법률에서 명확히 정하는 방향으로 입법을 추진하는 것도 가능하다. 이러한 방식은 현재의 입법체계를 유지하면서 예외적 허용을 하기 때문에 법체계를 크게 변화시키지 않는 장점이 있다. 그러나 이러한 예외적 허용 방식의 규정방식은 비식별화처리에 대해서는 예외를 허용할 수 있지만, 개인정보의 광범위한 범위로 인한 문제점을 모두 해결할 수 없는 한계가 있다.

196) 김민호, "개인정보처리자에 대한 연구", 「성균관법학」, 제26권 제4호(2014.12.), 246면. 이 견해(256-262면)는 개인정보의 개념이 아니라 개인정보처리자 개념의 재정립으로부터 문제점을 해결하자고 주장한다.

197) 최경진, "빅데이터·사물인터넷 시대 개인정보보호법제의 발전적 전환을 위한 제언", 개인정보보호위원회가 개최한 개인정보보호법 시행 4주년 기념 세미나(2015.10.8.) 자료집, 18-19면.

198) 최근 개정된 일본의 개인정보보호법 상의 익명가공정보의 개념을 신설한 것도 개인정보 개념 확정의 어려움으로부터 수범주체에게 기대가능성을 부여하기 위한 것으로 보인다.

3. 비식별화를 위한 절차적 해결방안 및 제3의 신뢰기관의 필요성

비식별화를 통한 개인정보의 적법한 처리를 위하여 실체적으로 비식별화의 개념을 정의하고, 합법적 처리가 가능한 비식별화 구간을 설정하더라도 실제에 있어서는 어느 정도가 비식별화된 것인지를 판단하는 구체적 과정을 거칠 수밖에 없다. 매우 다양한 상황 하에서 어떠한 경우에도 불변하는 실체적인 기준을 법령에 규정하거나 설정하는 것은 거의 불가능하고, 다양한 가능한 상황에 따라 비식별화 정도의 판단을 개별적으로 할 수 밖에 없다. 따라서 비식별화된 개인정보의 합법적 처리를 허용하는 방식으로 입법이 이루어지든 그렇지 않든 관계없이 절차적 측면에서 개인정보의 비식별화를 판단하고 그에 합법성을 부여하는 방안을 마련하는 것이 더 현실적이라고 할 수 있다. 따라서 개별 상황 하에서 합리적으로 판단하여 비식별화되었는지를 판단할 중립적 · 객관적 절차를 마련하고 그 절차에 따라 적절하게 판단된 경우에는 합법적 처리를 허용해주는 방안으로의 제도 개선 혹은 운영이 필요하다. 이러한 비식별화 판단 절차의 정당성을 확보하기 위한 보충적인 수단으로서 제3의 신뢰기관을 인정하는 것도 고려해볼 방안이다.

그런데 비식별화의 적정성 혹은 합리성 판단을 절차적으로 보장하는 경우에 실제 비식별화에 대한 관리와 인증을 하는 기관에 관해서는 별도의 상세한 검토가 필요하다. 위에서 본 것과 같이, 비식별화의 방법론에 대해서는 비식별화가 진행되는 맥락과 상황에 따라 각기 다른 방법이 적용될 수 있고, 비식별화가 진행된 후에도 재식별의 리스크에 대한 주기적 재검토 및 데이터 관리상황에 대한 모니터링이 필요할 수 있다. 이처럼 비식별화의 초기 단계에서부터 사후적 관리에 이르기까지 규범적 통제가 필요한 면이 있고, 이에 관해 관리하는 제3의 기관이 필요할 수 있다.

이러한 제3의 기관이 구체적으로 담당하게 되는 역할에 따라서 정부나 공공기관이 그 역할을 맡을 수도 있고, 민간의 독립적인 기관이 그 역할을 맡을 수도 있을 것이다. 민간의 경우에도 영리기관이 그 역할을 맡을 수도 있고,

연구소 등 비영리기관이 그 역할을 맡을 수도 있다. 각각의 경우에 각기 다른 장단점이 있을 것이다.[199]

이에 관한 판단 및 제도설계에 있어 중요한 요소는, 우선 적절한 비식별화 방법론을 사용하고 적절한 수준의 비식별화 조치를 이행하도록 유도하는 것에 있다. 그리고 일단 비식별화 조치를 하고 난 후 사후적 관리를 철저히 하고 이에 관해 모니터링 하는 것도 매우 중요하다. 이는 요컨대 데이터의 '라이프사이클' 을 설정하여, 전체 기간 동안 해당 데이터가 어떤 형태로 존재하게 될지, 누가 보유하게 될지, 어떤 유형의 분석 대상이 될지 등에 대해 단계별로 구분하여 정리하고 관리하는 역할이 될 것이다.

그렇다고 하여, 이러한 검증과 관리 및 인증 등을 수행하는 기관이 비식별화의 적정성에 대한 궁극적인 판단을 해줄 수 있는 것은 아니다. 또한 이러한 검증과 관리 및 인증이 1회성 판단이 아니라, 지속적이고 반복적으로 이루어져야 제도의 유용성이 확보될 수 있다.

4. 비식별화 관련 법령의 개정방향

중장기적으로는 비식별화 관련 법령의 개정이 이루어질 필요가 있다. 비식별화 맥락에서 간과해서는 안 되는 점은, 식별-비식별 여부에 대해 이분법적 또는 흑백논리적 구분을 암묵적으로 전제해서는 논의가 진전되기 어렵다는 것이다. 현행 법령을 보수적으로 해석한다면, 그리고 개인정보의 개념에 대한 몇몇 판례에서 보이고 있는 태도와 더불어 바라본다면, 비식별화된 정보는 대부분의 경우 지속적으로 개인정보라고 판단될 가능성이 있게 된다. 그러므로 개인정보의 개념 자체에 대한 재정립이 필요하다 할 수 있다. 특히 비식별화에 관련된 규정을 좀 더 명확하게 할 필요가 있다. 이는 몇 가지 방법을 통해 달성될 수 있을 것이다. 예를 들면, 개인정보의 정의 자체에 변화를 주는 방식이 될 수도 있고, 개인정보 보호법상 예외규정에 해당하는 내용(가령 제18조 등)을 더욱 명확하게 하는 방식이 될 수도 있다.

199) 이러한 기관의 역할을 구상할 때 APEC CBPR 제도에 이미 도입되어 있는 책임기관(Accountability Agent)의 개념이 일정한 시사점을 제공해 줄 수 있다.

다른 한편, 미국 HIPAA에 도입되어 있는 세이프하버 방식과 유사한 방식을 국내 개인정보 보호법에 도입하는 방안을 고려해 볼 수도 있다. 이는 비식별화된 개인정보의 개념이나 범위에 대해 명확히 할 현실적인 필요가 절박한 상황을 감안하면, 실무적으로 유용하며 간단한 방법이 될 수 있다. 이를 통해 수범자 입장에서는 예측가능성이 높아지게 될 것이고 법집행의 명확성과 실효성도 확보되게 될 것이다. 다만 HIPAA에 나열되어 있는 식별자 및 준식별자들에 대해 ① 국내 데이터 환경을 고려하고 ② 보건의료분야를 벗어나 여러 영역으로 확장 응용하여 국내의 다양한 상황에 적용될 규칙을 정하는 과정에서, 구체적으로 어떤 식별자와 준식별자를 포함할 것인지에 관해서는 별도의 연구가 필요하다. 경우에 따라서는, HIPAA의 경우와 유사하게 특정 영역의 데이터에 대해서만 적용되는 제한적인 방식의 규칙을 도입해 보는 것을 생각해 볼 수도 있다.

일반론적으로, HIPAA가 적용되는 미국의 데이터 환경과 국내 데이터 환경의 주요한 차이점은 다음 3가지로 요약해볼 수 있다. 우선 ① 미국은 시민이 일괄적으로 가입하게 되는 건강보험이 없다. 시민 개개인이 건강보험이나 의료보험에 가입되어 있을 수도 있고, 그렇지 않을 수도 있다. 설령 이들이 보험에 가입되어 있더라도, 보험은 여러 보험회사를 통해 제공될 수 있으므로 보험가입정보 등이 국가적인 차원으로 집중되지는 않는다. 다음으로 ② 미국에서는 일반인이 투표자등록부 기록을 무료 또는 저렴한 가격으로 구입할 수 있다. 그 결과 여기에 담겨 있는 정보가 준식별자로 사용될 수 있다. 이처럼 데이터 링크(link)를 가능하게 해주는 잠재적 데이터세트가 국내에 존재할 것인지에 대해서는 별도로 살펴보아야 할 것이다. 마지막으로 ③ 국내에서는 주민등록번호가 여러 데이터세트에 담긴 개인정보의 링크를 가능하게 해주는 식별자로 활용되기 쉬운 환경이 조성되어 있다. 주민등록번호가 최근까지 매우 광범위하게 수집 및 이용되어 왔기 때문이다. 이를 포함하여 데이터 환경의 차이에 대하여 상세한 검토와 고려를 할 필요가 있다.

법령 개정의 또 다른 방향은 위에서 언급한 제3의 인증 및 관리기관의 역할

을 조금 더 적극적으로 고려하여 이를 법령에 명시화하는 것이 될 것이다. 인증 및 관리 기관에 요구되는 노하우나 경험 등을 직접 규정할 수도 있고, HIPAA expert determination rule의 경우와 유사하게 자격요건에 관한 개략적인 기준을 제시하는 방법이 될 수도 있을 것이다. 또한 인증 및 관리기관이 부담하게 될 수 있는 재식별 관련 법적 책임의 한계에 대해 규정함으로써, 법적 책임의 소재를 명확히 하는 것이 필요할 수도 있다. 법적 책임의 소재를 합리적으로 규정하고 명확하게 해주는 것은 제3의 인증 및 관리기관이 수립되고 운영되는 데에 있어 큰 도움이 될 것이다.

