최종연구보고서 | KISA 2018-0272

해외 비식별 조치 가이드라인 등에 대한 비교·분석

주관기관 : 한국인터넷진흥원
수탁기관 : 서울대학교 산학협력단

2018년 10월

제 출 문

한국인터넷진흥원 원장 귀하

본 보고서를 “해외 비식별조치 가이드라인 등에 대한 비교·분석”의 최종연구개발 결과보고서로 제출합니다.

2018년 10월 30일

수탁 기관 : 서울대학교 산학협력단
연구책임자 : 교 수 고학수 (서울대학교 법학전문대학원)
참여연구원 : 교 수 이동진 (서울대학교 법학전문대학원)
교 수 신수용 (삼성융학의과학원)
연 구 원 박미정 (서울대학교 의과대학)
연 구 원 김은수 (서울대학교 법학대학원)
보조연구원 박지훈 (서울대학교 법학대학원)

목 차

제 1 장 서론 ····· 1
제 1 절 연구의 필요성 ····· 1
제 2 절 연구의 목표 및 범위 ····· 4

제 2 장 개념 정의 및 비교 ····· 6
제 1 절 개념 정의 ····· 6
1. 비식별정보 ····· 6
2. 익명정보 ····· 10
3. 가명정보 ····· 11
제 2 절 비교 및 정리 ····· 15

제 3 장 비식별 조치의 이론적 배경 및 활용사례 ····· 20
제 1 절 비식별 조치의 이론적 배경 ····· 20
1. 기술적 방법론 ····· 22
2. 관리적 방법론 ····· 37
제 2 절 비식별정보의 활용사례 ····· 39
1. 국내 사례 ····· 39
2. 해외 사례 ····· 46

제 4 장 비식별 조치에 대한 해외의 제도 현황 ····· 51
제 1 절 유럽 ····· 51
1. EU ····· 51
2. 독일 ····· 63
3. 프랑스 ····· 68
4. 핀란드 ····· 84
5. 영국 ····· 99
제 2 절 북미 ····· 115
1. 미국 ····· 115
2. 캐나다 ····· 134
제 3 절 아시아 ····· 146
1. 일본 ····· 146
2. 중국 ····· 161
3. 싱가포르 ····· 169
제 4 절 호주 ····· 177

제 5 장 비식별 조치에 대한 국내 법제도 분석 및 정책적 제언 ····· 185
제 1 절 국내 법제도에 대한 분석 ····· 185
1. 현황 ····· 185
2. 문제점 ····· 189
제 2 절 정책적 제언 ····· 190
1. 해외 법제도의 시사점 ····· 190
2. 국내 규제 환경하에서 가명처리 개념의 도입가능성 ····· 192

제 6 장 결론 ····· 194

참 고 문 헌 ····· 198

그림 목차

(그림3-1) 비식별 조치 강도에 따른 개인정보보호와 데이터 활용성의 상관관계 ····· 20
(그림3-2) 식별자의 종류 ····· 22
(그림3-3) 비식별 조치 기법적용 결과 ····· 26
(그림3-4) 비식별 조치 기법 소개 ····· 28
(그림3-5) 비식별 조치 용어 정의의 모호성 ····· 35
(그림3-6) 정보, 개인정보, 개인식별정보, 민감정보와의 관련성 ····· 36
(그림3-7) 국민건강보험공단 국민건강 알람서비스 화면 ····· 40
(그림3-8) 서울아산병원 개인식별정보 정의 ····· 43
(그림4-1) 익명화된 데이터와 가명화된 데이터의 차이 ····· 67
(그림4-2) Dichotomy between “Personal Data” and “Anonymized Data” ····· 72
(그림4-3) Blurred line between anonymized data and personal data ····· 73
(그림4-4) Anonymized data ≠ Pseudonymized data ····· 74
(그림4-5) 익명화 프로세스 ····· 77
(그림4-6) Evaluation of the anonymization techniques ····· 78
(그림4-7) PIA 수행을 위한 반복프로세스 ····· 81
(그림4-8) 라이선스 및 데이터를 얻는 방법 ····· 95
(그림4-9) 익명화된 데이터 해당여부 판단 메커니즘 ····· 102
(그림4-10) 데이터 보관 주체가 2인 경우 ····· 106
(그림4-11) 외부로 완전 공개된 형태 ····· 106
(그림4-12) 전문가판단 방법이 실제로 진행되는 과정 ····· 119
(그림4-13) 익명가공정보의 활용이 이루어지는 전체적인 흐름 ····· 150
(그림4-14) 개별사업자들이 익명가공정보와 관련해서 준수해야 할 법적 의무 ····· 155
(그림4-15) 비식별 조치 과정 ····· 163
(그림4-16) 익명화 전반적인 과정 ····· 172

표 목차

[표3-1] 비식별 조치 방법 비교 ····· 24
[표3-2] 재식별 공격 분류 ····· 30
[표3-3] 재식별 상황 분류 ····· 30
[표3-4] 재식별 된 데이터 분류 ····· 31
[표3-5] 기밀성 모델 ····· 32
[표3-6] 비식별 조치를 위한 관리적 보호 조치 사항 ····· 37
[표3-7] 재식별 가능성 모니터링 점검 항목 ····· 38
[표3-8] 라인웍스가 활용중인 비식별 데이터 ····· 41
[표4-1] 가명화 인센티브로서 데이터보호 규칙의 유연성 ····· 54
[표4-2] GDPR 비식별 데이터의 정의와 수준 ····· 56
[표4-3] GDPR 제89조 ····· 61
[표4-4] CNIL 개인정보 보안가이드라인의 구성 ····· 69
[표4-5] CNIL 가이드라인이 권고하는 기업의 리스크 관리 단계 ····· 70
[표4-6] 조직의 개인정보 보안수준을 판단하기 위한 고려사항 ····· 71
[표4-7] 관련 부문별 법률 목록 ····· 86
[표4-8] 식별자 유형과 익명화 방법 ····· 88
[표4-9] UKAN 보고서의 익명화 체계 ····· 107
[표4-10] 익명화 된 데이터의 활용 사례 ····· 113
[표4-11] HIPPA 프라이버시 규칙이 나열한 18가지 식별자 ····· 121
[표4-12] PIPEDA의 9단계 비식별 조치 모형 ····· 136
[표4-13] 재식별행위에 해당 사례와 그렇지 않은 사례 비교 ····· 152
[표4-14] 일본 보고서 3장, 4장, 5장의 목차 ····· 154
[표4-15] 익명가공정보의 실제 사용의 예시 ····· 160
[표4-16] 중국 사이버보안법 목차 ····· 161
[표4-17] OAIC 2016의 다섯 가지 안전성 모형 ····· 180
[표5-1] 지역특화발전특구에 대한 규제특례법(대안) ····· 186

제 1 장 서 론

제 1 절 연구의 필요성

오늘날 개인정보 보호법은, 한편으로는 개인정보의 보호를 강화하면서 다른 한편으로는 정보의 활용 가능성을 모색하여야 하는 이중의 과제에 직면하고 있다. 한쪽에서는 강력해 보이는 규제 프로그램에도 불구하고 실제 법 집행은 미미한 수준이고, 여전히 불법적인 정보 이용이 만연해 있으며, 정보 제공에 동의하지 않을 수 없는 소비자 등의 처지를 이용하여 별 고민 없이 지나치게 많은 정보를 수집하는 관행이 사실상 개선되지 아니하고 있다는 비판이 끊이지 않고 나타난다. 동시에 다른 한쪽에서는 개인정보 보호법상 규제가 비현실적으로 엄격하여 데이터 확보가 불가능에 가깝고 그 결과 데이터 산업이 진전하지 못하고 있다는, 그리하여 우리나라가, 우리 기업이 데이터 기술과 산업을 둘러싼 국제적 경쟁에서 밀릴 위험에 처해있다는 호소도 만만치 아니하다. 4차 산업혁명위원회 주관으로 2018년 2월과 4월에 열린 개인정보 관련 해커톤은 이러한 상황을 해결하기 위한 상호이해와 조정, 타협의 시도였다고 할 수 있다.

수년 전부터 개인정보 비식별화 내지 익명화는 개인정보의 보호와 활용의 조화를 꾀할 대표적인 묘책으로 주목받아 왔다. 2016년 관련 정부부처들이 합동으로 발표한 「개인정보 비식별 조치 가이드라인」도 현행 법제에서는 비식별화 내지 익명화가 가장 현실적이고도 가장 잠재력이 큰 방법이라는 판단을 배경으로 한다. 그러나 위 2016년 「개인정보 비식별 조치 가이드라인」에 대하여는 법적 근거가 없다는 비난이, 비식별화 내지 익명화 자체에 대하여는 다른 정보와 결합하면 손쉽게 정보주체를 식별해낼 수 있어 사생활 등 침해 우려가 크다는 비판이 제기되고

있기도 하다. 비식별정보의 활용이 적법한지를 둘러싼 법적 분쟁도 그 사이 발생하기 시작하였고, 아직도 명확히 해결되지 못하고 있다.

반면 해외에서는 최근까지도 개인정보 보호법 일반과 특히 비식별화 내지 익명화 등에 관하여 일정한 진전이 이루어지고 있다. 무엇보다도 지난 5월 유럽연합의 일반정보보호규정(General Data Protection Regulation; GDPR)이 발효되었다. 일반정보보호규정은 유럽연합 회원국 내에서 직접적 효력을 가져 회원국의 국내법을 사실상 상당 부분 대체하였고, 이에 발맞추어 주요 회원국들이 자국 정보보호법을 개정하였다. 2016년 유럽연합 일반정보보호규정은 기존의 1995년 정보보호지침(Data Protection Directive) 서문(recital)의 언급을 이어받아 익명화(anonymisation)를 인정할 뿐 아니라, 새로 가명화(pseudonymisation) 개념을 도입하고, 본문 여러 곳에서 정보보안과 관련하여 가명처리를 규정하거나 가명처리에 일정한 혜택을 부여하는 등 이를 활성화하기 위한 유인책을 포함시켰다.

유럽연합 회원국 중에서는 최초로 독일이 정보보호법을 전면 개정하는 외에 가명처리에 관하여는 별도의 백서를 발표하기도 하였다. 프랑스는 물론, 브렉시트(Brexit) 와중의 영국도 일반정보보호규정 시행에 맞추어 자국 정보보호법을 개정하였는데, 이들도 익명화와 가명화를 예정하고 있다. 특히 영국에는 정보보호 책임기관인 ICO 등이 일찍부터 발전시킨 익명화 가이드라인이 익명화의 여러 측면을 고려한 판단기준과 판단의 틀을 제시하였고, 프랑스 정보보호 책임기관인 CNIL도 익명화와 가명화를 포함하는 정보 보안에 관한 접근 틀을 제안하였다. 핀란드는 개인정보의 활용과 공공데이터의 개방을 위하여 구체적인 익명화 및 가명화 가이드라인을 발표하였다.

유럽 이외의 지역을 보면, 미국에서는 NIST가 비식별 조치 기법에 관한 보고서를 계속 발간하고 있고, 특히 올해에는 IT 및 데이터 산업이 특히 발달한 캘리포니아(California) 주가 매우 포괄적이고 강력한 정보보호법을 입법하는 일도 있었다. 캐나다는 포괄적인 연방 법령 외에 특히 온타리오(Ontario) 주에서 상세한 비식별화 가이드라인을 발간하고

있다. 정보 활용이 너무 저조하다는 문제의식 하에 일본은 개인정보 보호법을 전면 개정하여 익명가공정보라는 새로운 개념을 도입하고 익명가공정보 작성에 있어 따라야 할 절차와 의무를 명확히 하는 한편, 익명가공정보의 작성에 관한 가이드라인도 발표하였다. 그 밖에 중국, 싱가포르, 호주 등에서도 입법과 함께, 특히 비식별 조치에 관한 최신의 가이드라인들이 발표되었다. 이들을 묶어 놓고 보면 오늘날 비식별 조치 가이드라인들이 전 세계적으로 서로 조응하면서 계속 발전하고 있음이 드러난다.

개인 간, 기업 간, 국가 간 경계를 넘나드는 연결이 일상화되어 있는 오늘날의 세계에서 개인정보의 보호와 이용에 관한 해외 동향은 우리의 문제를 해결하는 하나의 지침, 참고자료이자, 동시에 우리가 맞닥뜨리고 조정해야 할 상대 법제이기도 하다. 우리 개인정보 보호법제를 더욱 발전시켜 정보보호와 그 활용 모두를 진전시키기 위하여 해외의 경험과 전략, 그 성패를 살펴볼 필요가 있는 것이다.

제 2 절 연구의 목표 및 범위

이 보고서는 이러한 관점에서 해외 비식별 조치 가이드라인의 동향을, 그 배경이 되는 법과 함께 상세히 소개하고 분석하는 것을 주된 목적으로 한다. 특히 그 동안 여러 기관에서 이러한 연구보고가 부분적으로 이루어져왔다는 점을 고려하여 해외의 동향 중에서도 최근에 이루어진 진전에 초점을 맞추고자 한다. 정보통신기술이 일반적으로 그러한 것처럼 개인정보 비식별화, 익명화, 가명화 등에 관한 기술도 하루가 다르게 발전하고 있다. 비식별화의 활용과 그 구체적인 규율 방식, 체계 등은 해외에서도 여전히 논쟁 중인 문제로써, 현재의 논의수준을 파악하기 위해서는 늘 추적 연구가 필요하다. 유럽연합은 최근 개인정보 보호법제에 근본적인 변화를 주었고, 그와 관련하여 가명화 개념을 도입하기까지 하였다. 비식별화, 익명화와 가명화는 서로 다른 기능과 법적 효과를 가지나, 그 요건에 있어서는 기본적으로 정도의 차이가 있을 뿐이다. 그러므로 이러한 논의는 종래의 비식별화, 익명화에도 영향을 줄 수밖에 없다. 최근 일본의 법률 개정은 무엇보다도 우리와 비슷한 문제 상황을 타개하기 위한 것으로 익명화에 초점을 맞추고 있다. 이들을 포괄적으로 다루는 것이 연구의 목표이다.