5. 합법적 개인정보처리로서의 비식별화의 판단기준

현행 법제를 유지하는 경우이든 위에서 제시한 입법적 개선을 추진하는 경우이든 일정한 범위의 비식별화를 합법적인 개인정보처리로서 인정하기 위한 법제도적 보완이 필요하다. 비식별화된 개인정보를 별도의 개념으로 입법화하여 법적으로 허용하면 일부 문제가 해결되지만 그러한 경우에도 비식별화의 정도 또는 재식별 가능성의 정도에 따라서 다시 그러한 비식별화된 개인정보 포함될 수 있을 것인지는 판단이 필요하다. 이 경우에 기술적으로는 재식별화 0%에서 100%까지, 반대로 비식별화도 0%에서 100%까지 가능하며, 기술적인 측면에서는 완전한 비식별화를 통한 재식별화 가능성 0%의 상태를 만드는 것이 가능하도고 할 수도 있고 이러한 경우에야 비로소 개인정보성을 제거할 수 있다고 판단할 수도 있다.[200] 그러나 개인정보보호는 규범적 측면에서 개인을 어느 정도로 보호할 것인가의 문제이다. 규범적 판단에 기술적인 방법이 활용될 수는 있지만 기술적 판단이 법규범적 판단의 전부일 수는 없다. 개인정보를 보호하는 목적은 식별가능성이 0%를 넘어 약간이라도 있다면 무조건 보호하자는 것이 아니라 사회적·경제적·규범적 보호의 필요성이 있는 범위에서의 법에 의한 강제를 동반하여 보호를 꾀하자는 것이기 때문이다. 합법과 불법의 경계는 기술적으로만 정해질 수 있는

200) 재식별 확률이 0%라 표현한 것은 재식별이 전혀 불가능한 상황을 가리키는 것이고, 재식별 확률이 100%라 표현한 것은 재식별이 반드시 발생할 경우를 가리키는 것이다. 재식별 가능성에 관해 별도의 수치로 측정 가능한 개념설정을 하지 않고는 확률의 형태로 표현하기는 어렵다. 다만 미국 HIPAA에서 볼 수 있는 것과 같이, 그 가능성이 매우 낮다(very small)거나 그 반대로 매우 높다는 방식으로 표현하는 것은 가능할 것이다.

것이 아니고 수범주체의 기대가능성이나 예측가능성, 현실적 실현 가능성, 사회의 보편적 기대수준 등을 종합적으로 고려하여 판단할 수 있는 것이다.

비식별화 판단시에 어느 정도의 비식별화를 규범적으로 허용되는 비식별화로 볼 것인지의 판단은 하나의 기준 혹은 몇 가지 획일적이고 기계적인 기준만으로 정하기 어렵다. 개인정보 혹은 부가정보가 결합되어 있는 모습이나 데이터세트에 포함된 개인정보의 양이나 배열순서, 데이터세트의 구성이나 활용 행태, 개인정보 활용 목적이나 사용 범위 등 개인정보가 수집·이용되는 환경과 비식별화에 사용되는 기술 등에 따라 비식별화 판단결과가 달라질 수 있기 때문에 비식별화의 실무적 판단기준에 대하여는 유연한 접근이 필요하다. 다만, 최소한 고려하여야 할 판단요소나 판단기준을 추상적으로 설정하고, 비식별화의 판단에 있어서도 합리적 판단 요소를 도입할 필요가 있다.

또한 일단 비식별화된 데이터의 사후적 관리가 어떻게 이루질 것인지에 관해서도 규범적 통제가 필요하다. 비식별화된 데이터에 누가 접근(access) 가능하게 될지, 접근에 시간이나 공간적 제한이 있는지, 쿼리나 연산에 제한이 있는지 여부, 링크(link) 가능한 데이터의 추가적 확보 가능성 등 여러 가지의 요소에 따라서 사후적으로 재식별 가능성이 크게 달라질 수 있기 때문이다. 예컨대, 별도의 통제 없이 일반 공개(소위 ‘release and forget’)가 되는 유형의 데이터에 대해서는 애초에 비식별화를 하는 단계에서 매우 높은 수준의 비식별화를 할 필요가 있고, 일반 공개에 앞서 비식별화 수준에 관하여 매우 면밀한 검토를 할 필요가 있을 것이다.

이와는 달리, 예를 들어, 특정한 연구실 공간에 인터넷 연결이 되지 않는 상태의 독립적 컴퓨터를 통해 데이터를 공개하고, 이 때 컴퓨터를 이용할 수 있는 시간에도 많은 제약을 두고, 원데이터(raw data) 자체는 볼 수 없고 연산 결과만 볼 수 있게 하는 등의 방식으로 매우 제한적인 접근만 허용하는 정보 공개의 경우라면, 비식별화의 강도를 매우 높게 잡을 필요는 없을 수도 있다. 다만, 이처럼 제한적인 접근만을 허용하는 경우라면, 접근통제

(access control)의 관점에서 통제의 내용과 절차에 대해 명확하게 규정하고 집행할 필요가 있을 것이다.

제6장 결론 및 정책적 제언

빅데이터 산업으로 대변되는 데이터 분석 시장이 확대되고 데이터 분석의 경제적 가치가 지속적으로 높아지고 있다. 이와 같이 데이터의 활용성에 대한 사회적 수요가 높아지는 상황에서 데이터에 포함되어 있는 개인정보를 어떻게 효과적으로 보호할 수 있는지에 대한 사회적 관심도도 높아지고 있다. 데이터의 활용성과 개인정보의 보호는 일종의 상충관계(trade-off)에 있는 가치이기 때문에 이 두 가치들 사이의 균형점을 찾는 것이 개인정보의 규제 체제가 고려해야 할 중요한 정책적 목표가 된다. 특히, 데이터의 활용 가능성과 개인정보 보호 사이에 상충관계가 있다고 해서, 서로 '제로섬' 게임과 같은 관계인 것은 아닌 것임에 유의할 필요가 있다. 즉 데이터 활용 가능성이 높아짐에 따라 그에 정확히 비례하여 개인정보 보호의 수준이 낮아지는 것은 아니고 또 그래야 할 이유도 없다는 것이다.

정책적으로 중요한 목표는, 사회적으로 최적 수준의 개인정보 보호를 달성하면서 그와 동시에 정보의 안전하고 유용한 활용가능성을 확보할 수 있는 방법을 모색하는 것이 된다. 이런 목표를 달성할 수 있는 접근법으로 데이터에서 특정한 개인을 식별할 수 있는 정보의 식별성을 제거하는 방법인 비식별화의 유용성이 강조되고 있다. 이 비식별화 개념이나 기법은 기술적인(technical) 개념이고 당연히 기술적인 측면이 중요하지만, 그와 함께 기술적인 방법론이 어떤 환경에서 어떤 절차에 따라 어떤 기준으로 활용될 수 있는지에 관한 규범적 영역에서의 규제에 대한 고민과 논의가 동시에 진행되어야 한다.

비식별화와 관련된 국내에서의 논의는 개인정보 보호법 제정 이전에는 정보통신망법을 중심으로 논의가 이루어졌다. 2011년 개인정보 보호법이 제정된 이후 논의는 개인정보 보호법이 비식별화를 어떻게 바라보고 있는지를 위주로 전개되었다. 개인정보 보호법이 지나치게 엄격하다는 비판을 의식하여 논의 초기에는 주로 규제완화 담론이 주류적이었다. 이후 정부3.0의 기조와 발

맞추어 공공데이터의 제공 및 이용활성화에 관한 법률이 제정되었다. 이를 통해 최초로 공식적인 법령을 통하여 비식별화에 대한 구체적인 방법론이 제시되었다. 이후 방송통신위원회는 빅데이터 개인정보보호 가이드라인(안)을 발표하여 개인정보 보호와 빅데이터 이용 간 균형을 모색하였으나 시민단체의 반발에 부딪혔고 개인정보보호위원회의 권고를 받아 수정된 행정규칙으로 빅데이터 개인정보보호 가이드라인을 2014년에 발표했다. 이후 2015년 이래로 변화하는 ICT 현실을 감안하여 개인정보 보호와 이용의 균형을 모색하고자 개인정보 보호법, 정보통신망 이용촉진 및 정보보호 등에 관한 법률 등 여러 개의 개정안이 국회에 발의되어 있는 상태이다.

비식별화에 대해 해외에서는 미국, 유럽, 일본을 중심으로 활발한 논의가 이루어지고 실무경험이 축적되고 있다. 미국에서는 주로 HIPAA 프라이버시 규칙의 제정과 이 규칙의 실무적인 적용을 두고 논의가 벌어져 왔다. 미국에는 개인정보의 보호를 일반적으로 규율하는 연방법이 없다. 개인정보 보호에 대한 규제는 사안별 또는 분야별로 이루어지는 형태를 가지고 있다. 이런 개별적 규제 체제 하에서 개인정보의 비식별화에 대한 규제 또한 분야에 특정되는 개별적인 규제의 형태를 가지고 있다. 이에 따라, 교육부가 비식별화된 학생기록에 대한 관계 법률의 적용 가능성을 판단하기 위해 시도했던 경험도 있고, HIPAA 프라이버시 규칙을 통하여 개인정보의 비식별화에 대해서 규범화를 한 바도 있다. 미국에서의 경험에서 출발하여, 개별적인 분야를 규제 대상으로 하는 법령들에서 각각 채택된 비식별화에 대한 접근 방식이 다른 분야들에도 일반적으로 통용될 것이라고 결론짓는 것은 성급한 판단이다. 개별 영역의 개인정보가 지닌 특수한 맥락 등을 고려하여 신중하게 접근할 필요가 있는 것이다. 그럼에도 불구하고 미국 의료개인정보에 대해서는 비식별화에 대한 규칙이 상세하게 규정되어 있는바, 이에 대해 집중적으로 논의가 전개되어 왔다.

미국에서의 논의 특히 HIPAA 프라이버시 규칙 및 그 적용은 미국의 보건의료 뿐 아니라 해외 여러 나라에서도 주목을 하는 데, 그 중요한 이유는, 데이터의 비식별화를 통한 활용가능성을 명시적으로 열어두고 그 기준이나 방

법에 대해 구체적으로 제시하고 있기 때문이다. 이에 따라 이 규칙에 기초하여 실제로 비식별화를 하게 되는 사례가 지속적으로 나타나고 있고, 관련된 연구역량과 방법론 또한 계속해서 개발되고 있다. HIPAA 프라이버시 규칙이 명시적으로 제시하고 있는 18개의 식별자 및 준식별자에 대해서는 그 타당성 등에 관하여 이론적인 논박은 계속되고 있지만, 적어도 아직까지는 이러한 방식으로 식별자와 준식별자를 특정하여 비식별화를 허용한 것으로 인해 사후적 재식별로 인한 심각한 문제가 발생되었다는 주장이나 보고는 나타나고 있지 않다. 그런 점에서 HIPAA 프라이버시 규칙은 정책적 목표를 일정 수준 이상 달성하고 있는 것으로 평가될 수 있을 것이다.