이 보고서는 이러한 관점에서 다음과 같이 구성된다. 제2장에서는 국내에서 크게 다투어진 비식별정보/비식별화/비식별 조치/비식별 처리, 익명정보/익명화, 가명정보/가명화/가명처리의 개념을 규정하고 이들 사이의 관계를 정리한다. 제3장에서는 비식별 조치에 쓰이는 기술적 및 관리적 조치에 대하여 특히 국제표준화기구의 제안을 참조하여 소개, 정리하고, 국내·외에서 비식별정보를 활용한 사례를 소개하여 그 현실적 적용이 어디까지 진전하고 있는지 가늠할 수 있게 한다. 보고서의 중심인 제4장은 해외 비식별 조치 관련 법령과 가이드라인의 소개이다. 유럽에서는 유럽연합 전체 수준과 그 회원국인 독일, 프랑스, 핀란드, 영국의 국내 상황을, 북미에서는 미국, 캐나다를 연방과 주 수준에서,

아시아에서는 일본, 중국, 싱가포르, 호주를 각각 살펴본다. 각각에 대하여 개인정보 보호법제의 개요와 비식별화, 익명화, 가명화에 관한 법적 규율, 이들에 관한 가이드라인 등의 현황을 다룬다. 제5장은 위 내용과 그간 이루어진 각계의 비판, 2018년 2월과 4월에 있었던 제2차 해커톤의 논의를 참고하여 개인정보 보호법의 개정 방향과 2016년 「개인정보 비식별 조치 가이드라인」의 개선 방향에 관하여 간단히 살펴보기로 한다. 제6장은 결론으로 이상의 논의를 요약한다.

제 2 장 개념 정의 및 비교

제 1 절 개념정의

1. 비식별정보

우리 법질서에서 개인정보는 일반법인 「개인정보 보호법」과 특별법인 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「신용정보의 이용 및 보호에 관한 법률」, 「위치정보의 보호 및 이용 등에 관한 법률」에 의하여 보호된다. 이들 법률은 현대 정보처리장치의 활용과 관련하여 개인정보의 수집, 이용이 갖는 여러 위험을 통제하기 위한 것으로, 다분히 기술적으로 규정된 '개인정보'를 그 적용대상으로 한다.

'개인정보'는 '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)'로 정의된다. 개인정보 보호법 제2조 제1호가 그와 같이 규정하고 있고, 표현은 조금 다르나 정보통신망 이용촉진 및 정보호보 등에 관한 법률 제2조 제6호의 '개인정보'도 사실상 같다. 정보의 귀속주체(정보주체, 개인정보 보호법 제2조 제3호)를 '알아볼 수 있는' 정보를 말하는 것이다. 개념은 두 부분으로 나뉜다. 살아 있는 특정 정보주체에 관한 정보여야 하고, 동시에 정보주체를 알아볼 수 있는 내용도 포함되어야 한다. 그중 중요한 것은 뒷부분이다. 여기에서 다시 두 유형을 나누어볼 수 있다. 하나는 (일정한 제한된 상황에서) 이름, 주민등록번호, 여권번호 등 그 정보만으로 정보주체를 알아볼 수 있는 정보(강학상 이를 '식별정보'라 한다)이고, 다른 하나는 그 정보만 가지고 직접 정보주체를 알아볼 수는 없으나 다른 정보와 결합하면 정보주체를 알아

볼 수 있는 정보(강학상 이를 '식별가능정보'라 한다)이다. 어떤 목록에 진료를 받은 환자의 성별, 연령과 진료일, 임신 여부가 기재되어 있다 하여 임신한 10대 여성이 누구인지 바로 특정할 수는 없으나, 그 목록이 특정 산부인과의원의 환자 목록이고 누군가 그 환자 목록에 기재된 날 진료 받은 환자 모두를 알고 있다면 두 정보를 결합하여 임신한 10대 여성을 구체적으로 특정할 수 있는 것이 보통일 것이다. 이때 '임신하였다'는 정보주체에 관한 정보이고, 임신한 사람이 '○○○'이라는 점은 그 정보주체를 알아보는 것이며, 앞의 목록 자체에는 식별정보가 없다고 할 수 있으나 뒤의 정보와 결합하면 식별정보가 되고, 따라서 앞의 목록만으로도 식별가능정보가 될 여지가 있는 것이다.

이러한 개인정보 개념은 특별법상 개인정보의 기초이기도 하다. 가령 신용정보의 이용 및 보호에 관한 법률 제2조 제1호는 그 적용대상인 '신용정보'를 '특정 신용정보주체를 식별할 수 있는 정보'와 '정보주체의 거래내용, 신용도, 신용거래능력을 판단할 수 있는 정보'로 구체화하고, 위치정보의 보호 및 이용 등에 관한 법률도 적용대상인 '개인위치정보'를 '특정 개인의 위치정보(위치정보만으로는 특정 개인의 위치를 알 수 없는 경우에도 다른 정보와 용이하게 결합하여 특정 개인의 위치를 알 수 있는 것을 포함한다)'로 규정한다. 이들은 일반적 개인정보에 신용과 위치라는 한정을 붙인 것이다.

개인정보 보호법은 개인정보처리자가 개인정보를 처리함에 있어서는 처리목적에 필요한 범위에서 최소한의 정보만을 수집·처리하여야 하고(개인정보 보호법 제3조 제1항, 제2항), 원칙적으로 정보주체의 결정, 즉 필요한 설명을 듣고 한 사전 동의(informed consent)에 기초하여 처리하여야 한다고 정한다(개인정보 보호법 제4조 제1호, 제2호). 개인정보 보호법에는 일정한 사유가 있는 경우에 동의 없는 목적 외 이용·제공 등을 허용하는 규정들이 있으나(개인정보 보호법 제18조 제2항), 그러한 사유는 제한적이다. 나아가 '개인정보'에는 식별정보뿐 아니라 식별가능정보도 포함되어 있다. 그 때문에 현실적으로 수집·활용하는 정보 중 매우 많은 부분이 '개인정보'의 범위에 포함되고, 목적구속과

사전동의의 원칙의 적용을 받게 된다. 물론 처음부터 처리목적을 명확히 하고 사전동의를 받아 처리할 수 있지만, 이미 수집한 데이터, 제3자가 다른 목적으로 수집한 데이터를 이용하여 새로운 가치를 창출할 수 있는 많은 경우에 위 개인정보 보호법의 원칙들은 그러한 이용을 사실상 차단하는 셈이 된다.

비식별정보는 이러한 맥락에서 대두된 개념이다.

'개인정보'는 특정 개인에 관한 정보로서 정보주체를 알아볼 수 있는, 즉, '식별(識別)할 수 있는' 정보이다. 개인으로부터 정보를 수집하면서 처음부터 그 개인을 식별할 수 있는 정보는 전혀 수집하지 아니한다면 개인정보 보호법이 적용되지 아니한다는 데 별 의문이 없다. 그렇다면 수집할 당시에는 개인정보였지만 본래의 목적에 따른 이용이 이루어진 뒤에 일정한 가공을 거쳐 당해 정보로부터 식별가능성을 제거하였을 때에도 더는 개인정보가 아니게 되므로 목적구속 및 사전동의를 비롯한 개인정보 보호법의 원칙들의 제한을 받지 아니하고 이를 재이용할 수 있을 것이다. 이처럼 본래 개인정보였지만 이후 가공을 거쳐 식별가능성이 제거되어 더는 개인정보라고 할 수 없는 정보를 '비식별정보'라고 하고, 개인정보를 비식별정보로 가공하는 것을 '비식별화' 또는 '비식별조치'라고 한다.

한 가지 주의할 점이 있다. 세밀하게 살펴보면 비식별화 내지 비식별조치가 약간 의미를 달리하는 두 방식으로 쓰인다는 것이다. 어떤 정보에서 식별가능성을 제거하는 작업 자체와 그 결과 개인정보 보호법의 의미의 '개인정보'에서 제외되는 것이 그것이다. 전자를 기술적(記述的; descriptive)인 의미의 비식별정보, 비식별화, 비식별 조치라고 한다면 후자를 규범적(normative) 의미의 비식별정보, 비식별화, 비식별 조치라 할 수 있다. 전자의 의미의 비식별화가 이루어졌다 하여 당연히 후자의 의미의 비식별정보가 되는 것은 아니다. 식별성뿐 아니라 식별가능성도 제거하여야 하는데, 식별가능성이 없어졌는지 여부는 여러 사정을 고려하여 따로 평가할 일이기 때문이다.

비식별정보, 비식별화, 비식별 조치는, 적어도 우리나라에서는, 실정법

상의 용어는 아니다. 법률 수준에서 이 표현을 쓴 예는 아직까지 찾아 볼 수 없다. 이 용어는 주로 강학상 그리고 실무상 사용되어오다가 2016. 6. 30. 관계부처(국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부) 합동으로 발표한 「개인정보 비식별 조치 가이드라인 - 비식별 조치 기준 및 지원·관리체계 안내 -」(이하 '2016 비식별 조치 가이드라인'이라 한다)에서 정면으로 채택되었고, 법령 수준에서는 위 2016 비식별 조치 가이드라인에 터 잡아 제정된 국가정보자원관리원 훈령 제256호 「국가정보자원관리원 개인정보 비식별 조치 운영지침」(2018. 3. 12. 개정·시행, 비식별 조치 가이드라인에 터 잡은 훈령으로, 최초의 제정은 2016. 11. 14.이었는데, 그때에도 '비식별'이라는 용어를 사용하였다.)과 몇몇 지방자치단체(대구, 경북, 포항, 부산, 창원, 전주, 광주, 창원, 충남, 충북 등) 빅데이터 관련 조례의 몇몇 조문에 등장할 뿐이다. 그러나 '개인정보' 개념이 식별성·식별가능성에 기초하고 있는 이상 그 반면으로 비식별정보, 비식별화, 비식별 조치라는 용어법이 법적 근거가 없다거나 법령과 유리(遊離)되어 있다고 할 것은 아니다. 나아가 이 용어는 이미 하급심 재판례에서 채택된 용어이기도 하다. 우리 법원에서 비식별정보가 문제된 최초의 사건인 이른바 약학정보원 사건에 관하여 서울중앙지방법원 2017. 9. 11. 선고 2014가합508066, 538302 판결은 다음과 같이 설시하고 있다:

"개인정보는 해당 정보를 처리하는 자의 입장에서 특정 개인을 식별할 수 있는(identifiable) 정보이므로, 개인정보에 암호화 등 적절한 비식별화(de-identification) 조치를 취함으로써 특정 개인을 식별할 수 없는 상태에 이른다면 이는 식별성을 요건으로 하는 개인정보에 해당한다고 볼 수 없고, 따라서 정보주체의 동의 없이 통계작성 등의 용도로 이용되거나 제3자에게 제공되더라도 개인정보 보호법을 위반한 것이라고 볼 수 없다. 다만, 비식별화 조치가 이루어졌다고 하더라도 재식별 가능성이 현저하다면 적절한 비식별화 조치가 이루어지지 않은 것이므로 여전히 개인정보 보호법이 적용되는 개인정보에 해당한다고 할 것이고 적절한 비식별화 조치가 이루어진 것인지 여부는 원본 데이터의 특성, 비식별화된 정보가 사용된 특정한 맥락이나 상황, 비식별화 조치에 활용된 기법·세부

기술의 수준, 비식별화된 정보를 제공받은 자의 이용목적 및 방법, 이용기간, 전문지식이나 기술력·경제력에 따른 재식별화 능력, 비식별화된 정보를 제공받은 자가 재식별화로 얻을 수 있는 이익의 유무, 비식별화된 정보를 제공받은 자의 개인정보 보호 수준, 비식별화된 정보와 외부 정보 사이의 결합 가능성, 비식별화된 정보를 제공한 자와 제공받은 자의 관계, 비식별화된 정보에 대한 접근권한 관리 및 접근통제 등을 종합적으로 고려하여 판단해야 할 것이다."

이러한 설시는 물론 당사자의 주장에 대한 판단으로 쌍방 당사자가 '비식별화'라는 표현을 써 주장을 폈다는 사정과 전혀 무관하다고 할 수는 없다. 그러나 여전히 그 자체 법원의 판단의 일부라는 점도 부정할 수는 없는 것이다.

2. 익명정보

비식별화 이외에 '익명화', '익명정보'라는 표현도 쓰인다. 무엇보다도 대통령 직속 4 차산업혁명위원회에서 2018. 2. 1.~2. 2. 개최한 제2차 규제·제도혁신 해커톤에서 '개인정보의 보호와 활용의 균형 방안 마련'과 관련하여 토론결과 합의된 사항에 다음이 포함되어 있다:

"① 개인정보 관련 법적 개념체계 정비
개인정보와 관련된 법적 개념체계는 개인정보, 가명정보, 익명정보로 구분하여 정비하기로 하였다. 그리고 익명정보는 개인정보 보호법의 적용대상이 아니라고 합의하여 개인정보와 구분하기로 하였다.
② 익명정보 개념은 법에 명시하지 않음
'익명정보' 개념을 명확히 하기 위하여 '익명정보' 정의를 법에 명시하는 대신 EU GDPR 전문(26)을 참조하여 '개인정보'의 개념을 보완하기로 논의하였다."

위 합의사항에서도 드러나듯 '익명정보(匿名情報)'는 '개인정보' 개념에서 도출되는 개인정보 보호법의 적용제외 대상이다. 즉, 식별가능성이 제거된 정보를 말한다. 그러므로 이는 위 1.에서 본 비식별정보와 같은

대상을 가리키는 다른 용어에 불과하다. 또한 위 합의사항도 확인하듯 이는 법령상 용어가 아니고, 앞으로 법령상 용어가 될 것인지도 불분명하다. 결국 비식별과 익명은 용어 선택에 관한 두 대안인 셈이다.

3. 가명정보

반면 가명정보(假名情報)는 여전히 개인정보에 속할 수 있다. 위 제2차 규제·제도혁신 해커톤 합의사항도 개인정보와 관련된 법적 개념체계를 개인정보, 가명정보, 익명정보로 나누면서 익명정보가 개인정보 보호법의 적용대상이 아니라는 점만을 강조함으로써 간접적으로 이를 확인한다.

사전적(辭典的) 의미에서 가명(假名)은 본래의 이름이 아닌 다른 이름을 말한다. 다른 이름이 반드시 문자열이어야 하는 것은 아니다. 가령 숫자로 구성된 일련번호나 그림도 다른 이름이 될 수 있다. 그러므로 가명정보는 이름 기타 정보주체를 직접 특정하는 표지(이를 식별자 identifier 라고 한다)를 다른, 그 자체로 식별기능을 하지 못하는 기호 등으로 대체한 정보를 말한다. 예컨대 '김갑동'의 재산 보유 현황에 대한 정보를 기재해놓고 그중 '김갑동' 부분을 '479810AX-7'로 바꾸는 식이다. 개인정보를 가명화하면 정보주체를 직접 특정할 수 있는 정보가 삭제되어 그 정보만 보아서는 정보주체가 누구인지 알기 어려워진다.

그러나 어떤 정보가 가명화 되었다고 하여 곧바로 개인정보가 아니라고 할 수는 없다[GDPR Recital (26)은 이를 확인한다]. 식별정보뿐 아니라 식별가능정보도 개인정보에 해당하므로, 식별자를 다른 기호로 대체하여도 여전히 남아 있는 속성(attribute) 정보에 다른 정보를 결합하여 그 정보주체를 알아볼 수 있는 경우가 있다. 나아가 비식별정보 내지 익명정보가 당해 정보를 가공하여 식별가능성을 지속적으로, 즉 비가역적으로 제거하고자 하는 것임에 비하여 가명정보는 식별자를 대체한 '가명'의 부여방법('키'/'코드') 내지 가명과 실명을 매치(match)한

별도의 표가 있어 그러한 키/코드나 표를 갖고 있는 사람은 가명을 본래의 식별자로 복구할 수 있다. 결국 그러한 키나 표에 접근할 수 있는 사람은 언제나 가명정보로부터 정보주체를 식별해낼 가능성이 있어 가명정보도 여전히 개인정보가 된다.

반대로 가명정보가 늘 개인정보인 것도 아니다. 식별자를 제외하였을 때에는 정보량이 충분하지 아니하여 다른 정보와 결합하기 어렵고 그 자체 정보주체를 식별할 수 없으며, 당해 정보를 이용하거나 제공받는 사람이 키나 표에 접근할 가능성이 배제되어 있다면, 그러한 사람에 대하여는 가명정보도 비식별정보 내지 익명정보에 불과하고, 개인정보라고 할 수 없다.

그렇다면 가명정보 개념은 왜 필요한가.

가명정보 또한 우리 법이 쓰는 용어는 아니다. 그러나 그에 대응하는 개념 내지 제도는 이미 개인정보 보호법에 있다. 먼저 개인정보 보호법 제3조 제7항은 "개인정보처리자는 개인정보의 익명처리가 가능한 경우에는 익명에 의하여 처리될 수 있도록 하여야 한다"고 정한다. 이 규정은 '개인정보처리자'와 '개인정보'라고 하고 있으므로 이 규정에서 '익명처리'는 위 2.에서 언급한 의미의, 더는 개인정보가 아닌 익명정보일 수 없다. 이때 익명처리는 그 표현 그대로 이름 등을 지워야 한다는 뜻에 그치고, 그 결과 개인정보 보호법적 의미의 식별가능성이 제거되어야 한다는 뜻은 포함하지 않는 것으로 보아야 할 것이다. 이는 가명화에 다름 아니다. 다음 같은 법 제18조 제2항 본문은 "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"면서 그 제4호에서 "통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우"를 들고 있다. 여기에서 "특정 개인을 알아볼 수 없는 형태"가 비식별정보 내지 익명정보를 의미한다면 이 규정은 비식별화, 익명화를 하여 더는 개인정보가 아닌 때에도 목적 외 이용 및 제3자 제공을 위해

서는 통계목적 등 일정한 목적까지 갖추어야 한다는 취지가 된다. 그러한 해석은 지금까지 국내·외에서 이루어진 비식별화를 둘러싼 논쟁의 기본전제를 엎는 것으로 입법자의 의도라고 보기 어렵다. 따라서 여기에서 "특정 개인을 알아볼 수 없는 형태"는 개인정보의 식별가능성과 달리 다른 정보와의 결합을 전제하지 아니한 채 그 자체로 정보주체가 식별될 수 없는 상태로 만드는 것, 즉 식별자를 제거하는 것을 가리킨다고 해석하는 것이 합리적이다(이동진, 2016:262). 이는 가명화이다. 그밖에 생명윤리 및 안전에 관한 법률 제18조 제2항, 제38조 제2항도 그러한 예에 속한다. 같은 규정은 인간대상연구 및 인체유래물연구에서 대상자가 "개인식별정보를 포함하는 것에 동의"하지 아니한 이상 연구목적으로 수집한 정보를 제3자에게 제공할 때에는 "익명화"하여 제공하여야 한다고 하는데, 이때 "익명화"가 "개인식별정보", 즉 식별자를 제외한 채 제공하는 조치로 가명정보를 의미함은 전후 맥락상 분명하다.

이상의 논의로부터 가명정보가 문제되는 까닭도 간취할 수 있다. 개인정보 보호법은 오늘날 발달된 정보처리기술과 대규모로 집적된 데이터의 연결가능성을 고려하여 개인정보의 범위를 다소 넓게 파악하였다. 그러나 개인정보에 해당한다 하더라도 식별정보와 식별가능정보의 식별가능성에는 큰 차이가 있고, 식별가능정보 안에서도 구체적 식별가능성의 다과(多寡)에 차이가 있게 마련이다. 수집 및 처리 목적상 개인정보 보호법적 의미의 개인정보를 처리하는 것은 불가피하다 하더라도 가급적 식별자를 제거하면 구체적인 식별은 더 어려워지고, 따라서 더 잘 보호될 수 있다. 그러므로 굳이 이름 기타 식별자를 그대로 써야 하는 경우가 아닌 한 식별자를 가명으로 대체하고 추후 필요할 때를 위하여 키/코드 등은 별도로 보관하며 식별자가 포함되지 아니한 형태로 이용하게 하는 것은 본래의 이용 목적을 위한 처리과정에서 또는 목적 외 이용과 제3자 제공과정에서 오·남용될 위험과 해킹 등에 의하여 유출될 위험 모두를 감소시키는 데 기여하게 된다. 이것이 개인정보처리자에게는 기술적·관리적 조치의 일환이 되고 제3자 제공 등을 할 때에는 안전장치의 일환이 될 수 있는 까닭이다(위 서울중앙지방법원 판결이 보여

주듯 이러한 조치는 비재산적 손해를 부정하는 근거가 될 수도 있다).

그러므로 가명정보와 가명화, 가명처리 개념은 개인정보의 최소 수집·처리와 개인정보로 인한 위험을 통제하기 위한 기술적·관리적 조치 및 목적 외 이용과 제3자 제공을 위한 안전장치의 한 전형적인 방법으로 별도로 논의할 필요가 생긴 것이다.

제 2 절 비교 및 정리

비식별정보, 익명정보, 가명정보의 비교, 정리와 관련하여서는 다음 두 가지 점을 다룬다:

첫째, 비식별정보와 익명정보 개념의 이동(異同) 및 선택,
둘째, 비식별정보 내지 익명정보와 가명정보 사이의 경계.

먼저 비식별정보와 익명정보 개념의 이동(異同) 및 선택을 본다. 두 개념의 이동(異同)에 관하여는 크게 둘, 세분하면 세 가지 입장이 있다. 양자를 동의어로 쓰는 경우, 비식별화에 비하여 재식별이 어려운 경우를 익명화라고 하는 경우, 그리고 후자의 극단적인 예로 재식별이 절대적으로 불가능한 비식별화만을 익명화라고 하는 경우가 그것이다. 두 개념 사이의 선택에 관하여는 비식별화라는, 기존에 어느 정도 정착된 용어를 선호하는 입장과 비식별화라는 용어에 비판적인 태도를 취하고 익명화를 선호하는 입장[대표적으로 이은우, 2015]이 있다.

비식별화는 de-identification의, 익명화(匿名化)는 anonymisation의 역어(譯語)이다. 전자는 주로 미국에서 사용하는, 후자는 주로 유럽 및 그 영향을 받은 일본 등에서 사용하는 용어이다. 가령 미국의 건강정보 보호에 관한 법령인 HIPAA Privacy Rule 164.514 (45 CFR 164.514) (a)는 "표준: 보호되는 건강정보의 비식별화 (Standard: De-identification of protected health information)"라는 표제 하에 "개인을 식별하지 아니하고 그 정보가 개인을 식별하는 데 쓰이리라고 믿을 만한 합리적 근거가 없는 건강정보는 개인을 식별할 수 있는 건강정보가 아니"라고 규정하는 반면, 유럽연합 1995년 개인정보 보호지침 전문(26) 후단은 "보호 원칙들은 정보주체가 더는 특정가능하지 아니하도록 익명화된 정보에는 적용되지 아니하고; 제27조의 행동강령이 정보주체의 특정이 더는 불가능한 형태로 정보를 익명화하는 방법에 관한

지침을 제공하는 유용한 도구가 될 수 있[다]"고 정한다. 때문에 법제가 다른 미국의 용어를 사용한 것은 잘못이라는 비판도 있다(이은우, 2015).

개인정보 보호에 관하여 미국과 유럽의 법제가 상당한 차이를 보이고, (법제도의 집행체계나 다양한 관련된 법령 사이의 관계 등에 관한 복잡한 문제를 제외하고) 우리 개인정보 보호법의 내용 위주로 보면 개인정보 보호법이 상대적으로 유럽 법제와 가까운 것은 사실이다. 그러나 양 법제의 차이는 특히 규제가 필요한 병원, 금융기관, 학교, 유선방송가입내역, 비디오대여기록, 자동차운전면허기록 등을 특정하여 영역별 입법으로 대응하고 그 이외의 영역은 원칙적으로 일반 법리에 맡길 것인가, 아니면 포괄적 개인정보 보호 입법을 할 것인가, 그리고 그 결과 개별 영역별로 개인정보를 보호하는 방식(가령 동의원칙에 관한 opt-in과 opt-out 사이의 선택, 법적 규제와 시장에 의한 규제)에 편차가 생기는가 하는 점에 있을 뿐이다. 미국과 유럽 모두 개인정보를 '개인을 식별할 수 있는(identifiable)' 정보인지 여부를 주요 기준으로 하여 정의하고, 다른 정보와의 결합가능성을 고려하는 이상, 그러한 식별가능성을 제거하는 작업으로서 비식별화와 익명화의 의미와 기능에는 차이가 없다. 즉, 비식별화나 익명화의 개념구분에 관한 한, 우리 법제가 유럽 법제와 가까운지 여부는 중요하지 않다.

그럼에도 불구하고 비식별화와 익명화를 구분하면서 후자의 용어법을 채택하여야 한다고 주장하는 이는 대체로 다음 둘 중 하나 또는 둘 다에 기대고 있다. 비식별화를 기술적(記述的) 개념으로 제한하여 이해하거나, 앞서 본 바와 같이, 비식별화보다 익명화의 기준이 더 엄격하다고 바라보는 관점이다. 차례로 본다.

먼저 익명화를 채택하여야 한다는 또는 익명화만이 허용된다는 주장의 중요한 근거는 비식별 조치를 취하더라도 재식별(re-identification) 가능성이 있으면 개인정보가 된다는 점이다(이은우, 2015). 여기에서 비식별 조치는 식별가능성이 제거된 결과 내지 규범적 평가가 아니라 식별가능성을 제거하기 위하여 한 조치를 기술(記述) 내지 묘사하는 단어

로 쓰이고 있다. 이러한 용어법은 다른 나라에서도 보인다. 가령 유럽연합 1995년 개인정보 보호지침 제29조와 관련하여 제28조 작업반(Article 29 Working Party)이 2014년 발간한 '익명화 기술에 대한 의견서'는 익명화를 "비가역적으로 비식별화하기 위하여 개인정보에 적용된 기법"이라고 설명하고(WP 29, 2014:7), 미국 국립표준기술연구소(National Institute of Standards and Technology; NIST)가 2010년 발간한 '개인정보의 비밀유지를 위한 안내'는 '익명화'를, '이전에는 식별 가능하였던 정보로서 비식별화 처리가 되고, 재식별을 위한 코드 기타 관련요소가 더는 존재하지 아니하는 경우'로 규정한다(NIST, 2010:4).

그러나 이러한 용어법이 일반적으로 인정된 것이라고 할 수는 없다. 위 문건들도 그러한 용어법이 타당함을 전제하여 위와 같이 설명하였다기보다는 개인정보에서 식별가능성을 제거하여 개인정보 보호법제의 적용범위 밖으로 옮기는 작업과 그 결과를 지칭하는 용어가 필요하고, 경우에 따라서는 전자의 작업은 이루어졌으나 후자의 결과에는 이르지 못한 것으로 평가되는 일도 있을 수 있으므로 이를 지칭하기 위하여 두 용어를 달리 사용하였을 뿐인 것이라고 보인다. 미국과 우리나라에서 비식별정보를 이용할 수 있다고 할 때 비식별정보는 당연히 결과로서 식별가능성을 제거한 정보를 가리킨다. 두 용어는 별 구분 없이 논자에 따라 혼용되는 경향이 있다. 선험적으로 또는 관용적으로 어느 한 용어를 선호할 별도의 근거도 없다.

두 용어는 모두 매우 기술적(技術的)인 전문용어이다. 일상어로서의 적합성을 따진다면 오히려 익명화가 의미를 호도할 위험이 크다. 익명(匿名)은 이름 기타 신원을 알리지 아니하는 것을 가리키므로 식별자를 모두 삭제하면 되고 나아가 나머지 속성정보에 의한 식별가능성까지 따질 필요는 없다고 읽힐 수 있기 때문이다. 실제로 '비식별'과 달리 '익명'은 우리 법에서 제법 쓰이고 있는데, 대부분 위와 같은 의미의, 즉 자신의 이름을 알리지 아니한 채 하는 기부나 (공익목적) 신고 등에, '가명'과 함께, 붙는다(가령 2018 평창 동계올림픽대회 및 동계패럴림픽

대회 지원 등에 관한 특별법 시행령 제2조의2, 정치자금법 제11조, 후천성면역결핍증 예방법 제8조 등). 개인정보 보호법과 생명윤리 및 안전에 관한 법률도 '익명처리'와 '익명화'를 가명화의 의미로 쓴다.