HIPAA를 포함하여 더 넓게 볼 때 미국에서 보호대상이 되는 개인정보는 일반적으로 '개인적으로 식별가능한 정보'(personally identifiable information, PII)로 규정되어 있다. 개별 법령에서 PII는 예시를 나열하는 형태로 규정되는 경우도 있고 개념 정의를 통해 규정하는 경우도 있다. 종합적으로는 PII는 '어떤 기관이 보유하고 있는 개인에 대한 정보'로서 크게 다음과 같은 유형의 정보를 포함하는 것으로 볼 수 있다. (1) 개인의 신분을 구별하거나 추적할 수 있는 용도로 이용될 수 있는 정보, 그리고 (2) 개인과 연결되거나 연결될 수 있는 기타 정보가 그것이다. 이와 같은 개념규정을 구체적으로 생각해 보면 보호되는 개인정보가 상당히 광범위하게 정의될 수 있음을 알 수 있다. 특히 개인정보 개념의 확장성은 (2) 부분에 해당하는 개인과의 연결가능성에서 두드러지게 나타난다. 비식별화 처리는 이러한 연결가능성을 사전적으로 막는 방식으로서 처음부터 처리하려는 데이터에 개인적으로 식별가능한 정보를 포함되지 않게 하는 방법이라는 점에서 그 유용성을 지니게 된다. 이런 배경에서 비식별화 방식들에 대한 다양한 연구들이 수행되었고 수학적 모형을 적용해서 통계학적으로 비식별화 문제에 접근하려는 시도들이 나타나게 되었다.

하지만 이런 다양한 통계학적 접근법들의 어느 방식도 비식별화의 가장 큰 쟁점인 재식별의 위험성을 완벽하게 해거해 주지 못하는(또는 제거되었다고 확실하게 보장할 수는 없다는) 한계를 지니고 있다. 그럼에도 불구하고 빅데

이터 산업의 활성화 정책 및 기술동향 등과 맞물려 정부기관들은 데이터를 대중에 공개해야 할 필요가 생겼고 그로인해 개인을 식별화할 수 없게 하는 비식별화 조치의 필요성은 더욱 커지고 있다. 이를 배경으로 미국의 표준화 기관인 NIST는 비식별화에 대한 최근 20년간의 논의들을 보고서의 형태로 정리하여 발간하였다. 동 보고서는 기본적으로 모든 데이터 비식별화 기술이 재식별의 위험성을 약간이라도 가지고 있다는 것을 기본 전제로 하고 있다. 다만 주어진 맥락에서 어떤 비식별처리 방식이 적용되었는지에 따라 그 위험성의 정도에 차이가 나타날 수 있다고 설명하고 있다. 그리고 이러한 비식별처리 방식 중 절대적으로 우위에 있다고 할 수 있는 방식은 없고, 맥락에 따라 비식별후 재식별의 위험성이 줄어드는 정도가 달라질 것으로 보고 있다. 이러한 점을 감안하면 비식별화를 위한 기술적 방식뿐만 아니라 여러 가지의 사후적 보안 통제책을 적절하게 적용하는 관리적 방식도 함께 고민하여 집행할 필요가 있다.

이러한 미국의 비식별 논의는 HIPAA 프라이버시 규칙과 이를 둘러싼 논의에 주로 바탕을 두고 있다. HIPAA는 의료 분야의 개인정보 보호에 대해 규정하는 연방법으로서 HIPAA 프라이버시 규칙은 HIPAA의 내용을 보충하는 행정규칙의 성격을 지닌다. HIPAA 프라이버시 규칙과 이 규칙의 내용을 보충적으로 설명해주는 안내서들에는 비식별화 기준들이 소개되어 있다. 이는 비식별화된 의료개인정보도 데이터로서의 효용성을 여전히 가지고 있다는 인식하에 비식별화 방식으로 두 가지를 제시한다. 하나는 절차적인 방식에 의한 접근법으로서 관련전문가의 개별적인 판단에 의하는 방식이다. 다른 하나는 내용적 방식에 의한 접근법으로서 특정한 개인 식별자들 또는 준식별자들이 데이터에서 제거되면 비식별화가 된 것으로 간주하는 방식이다. 앞의 방식은 전문가활용법(Expert Determination Rule)이라고 불리고, 후자의 방식은 세이프하버 방식(Safe Harbor Rule)이라고 불린다. HIPAA 프라이버시 규칙의 수범자인 의료정보관리기관은 이 두 가지 방식들 중에 원하는 방식을 선택해서 의료개인정보를 비식별화할 수 있다.

유럽에서는 영국에서 발표된 행동강령(Code of Practice)과 이후 발표된

Article 29 Data Protection Working Party의 비식별화 의견서(Opinion)이 대표적이다. 규정상, 유럽 역내 개인정보를 보호하는 일반규정에 해당하는 EU Directive에서 익명화된 데이터는 규제대상이 되는 개인정보가 아니다. 하지만 이러한 익명화에 대한 언급은 EU Directive의 서문에만 명시되어 있을 뿐, EU Directive의 본문 조항들에는 여기에 대한 구체적인 언급이 없다. 다만 EU Directive는 익명화 방식에 대한 구체적인 기준을 규정하지 않고 이를 EU 회원국들에게 행동강령 형식으로 위임하고 있다. 영국을 필두고 익명화에 대한 행동강령이 등장하고 있다. 이후 EU 정보보호 워킹그룹 29는 오피니언을 발표하였다. 현재는 EU 개인정보 보호법제에 가장 중요한 변화가 될 수 있는 일반데이터보호규칙(GDPR) 도입 논의가 한창이다. 비식별화 관련 쟁점은 주로 개인정보의 정의와 연관성이 높은 부분이며, 논의중인 GDPR 내용에는 기존의 EU Directive 내용과 다르게 익명화에 대한 자세한 사항을 행동강령에 위임할 수 있다고 명시되어 있던 부분이 삭제되었다는 점이 가장 두드러진 차이점이다. 현재는 GDPR 안에 개연성 내지 가능성(reasonably likely to be used to identify the individual)이라는 개념을 도입할 것인지를 둘러싸고 첨예하게 논의 중이다.

영국의 정보보호기관(ICO)이 2012년에 발간한 행동강령에는 EU Directive 체제 하의 익명화에 대한 전반적인 내용들이 담겨 있다. 동 강령은 익명화에 대한 구체적인 사항들은 행동강령의 형태로 개별 국가들이 마련할 수 있다는 규정(EU Directive 서문 26조)에 근거하여 마련된 최초의 행동강령이라는 점에서 그 의의가 있다. 기본적으로 이 행동강령에서 허용되는 익명화의 정도는 식별위험이 없는 수준이 아니라는 것이 중요하다. 식별의 위험성을 판단하는 기준으로 합리적 가능성(reasonably likely test) 기준을 채택함으로써 식별의 위험성이 합리적으로 존재할 경우 규제의 대상이 되는 개인정보에 해당하게 되고, 그 반대로 합리적인 기준상 식별의 위험성이 없다면 개인정보 규제틀을 벗어날 수 있게 된다. 동 행동강령의 가장 두드러진 특징은 재식별 위험성과 관련하여 현실적인 메커니즘을 고려하여, “정보유형에 따른 정보공개의 결과물”과 연결된 “잠재적인 공격자의 의도”를 고려한 “관리적인 측면”에서 재식별 위험성 판단기준을 설정했다는 점이다. 이를

통해 어떤 의미에서는 기술적인 접근법보다 오히려 관리적인 접근방법을 조금 더 중요한 것으로 강조하고 있다. 이러한 잠재적인 공격자의 측면을 고려한 판단기준은 “의도적인 공격자”(motivated intruder) 기준이라고 불린다. 이 기준은 법률에는 규정되어 있지 않았기 때문에 행동강령을 통해 새롭게 규정된 기준에 해당한다. 이 개념은 재식별의 가능성을 평가함에 있어 기술전문가 등에 의한 공격가능성은 고려하지 않게 함으로써 실질적으로는 재식별의 가능성이 무한대로 확장되는 것을 방지해 주는 역할을 한다.

다른 한편, 2014년에 발간된 익명화 기술에 대한 EU 제29조 작업반의 의견서(Opinion)는 익명화의 범위를 개인정보의 개념 정의에 비추어 명확히 인식할 필요가 있다고 한다. 이 의견서는, 개인정보 중에서 개인식별정보(PII)만을 제거한 정보를 익명화된 정보에 해당한다고 오해하는 경우도 있다고 지적하고, 또한 가명화와 익명화는 구분되는 개념임을 명확히 한다. 이에서 출발하여 이 의견서는 데이터가 익명화되었는지를 판단할 수 있는 방법으로 두 가지 옵션을 제시한다. 하나는 데이터세트에서 일정한 속성, 즉 식별성, 연결가능성, 추론가능성이 제거되었는지 확인하는 것이다. 또 다른 하나는 재식별 리스크에 대한 분석을 받는 것이다. 첫 째 옵션에서 유의할 점은 하나의 익명화 기법만으로는 식별성, 연결가능성, 추론가능성 등 모든 속성들을 제거하기 어렵다는 점이다. 이와 같은 평가 등을 전제로 하여 빅데이터 분석의 가능성을 살린 채 익명화를 수행할 수 있다고 한다. 이 의견서는 EU의 공식 기구를 통해 익명화에 대해 직접적인 분석이 이루어지고 의견이 제시되었다는 데에 커다란 의의가 있다. 하지만 다른 한편으로는 실질적으로 유용한 지침은 제시되지 못하고 원론적인 내용의 나열에 그칠 뿐이어서 실무적으로는 불확실성이 계속 남아있을 뿐이라는 비판에 직면하기도 하였다.

일본에서는 2015년 9월 3일 개인정보 보호법 개정안이 통과되었다. 그리고 이 개정안을 준비하는 과정에서 2년 동안 개인 데이터에 관한 검토회를 통해 논의를 하였고, 그 기간 중 6개월 동안 기술 검토 워킹 그룹을 운영하며 내부 보고서를 발간하였다. 그 결과 2003년 제정된 이래로 12년 만에 개정된 개인정보 보호법에는 익명가공정보라는 개념이 새롭게 만들어져 반영되

게 되었다. 기술 검토 워킹 그룹은 기술의 발전 및 외부 정보의 증가에 따라 다른 정보와 비교하는 과정에서 프라이버시 침해가 초래될 수 있다는 관점에서 개인데이터에 대한 개인식별 문제에 대처하기 위해 “특정”과 “식별”이라는 기준을 도입하였다. 이를 바탕으로 “준개인정보”와 “개인 특정성 감소 데이터” 개념을 전제하며 개인의 권익침해 가능성을 고민했다. 이를 바탕으로 현행법상 개인정보의 해석론과 비교하여 연구하였고 그 결과 개인정보와 익명정보 사이에 새로운 정보유형을 추가해야 할 필요성을 제기하였다. 실제로 의회를 통과한 개정법에는 “익명가공정보”의 개념이 새로운 유형의 정보로 포함되어, 향후 빅데이터 분석 등이 진행될 수 있는 가능성을 명시적으로 열어주게 되었다. 이러한 개정안을 두고 논의를 하는 과정에서 보건의료 등의 민간 영역에서는 자체적으로 비식별 가이드라인을 발표하기도 하였다.

개인정보 보호의 규제 대상이 되는 개인정보의 범위가 매우 광범위한 국내의 현행 법령과 규제 체제에 비추어 보면, 위와 같은 외국의 비식별화 접근법들을 국내 상황에 적용하는 것에는 커다란 한계가 있을 수밖에 없다. 특히, 이와 같이 광범위하게 규정된 규정들로 인해 국내기업들이 개인정보의 비식별화를 시도하는 것 자체가 어렵다. 외국의 규제 체제들의 국내 상황에 대한 적합한 적용을 통해 현재의 상황을 타개하기 위해서는 법령의 해석에 대한 불확실성을 제거하고 향후 법개정을 도모하는 것을 포함해서 현재의 규제 체제들을 합리화하려는 노력이 필요하다.

그러한 맥락에서, 비식별화 관련한 향후 국내의 규제체계에 대한 방향 제시를 요약하면 다음과 같다. 첫째는, 법령상 개인정보 정의의 해석에 대한 지침의 제시이다. 현행의 법령상 개인정보의 개념은 실무적인 관점에서 볼 때 명확성이 높지 않은 편이다. 그로 인해 불확실성이 야기되고 있으며 또한 개인정보의 개념을 둘러싼 기존의 법원의 1심 판례들을 감안하면, 살아있는 개인과 관련된 거의 모든 정보를 개인정보로 파악하여 취급해야 하는 것으로 인식될 수 있다. 실무계에서의 불필요한 오해와 불확실성을 줄이기 위해서는, 법령에 제시된 개인정보의 개념을 좀 더 명확히 해석하는 지침을 마련하

는 것이 도움이 될 수 있을 것이다.

두 번째로는 법령의 개정을 추구하는 것이다. 현행 법령에 대한 보수적 해석과 개념적 불확실성을 결합하면, 아무리 높은 수준의 비식별화 조치를 하더라도 향후 재식별의 가능성이 조금이라도 남아있으면 곤란한 것으로 해석될 수 있다. 비식별화에 대한 이론적 논의와 외국의 법령 사례를 살펴보면, 비식별화 조치를 한 후 정보의 이용을 허용하는 것은 재식별의 가능성을 매우 낮은 수준으로 할 것을 전제로 하는 것이지 재식별이 전혀 불가능할 것을 요구하는 것은 아님을 알 수 있다. 실제로 비식별화 작업을 하는 단계에서는 향후에 재식별이 발생하지 않도록 그 가능성을 완벽하게 차단하고 이에 관해 보장을 하는 것은 실질적으로 불가능하다. 이와 같은 이론적, 현실적인 측면을 감안하여 법령 개정을 생각할 필요가 있다.

세 번째, 비식별화 조치에 관한 실체적이고 기술적(technical) 방법론의 개발과 더불어 절차적이고 관리적인 측면에 대해 고민하고 적절한 장치 또는 기구를 마련할 필요가 있다. 비식별화의 방법이나 수준 등은 매우 다양하고, 이를 일의적으로 규정할 수도 없다. 개별 상황 및 맥락에 따라 다른 방법론이 이용될 수밖에 없다. 이는 기술적 방법론에 관한 문제이기도 하지만 동시에 규범적 통제가 필요한 영역이기도 하다. 일단 비식별화가 진행된 후에도, 해당 데이터의 내용은 어떠한지, 데이터가 어떤 형태로 누구에게 공개되는지, 기술적 방법론이 어떻게 변화하는지 등 여러 요소에 따라 재식별 가능성과 관련된 리스크는 달리 평가될 수밖에 없다. 따라서 비식별화 후에도 절차적 관리는 매우 중요한 요소가 되고, 경우에 따라서는 비식별화 수준에 대한 주기적 재평가가 필요할 수도 있다. 이러한 측면을 관리하는 신뢰할만한 장치나 기구가 필요할 수 있다.

개인정보를 토대로 한 빅데이터 분석 등 여러 가지 형태의 정보 수집 및 활용은 중요하고 불가피한 시대적 흐름이다. 무분별한 개인정보의 수집과 분석은 당연히 제어되어야 한다. 다른 한편, 비식별화의 기준을 비현실적으로 까다롭게 하여 정보의 유용한 활용 가능성을 차단하는 것은 시대의 흐름에 역

행하는 결과를 가져올 것이다. 비식별화의 기준을 합리적으로 설정하고 적절한 규범적 통제를 하는 것은, 개인정보를 보호하는 동시에 유용한 정보의 활용가능성을 열어주는 유익한 기능을 할 것이다.

<Appendix>

[별첨 1] 프라이버시 보호 모델 비교표

	k-익명성	l-다양성	t-근접성
제안자	L. Sweeney	A. Machanavajjhala D. Kifer J. Gehrke M. Venkitasubramaniam	N. Li T. Li S. Venkatasubramanian
제안시점	2002년 경	2006년 경	2007년 경
목적	공개데이터에 대한 연결공격의 방어	k-익명성의 문제점 보완	k-익명성과 l-다양성의 문제점 보완
정의	주어진 데이터 집합에서 준식별자 속성값들이 동일한 레코드가 최소 k개가 존재해야 함	주어진 데이터 집합에서 레코드들은 동질 집합* 내에서는 최소 l개의 서로 다른 민감한 정보를 가져야 함	동질 집합 내에서의 민감한 정보의 분포와 전체 데이터 집합에서 민감한 정보의 분포가 t 이하여야 함
의미	• 데이터 집합의 일부를 수정해서 각각의 레코드가 서로 구별되지 않는 k-1개의 레코드를 가지도록 함 • 공격자 입장에서 정확히 어떤 레코드가 공격 대상인지 모르게 하는 접근법	• 익명화하는 과정에서 l개 이상의 서로 다른 민감한 정보를 가질 수 있도록 동질 집합을 구성함 • 민감한 정보가 충분한 다양성을 가지므로 다양성의 부족으로 인한 공격과 배경지식에 의한 연결성 공격에 어느 정도의 방어가 가능	• 동질 집합 내에서 민감한 정보의 분포가 전체 데이터 집합의 분포와 비교해서 지나친 특이성을 제거하려는 접근법 • 이 민감한 정보의 분포를 조종해서 민감한 정보가 특정한 수치에 집중되거나, 유사한 수치 주위에 편중되는 경우를 방지함
평가	'동질성 공격'**과 '배경지식에 의한 공격'***에 취약하다는 비판을 받음	'쏠림 공격' ****과 '유사성 공격' *****에 취약하다는 비판을 받음	민감한 정보의 의미까지 고려를 해서 k-익명성과 l-다양성 모델에 대한 4가지 비판을 해결하려는 시도

* 동질집합(Equivalence class): 동일한 준식별자의 속성값들로 익명화된 레코드들(records)의 모임

** 동질성 공격(Homogeneity attack): 데이터 집합에서 동일한 민감한 정보를 이

용해서 공격 대상의 민감한 정보를 파악하는 공격

*** 배경지식에 의한 공격(Background knowledge attack): 주어진 데이터 이외의 공격자의 배경 지식을 통해서 공격 대상의 민감한 정보를 알아내는 공격

**** 쏠림 공격(Skewness attack): 민감한 정보가 특정한 값에 집중되는 경우에의 공격

*****유사성 공격(Similarity attack): 민감한 정보가 서로 유사할 경우의 공격

[별첨 2] GDPR 논의 중 비식별화 관련 쟁점들[201][202]

Whereas (23)		
European Commission (Original version)	European Parliament (Text adopted by Parliament)	Council of European Union (Consolidated text of the Commission and Council)
The principles of protection should apply to any information concerning an identified or identifiable person. To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the individual. The principles of data protection should not apply to data rendered anonymous in such a way that the data subject is no longer identifiable.	The principles of ***data*** protection should apply to any information concerning an identified or identifiable ***natural*** person. To determine whether a person is identifiable, account should be taken of all the means ***reasonably likely*** to be used either by the controller or by any other person to identify ***or single out*** the individual ***directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development.*** The principles of data protection should ***therefore*** not apply to ***anonymous*** data, ***which is information that does not relate to an identified or identifiable natural person. This Regulation does therefore not concern the processing of such anonymous data, including for statistical and research purposes.***	The principles of data protection should apply to any information concerning an identified or identifiable natural person. Data including pseudonymised data, which could be attributed to a natural person by the use of additional information, should be considered as information on an identifiable natural person. To determine whether a person is identifiable, account should be taken of all the means reasonably likely to be used either by the controller or by any other person to identify the individual directly or indirectly. To ascertain whether means are reasonable likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development. The principles of data protection should therefore not apply to anonymous information, that is information which does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data

201) 표 보는 방법

Comments on the Parliament text
• ***Bold cursive*** : Changes from Commission proposal
• [not amended] : This paragraph or title was not amended
Comments on the Council text:
• Underline : Text included by the COUNCIL compared with Draft COM
• ~~Crossed~~ : Text deleted by the COUNCIL compared with Draft COM
• 궁서체 : Text COUNCIL agreed
• **Bold** : Changes after Draft COUNCIL (No. 15395/14, 31.12.2014 (Italian))
• not agreed

202) https://edri.org/files/EP_Council_Comparison.pdf

		subject is not or no longer identifiable. This Regulation does therefore not concern the processing of such anonymous information, including for statistical and research purposes. The principles of data protection should not apply to deceased persons, unless information on deceased persons is related to an identified or identifiable natural person.