비식별화보다 익명화를 선호하는 입장의 또 다른 근거 내지 전제는 비식별화는 어느 정도 재식별가능성을 남기는 것을 전제하나 익명화는 그렇지 아니하다는 것이다(이은우, 2015). 그러나 이 또한 근거가 없다. 재식별가능성은 비식별화 내지 익명화가 식별가능성이 있었던 개인정보로부터 식별가능성을 제거하였다는 이유에서 붙인 이름일 뿐 식별가능성에 다름 아니다. 개인정보의 요건으로서 정보주체의 식별가능성에 대하여는 논란이 있지만, 누구에 대하여도 계속적 내지 영구적으로 식별이 불가능할 정도에 이르러야 한다는 입장(절대설)은 오늘날 현실성이 없어 거의 일치하여 거부되고 있다. 법문언도 개인정보의 요건으로서 식별가능성과 관련하여 '다른 정보와 쉽게 결합하여 알아볼 수 있'을 것을 요구함으로써 다른 정보와 결합하여 알아보는 것이 가능하여도 그것이 '어려울' 때에는 개인정보가 아님을 분명히 한다. (재)식별가능성이 절대적으로 배제되는 경우를 익명화, 익명정보라고 한다면, 정보로서의 가치를 보존하면서 그러한 조치를 취하는 것이 도대체 가능한지, 그러한 정보가 존재하는지도 문제되지만, 그러한 정보가 존재한다 하더라도 개인정보 보호법상으로는 독자적 의미가 전혀 없다. 개인정보 보호법적으로는 '쉽게 결합하여 알아볼 수 있'는지 여부만이 문제되기 때문이다.

결국 전문적 또는 기술적 용어로서 비식별화와 익명화 사이의 선택은 어차피 용어의 문제에 불과하고, 법령상 근거가 없는 상황에서는 더욱 그러하다.

다음 비식별정보 내지 익명정보와 가명정보 사이의 경계를 본다. 근래 반복하여 지적되는 바와 같이 절대적인 비식별화, 익명화가 많은 경우 불가능하거나 현실적이지 아니하고, '쉽게 결합하여 알아볼 수 있'는지 여부가 기준이 되어야 한다면, 개인정보와 비식별정보 내지 익명정보의 경계는 다소간 유동적일 수밖에 없다. 한편으로는 '쉽게 결합하여 알아볼 수 있'는지 여부를 따질 때 누구의 입장에서 어떠한 사정, 가령 어떠

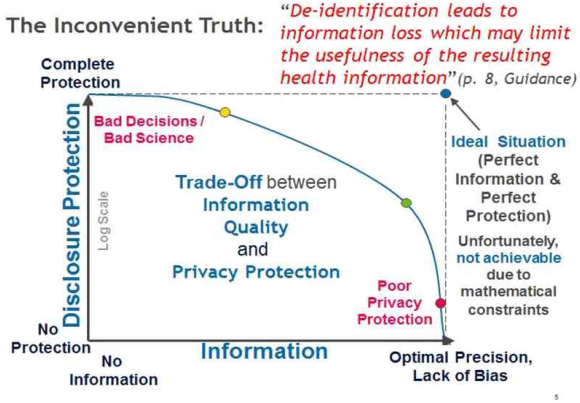
한 결합정보를 전제하여야 하는지가 문제되고, 다른 한편으로는 시간이 지남에 따라 입수할 수 있는 결합정보가 증가할 뿐 아니라 결합 내지 재식별기술도 발전하게 된다는 점을 고려하여야 하기 때문이다. 어떤 정보가 그 정보를 보유하고 있는 관리자의 입장에서는, 그 자신이 입수할 수 있는 정보와 잠재적 공격자가 입수할 수 있는 정보, 이들 각각의 재식별동기와 재식별능력에 비추어볼 때 충분히 비식별화 되었다고 할 수 있으나, 일단 다른 사람의 손에 넘어간 뒤에는 그가 갖고 있는 결합정보와 재식별동기, 재식별능력에 비추어 개인정보가 된다든지, 당초 비식별 조치 내지 익명처리를 하였을 때에는 그 수준으로 충분하여 더는 개인정보가 아니고 비식별정보, 익명정보였으나 그 뒤에 데이터의 축적과 기술발전으로 재식별가능성이 증가하여 다시 개인정보가 되는 일이 생길 수 있다.

이러한 특징은 비식별정보, 익명정보와 가명정보의 구별과 관련하여서도 의미가 있다. 어느 한 정보가 가명정보라 하더라도 일반적으로 입수할 수 있는 결합정보를 활용하여서는 정보주체를 식별할 수 없고 가명처리에 쓰인 '키'/'코드', 표 등을 이용하여 식별자를 복구하여야 비로소 식별가능성이 인정되는 경우가 있다. 이때에도 그 '키'나 표에 접근할 권한이 있는 한 그 정보처리자에게 당해 정보가 개인정보임은 부정할 수 없다. 그러나 그 '키'나 표에 접근할 수 없는 제3자에게 당해 정보는 비식별정보 내지 익명정보가 될 수도 있다.

제 3 장 비식별 조치의 이론적 배경 및 활용사례

제 1 절 비식별 조치의 이론적 배경

비식별 조치 기술은 데이터의 사용·저장·공유 과정에서 정보주체를 식별하지 못하게 처리하는 일련의 방법들을 의미한다. 비식별 조치 기술을 통하여 만들어지는 데이터는 비식별 조치의 강도가 높아질수록 보호의 정도가 높아진다. 그러나 그 대신 데이터의 활용도는 일반적으로 낮아지는 특징을 가짐을 명심하여야 한다. 즉, 개인정보를 완벽하게 보호하면서 정보의 활용도 또한 무한정 보장하는 기술은 존재할 수 없다.



(그림3-1) 비식별 조치 강도에 따른 개인정보보호와 데이터 활용성의 상관관계
출처: Twitter
https://twitter.com/dbarthjones/status/681572627455029248

비식별 조치와 관련하여 국제표준화 기구, 특히 ISO(International Organization for Standardization)에서 발표한 국제표준으로는 다음이 있다:

▹ 의료정보 관련 표준을 제정하는 ISO TC/215 Health informatics에서 2017년에 개정한 ISO 25237 Health informatics - Pseudonymization 표준; 그리고

▹ 정보보안 관련 국제표준을 제정하는 ISO/IEC JTC 1/SC 27에 개발하고 있는 ISO/IEC 20889 Information technology - Security techniques - Privacy enhancing data de-identification terminology and classification of techniques 표준.

3대 국제표준화 기구인 ISO, IEC (International Electrotechnical Commission), ITU (International Telecommunication Union) 중에서 ISO와 IEC가 공동으로 출판한 ISO/IEC 20889는 IT 전반에 관한 내용을 포함하고 있으므로, 여기에서는 해당 문건에 주로 기초하여 설명하고자 한다.

ISO/IEC 20889는 비식별 조치 절차(de-identification process)를 "일련의 식별 속성과 데이터 주체 사이의 연관성을 제거하는 절차"로 정의한다. 제2장의 개념으로 설명하자면 "정보주체와 해당 정보주체를 알아볼 수 있는 식별정보 및 식별가능정보들 사이의 연결성(식별성)을 제거하는 절차"라고 할 수 있다. 또한, 재식별화(re-identification)는 "비식별 조치된 데이터 집합을 원본 데이터 주체와 연관시키는 과정"으로 정의한다. 즉, "비식별 조치된 데이터를 다시 정보 주체와 연결(Linking)시키는 절차"이다. 20889 표준 문서에서는 식별자(identifier)를 기술적으로 상세히 구분하고 있는데, 그림 3-2에 상세히 설명되어 있다. 식별자는 크게 직접식별자(direct identifier)와 간접식별자(indirect identifier)로 나뉜다. 직접식별자는 "데이터 주체를 고유하게

식별할 수 있도록 해주는 속성", 간접식별자는 "데이터 집합에 포함되어 있거나 외부에 속한 속성과 함께 특정 운영 환경에서 데이터 주체의 고유 식별을 가능하게 하는 속성"을 의미한다. 우리나라 개인정보 보호법의 정의에 따르면 직접식별자는 "개인을 알아볼 수 있는 정보"이고, 간접식별자는 "해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것"이라고 할 수 있다. 이 분류를 좀 더 상세히 구분하여, 주어진 데이터 집합에 속해있는 직접식별자는 "고유식별자(unique identifier)"라고 별도로 정의하고, 간접식별자 중에서 데이터 집합에 속해있는 것들은 "준식별자(quasi-identifier)"라고 정의하였다.

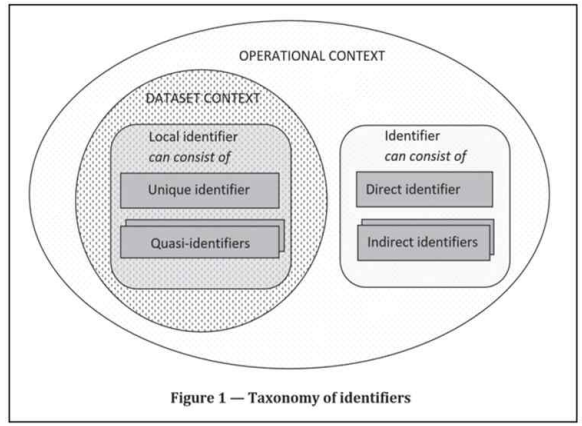


Figure 1 — Types of identifier

(그림3-2) 식별자의 종류

출처: ISO/IEC 20889

1. 기술적 방법론

(1) 비식별 조치의 분류

비식별 조치는 크게 무작위화(randomization)와 일반화(generalization)로 나뉜다. 무작위화는 데이터의 일부를 임의의 값으로 변경하거나[가명화(psuedonymization)를 포함한다] 임의의 오차를 추가하여 특정 개인을 식별할 수 있는 연결성을 제거하는 방식이다. 예를 들면, 환자 번호를 무의미한 일련번호로 대체하거나, “김철수, 경기 거주, 대한대 재학”이라는 정보를 “홍길동, 경기 거주, 민국대 재학”과 같이 이름과 학교명을 변경하는 것을 포함한다. 일반화는 데이터의 값을 상위 범위의 값으로 범주화하여 특정 개인을 식별할 수 없도록 하는 방식을 의미한다. 나이를 20대, 30대, 40대 등으로 변경하는 것이 대표적이다.

(2) 비식별 조치 기술 소개

무작위화와 일반화를 구현하는 기술들에 따라 구분하면 통계적 방법(샘플링, 총계처리 등)과 암호화 방법으로 나눌 수 있다. 다만, 암호화 방법은 비식별 조치 기술의 효율 향상(보완책) 혹은 도구의 일부로 사용이 가능하다. 즉, 일반적인 암호화는 그 자체만으로는 통상적으로 비식별화를 의미하지 아니한다는 점을 명심하여야 한다. 실제 데이터를 비식별 조치를 할 때에는 데이터의 특성과 향후 추가될 것으로 예상되는 데이터 등을 고려하여 두 가지 방법을 적절히 조합해서 진행하여야 한다.

1) 비식별 조치 가이드라인에 소개된 비식별 조치 방법

먼저 2016년에 정부부처 합동으로 발표된 비식별 조치 가이드라인에 소개되어 있는 비식별 조치 방법들을 개념과 대상, 장·단점을 비교 정리하면 표 3-1과 같다.

[표3-1] 비식별 조치 방법 비교

기법	개념	대상	장점	단점
가명처리	개인 식별이 가능한 데이터를 직접적으로 식별할 수 없는 다른 값으로 대체	성명, 기타 고유특징 (출신학교, 근무처 등)	데이터의 변형 또는 변질 수준이 적음	대체 값 부여 시에도 식별 가능한 고유 속성이 계속 유지
총계처리	통계값(전체 혹은 부분)을 적용하여 특정 개인을 식별할 수 없도록 함	개인과 직접 관련된 날짜 정보(생일, 자격취득일), 기타 고유특징(신체정보, 진료기록, 병력정보, 특정소비기록 등 민감한 정보)	민감한 수치정보에 대하여 비식별 조치가 가능하며, 통계 분석용 데이터 셋 작성에 유리함	정밀 분석이 어려우며, 집계 수량이 적을 경우 추론에 의한 식별 가능성 있음
데이터 삭제	개인 식별이 가능한 데이터 삭제 처리	개인을 식별 할 수 있는 정보(이름, 전화번호, 주소, 생년월일, 사진. 고유식별정보, 생체정보, 등)	개인 식별요소의 전부 및 일부 삭제처리가 가능	분석의 다양성과 분석결과의 유효성·신뢰성 저하
데이터 범주화	특정 정보를 해당 그룹의 대푯값으로 변환하거나 구간값으로 변환(범주화)하여 개인식별을 방지	개인을 식별할 수 있는 정보(주소, 생년월일, 고유식별정보, 기관·단체 등의 이용자 계정	통계형 데이터 형식이므로 다양한 분석 및 가공 가능	정확한 분석결과 도출이 어려우며, 데이터 범위 구간이 좁혀질 경우 추론 가능성 있음

데이터 마스킹	데이터 전부 또는 일부분을 공백, 노이즈 등으로 변환	상동	개인 식별 요소를 제거하는 것이 가능하며, 원 데이터 구조에 대한 변형이 적음	마스킹을 과도하게 적용할 경우 데이터 필요 목적에 활용하기 어려우며 마스킹 수준이 낮을 경우 특정한 값에 대한 추론 가능

자료 : 개인정보 비식별 조치 가이드라인 (연구진 재구성)

위 비식별 조치 기법들을 적용한 사례 역시 가이드라인에 잘 설명되어 있는데 대표적인 사례는 그림 3-3과 같다.