Whereas (122a)		
European Commission (Original version)	European Parliament (Text adopted by Parliament)	Council of European Union (Consolidated text of the Commission and Council)
	A professional who processes personal data concerning health should receive, if possible, anonymised or pseudonymised data, leaving the knowledge of the identity only to the General Practitioner or to the Specialist who has requested such data processing.	

Article 5 Principles relating to personal data processing		
European Commission (Original version)	European Parliament (Text adopted by Parliament)	Council of European Union (Consolidated text of the Commission and Council)
Personal data must be: (a) processed lawfully, fairly and in a transparent manner in relation to the data subject; (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes; (c) adequate, relevant, and limited to the minimum necessary in relation to the purposes for which they are processed; they shall only be processed if, and as long as, the purposes could not be fulfilled by processing information that does not involve personal data; (d) accurate and kept up to date; every reasonable step must be taken to ensure that personal data that are in accurate, having regard to the purposes for which the y are processed, are erased or rectified without delay; (e) kept in a form which permits identification of data subjects for	Principles relating to personal data processing 1. Personal data ***shall*** be: (a) processed lawfully, fairly and in a transparent manner in relation to the data subject ***(lawfulness, fairness and transparency)***; (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes ***(purpose limitation)***; (c) adequate, relevant, and limited to the minimum necessary in relation to the purposes for which they are processed; they shall only be processed if, and as long as, the purposes could not be fulfilled by processing information that does not involve personal data ***(data minimisation)***; (d) accurate and, ***where necessary,*** kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for	Principles relating to personal data processing 1. Personal data must be: (a) processed lawfully, fairly and in a transparent manner in relation to the data subject; (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes; [further processing of personal data for archiving purposes in the public interest, statistical, scientific or historical purposes shall not be considered incompatible with the initial purposes] (c) adequate, relevant ~~and limited to the minimum necessary~~ not excessive in relation to the purposes for which ~~they are processed; they shall only be processed if, and as long as, the purposes could not be fulfilled by processing information that does not involve personal data~~; (d) accurate and, where necessary,

no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the data will be processed solely for historical, statistical or scientific research purposes in accordance with the rules and conditions of Article 83 and if a periodic review is carried out to assess the necessity to continue the storage; (f) processed under the responsibility and liability of the controller, who shall ensure and demonstrate for each processing operation the compliance with the provisions of this Regulation.	which they are processed, are erased or rectified without delay ***(accuracy).*** (e) kept in a form which permits ***direct or indirect*** identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the data will be processed solely for historical, statistical or scientific research ***or for archive*** purposes in accordance with the rules and conditions of Articles 83 ***and 83a*** and if a periodic review is carried out to assess the necessity to continue the storage, ***and if appropriate technical and organizational measures are put in place to limit access to the data only for these purposes (storage minimisation)***; ***(ea) processed in a way that effectively allows the data subject to exercise his or her rights (effectiveness); (eb) processed in a way that protects against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures (integrity);*** (f) processed under the responsibility and liability of the controller, who shall ensure and ***be able to*** demonstrate the ~~for each processing operation~~ compliance with the provisions of this Regulation ***(accountability)***.	kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay; (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the data will be processed solely ~~for historical, statistical or scientific research purposes in accordance with the rules and conditions of Article 83 and if a periodic review is carried out to assess the necessity to continue the storage~~; [for archiving purposes in the public interest, for statistical, scientific or historical purposes purposes subject to implementation of the appropriate technical and organisational measures required by the Regulation in order to safeguard the rights and freedoms of data subject ;] [Proposal GER: (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed, ~~personal data may be stored for longer periods insofar as the data will be processed solely for historical, statistical or scientific research purposes in accordance with the rules and conditions of Article 83 and if a periodic review is carried out to assess the necessity to continue the storage~~; especially by applying measures of pseudonymisation or anonymisation at the earliest possible stage;] (ee) processed in a manner that ensures appropriate security of the personal data. (f) ~~processed under the responsibility and liability of the controller, who shall ensure and demonstrate for each processing operation the compliance with the provisions of this Regulation~~. 1. The controller shall be responsible for compliance with paragraph 1. [Proposal GER: 2. Compliance with fair processing referred to in

		paragraph (1) (a) means (a) ... (a) (b) ... [Subject of another German Note coming soon] (f) the use of privacy-enhancing technologies, such as anonymisation and pseudonymisation applied at the earliest possible stage, having regard to available technology and the cost of implementation, in order to minimise the risk for the rights and freedoms of the data subject]

Article 6 Lawfulness of processing		
European Commission (Original version)	European Parliament (Text adopted by Parliament)	Council of European Union (Consolidated text of the Commission and Council)
1. Processing of personal data shall be lawful only if and to the extent that at least one of the following applies: (a) the data subject has given consent to the processing of their personal data for one or more specific purposes; (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; (c) processing is necessary f or compliance with a legal obligation to which the controller is subject; (d) processing is necessary in order to protect the vital interests of the data subject; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (f) processing is necessary for the purposes of the legitimate interests pursued by a controller, except where such interests are overridden by the interests or fundamental rights and freedom s of the data subject which require protection of personal data, in particular where the data subject is a child. This shall not apply to processing carried out by public authorities in the performance of their tasks. 2. Processing	Lawfulness of processing 1. Processing of personal data shall be lawful only if and to the extent that at least one of the following applies: (a) the data subject has given consent to the processing of their personal data for one or more specific purposes; (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; (c) processing is necessary for compliance with a legal obligation to which the controller is subject; (d) processing is necessary in order to protect the vital interests of the data subject; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (f) processing is necessary for the purposes of the legitimate interests pursued by *the* controller *or, in case of disclosure, by the third party to whom the data is disclosed, and which meet the reasonable expectations of the data subject based on his or her relationship with the controller*, except where such interests are overridden by the interests or fundamental rights and freedoms of the	[Lawfulness of processing 1. Processing of personal data shall be lawful only if and to the extent that at least one of the following applies: (a) the data subject has given unambiguous consent to the processing of their personal data for one or more specific purposes; (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; (c) processing is necessary for compliance with a legal obligation to which the controller is subject; (d) processing is necessary in order to protect the vital interests of the data subject; (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller; (f) processing is necessary for the purposes of the legitimate interests pursued by ~~a~~ the controller, or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. This subparagraph shall not apply to processing carried out by public

of personal data which is necessary for the purposes of historical, statistical or scientific research shall be lawful subject to the conditions and safeguards referred to in Article 83. 3. The basis of the processing referred to in points (c) and (e) of paragraph 1 must be provided for in: (a) Union law, or (b) the law of the Member State to which the controller is subject. The law of the Member State must meet an objective of public interest or must be necessary to protect the rights and freedoms of others, respect the essence of the right to the protection of personal data and be proportionate to the legitimate aim pursued. 4. Where the purpose of further processing is not compatible with the one for which the personal data have been collected, the processing must have a legal basis at least in one of the grounds ref erred to in points (a) to (e) of paragraph 1. T his shall in particular apply to any change of terms and general conditions of a contract. 5. The Commission shall be em powered to adopt delegated acts in accordance with Article 86 for the purpose of further specifying the conditions referred to in point (f) of paragraph 1 for various sectors and data processing situations, including as regards the processing of personal data related to a child.	data subject which require protection of personal data. This shall not apply to processing carried out by public authorities in the performance of their tasks. 2. Processing of personal data which is necessary for the purposes of historical, statistical or scientific research shall be lawful subject to the conditions and safeguards referred to in Article 83. 3. The basis of the processing referred to in points (c) and (e) of paragraph 1 must be provided for in: (a) Union law, or (b) the law of the Member State to which the controller is subject. The law of the Member State must meet an objective of public interest or must be necessary to protect the rights and freedoms of others, respect the essence of the right to the protection of personal data and be proportionate to the legitimate aim pursued. ***Within the limits of this Regulation, the law of the Member State may provide details of the lawfulness of processing, particularly as regards data controllers, the purpose of processing and purpose limitation, the nature of the data and the data subjects, processing measures and procedures, recipients, and the duration of storage.***	authorities in the p~~erformance of their tasks~~ exercise of their public duties. [Proposal GER: processing is necessary for the purposes of the legitimate interests pursued by the controller or by a controller to which the data are disclosed except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child. In assessing the interests it should be taken into account that the controller or personnel of the controller has taken effective measures of pseudonymisation of personal data in order to minimize the risk of the data subject. In such cases there is a refutable presumption that the subject' s interests and fundamental rights and freedoms do not override the controller's interests. This subparagraph shall not apply to processing carried out by public authorities ~~in the performance of their tasks~~] 2. Processing of personal data which is necessary for ~~purposes of~~ archiving purposes in the public interest , or for historical, statistical or scientific ~~research~~ purposes shall be lawful subject also to the conditions and safeguards referred to in Article 83. 3. The basis ~~of~~ for the processing referred to in points (c) and (e) of paragraph 1 must be [~~provided for in~~ established in accordance with:]8 (a) Union law, or (b) ~~the~~ national law of the Member State to which the controller is subject. ~~The law of the Member State must meet an objective of public interest or must be necessary to protect the rights and freedoms of others, respect the essence of the right to the protection of personal data and be proportionate to the legitimate aim pursued.~~ The purpose of

		the processing shall be determined in this legal basis or as regards the processing referred to in point (e) of paragraph 1, be necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. This legal basis may contain specific provisions to adapt the application of rules of this Regulation, inter alia the general conditions governing the lawfulness of data processing by the controller, the type of data which are subject to the processing, the data subjects concerned; the entities to, and the purposes for which the data may be disclosed; the purpose limitation; storage periods and processing operations and processing procedures, including measures to ensure lawful and fair processing, including for other specific processing situations as provided for in Chapter IX. 3a. In order to ascertain whether a purpose of further processing is compatible with the one for which the data are initially collected, the controller shall take into account, inter alia: (a) any link between the purposes for which the data have been collected and the purposes of the intended further processing; (b) the context in which the data have been collected; (c) the nature of the personal data; (d) the possible consequences of the intended further processing for data subjects; (e) the existence of appropriate safeguards. [Proposal GER: (f) whether measures of anonymisation or pseudonymisation have been applied to the data.] 4. Where the purpose of further processing is incompatible with the one for which the personal data have been collected, the further processing must have a legal basis at least in one of the grounds referred to in points (a) to (e~~f) of

		paragraph 1. ~~This shall in particular apply to any change of terms and general conditions of a contract.~~ 5. ~~The Commission shall be empowered to adopt delegated acts in accordance with Article 86 for the purpose of further specifying the conditions referred to in point (f) of paragraph 1 for various sectors and data processing situations, including as regards the processing of personal data related to a child.~~] [Proposal GER 5. Where personal data are pseudonymised, the re-identification of the data subject and further processing of these data shall only be allowed based on points (a), (b), (c), (d) or (e) of Article 6 (1), or if the controller demonstrates compelling legitimate grounds for the re-identification which override the interests or fundamental rights and freedoms of the data subject. The same applies to personal data which have been anonymized by the controller if they are attributable to a data subject again.]