● 〈 예시 〉 비식별 조치 기법 적용 ●

원본데이터

주민등록번호	성별	입원날짜	연령	병명
770914-1234567	남	2015/06/23	39	독감
850930-1234567	남	2015/10/01	31	독감
710119-2345678	여	2014/01/21	45	고혈압
770619-2345678	여	2014/09/23	39	고혈압
830425-1234567	남	2015/04/16	33	간염
860804-2345678	여	2014/11/11	30	간염

비식별 데이터

① 데이터 삭제(주민등록번호)

~~주민등록번호~~	성별	입원날짜	연령	병명
	남	2015/06/23	39	독감
	남	2015/10/01	31	독감
	여	2014/01/21	45	고혈압
	여	2014/09/23	39	고혈압
	남	2015/04/16	33	간염
	여	2014/11/11	30	간염

② 데이터 마스킹(주민등록번호, 입원날짜), 총계처리(평균 연령)

주민등록번호	성별	입원날짜	연령	병명
7*****-1******	남	2015/**/**	35	독감
8*****-1******	남	2015/**/**	35	독감
7*****-2******	여	2014/**/**	35	고혈압
7*****-2******	여	2014/**/**	35	고혈압
8*****-1******	남	2015/**/**	35	간염
8*****-2******	여	2015/**/**	35	간염

(그림3-3) 비식별 조치 기법적용 결과
출처: 개인정보 비식별 조치 가이드라인

2) ISO/IEC 20899에 소개된 비식별 조치 기법

앞에서 소개한 ISO/IEC 20899에서 소개하고 있는 비식별 조치 기법을 요약하면 아래 그림 3-4와 같다. 기본적인 사항은 동일하지만, 동형 암호화(homomorphic encryption)와 차분 프라이버시(differential privacy) 등의 고급기법이 소개되어 있다. 동형 암호화는 무작위 암호화 기법 중 하나로, 값을 해독하지 않고 암호된 값의 연산을 지원하여 암호화된 상태로 데이터 분석을 가능하게 하는 기법이고, 차분 프라이버시는 통계 분석 출력의 확률분포가 지정된 특정값을 초과하지 않도록 보장하여 개인정보를 보호하는 기법이다.

Table A.1 — Properties of de-identification tools, techniques and models

Technique Name	Data truthfulness at record level	Applicable to types of values	Applicable to types of attributes	Reduces the risk of		
				Singling out	Linking	Inference
Statistical tools						
Sampling						
Aggregation	N.A.	Continuous, discrete	All attributes	Yes	Yes	Yes
Cryptographic tools	Yes					
Deterministic encryption	Yes	All	All attributes	No	Partially	No
Order-preserving encryption	Yes	All	All attributes	No	Partially	No
Homomorphic encryption	Yes	All	All attributes	No	No	No
Homomorphic secret sharing	Yes	All	All attributes	No	No	No
Suppression	Yes					
Masking	Yes	Categorical	Local identifiers	Yes	Partially	No
Local suppression	Yes	Categorical	Identifying attributes	Partially	Partially	Partially
Record suppression	Yes	N.A.	N.A.	Partially	Partially	Partially
Sampling	Yes	N.A.	N.A.	Partially[a]	Partially	Partially
Pseudonymization	Yes	Categorical	Direct identifiers	No	Partially	No
Generalization	Yes	All, subject to meaning	Identifying attributes			
Rounding	Yes	Continuous	Identifying attributes	No	Partially	Partially
Top/bottom coding	Yes	Continuous, ordinal	Identifying attributes	No	Partially	Partially
Randomization	No		Identifying attributes			
Noise Addition	No	Continuous	Identifying attributes	Partially	Partially	Partially
Permutation	No	All	Identifying attributes	Partially	Partially	Partially
Micro aggregation	No	Continuous	Indirect identifiers, and all attributes	No	Partially	Partially
Differential privacy	No	All	Identifying attributes	Yes	Yes	Partially
K-anonymity	Yes[b]	All	Quasi identifiers	Yes	Partially	No

[a] If the data principal record is not included in the sample.

[b] Unless *K*-anonymity is implemented using microaggregation.

(그림3-4) 비식별 조치 기법 소개
출처: ISO/IEC 20889

(3) 기밀성 모델(confidentiality model)

ISO/IEC 20899에서는 단순히 비식별 조치 기법만을 소개하는 것이 아니라, 해당 기법들의 재식별 위험도를 계산하는 방법인 기밀성 모델(confidentiality model)도 소개하고 있다.

기본적으로 비식별 조치 기술은 재식별 위험을 줄이도록 개발되었다. 따라서 적용한 비식별 조치 및 해당 데이터 집합을 고려한 재식별 위험도를 정량화하기 위해서 여러 가지 기밀성 모델이 고안되었다. 기밀성 모델을 이용하여 재식별 위험도를 계산하려면 기본적으로 재식별 공격(re-identification attack)을 수행하는 시나리오를 전제하여야 한다.

1) 재식별 공격(re-identification attack)

재식별 공격이란 비식별 조치된 데이터를 대상으로 특정 개인을 알아내기 위해 수행하는 행위로 크게 다섯 가지로 구분된다(표 3-2). 크게 분류하면 데이터 주체가 누구인지 알아내는 경우(Prosecutor attack, Journalist attack)와 누구인지는 모르나 해당 비식별 조치 데이터 안에 포함되어 있는지를 알아내는 경우(Marketer attack, Data membership attack)등이 있다. Inference attack은 누구인지 알아내거나, 포함여부를 알아내는 과정 모두에서 추론 과정이 들어가는 경우라고 할 수 있다.

[표3-2] 재식별 공격 분류

재식별 공격	특징
Prosecutor attack	기존 지식을 사용하여 특정 데이터 주체에 속한 데이터를 재식별 ex) 유명인이나 친구, 친척 등을 찾아내는 것
Journalist attack	기존 지식을 사용하여 특정 데이터의 데이터 주체를 재식별 ex) 공개된 DB를 활용 (US voting registry 활용 등)
Marketer attack	기존 지식을 사용하여 가능한 많은 레코드에 상응하는 주체에 관하여 재식별 ex) 고객군 분류 등
Data membership attack	데이터 세트에서 특정 데이터 주체의 존재를 확인 ex) 특정 개인이 해당 데이터 집합에 있는지를 확인
Inference attack	다른 속성 그룹과 연관된 민감한 속성을 추론 ex) 서로 다른 데이터 집합 연계를 위한 특정 값 추론

자료 : 연구진 작성(ISO/IEC 20889 재구성)

2) 재식별 상황 분류

또한, 재식별 공격을 통해 재식별되는 경우를 표 3-3과 같이 두 가지로 구분한다. Exact disclosure 과정에서 공격자가 통계를 사용하지 않고 비식별 조치된 데이터 집합을 원본 데이터 주체와 바로 연관시키는 것을 deterministic re-identification이라고도 한다.

[표3-3] 재식별 상황 분류

재식별 상황	설명
Exact disclosure	침입자가 데이터 주체의 속성값을 정확하게 결정할 때 발생
Statistical disclosure	공격자가 합계 데이터(aggregated data)가 존재하는 경우에 속성값을 더욱 정확히 예상 가능할 때 발생

자료 : 연구진 작성(ISO/IEC 20889 재구성)

더 나아가, 재식별을 통해 식별된 데이터들의 구분은 표 3-4와 같다. 예를 들어 Single out은 특정 개인을 명확히 알아내는 경우이고, Linking은 누구인지 알 수는 없으나 서로 다른 데이터 그룹에 동일한 데이터 주체가 있을 경우이다. Inference는 주로 속성값을 예측할 때에 쓴다. 예를 들어, 환자 데이터를 보고 다른 값들을 통해 특정 환자의 질환을 예측하여, 해당 환자가 누구인지 식별해 내는 것이다.

[표3-4] 재식별 된 데이터 분류

재식별 데이터 분류	설명
Single out	해당 데이터 주체를 고유 식별하는 특성 집합을 관찰하여, 데이터 세트 내부의 데이터 주체에 속한 데이터의 일부 또는 전부를 격리
Linking	동일한 데이터 주체 혹은 데이터 주체 그룹과 관련된 데이터를 별도의 데이터 세트에 연결
Inference	무시할 수 없는 확률로 다른 속성 집합의 값을 추론. 다만 항상 재식별 공격이 아닌 데이터 분석의 side-effect로 발생할 수 있음 (ex. 의료 데이터 분석 등)

자료 : 연구진 작성(ISO/IEC 20889 재구성)

재식별 공격 기법과 연계하여 설명하면, Prosecutor attack과 Journalist attack은 Exact disclosure를 찾는 방법으로, 배경지식 혹은 추가정보를 이용하여 Linking을 하여 Single out을 하는 것이다.

3) 기밀성 모델 종류

2016년 정부부처 합동으로 발표된 개인정보 비식별 조치 가이드라인에 소개된 대표적인 기밀성 모델로는 k-익명성(k-anonymity), l-다양성(l-diversity), t-근접성(t-closeness) 등이 있고, k-익명성 모델의 이용을 권장하고 있다. 세 가지 모델을 합쳐서 KLT라고 통칭하기도 한다.

그런데, KLT는 기본적인 기밀성 모델로 한계가 있다. 따라서 이를 보완하기 위한 여러 다양한 기밀성 모델들이 제안되어 있다. 표 3-5에 KLT를 포함하여 대표적인 기밀성 모델을 소개하였다. 최근의 추세는 단순히 k값으로 기밀성을 보장하는 것이 아니라, 전체 데이터의 크기와 분포를 고려하여 재식별 위험도를 계산하는 형식으로 진화하고 있다.

[표3-5] 기밀성 모델

기밀성 모델	특징
k-익명성	주어진 데이터 집합에서 같은 값이 적어도 k개 이상 존재하도록 하여 쉽게 다른 정보와 결합할 수 없도록 함
l-다양성	k-익명성의 취약점을 보완한 모델로, 주어진 데이터 집합에서 함께 비식별되는 데이터들은 (동질 집합에서) 적어도 l개 이상의 서로 다른 민감한 정보를 가지도록 함. 즉, 충분히 다양한(l개 이상) 서로 다른 민감한 정보를 가지도록 구성. 이로 인해 정보가 충분한 다양성을 가지므로 다양성 부족으로 인한 공격에 방어가 가능하고, 배경지식으로 인한 공격에도 일정 수준의 방어 능력을 가지게 된다.
t-근접성	l-다양성의 취약점(쏠림 공격, 유사성 공격 등)을 보완하기 위한 모델로, 동질 집합에서 특정 정보의 분포와 전체 데이터 집합에서 정보의 분포가 t 이하의 차이를 보이도록 함
Differential privacy	전체 데이터 집합에서 특정 데이터를 추출할 때 사전에 정의한 확률 분포에 따라 추출되는 데이터에 임의의 값을 추가하여 원래 데이터를 알 수 없게 만드는 기법
Linear sensitivity model	통계용 마이크로 데이터를 생성할 때 데이터 주체의 영향력이 어느 정도 되는지 계산하여 재식별 위험도를 예측하는 모델

자료 : 연구진 작성(ISO/IEC 20889 재구성)

위와 같은 기밀성 모델을 바탕으로 재식별 공격을 통해 비식별 조치

를 수행한 사례는 Khaled El Emam이 Heritage Health Prize를 위해 Heritage Provider Network이 가지고 있는 청구 데이터를 비식별 조치한 논문(El Emam et al., 2012)에 상세히 설명되어 있다. 해당 논문에 의하면 identity disclosure만 고려하였고, 공격자는 비식별 조치된 데이터 전부가 아닌 일부만 알고 있으면서, 아는 사람(유명인이거나 친구·친지 등으로 충분한 배경지식을 가지고 있는 사람)을 대상으로 해당 인물이 현 비식별 조치 데이터 집합에 있는지를 모르는 상태에서 해당 인물을 찾는 과정을 수행하는 것으로 정의했다. k-익명성을 기반으로 재식별 위험도를 계산하였는데, 다양한 비식별 조치(가명화, Top-coding 등)를 적용하여 총 113,000명의 청구자료 2,668,990개에서 9,556명의 청구자료를 수정하여 재식별 위험도가 0.0084%가 되도록 비식별 조치하였다.

(4) 대표적인 비식별 조치 공개 소프트웨어

이상과 같은 개인정보 비식별화를 위하여 이미 다양한 소프트웨어가 개발되어 있다. 오픈소스 소프트웨어로 다운로드할 수 있는 대표적인 것들만 소개하면 다음과 같다. 다만, 국내에서 개발되어 공개된 오픈소스 소프트웨어는 존재하지 않고, 해외에서 개발된 소프트웨어는 국내 적용시 한글처리에 문제가 있는 경우가 많다는 한계가 있다.

1) ARX Data Anonymization Tool (https://arx.deidentifier.org/)

Java 기반의 비식별화 소프트웨어로 2018년 8월에 3.7.1 버전이 공개되었다. 가장 유명하면서 널리 사용되는 비식별화 소프트웨어로 독립된 소프트웨어로 사용되거나 API를 통해 다른 소프트웨어의 일부로도 사용될 수 있다.

2) UDT Anonymization Toolbox
(http://cs.utdallas.edu/dspl/cgi-bin/toolbox/)

UT Dallas의 Data Security and Privacy Lab에서 개발한 비식별화 소프트웨어이다. 이 소프트웨어 또한 독립된 소프트웨어로 사용되거나 다른 소프트웨어의 기능으로도 사용될 수 있다는 장점이 있고, 6가지의 서로 다른 비식별화 기법을 제공하고 있다.

3) Cornell Anonymization Toolkit
(https://sourceforge.net/projects/anony-toolkit/)

코넬대에서 개발하여 공개한 소프트웨어로, 다양한 공격자 모델에 대응할 수 있게 한 대화형 디자인이 특징이다.

4) Open Anonymizer
(https://sourceforge.net/p/openanonymizer/code/HEAD/tree/)

k-익명성 개념을 기반으로 데이터를 일반화하여 비식별화를 수행하는 소프트웨어이다.

5) μArgus (http://neon.vb.cbs.nl/casc/mu.htm)
개인정보를 보호하는 마이크로 데이터를 생성하기 위한 비식별화 소프트웨어이다.

6) sdcMirco
(http://cran.r-project.org/web/packages/sdcMicro/index.html)

R 기반의 비식별화 프로그램이다.

(5) 용어 정의와 관련된 모호성 존재

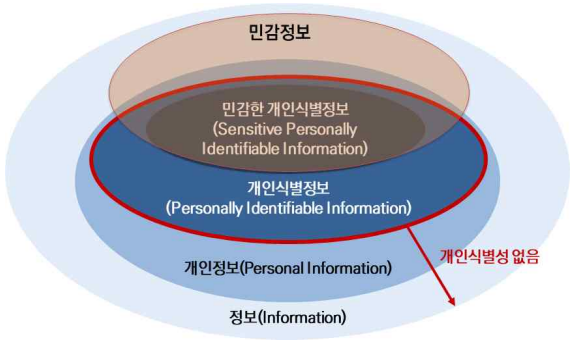
비식별 조치와 관련된 여러 용어들은 그림 3-5처럼 기술문서들인 국제표준문건에도 용어의 일치가 이루어지지 않은 상태이다.

Table B.1 — Mapping of de-identification terminology to the prior art

Term used in this document	ISO 25237 2017[26],	ISO 29100 2011	ICO 2012[19],	Article 29 2014[1],
De-identification	De-identification, Anonymisation	Anonymisation	Anonymisation	N/A
Masking	N/A	N/A	Anonymisation	N/A
Pseudonymization with controlled re-identification	Pseudonymization reversible	Pseudonymisation	Anonymisation	Pseudonymisation
Pseudonymization without controlled re-identification	Pseudonymization irreversible	Anonymisation	Anonymisation	Pseudonymisation
Randomization	N/A	N/A	Anonymisation	Anonymisation
Generalization	N/A	N/A	Anonymisation	Anonymisation
Differential Privacy	N/A	N/A	N/A	Anonymisation

(그림3-5) 비식별 조치 용어 정의의 모호성
출처: ISO/IEC 20889

이러한 모호성을 해결하기 위해서 제2장에서 소개한 사항들을 바탕으로 국내 법·제도에서 비식별 조치와 관련된 개인(식별)정보, 비식별 조치, 가명조치 등 용어를 정확히 정의할 필요가 있다. 다만 기술적인 사항을 고려할 때, 그림 3-6과 같이 개인과 관련된 개인정보 중에서도 식별성을 가지지 않는 것들이 있다는 것을 감안해야 한다. 식별성이 나타나는 것은 특정 개인에 대한 배경지식의 존재여부에 의존하는 경우가 많기 때문이다. 즉, 배경 지식과 특정 개인에 대해 연결(linking)할 수 있느냐에 따라 식별성이 결정되는 경우가 많다. 민감정보가 모두 개인식별정보가 아니라는 것도 고려하여야 한다. 가령 특정 사이트의 로그인을 위한 패스워드는 민감정보라 할 수 있겠지만 개인식별성은 가지지 않는다.



(그림3-6) 정보, 개인정보, 개인식별정보, 민감정보와의 관련성
출처 : BioINpro 2018 Vol. 49 (그림 수정)

이 외에도 식별성에 대해서도 명확히 정의를 해야 한다. 특정 개인을 구분해 내는 identification과 여러 데이터 중에서 하나를 구분해 내는 individualization은 전혀 다른 개념이다. Individualization을 하더라도 그 개인이 누구인지 알아내기 위해서는 별도의 정보를 연결하는 과정이 반드시 필요하다. 개인을 특정(identification)하기 위한 부가정보 연결이 불가능한 상황에서 individualization을 한 것만으로 개인이 재식별화된 것으로 보아야 할 것인지에 관해서는 명확히 정해진 바 없다.

2. 관리적 방법론

(1) 관리적 보호 조치

비식별 조치된 데이터는 기본적으로 정보보호관리시스템(Information Security Management System, ISMS)을 통하여 필요한 물리적, 기술적 및 관리적 보호 조치를 준수하여야 한다. 예를 들어, 방송통신위원회 고시 제2015-3호로 정하여 2015. 5. 19.부터 시행되고 있는 "개인정보의 기술적·관리적 보호조치 기준"에 있는 사항 중 필요한 사항들을 준수하여야 한다. 이에 추가로 재식별위험을 관리하기 위해서, 비식별 조치 가이드라인에 제시되어 있는 관리적 보호 조치 사항을 참조할 수 있다.

[표3-6] 비식별 조치를 위한 관리적 보호 조치 사항

- 비식별 정보파일 관리담당자 지정
- 비식별 정보파일 대장관리
- 원본정보 관리부서(기관)와 비식별 정보 관리부서(기관) 간 비식별 조치 관련 정보공유 금지
- 이용목적 달성시 지체없이 파기
- 비식별 정보파일 유출시 대응계획 수립

출처 : 비식별 조치 가이드라인

(2) 재식별 가능성 모니터링

그 이외에도 재식별 가능성을 주기적으로 모니터링 하여야 한다. 비식별 조치 가이드라인에 제시되어 있는 점검항목이 참고가 될 수 있는데, 이를 요약하면 표 3-7과 같다. 주목할 점은 내부 요인의 변화와 외부 환경의 변화 모두 점검하여야 한다는 것이다. 특히 새로운 기술개발로 인하여 재식별이 가능한 경우가 생길 수 있으므로 비식별 조치는 일회로 끝나는 것이 아니라 지속적, 반복적으로 점검하면서 기술 및 데이터

환경의 변화에 대응하여야 한다는 것이 핵심이다.

만일 비식별 정보를 제공·위탁한 자가 재식별 가능성을 발견하였다면, 그는 이를 즉시 그 정보를 처리하고 있는 자에게 통지하고 처리 중단 요구 및 해당 정보를 회수·파기하는 등의 필요한 조치를 하여야 한다.

[표3-7] 재식별 가능성 모니터링 점검 항목

구분	점검 항목
내부 요인의 변화	비식별 조치된 정보와 연계하여 재식별 우려가 있는 추가적인 정보를 수집하였거나 제공받은 경우
	데이터 이용과정에서 생성되는 정보가 비식별정보와 결합하여 새로운 정보가 생성되는 경우
	이용부서에서 비식별 정보에 대한 비식별 수준을 당초보다 낮추어 달라고 하는 요구가 있는 경우
	신규 또는 추가로 구축되는 시스템이 비식별 정보에 대한 접근을 관리·통제하는 보안체계에 중대한 변화를 초래하는 경우
외부 환경의 변화	이용 중인 데이터에 적용된 비식별 조치 기법과 유사한 방법으로 비식별 조치한 사례가 재식별 되었다고 알려진 경우
	이용 중인 데이터에 적용된 비식별 기법과 기술을 무력화하는 새로운 기술이 등장하거나 공개된 경우
	이용 중인 데이터와 새롭게 연계 가능한 정보가 출현하거나, 공개된 것으로 알려진 경우

자료 : 연구진 작성(비식별 조치 가이드라인 참조)

제 2 절 비식별정보의 활용사례

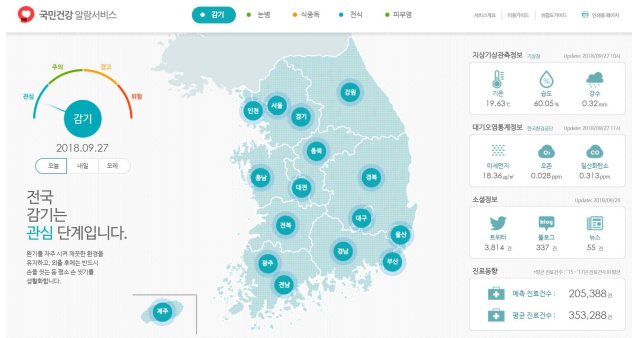
1. 국내 사례

(1) 바이오·헬스 분야

1) 국민건강보험공단

국민건강보험공단(이하 '공단'이라 한다)에서는 국민건강 알람서비스(http://forecast.nhis.or.kr/)를 2013년부터 운영해왔고, 2015년 고도화를 진행하였다. 공단과 식품의약안전처가 갖고 있는 보건의료데이터에 기상청의 기상데이터, 국립환경과학원의 환경데이터, 뉴스, 블로그, 검색 트렌드 등의 SNS 데이터를 연계하여 분석하였다. 현재, 5대 유행성 질병인 감기, 눈병, 식중독, 천식, 피부염을 선정하여 알람 서비스를 제공하고 있다. 질병 별로 지역과 연령을 구분해 관심, 주의, 경계, 위험의 4단계로 위험도를 표시하여 사전에 주의사항을 안내하고 사전예방 캠페인을 진행하고 있다. 국민건강 알람서비스는 기간별 진료데이터 및 SNS상 키워드 빈도를 근거로 질병을 예측하는 시스템으로, 5개 질환에 대해서 모두 0.9 이상의 상관관계를 보여주고 있다[1].

1) 공공데이터포털, "국민건강 알람서비스 고도화" 완료보고, https://www.data.go.kr/comm/file/download.do?atchFileId=FILE_00000000131 5675&fileDetailSn=0



(그림3-7) 국민건강보험공단 국민건강 알람서비스 화면
출처: 국민건강 알람서비스

2) CJ헬스케어

건강보험심사평가원(이하 '심평원'이라 한다)은 2015년부터 운용하고 있는 보건의료빅데이터 개방시스템을 통해서 전 국민의 진료정보, 의약품 처방정보, 의약품 안심사용 정보 등을 제공하고 있다. CJ헬스케어는 위식도역류질환 치료제 신규물질 개발에 이러한 심평원의 비식별화된 빅데이터를 활용하였다. CJ헬스케어는 심평원에서 제공하는 표본데이터 집합을 이용해 신약 개발의 타당성, 신규 복합제 발굴, 시장 분석, 처방, 패턴 분석 및 환자군 분석 등을 진행하였다. 심평원 데이터를 분석한 결과, 소화불량의 경우 더부룩한 경우가 많고 소화불량 환자의 80%가 역류성 식도염을 같이 가지고 있다는 것을 확인하여 신규약물 개발에 활용하였다. 이 과정을 통해 개발된 약물은 기존의 약물을 뛰어넘는 신개념 소화성 궤양용제로 기대를 받고 있으며, 2017년 식품의약안전처에 품목허가를 신청하였고, 현재 3500억 규모의 시장에 진출하여 기존 수입약품을 대체할 것으로 전망하고 있다[2]. 2018년 7월 해당 약물은

2) 뉴스투데이, CJ헬스케어, 빅데이터로 '니즈'를 찾아내 신약개발, http://www.news2day.co.kr/92887

상품명 “케이캡정”으로 우리나라 30호 국산 신약으로 허가를 받았고, 2019년 2월 출시될 예정이다[3].

3) 라인웍스

라인웍스는 심평원에서 제공하는 공공데이터인 의료명세서 등을 분석하여 특정 질병이 얼마만큼의 비용을 초래하고 있는지를 나타내는 질병부담도 자료를 홈페이지에 무료로 공개하고 있으며, 관련 기업체 등을 대상으로는 맞춤형 분석 데이터를 제공하고 있다[4]. 특히 최근에는 병원 전자의무기록을 활용하여 환자의 30일 이내 재입원율을 예측하는 모델을 개발하기도 하였다[5]. 라인웍스가 사용하는 비식별처리된 공공데이터와 서비스들은 다음과 같다.

[표3-8] 라인웍스가 활용중인 비식별 데이터

서비스 기능 및 특징	활용 데이터
- 심평원에서 공개한 전체환자표본데이터를 바탕으로 정확한 의료수요 정보 분석 - 사용자 목적에 따라 맞춤형으로 의료수요정보 분석 제공 - 2010년~2016년까지의 의료 수요 시계열 분석 - 환자 인구통계학적 분석 - 질병, 의료행위, 의약품 주성분, 치료재료 등을 기준으로 한 의료수요 분석 - 지역, 요양기관종별, 진단과 별로 분석된 자료의 세부 분석	건강보험심사평가원 - 환자데이터 - 병의원정보 및 평가정보 통계청 - 국가통계포털 식품의약안전처 - 온라인의약도서관 - 의료기기 민원창구

3) 한국일보, CJ헬스케어, 위식도 역류질환 신약 ‘케이캡정’허가받아, http://hankookilbo.com/v/b184f001cc1246a18c97ed641ba8a8ee
4) 공공데이터포털 기업탐방인터뷰, “공공데이터 활용한 창업으로 현대판 나이팅게일을 꿈꾸는 남자”, https://www.data.go.kr/useCase/interview/1000666/show.do
5) 청년의사, 라인웍스, 의료 빅데이터 기계학습을 통해 재입원 예측, http://www.docdocdoc.co.kr/news/articleView.html?idxno=1058491

자료 : 연구진 작성(공공데이터 포털 기업탐방리포트)

4) 서울아산병원

서울아산병원은 2013년부터 ABLE (Asan BiomedicaL research Environment)라는 비식별 조치된 의료정보를 병원 내부의 연구자들에게 제공하는 시스템을 개발하여 운영하고 있다. 이를 위해서 병원 내규로 그림 3-8과 같이 21개의 개인식별정보를 정의하였으며, 직접 식별자는 제거하고, 간접식별자는 마스킹 처리하여 연구자들에게 제공하고 있다. 서울아산병원 개원 이후 모든 환자들의 전자의무기록이 포함되어 있고, CT/MR 영상 등 이미지 데이터도 비식별 조치되어 제공되고 있다[6].

매년 기관윤리심의위원회(IRB)에서 비식별 조치 성능을 평가받으면서 비식별 조치의 정확도와 민감도를 96 ~ 98% 사이에서 유지하여 운용을 하고 있는 중이다.

6) Shin, S.-Y., et al., “A De-identification for Bilingual Clinical Texts of Various Note Types”, Journal Korean Medical Science, Vol. 30, No. 1, pp. 7-15 (2015). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4278030/

No	개인식별정보
1	이름 (의료진 이름 제외)
2	읍/면/동 이하 상세 주소
3	전화번호 일체(Fax번호 포함)
4	이메일주소
5	주민등록번호
6	외국인등록번호
7	여권번호
8	건강보험증번호
9	은행계좌번호
10	신용카드번호
11	자격증번호/면허번호/학번
12	차량번호
13	환자등록번호
14	회원ID (홈페이지, ARC 등)
15	사번
16	IP 주소
17	URLs
18	바이오정보: 지문, 홍체, 정맥, 음성, 필적, 개인식별이 가능한 유전 정보 등
19	얼굴의 전판 사진 또는 이에 상응하는 이미지
20	기타 개인식별이 가능한 정보(예: 병리번호)
21	생년월일 (생년월까지 허용)

(그림3-8) 서울아산병원 개인식별정보 정의
출처: Shin, S.-Y., et al.

(2) 금융 분야

1) 렌딧(LENDIT)

렌딧은 국내 P2P(Peer to Peer; 개인간) 대출 플랫폼 회사로, 개인신용대출 부분에서 업계 1위(점유율 45%)인 회사이다. 렌딧은 비식별 조치된 신용 데이터로 모형을 설계하였다. 특이한 점은 렌딧 웹사이트에서 투자설명서를 읽는 동작 패턴, 즉 얼마나 오래 머물면서 약관을 꼼꼼히 읽고 신청했는지 등을 상환의지에 대한 평가에 포함하여 대출시 중요한 요소로 사용하였다는 것이다. 이러한 비금융 정보들은 채권 100개 중 5개 정도의 신용평가에 큰 영향을 준다고 한다[7]. 이러한 고유의 신용평가 모델인 “렌딧 스코어링 시스템(Lendit Scoring System)”을 개발하는 데 미국의 P2P업체인 렌딩클럽을 참고했다고 한다. 이를 통해 2015년 창립 이후 현재까지 3년 동안 대출 이용자들에게 100억원에 가까운 이자액을 절감시켜준 것으로 평가받는다[8].