Article 81 Processing of personal data concerning health		
European Commission (Original version)	European Parliament (Text adopted by Parliament)	Council of European Union (Consolidated text of the Commission and Council)
1. Within the limits of this Regulation and in accordance with point (h) of Article 9(2), processing of personal data concerning health must be on the basis of Union law or Member State law which shall provide for suitable and specific measures to safeguard the data subject's legitimate interests, and be necessary for: (a) the purposes of preventive or occupational medicine, medical diagnosis, the provision of care or treatment or the management of health-care services, and where those data are processed by a health professional subject to the obligation of professional	Processing of personal data concerning health ***1. In accordance with the rules set out in*** this Regulation, ***in particular*** with point (h) of Article 9(2), processing of personal data concerning health must be on the basis of Union law or Member State law which shall provide for suitable, consistent, and specific measures to safeguard the data subject's interests ***and fundamental rights, to the extent that these are necessary and proportionate, and of which the effects shall be foreseeable by the data subject,*** for: (a) the purposes of preventive or occupational	~~Processing of personal data for health-related purposes 1. Within the limits of this Regulation and in accordance with points (g) and (h) of Article 9(2), processing of personal data concerning health must be referred to in Article 9(1) may be processed on the basis of Union law or Member State law which shall provides for suitable and specific measures to safeguard the data subject's legitimate interests, and be when necessary for: (a) the purposes of preventive or occupational medicine, medical diagnosis, the provision of care or treatment or the management of~~

secrecy or another person also subject to an equivalent obligation of confidentiality under Member State law or rule s established by national competent bodies; or (b) reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety, inter alia for medicinal products or medical devices; or (c) other reasons of public interest in areas such as social protection, especially in order to ensure the quality and cost-effectiveness of the procedures used for settling claims for benefits and services in the health insurance system. 2. Processing of personal data concerning health which is necessary for historical, statistical or scientific research purposes, such as patient registries set up for improving diagnoses and differentiating between similar types of diseases and preparing studies for therapies, is subject to the conditions and safeguards referred to in Article 83. 3. The Commission shall be em powered to adopt delegated acts in accordance with Article 86 for the purpose of further specifying other reasons of public interest in the area of public health as referred to in point (b) of paragraph 1, as well as criteria and requirements for the safeguards for the processing of personal data for the purposes referred to in paragraph 1. A	medicine, medical diagnosis, the provision of care or treatment or the management of healthcare services, and where those data are processed by a health professional subject to the obligation of professional secrecy or another person also subject to an equivalent obligation of confidentiality under Member State law or rules established by national competent bodies; or (b) reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety, inter alia for medicinal products or medical devices, ***and if the processing is carried out by a person bound by a confidentiality obligation; or*** (c) other reasons of public interest in areas such as social protection, especially in order to ensure the quality and cost-effectiveness of the procedures used for settling claims for benefits and services in the health insurance system ***and the provision of health services. Such processing of personal data concerning health for reasons of public interest shall not result in data being processed for other purposes, unless with the consent of the data subject or on the basis of Union or Member State law. 1a. When the purposes referred to in points (a) to (c) of paragraph 1 can be achieved without the use of personal data, such data shall not be used for those purposes, unless based on the consent of the data subject or Member State law. 1b. Where the data subject's consent is required for the processing of medical data exclusively for public health purposes of scientific research, the consent may be given for one or more specific and similar researches. However, the data subject***	~~health-care services, and where those data are processed by a health professional subject to the obligation of professional secrecy under Union or Member State law or rules established by national competent bodies to the obligation of professional secrecy or by another person also subject to an equivalent obligation of confidentiality secrecy; under Member State law or rules established by national competent bodies; or (b) reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety, inter alia for of health care and of medicinal products or medical devices; or (c) other reasons of public interest in areas such as social protection, especially in order to ensure the quality and cost effectiveness of the procedures used for settling claims for benefits and services in the health insurance system. 2. Processing of personal data concerning health which is necessary for historical, statistical or scientific research purposes, such as patient registries set up for improving diagnoses and differentiating between similar types of diseases and preparing studies for therapies, is subject to the conditions and safeguards referred to in Article 83 Articles 83a to 83d. 3. The Commission shall be empowered to adopt delegated acts in accordance with Article 86 for the purpose of further specifying other reasons of public interest in the area of public health as referred to in point (b) of paragraph 1, as well as criteria and requirements for the safeguards for the processing of personal data for the purposes referred to in~~

	may withdraw the consent at any time. 1c. For the purpose of consenting to the participation in scientific research activities in clinical trials, the relevant provisions of Directive 2001/20/EC of the European Parliament and of the Council1 shall apply. 2. Processing of personal data concerning health which is necessary for historical, statistical or scientific research purposes *shall be permitted only with the consent of the data subject, and shall be* subject to the conditions and safeguards referred to in Article 83. *2a. Member States law may provide for exceptions to the requirement of consent for research, as referred to in paragraph 2, with regard to research that serves a high public interest, if that research cannot possibly be carried out otherwise. The data in question shall be anonymised, or if that is not possible for the research purposes, pseudonymised under the highest technical standards, and all necessary measures shall be taken to prevent unwarranted re-identification of the data subjects. However, the data subject shall have the right to object at anytime in accordance with Article 19.* 3. The Commission shall be empowered to adopt, *after requesting an opinion of the European Data Protection Board,* delegated acts in accordance with Article 86 for the purpose of further specifying public interest in the area of public health as referred to in point (b) of paragraph 1 *and high public interest in the area of research as referred to in paragraph 2a. 3a. Each Member State shall notify to the Commission those provisions of its law which it adopts pursuant to paragraph 1, by the date specified in Article 91(2) at the latest and, without delay, any*	~~paragraph 1.]~~

	subsequent amendment affecting them.	

[별첨 3] 일본 개정 개인정보 보호법 개정안의 요강(2015.9.3.)[203]

[1]	개인정보 보호에 관한 법률 일부 개정 관계
1	목적에 관한 것 이 법률은 개인정보의 적정하고 효과적인 활용으로 새로운 산업의 창출 및 활력있는 경제 사회 및 풍부한 국민생활 실현에 이바지하고 기타 개인정보의 유용성을 증진하면서 개인의 권익을 보호하는 것을 목적으로 하도록 한다
2	정의에 관한 것 (1) 이 법률에서 "개인정보"란 생존하는 개인에 관한 정보이며, 다음 중 하나에 해당하는 것이다. 1) 해당 정보에 포함된 이름, 생년월일 기타 기술 등으로 특정 개인을 식별할 수 있는 것(다른 정보와 쉽게 조회 비교할 수 있으며 그것으로 특정 개인을 식별할 수 있는 것을 포함한다.) 2) 개인식별부호가 포함될 것 (2) 이 법률에서 "개인식별부호"는 다음 중 하나에 해당하는 문자 번호 기호 기타 부호 중, 정령으로 정하도록 한다. 1) 특정 개인의 신체 일부의 특징을 전자계산기 용으로 제공하기 위하여 변환한 부호이며, 해당 특정 개인을 식별할 수 있는 것 2) 개인에게 제공되는 역무의 이용 혹은 개인에게 판매되는 상품의 구입에 대해 할당되거나 개인에 발행되는 카드 기타 서류에 기재되거나 전자적 방식에 의해 기록된 부호로서 그 이용자나 구매자 또는 발행을 받는 사람마다 다르게 배정 받아 또는 기재되거나 기록됨으로써 특정 이용자나 구매자 또는 발행을 받는 사람을 식별할 수 있는 것 (3) 이 법률에서 "배려가 필요한 개인정보"라 함은 본인의 인종, 신조, 사회적 신분, 병력, 범죄 경력, 범죄로 인한 해를 입은 사실 기타 본인에 대한 부당 차별, 편견 기타 불이익이 생기지 않도록 그 취급에 특히 배려가 필요한 것으로 정령으로 정하는 기술 등이 포함되는 개인정보를 말한다. (4) " 개인정보 데이터베이스 등 " 의 정의에서 이용방법을 보아 개인의 권익을 해칠 우려가 적은 것으로 정령으로 정하는 것은 제외한다. (5) 개인정보취급사업자의 정의에서 취급하는 개인정보의 양 및 이용 방법으로 보아 개인의 권익을 해칠 우려가 적은 것으로 정령으로 정하는 자를 제외한다는 취지의 규정을 두도록 한다. (6) 이 법률에서 "익명가공정보"란 특정 개인을 식별할 수 없도록 개인정

203) 다음 내용은 아래 사이트에 제시된 내용을 요약번역한 것이다.
http://www.cas.go.jp/jp/houan/150310/siryou2.pdf