2) 삼성카드

삼성카드는 맞춤형 할인 혜택을 주는 링크(LINK)서비스를 2014년에 개발하였다. 고객의 개인별 소비패턴을 분석하고, 선호 업종, 활동 지역, 가맹점 인기도 등을 고려해 개인별 맞춤 혜택을 제공하는 서비스로, 카드사가 빅데이터를 이용해 고객에게 혜택을 제공한 국내 첫 사례로 알려져 있다. 전략 가맹점 마케팅 테스팅 시 기존의 무작위적 문자서비스에 대한 고객 반응율이 5%였던 반면, 링크 서비스 이용 고객의 반응율은 24%를 상회하였고, 삼성 카드의 분석에 의하면 링크 서비스를 통

7) 매일경제, http://news.mk.co.kr/newsRead.php?year=2017&no=72979
8) 매일경제, 매경 핀테크 어워드 2018 / 최우수상, http://news.mk.co.kr/newsRead.php?year=2018&no=544729

하여 할인혜택을 제공한 외식 가맹점 업체 한 곳은 유입된 고객의 86%가 신규 고객으로 나타나는 등 마케팅 측면에서 큰 효과를 보이고 있다고 한다[9].

3) 신한카드

신한카드는 빅데이터에 기반한 소비 분류방식인 "코드나인(code9)"을 개발하였다. 이를 통해 신한 카드 고객 2,200만 고객의 카드 사용 실적을 바탕으로 남녀 각 9개의 고객군으로 구분하여 총 18개 군으로 나누어서 카드를 출시하였다. 해당 상품은 기존 주력카드보다 평균 10% 이상 이용률이 높다고 하며, 출시 2년 만에 500만매를 돌파할 정도로 인기를 얻고 있다[10].

4) 삼성화재

삼성화재는 비식별화된 보험계약, 보험 정보 등 방대한 데이터를 활용하여 IFDS (Insurance Fraud Detection System)를 개발하여 사기 고위험군을 예측하는데 사용하고 있다. 접수된 사고에 대해, 데이터를 토대로 보험사기와 관련해 일정 기준 이상의 점수로 산출되면 자체기준에 따라 F1, F2, F3 등으로 분류하고, 보험사기 가능성이 가장 높은 F1 사고에 대해서는 별도의 조사에 착수한다. 해당 시스템을 통해 다양한 보험사기를 찾아내고 있는데, 이를 통해 고급승용차 분실 사고 신고인데 차량 담보대출 이후 이를 갚지 않기 위해서 허위 신고를 한 사례, 음주운전사고 신고이나 음주 사실을 숨긴 사례 등을 찾아낼 수 있었다[11].

9) 디지털타임스, 삼성카드, 빅데이터 기반 '링크'서비스,
http://www.dt.co.kr/contents.html?article_no=2014101502100558785001
10) 신한카드블로그, 신한카드 코드9 시리즈, 500만매 돌파 - 빅데이터 정확성의 결과,
http://www.shinhancardblog.com/291
11) 중앙일보, 보험금 청구 즉시 빅데이터로 "사기" 적발,

(3) 통신 분야

1) 서울시-KT

서울시는 KT와 협력하여 2013년 4월부터 심야버스(N버스)의 노선을 결정하였고, 현재 9개 노선이 운행 중이다. 해당 노선은 시가 보유하는 데이터와 KT의 이동통신망 데이터를 분석해 활용했다. 서울을 1km반경의 1천250개 구역으로 분할하고, KT가 제공한 비식별 조치된 가입자의 심야시간(0~5시) 통화 기지국 위치와 청구지 주소 데이터의 통계치를 가지고 유동인구 밀집도를 계산했다. 밀집도에 따라 색상을 시각화하고, 노선, 시간, 요일별 패턴도 같이 분석하였으며, 노선 부근의 유동인구 가중치를 계산해 노선을 최적화했다. 한 달간의 데이터가 약 30억 건에 달했으며, 스마트카드를 통한 택시 승하차정보 1주일 치 데이터도 같이 분석하였다.[12]

2. 해외 사례

(1) 바이오·헬스 분야

1) Google Flu Trends

Google Flu Trends는 공단의 국민건강 알람서비스의 원형으로 2008년부터 제공되었다. 구글에서 검색한 독감과 관련이 있는 검색어의 검색 빈도를 이용하여 인플루엔자 유행 시기를 미국 질병관리본부(CDC)보

https://news.joins.com/article/11536647
12) ZDNet Korea, 서울시, 심야버스노선 어떻게 만드나,
http://www.zdnet.co.kr/news/news_view.asp?artice_id=20130702115100

다 1주일 먼저 예측할 수 있음을 증명하였다[13]. 미국을 시작으로 하여 전 세계로 해당 서비스를 확장하였다. 추후 뎅기열로도 해당 서비스를 확장하기도 하였다. 해당 서비스는 현재 중단된 상태이지만, 지속적인 연구를 위해서 콜롬비아 대학, 보스톤 아동 병원, 미국 질병관리본부에 필요한 정보가 제공되었다[14].

2) IBM Watson의 의료분야 활용

질의응답(Question and Answering) 시스템인 IBM Watson은 방대한 의료데이터를 분석하여 환자를 진단하여 의사에게 치료법을 추천해 준다. 특히 Watson for Oncology는 미국의 메모리얼슬로운케터링 암 병원의 비식별화된 의료정보와 의료서적 등을 학습하여, 환자에 대해서 가장 확률이 높은 병명과 성공 가능성이 높은 치료법을 의사에게 제시한다.

3) Heritage Provider Network

헤리티지 헬스는 미국 서부에서 병원과 보험회사를 운영하는 회사로, 3백만 달러의 상금을 걸고 청구 데이터를 이용, 그 다음해의 입원기간을 예측하는 경진대회를 2011년부터 2013년까지 진행하였다. 그 결과 주최 측이 요구하는 성능 기준을 달성한 참가팀이 없어 1등을 한 팀은 없었지만, 참가팀 중에서 가장 우수한 성적을 거둔 팀은 오십만 달러의 상금을 획득하였다. 그리고 해당 데이터는 현재 Kaggle을 통해 연구자들에게 공개되어 있다[15].

13) Ginsberg, J. et al., "Detecting influenza epidemics using search engine query data", Nature, Vol. 457, pp. 1012-1014 (2009).
14) mobihealthnews, Google Flu Trends website shuts down; will send data to Boston Children's, Columbia, CDC. https://www.mobihealthnews.com/46248/google-flu-trends-website-shuts-down-will-send-data-to-boston-childrens-columbia-cdc
15) https://www.kaggle.com/c/hhp

4) Github Open Data

오픈소스 공유 사이트인 Github에는 다양한 데이터가 공개되어 있다. 그중 https://github.com/beamandrew/medical-data에는 상당히 다양한 의료 데이터들이 비식별 조치되어 공개되어 있다. 22종의 의료 영상 데이터들과 19종의 각종 경진대회 데이터들, 4종의 전자의무기록 데이터, 5건의 미국 국가 데이터를 포함하여 다양한 의료 데이터들이 제공되어 있다. 이 외에도 Github에는 다양한 의료 공개 데이터들이 제공되고 있다.

(2) 금융 분야

1) 아멕스

아멕스(AMEX)에서는 위치 기반 소셜네트워크 정보를 활용한 고객별 맞춤형 마케팅을 실시해 고객들로부터 긍정적인 반응을 얻고 있다. AMEX Sync 프로그램은 제휴를 맺은 소셜 플랫폼의 고객 계정을 AMEX카드와 연동시켜 고객에게 맞춤형 할인 혜택을 준다. 예를 들면 페이스북이나 트위터에서 특정 상품 및 레스토랑에 '좋아요'를 클릭하면 할인쿠폰 및 관련 정보를 미리 제공해줌으로써 기존의 타깃 마케팅보다 큰 효과를 얻을 수 있었다[16].

2) 프로그래시브

미국의 보험회사인 프로그래시브(Progressive)사는 자동차에 부착된

16) A Complete Guide To Amex Sync (American Express Sync), https://milestomemories.boardingarea.com/a-complete-guide-to-amex-sync-american-express-sync/

기기가 전송하는 데이터를 바탕으로 고객의 운전 패턴을 분석하고 미래의 사고 가능성을 예측하는 Snapshot이라는 프로그램을 개발하였다. 이를 통해 자동차 보험료를 산정하는 Pay as You Drive 시스템을 운영하고 있다. 즉, 상대적으로 덜 위험한 방법(속도, 운전습관)으로, 덜 위험한 시간대 및 지역에서 운전하는 고객일수록 더 적은 보험료를 낸다. 이는 고객들의 안전 운전을 유도하는데 도움을 준다[17].

3) Lenddo

미국 기업인 Lenddo는 신용평가 알고리즘을 개발하면서, 온라인상 대출자의 평판에 대한 비정형 데이터를 추출하여 신용도를 평가하였다. 기존의 대출회사처럼 담보물이나 보증인, 기존 신용평가 등을 요구하지 않고, 온라인 소셜네트워크 기록을 상세히 분석하여 신용평가를 한다. 예를 들어, Lenddo와 연결된 개인의 SNS 계정 수, 연결된 계정의 과거 이력, 각 SNS에서의 친구나 팔로어 수 등을 분석한다. 또한, SNS 친구 중 연체자가 있거나 '자동차사고' 및 '실직' 같은 부정적인 단어가 많이 나오면 신용점수가 낮아진다[18].

4) Carpe Data

Carpe Data는 SNS를 포함한 온라인 활동, 웨어러블 기기, IoT 기기 등을 통해 생성된, 기존 보험회사 등이 고려하지 않던 데이터를 이용하여 사람들이 특정 보험 상품을 구입할 용의가 있을지를 판단한다. 기존의 신용 점수, 각종 병원 검사 결과 등은 보험 가입자의 생활 습관을 충분히 반영하는데 실패하였기 때문에 이에 대한 보완적인 역할을 할

17) The Balance, Progressive Snapshot Review, https://www.thebalance.com/progressive-snapshot-review-4141266
18) GIGAOM, Credit scores, with a little help from your friends, https://gigaom.com/2012/05/16/credit-scores-with-a-little-help-from-your-friends/

수 있도록 고안된 것이다[19]. 보험 심사에 전통적으로 사용되던 데이터 이외의 새로운 데이터들인 인터넷에 공개된 건강, 금융, SNS, 공공 데이터들을 연결하여 보험회사들에게 새로운 형태의 데이터를 제공하고 있다.

(3) 통신 분야

1) T-Mobile

통신사들은 전통적으로 대량의 비식별화된 데이터(얼마나 오래 혹은 언제 전화를 하는지, 문자는 언제 많이 보내는지, 인터넷 사용량은 어떻게 되는지 등)를 수집하고 있다. 미국의 통신사인 티모바일은 타 통신사로 회선을 옮기는 고객의 이용패턴(청구 내역, 통화 실패율, 고객 감정 등)을 분석하는 "Quick View"라는 시스템을 구축하였다. 이 시스템을 이용하여 고객이탈(타 통신사로 이동)을 감지하였고, 이 시스템을 시행한 결과, 2011년 1사분기 이탈 고객 수가 약 10만 명이었으나 2사분기에서는 5만 명으로 감소하였다[20].

19) PR Newswire, Carpe Data: The Next Generation Insurance Data Company Announces Launch, https://www.prnewswire.com/news-releases/carpe-data-the-next-generation-insurance-data-company-announces-launch-300337756.html
20) DataFloq, T-Mobile USA Cuts Downs Churn Rate By 50% With Big Data, https://datafloq.com/read/t-mobile-usa-cuts-downs-churn-rate-with-big-data/512

제 4 장 비식별 조치에 대한 해외의 제도 현황

제 1 절 유럽

1. EU

(1) 제도적 개괄

1) 개인정보보호 규제 일반

일반정보보호규정(General Data Protection Regulation; 이하 'GDPR'이라 한다)은 2016년 공포된 유럽연합 내 개인정보의 처리에 관한 유럽연합의 규정(Regulation)이다. GDPR은 2년의 경과기간을 거쳐 2018. 5. 25. 유럽연합 모든 회원국에서 발효하였다. 유럽연합 규범체계상 지침(Directive)이 아닌 규정(Regulation)이므로 별도의 (국내) 이행입법 없이 모든 회원국 내에서 직접 효력을 갖는다. 그러나 회원국은 자국의 국내 정보보호법을 추가로 입법할 수 있다. 그밖에 회원국 간의 해석 및 집행이 일관적이지 않을 가능성은 남아 있다.

유럽연합 내에서 GDPR 전 이 문제를 규율한 것은 1995년 개인정보보호지침(Directive 95/46/EC)이었다. 이는 지침이므로 국내 이행입법이 필요하였고, 자세한 내용은 행동강령(code of conduct) 형식으로 각 회원국에게 맡겨져 있었으며, 각국의 해석과 집행이 달라질 소지도 더 컸다. 식별 가능한 데이터의 익명화(anonymous)에 관하여는 그 서문(Recital) (26)에 언급되어 있었다. 서문 (26)은 익명화가 정보주체를 식별할 수 없게 하는 방식으로 이루어져야 하고, 익명화되면 지침의 적

용대상이 아니라는 점을 밝혔을 뿐이고, 데이터가 익명화되는 구체적 방식에 대하여 규정하지는 아니하였다. 그밖에 본문에 익명화, 가명화 등에 대한 규정을 별도로 두지는 아니하였다.

GDPR은 개인정보처리에 대한 법적 규율을 전반적으로 강화하였다. 개인정보처리자는 개인정보처리에 관한 6대 원칙을 준수하여야 한다(GDPR 제5조). 6대 원칙이란 ① 적법성·공정성·투명성(Lawfulness, fairness and transparency), ② 목적제한(Purpose limitation), ③ 개인정보처리의 최소화(Data minimization), ④ 정확성(Accuracy), ⑤ 보관기간의 제한(Storage limitation), ⑥ 무결성·기밀성(Integrity and confidentiality)을 가리킨다. 또한 적법처리의 6대 기준 중에 최소한 한 개를 충족해야 한다. 적법처리 6대 기준이란 정보주체의 동의, 계약의 이행, 법률상 의무 준수, 중대한 이익 보호, 공익을 위한 또한 공공기관의 처리, 정당한 이익을 말한다.