	보를 가공하여 얻는 개인에 관한 정보이며, 해당 개인정보를 복원할 수 없도록 한 것을 말한다. (7) 이 법률에서 "익명가공정보 취급사업자"란 특정의 익명가공정보를 전자 계산기를 이용하여 검색할 수 있도록 체계적으로 구성한 것 등을 사업용으로 제공하는 자를 말한다.
3	국가 및 지방 공공 단체의 책무 등에 관한 것 정부는 국제기관 기타 국제적 활동에 협력을 통해서, 각국 정부와 공동으로 국제적으로 조화를 이룬 개인정보에 관련 제도를 구축하기 위해서 필요한 조치를 강구한다.
4	개인정보 취급 사업자의 의무에 관한 것 (1) 이용목적의 특정 개인정보취급사업자는 이용목적을 변경하는 경우에는 변경 전의 이용목적과 관련성을 갖는다고 합리적으로 인정되는 범위를 넘어 가서는 안 된다. (2) 적정한 취득 개인정보취급사업자는 일정 경우를 제외하고 미리 본인의 동의를 받지 않고 배려가 필요한 개인정보를 취득해서는 안 된다. (3) 데이터 내용의 정확성 확보 등 개인정보취급사업자는 개인 데이터를 이용할 필요가 없어진 경우 해당 개인 데이터를 지체 없이 삭제하도록 노력해야 한다. (4) 제3자 제공의 제한 1) 일정한 경우 미리 본인의 동의를 받지 않고 해당 본인을 식별하는 개인 데이터를 제3자에게 제공할 수 있는 취지의 규율에 대해서, 해당 규율의 대상이 되는 개인 데이터로부터 배려가 필요한 개인정보를 제외하고 해당 규율에 의한 개인 데이터를 제공하려면 일정한 사항을 개인정보보호위원회 규칙에서 정하는 바에 의하여, 미리 본인에게 통지하거나 본인이 쉽게 알 수 있는 상황과 함께 개인정보보호 위원회에 신고해야 한다. 2) 개인정보보호위원회는 신고가 있을 때, 개인정보보호위원회 규칙으로 정하는 바에 의해 해당 신고에 관한 사항을 공표해야 한다. (5) 외국에 있는 제3자에게의 제공 제한 개인정보취급사업자는 외국에 있는 제3자에게 개인 데이터를 제공할 경우에 일정한 경우를 제외하고 미리 외국에 있는 제3자에게의 제공을 인정하는 취지의 본인 동의를 얻어야 한다. (6) 제3자 제공에 관한 기록의 작성 등 개인정보취급사업자는 개인 데이터를 제3자에게 제공한 때는 개인정보보호 위원회규칙으로 정하는 바에 의한 해당 개인 데이터를 제공한 연월

	일, 해당 제3자의 이름 등의 기록을 작성하여 일정 기간 보존해야 한다. (7) 제3자 제공을 받을 때 확인 등 개인정보취급사업자는, 제3자로부터 개인 데이터의 제공을 받을 때에는 개인정보보호위원회 규칙으로 정하는 바에 의해 해당 제3자에 의한 해당 개인 데이터의 취득경위 등을 확인하는 동시에, 해당 개인 데이터의 제공을 받은 연월일 등의 기록을 작성하여 일정 기간 보존해야 한다. (8) 공개 등 1) 본인은 개인정보취급사업자에 대해 해당 본인을 식별할 수 있는 개인 데이터의 공개를 청구할 수 있으며 일정한 경우 해당 개인 데이터의 내용의 정정, 추가 또는 삭제, 이용 정지 혹은 제거 또는 제3자에게의 제공 중단을 청구할 수 있다. 2) 본인은 (1)에 의한 청구에 관한 소송을 제기하려고 할 때는 일정한 경우를 제외하고, 그 피고가 될 사람에게 미리 해당 청구를 행하고, 그 도달한 날로부터 2주일을 경과한 후가 아니면 그 소송을 제기할 수 없다.
5	익명가공정보 취급사업자 등의 의무에 관한 것 (1)익명 가공 정보의 작성 등 개인정보취급사업자는 익명가공정보의 작성 등에 대해서 다음과 같이 한다. 1) 익명가공정보를 작성할 때는 특정 개인을 식별하는 것 및 그 작성에 이용하는 개인정보를 복원할 수 없도록 하기 위해서 필요한 것으로서 개인정보보호위원회 규칙으로 정하는 기준에 따르고 해당 개인정보를 가공해야 한다. 2)익명가공정보를 작성했을 때는 가공방법에 관한 정보 등의 누설을 방지하기 위해서 필요한 것으로서 개인정보보호위원회 규칙으로 정하는 기준에 따르고 이들 정보의 안전관리를 위한 조치를 강구하지 않으면 안 된다. 동시에, 개인정보보호위원회 규칙으로 정하는 바에 따라 해당 익명가공정보에 포함된 개인에 관한 정보 항목을 발표해야 한다. 3) 익명가공정보를 작성하고 스스로 해당 익명가공정보를 취급하는데 있어서는 해당 익명가공정보의 작성에 사용된 개인정보에 관련된 본인을 식별하기 위해서 해당 익명가공정보를 다른 정보와 대조해서는 안 된다. (2) 익명가공정보의 제공 익명가공정보취급사업자(익명가공정보를 작성한 개인정보취급사업자를 포함한다. (4)에서도 동일)는 익명가공정보를 제3자에게 제공할 때 개인정보보호위원회 규칙에서 정하는 바에 의하여, 미리, 제3자에게 제공되는 익명가공정보에 포함된 개인에 관한 정보항목 및 그 제공방법에 대해 공표하는 동시에, 해당 제3자에 대해서 해당 제공에 관한 정보가 익명가공

	정보인 사실을 명시해야 하도록 한다. (3) 식별행위의 금지 익명가공정보취급사업자는 익명가공정보(개인정보를 가공하여 작성한 것을 제외)를 취급하는데 있어서는 해당 익명가공정보의 작성에 사용된 개인정보에 관련된 본인을 식별하기 위해서, 가공 방법에 관한 정보 등을 취득하거나 해당 익명가공정보를 다른 정보와 대조해서는 안 된다. (4) 안전 관리 조치 등 익명가공정보취급사업자는 익명가공정보의 안전관리 때문에 적절한 조치, 익명가공정보의 취급에 관한 민원처리 기타 익명가공정보를 적정히 취급하는 데 필요한 조치를 스스로 강구하고 해당 조치의 내용을 공개하도록 노력해야 한다.
6	감독에 관한 것 (1) 감독의 주체 및 실시 개인정보취급사업자의 감독을 실시하는 주체를 주무대신으로부터 개인정보보호위원회에 고치고, 익명가공정보취급사업자의 감독을 개인정보보호위원회가 실시하도록 한다. (2) 보고 및 출입 검사 개인정보보호위원회는 일정한 경우에 있어서 개인정보취급사업자 또는 익명가공정보취급사업자(이하"개인정보취급사업자 등"이라 한다.)에 대한 개인정보 또는 익명가공정보(이하"개인정보 등"이라 함)의 취급에 관하여 필요한 보고 혹은 자료의 제출을 요구하고 또는 그 직원에 해당 개인정보취급사업자 등의 사무소 기타 필요한 장소에 출입시키고 검사하는 등 할 수 있도록 한다. (3) 지도 및 조언 개인정보보호위원회는 일정한 경우에 있어 개인정보취급사업자 등에 대한 개인정보 등 취급에 관해 필요한 지도 및 조언을 할 수 있다. (4) 권한의 위임 개인정보보호위원회는 긴급하고 중점적으로 개인정보 등의 적절한 취급 확보를 도모할 필요가 있음 기타 정령에서 정하는 사정이 있으므로 필요가 있다고 인정될 때는 정령으로 정하는 바에 의해, (2)에 의한 권한을 사업 소관 장관에게 위임할 수 있다. (5) 사업 소관 장관의 청구 사업 소관 장관은 개인정보취급사업자 등에 의한 개인정보 등의 적정한 취급을 확보하기 위해서 필요하다고 인정될 때는 개인정보보호위원회에게 이 법률의 규정에 따르는 적당한 조치를 취하도록 요구할 수 있다.
7	민간단체의 개인정보 보호의 추진에 관한 것 (1) 인정개인정보보호단체의 인정 및 감독을 실시하는 주체를 주무 대신

	으로부터 개인정보보호위원회로 고친다. (2) 개인정보보호지침 1) 인정개인정보보호단체는 대상 사업자의 개인정보 등의 적절한 취급 확보 때문에 소비자의 의견을 대표하는 자 기타 관계자의 의견을 듣고 이 법률의 규정취지에 따른 지침(이하"개인정보보호지침"이라 칭함)을 작성하도록 노력해야 한다. 2) 인정개인정보보호단체는 1)에 의한 개인정보보호지침을 작성한 때에는 개인정보보호위원회 규칙으로 정하는 바에 따라 지체없이 해당 개인정보보호지침을 개인정보보호위원회에 신고해야 한다. 3) 개인정보보호위원회는 2)에 의한 개인정보보호지침의 신고가 있을 때는, 개인정보보호위원회 규칙으로 정하는 바에 의해 해당 개인정보 보호 지침을 발표해야 한다.
8	개인정보 보호 위원회에 관한 것 (신설) (1) 설치 내각부 설치 법 규정에 의거, 내각 총리대신 소관에 속하는 개인정보보호위원회를 둔다. (2) 임무 위원회는 개인정보의 적정하고 효과적인 활용으로 새로운 산업의 창출 및 활력있는 경제사회와 풍부한 국민생활 실현에 이바지하며 기타 개인정보의 유용성을 배려하는 한편 개인의 권익보호를 위해 개인정보의 적절한 취급확보를 도모함(개인번호이용사무 등 실시자에 대한 지도 및 조언 기타 조치를 강구하는 것을 포함)을 임무로 한다. (3) 소관 사무 위원회는 (2)의 임무를 달성하기 위하여 다음과 같은 사무를 관장한다. 1) 기본 방침의 책정 및 추진 2) 개인정보 및 익명가공정보의 취급에 관한 감독 및 민원 신청에 대한 필요한 알선 및 그 처리를 실시하는 사업자의 협력 3) 인정 개인정보 보호단체 4) 특정 개인정보 취급에 관한 감시 또는 감독 및 민원 신청에 대한 필요한 알선 및 그 처리를 실시하는 사업자의 협력 5) 특정개인정보 보호 평가 6) 개인정보보호 및 적정하고 효과적 활용에 대한 홍보 및 계발 7) 1)에서까지에 제시하는 사무를 실시하기 위해서 필요한 조사 및 연구 8) 소관 사무에 관한 국제 협력 9) 기타 법률에 의거 위원회에 속하게 된 사무 (4) 조직 등

	1) 위원회는 위원장 및 위원 여덟 명을 가지고 조직한다. 2) 위원장 및 위원은 의원의 동의를 얻어 내각 총리대신이 임명한다. 3) 위원회에 전문 사항을 조사할 전문 위원을 둘 수 있다. 4) 위원회는 그 소관 사무에 대해서, 법률 혹은 정령을 실시하기 위한 또는 법률이나 정령의 특별한 위임에 근거하여 개인정보 보호 위원회 규칙을 제정할 수 있다.
9	기타 사항 에 관한 것 (1) 적용 범위 이 법률의 일정한 규정은 국내에 있는 자에게 물품 또는 역무의 제공과 관련해서 그를 본인으로 하는 개인정보를 취득한 개인정보취급사업자가 외국에서 해당 개인정보 또는 해당 개인정보를 이용하여 작성한 익명가공정보를 취급하는 경우에 대해서도 적용한다. (2) 외국 집행 당국에 대한 정보 제공 개인정보보호위원회는 이 법에 해당하는 외국의 법령을 집행하는 외국당국에 그 직무수행에 이바지한다고 인정하는 정보의 제공을 실시할 수 있다.
10	벌칙에 관한 것 개인정보취급사업자(그 자가 법인인 경우에는 그 임원, 대표자 또는 관리인)혹은 종업자 또는 이들이었던 자가 그 업무에 관해서 다룬 개인정보데이터베이스 등을 자기 또는 제3자의 부당이득을 도모할 목적으로 제공하거나 도용했을 때는 1년 이하의 징역 또는 50 만엔 이하의 벌금에 처한다.
11	기타 기타 필요한 규정을 정비한다.

<참고 문헌>

1. 국내 자료

[1] 개인정보보호위원회,「해외 개인정보보호 집행체계 및 개인정보보호 주요 동향조사」, 2012.

[2] _____,「개인정보 범위에 관한 연구」, 2014.

[3] 경실련 소비자정의센터,「빅데이터활용과 다가올 위험-개인정보 비식별화의 문제를 중심으로-」, 토론회발표자료, 2015.

[4] 김민호, “개인정보처리자에 대한 연구”, 「성균관법학」, 제26권 제4호(2014.12.)

[5] 김선남 · 이환수,「빅데이터 개인정보보호 가이드라인(안)의 개선 방향에 관한 연구」,『정보화정책』 제21권 제4호, 2014.

[6] 문재완, “개인정보의 개념에 관한 연구”, 「공법연구」, 제42집 제3호(2014.2.)

[7] 미래창조과학부 · 한국정보화진흥원 · K-ICT빅데이터센터,「빅데이터 활용을 위한 개인정보 비식별화 기술활용 안내서 Ver1.0」, 2015.

[8] 미래창조과학부 · 한국정보화진흥원,「빅데이터 활용을 위한 개인정보 비식별화 사례집」, 2014.

[9] 박원환,「빅데이터 분석을 위한 개인정보의 비식별화 방법」, 발표자료, 2015.

[10] 박원환 · 황조연,「통계자료의 비밀보호를 위한 익명화 방법들」,『통계연구』, 제9권 제2호, 2004.

[11] 방송통신위원회, 「빅데이터 개인정보보호 가이드라인(안)」, 2013.

[12] 오길영,「데이터 상업화 과정으로서의 개인정보 비식별화」,『민주법학』 제58호, 2015.

[13] 이시직,「공공데이터 제공 및 이용 활성화를 위한 법·제도적 개선방안」,『정보통신방송정책』제26권 제3호, 2014.

[14] 이인호,「개인정보 보호법상의 ‘개인정보’ 개념에 대한 해석론-익명화한 처방전 정보를 중심으로」,『정보법학』, 제19권 제1호, 2015.

[15] 전수연,「개인정보 비식별화 동향 및 시사점」, ISSUE&TREND, 2014.
[16] 정상조 외,「비식별개인정보의 보호 및 활용에 관한 연구」, 방송통신위원회, 2010.8.
[17] 정영철,「의료분야 빅데이터 활용을 위한 개인정보 비식별화 규정 현황과 과제」,『보건복지포럼』, 통권 제227호, 2015.
[18] 최경진, 「빅데이터 · 사물인터넷 시대 개인정보보호법제의 발전적 전환을 위한 연구」, 『중앙법학』, 제17집 제4호, 2015
[19] 프라이버시 정책연구 포럼,「개인정보보호법제 개선을 위한 정책연구 보고서」, 2013.
[20] 한국정보화진흥원, 「각국의 개인정보 보호법상 개인정보의 정의 및 해석」, 2014.
[21] 행정자치부 · 한국정보화진흥원,「개인정보 비식별화에 대한 적정성 자율평가 안내서」, 2014.

2. 외국 자료

[1] Bambauer, J., Muralidhar, K., and Sarathy, R., "Fool's Gold: An Illustrated Critique of Differential Privacy", *16 Vanderbilt Journal of Entertainment & Technology Law*, 2014, 16, pp. 701-755.
[2] Cavoukian, A. and Castro, D., Information and Privacy Commissioner, *Big Data and Innovation, Setting the Record Straight: De-identification Does Work*, 2014.
[3] Commission nationale de l'informatique et des liberties(CNIL), *WP 29 Opinion on anonymization techniques*, 2015.
[4] Committee on Strategies for Responsible Sharing of Clinical Trial Data, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks*, 2015, National Academics of Science.
[5] D'Orazio, V., Honaker, *J., and King, G., Differential Privacy for*

Social Science Inference. Sloan Foundation Economics Research Paper No. 2676160, 2015.

[6] El Emam, K., Arbuckle, L., Koru G., Eze, B., Gaudette, L., Neri, E., Rose, S., Howard, J., and Gluck, J., "De-identification Methods for Open Health Data: The Case of the Heritage Health Prize", *Journal of Medical Internet Research*, 2012, 14, pp. 1-18.

[7] El Emam, K. and Arbuckle L., *Anonymizing Health Data*, 2014, O'reilly.

[8] El Emam, K. and Alvarez, C., "A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques", *International Data Privacy Law*, 2015, 5, pp. 73-87.

[9] EU Commission, *ARTICLE 29 Data Protection Working Party Opinion 05/2014 on Anonymisation Techniques,* 2014.

[10] Garfinkel, S., National Institute of Standards and Technology, *De-Identification of Personally Identifiable Information*, 2015.

[11] Gellman, R., "The Deidentification Dilemma: A Legislative and Contractual Proposal", 21 *Fordham Intellectual Property, Media & Entertainment law Journal*, 2010, 21, pp. 33-61.

[12] Greens/European Free Alliance(EFA), European Parliament, *EU General Data Protection Regulation State of play and 10 main issues*, 2015.

[13] Information Commissioner's Office(ICO), *Anonymisation: managing data protection risk code of practice*, 2012.

[14] _____, *Determining what is personal data*, 2012.

[15] McCalister, E., Grance, T., and Scarfone, K., National Institute of Standards and Technology, *Guide to Protecting the Confidentiality of Personally Identifiable Information*, 2012.

[16] Narayanan A., "An Adversarial Analysis of the Reidentifiability of the Heritage Health Prize Dataset", 2011(manuscript).

[17] Office for Civil Rights(OCR), U.S. Department of Health & Human Services, *Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule*, 2012.

[18] Ohm, P., "Broken Promises of Privacy: Responding To The Surprising Failure of Anonymization", *UCLA Law Review*, 2010, 57, pp. 1701-1777.

[19] Oswald, M., Information Standards Board for Health and Social Care, *Anonymisation Standard for Publishing Health and Social Care Data Specification*, 2013.

[20] Romanosky, S. and Acquisti, A., "Privacy Costs and Personal Data Protection: Economic and Legal Perspectives", *Berkeley Technology Law Journal*, 2009, 24, pp. 1061-1099.

[21] Schwartz, P. and Solove, D., "The PII Problem: Privacy and a New Concept of Personally Identifiable Information", *New York University Law Review*, 2011, 86, pp. 1814-1894.

[22] Sweeney, L., *Simple Demographics Often Identify People Uniquely.* Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

[23] _____, "k-Anonymity: A Model for Protecting Privacy", *International Journal on Uncertainty, Fuzziness and Knowledge-based System*, 2002, 10, pp. 557-570.

[24] U.S. Department of Health & Human Services, *Protecting Personal Information in Research: Understanding the HIPAA Privacy Rule*, 2004.

[25] 高度情報通信ネットワーク社会推進戦略本部, パーソナルデータの利活用に関する制度見直し方針, 平成25年12月20日.

[26] 高度情報通信ネットワーク社会推進戦略本部, パーソナルデータの利活用に関する制度改正大綱, 平成26年 6 月24日.

[27]「個人情報の保護に関する法律及び行政手続における特定の個人を識別するための番号の利用等に関する法律の一部を改正する法律」，平成27年９月３日成立・同月９日公布.

[28] 伊藤，個人特定性低減のための加工について，平成26年 5月13日(第６回会合)[204].

[29] 岡村，準個人情報と特定性低減データの考察について，平成26年5月13日(第６回会合).

[30] 菊池，個人特定性低減データの考察，平成26年5月13日(第６回会合).

[31] 佐久間， 「準個人情報」及び「個人特定性低減データ」について，平成26年5月13日(第６回会合).

[32] 佐藤，技術WGによる検討作業の前提，平成26年5月13日(第６回会合).

[33] 高橋，非特定情報のリスクの考察，平成26年5月13日(第６回会合）.

[34] 松本， 「（仮称）準個人情報」及び「（仮称）個人特定性低減データ」の考察，平成26年5月13日(第６回会合).

[35] 森，特定情報による権利侵害と非特定情報による権利侵害，平成26年5月13日(第６回会合).

[36] 佐藤，準個人情報と低減データ，平成26年5月13日(第６回会合).

[37] 佐藤，「（仮称）準個人情報」及び「（仮称）個人特定性低減データ」.に関する技術的観点からの考察について（中間報告）【詳細版】， 平成２６年５月２０日(第９回検討会)[205]

[38] 佐藤，技術検討ワーキンググループ報告書＜中間報告からの変更点＞，平成２６年５月２９日(第１０回検討会)[206].

[39] 日本画像医療システム工業会(JIRA)，医療情報利活用における匿名化技術ガイド Ver1.0，2015.

[40] 内閣官房情報通信技術（IT）総合戦略室，個人情報の保護に関する法律及び行政手続における特定の個人を識別するための番号の利用等に関する法律の一部を改正する法律案(概要,要綱,法律案・理由,新旧対照表,参照条文)，第１８９回 通常国会，2015.

204) 平成26年 5月13日(第6回会合)는 回技術検討ワーキンググループ의 6차 회의를 의미, 이하 동일
205) 平成２６年５月２０日(第9回検討会)는 パーソナルデータに関する検討会의 9차 회의를 의미
206) 平成２６年５月２９日(第１０回検討会)는 パーソナルデータに関する検討会의 10차 회의를 의미

3. 참조 사이트

[1] http://ptac.ed.gov
[2] http://www.jmir.org
[3] https://www.huntonprivacyblog.com
[4] http://www.bna.com
[5] http://www.hhs.gov
[6] http://ukanon.net
[7] http://ec.europa.eu
[8] https://www.kantei.go.jp
[9] http://www.cas.go.jp