GDPR 제5조 제1항과 제6조 제1항은, 콘트롤러(controller)는 개인정보를 정보주체와 관련하여 적법하고 공정하며 투명한 방식으로 처리하여야 하고, 정보주체의 이익이나 기본권 및 자유를 우선하여야 한다고 규정한다. 그리고 이를 위하여 독립성과 전문성이 있는 정보보호 담당관(DPO, Data Protection Officer; 이하 'DPO'라 한다)을 지정하여야 한다. 특히 DPO 지정의무는 ① 유럽연합 내에 거점을 두고 있는 경우(Established in the EU), ② 유럽연합 역외에서 유럽연합에 위치하고 있는 정보주체의 개인정보를 처리하는 경우(Directed at Individuals in the EU), 또는 ③ 국제법에 따라 유럽연합의 회원국의 법률이 유럽연합의 역외에 적용되는 경우에 적용된다. DPO는 ① 전문성(다년간의 개인정보보호 실무경험 및 법적 지식 보유), ② 독립성(업무상 이익의 충돌이 발생해서는 안 되며, 업무로 인한 징계나 해직 불가) 및 ③ 협력 가능성(유럽연합 규제당국과 협력 가능한 경험 및 언어구사 또는 통역)을 갖추어야 한다.

그밖에 GDPR은 유럽연합 역내의 개인정보를 처리하는 경우 유럽연합 역외의 국가에도 GDPR을 적용하는 것을 포함하는 개인정보 국외 전송

체계를 확립하였다. 그 핵심은 개인정보가 '적정하게 보호' 될 수 있는 국가로만 역외 전송을 허가하는 것이다. 유럽연합 회원국 간, 또는 적정성 평가(adequacy decision)를 받은 국가로 전송하는 것은 별도 조치 없이 가능하다. 이러한 점에서 GDPR은 우리 법은 아니지만 그 적정성 평가는 유럽연합 역내 개인정보의 처리 및 전송이 불가피한 우리나라 내의 여러 활동에도 관계한다.

또한 GDPR은 정보주체의 권리를 확대하여, 개인정보 열람권, 정정권, 처리 제한권, 자동화된 개인정보 처리 대상이 되지 아니할 권리, 잊힐 권리(개인정보 삭제를 요청하는 경우 삭제처리), 이동권(개인정보를 정보주체 또는 그가 지정하는 다른 사업자에게 전송) 등을 신설하였다.

2) 비식별 조치

GDPR은 그 이전의 개인정보보호지침과 달리 익명화(anonymisation)와 가명화(pseudonymization)에 대하여 여러 규정을 두고 있다.

먼저 GDPR 서문 (26)은 익명처리(anonymisation)의 결과에 대하여 언급하고 있다. 익명정보는 개인정보의 반대되는 개념으로, GDPR의 적용을 받지 아니하며, 이는 처음부터 익명인 경우와 사후적으로 익명화되는 경우 모두에 그러함을 분명히 한다. 그러나 익명처리의 과정이나 정도에 관하여 별도로 규정하고 있지는 않다.

다음, GDPR은 식별을 곤란하게 하는 조치가 다양한 수준에서 존재함을 인정하고, 이와 관련하여 가명화(pseudonymisation) 개념을 명시적으로 추가하였다. 가명화된 정보는 추가정보를 통하여 간접적으로 개인을 식별할 수 있는 정보를 의미한다. 따라서 가명정보는 개인정보의 특수한 형태라 할 수 있다. GDPR 서문 (28)은, 가명화는 현재시점에서 효과적이고 안전한 데이터 보호를 위해 모든 상황에서 사용할 수 있는 기술임을 분명히 한다. 그리고 GDPR 서문 (29)는 콘트롤러와 프로세서가 데이터셋을 가명화할 '유인을 창출'할 것을 규정한다.

데이터 처리(processing)는 법적 절차에 따라 합법적으로 이루어져야

한다. 보안에 대한 기술적 수단을 제공하여 외부침입자로부터 데이터를 기술적으로 보호하여야 한다. 그런데 가명을 사용하면, 개인정보가 유출되었을 경우 데이터 보호규칙을 좀 더 유연하게 적용 할 수 있다. 가령 GDPR 제11조(2)는, 제15-20조는 콘트롤러가 정보주체를 식별 할 수 있는 입장에 있지 않음을 입증 할 수 있는 경우에는 적용되지 아니한다고 정한다. 즉 데이터 유출 문제에 대한 익명화 접근방식과 일관되게 데이터가 익명화된 상황에서 공개요구 사항을 완화하는 것이다.

가명화의 효과로 데이터 유출이 되어도 식별 위험이 높지 아니하다는 것은 가명처리의 유인으로 작용된다. GDPR 제34조(2)(a)는 데이터 유출 시 정보주체에게 그 사실을 통지할 의무[제34조(1)]가 암호화(encryption)를 포함하여 가명화가 이루어진 경우 완화될 수 있음을 보여준다. 이는 가명처리의 직접적 유인으로 작용할 수 있다.

[표4-1] 가명화 인센티브로서 데이터보호 규칙의 유연성

콘트롤러(controller)의 의무
▶ Article 11 1. 콘트롤러가 개인정보를 처리하는 목적 상 콘트롤러가 정보주체의 식별할 필요가 없거나 더 이상 없어지는 경우, 콘트롤러는 본 규정의 준 수 목적만으로 정보 주체를 식별하기 위한 추가 정보를 유지, 획득 또는 처리해야할 의무가 없다. 2. 본조 1항에서 언급된 경우에 있어 콘트롤러는 자신이 정보 주체를 식별할 위치에 있지 않음을 입증할 수 있는 경우, 콘트롤러는 가능한 한 정보 주체에게 이를 통지해야한다. 이 경우 제15-20조에 따른 자신의 권리 행사 목적 상 정보 주체가 자신의 식별을 가능케 하는 추가 정보를 제공하는 경우를 제외하고 제15-20조는 적용되지 않는다.
▶Recital 57 콘트롤러는 정보주체가 자신의 권리행사를 지원하기 위해 제공하는 추가정보를 거부할 수 없다. 예를 들면, 콘트롤러가 제시하는 온라인 서비스에 정보주체가 로그인하기 위해 사용하는 것과 같은 동일한 인증(credential)을 통한 디지털식별이 포함된다.

자료 : GDPR, 개인정보보호위원회 번역 (연구진 수정)

그리고 GDPR 서문 (78)은 간접식별자가 가명으로 직접식별자를 대체하는 것은 이미 데이터 최소화 원칙을 보장하기 위한 적절한 수단이 될 수 있다고 한다. 그러므로 가명을 사용하는 것은 일반적으로 데이터 처리 규정준수에 유리하다.

(2) 비식별 정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

GDPR은 가명화(pseudonymisation)의 중요성을 인식하여 가명화라는 용어의 정의뿐만 아니라 데이터 보안 및 보안처리를 보장하기 위한 기술적 수단으로서 가명화를 다루는 서문 및 본문 규정을 마련하고 있다. GDPR 제4조(5)는 “가명화란 추가정보를 사용하지 아니하는 한 특정 정보주체를 더는 식별할 수 없도록 데이터를 처리하는 것을 의미한다. 그러한 추가정보는 따로 보관된다. 개인적으로 식별되거나 식별가능한 자연인에게 귀속되지 아니하도록 기술적 및 조직적 조치를 취해야한다”고 규정한다. 한편 서문 (26)은 “부가정보의 사용으로 자연인을 식별할 수 있는 가명을 받은 데이터는 식별가능한 자연인에 대한 정보로 간주되어야 한다”고 명시하고 있다. 양자를 결합하여 보면 가명정보는 여전히 개인정보임을 알 수 있다.

가명(Pseudonyms)은 직접식별자를 제거하고 데이터 셋이 간접적으로 식별 가능한 데이터만 포함하도록 하는 유용한 개인정보 보호 강화기술이라고 할 수 있다. 간접식별자로서 가명은 정보주체를 식별 가능하게 하는 개인정보이다. 서문 (28)이 밝히는 바와 같이 가명화는 정보주체에 대한 위험을 줄일 뿐 상황에 따라 GDPR의 범위에서 제외되지 않는다.

만약 추가처리가 당초의 목적과 양립할(compatible) 수 있다면, GDPR 제6조(4)의 "개인정보의 수집한 목적 외 처리가 해당 개인정보를 수집한 당

초 목적과 양립될 수 있는지”를 고려하여야 한다. 최초 목적과 추가처리(further processing)의 목적이 양립할 경우, 제5조(1)(b)에 따라 합법적이다. 제89조에 의하여 과학적, 역사적 및 통계적 연구는 데이터 처리의 목적제한 원칙에 조응하면서 효과적인 예외를 허용하기 위하여 가명화에 기초한다. 이는 데이터보호 원칙, 특히 데이터 최소화 원칙을 보장하기 위한 적절한 기술적·조직적 조치로 간주된다.

[표4-2] GDPR 비식별 데이터의 정의와 수준

	Pseudonymised data	Anonymous data
정의	▶ Article 4(5) “추가정보의 사용 없이는 특정정보주체에게 더 이상 귀속될 수 없는 방식으로 개인정보를 처리하는 것을 말한다. 단 이러한 추가정보는 별도로 보관되며, 개인정보가 식별된 또는 식별가능한 개인에게 귀속되지 않도록 보장하기 위한 기술적 및 조직적 조치의 대상이 된다.”	▶ Recital 26 “식별되거나 식별 가능한 자연인 또는 개인정보와 관련이 없도록 정보주체를 더 이상 식별할 수 없도록 익명으로 처리된 정보”
수준	직접식별자들이 변환되고, 암호키가 적절한 저장 및 처리되도록 적절한 제어장치가 마련되어야 한다.	GDPR은 리사이틀 내의 익명화 정의만 있음. 기술이 지속적으로 개발되고 법률적 의미에서 효과적인 익명화에 대한 논의가 제기되었지만 새로운 데이터 보호 프레임워크는 익명화 프로세스에 대한 명확한 규정은 제공하지 않음. 익명화처리자체에 대한 조건보다는 익명화 결과에 초점을 맞춤.

	필요에 따라 데이터 내의 일부 간접식별자도 추가로 변환한다. 이 경우, 정보를 자연인에게 귀속시키는 것이 보다 더 어렵다.	효과적인 익명화의 적절한 수준을 결정할 때 해당정보가 더 이상 식별 가능한 개인을 언급하지 않는다는 것을 보여주기 위해 모든 객관적인 요소(미래의 새로운 기술 발전)를 합리적으로 고려해야 함.

자료 : 연구진 작성 (GDPR, 개인정보보호위원회 번역 수정)

가명화가 익명 데이터(anonymous data)로 이어질 수 있는 문제에 관하여는 제4조(5) 및 서문 (26)의 정의를 따를 수 있다. 일부 가명으로 처리된 데이터는 일반적으로 개인정보이다. '합리적 가능성테스트(reasonably-likely-test)' 하에서 사례별로 판단해야 한다. 익명처리 기술을 사용하는 콘트롤러는 재식별 위험을 평가하여야 한다. 가명화는 GDPR의 데이터보호원칙(데이터 최소화원칙)을 구현하고 준수하기 위한 적절한 기술적 방식이자 비식별화 정보 활용을 위한 안전조치라고 할 수 있다. 중요한 것은 가명화를 데이터 처리의 한 형태로 명확하게 정의하고 있다는 점이다.

2) 법령 이외의 형태

익명화에 관하여는 법령 이외에 제29조 작업반(The Article 29 Data Protection Working Party; 이하 'WP 29'라 한다)이 2014년에 발표한 Opinion 05/2014가 중요하다. 같은 의견 제7항은 익명화를 개인정보가 최초에 수집된 목적과 일치하며 따라서 추가적 법적 기초를 필요로 하지 아니한다는 점을 분명히 하였다. 즉 완전히 익명화된 정보는 GDPR의 적용범위를 벗어나며, 익명화가 정보처리의 본래 목적과 일치한다고 간주될 수 있다. 익명처리는 익명정보를 신뢰성 있게 생성하는 것을 가리킨다. 데이터가 효과적으로 익명 처리되었다면 GDPR의 범위를 벗어

나 추가 처리될 수 있다. 다만 이 의견서는 GDPR의 공표나 발효 이전 시기에 발표된 것이라는 한계가 있다.

(3) 비식별 정보, 익명정보, 가명정보 개념의 도입 과정

1) 배경 및 구체적 절차

GDPR은 개인의 신원을 직접 확인하거나 재식별을 불가능하게 하는 새로운 범주의 데이터로 가명을 도입하였다. 가명정보는 법적으로는 개인정보의 한 특수한 형태로 파악되는데, 실무적으로는 개인정보와 익명데이터 사이에 있는 세 번째 범주의 데이터로 취급될 수 있다(Bevott, 2016:2). 가명화(pseudonymisation)는 익명화(anonymisation)가 아닌 형태로 정보주체의 위험을 줄이고 일정 수준의 데이터 프라이버시를 보장하는 수단이 될 수 있다.

GDPR 제4조(2)는 '데이터처리(data processing)'를 정의한다. 이는 처리 형식을 구성할 수 있는 데이터의 수집, 저장, 조합 및 삭제를 포함한다. 기술이 계속 개발되는 상황에서 법적 의미의 효과적 익명화 방식에 대한 논의가 제기되었지만 GDPR에는 그에 관하여 기술적 방법을 정하는 규정은 존재하지 아니한다. 한편 가명화(pseudonymisation)에 관하여는 크게 '의무'와 '유인(incentive)'에 관한 규정이 있다. GDPR은 기술보호 조치로서 가명화를 사용하는데 유인을 제공한다. 콘트롤러(data controller)와 프로세서(data processor)가 실제 가명화 기술을 구현하고 보다 효과적인 데이터 보호를 위한 조치를 취하며 침해에 대응하게 하는 방식이다. 가명화에 대해서는 일부 법적 혜택을 제공하여 이를 유도하고 있다.

익명화(Anonymisation)는 식별을 위해 합리적으로 사용될 법한 모든 수단을 고려할 때 데이터 그 자체 또는 그 데이터와 다른 데이터를 조합한 것으로부터 개인을 식별할 수 없다는 뜻이다. 데이터가 더 이상

개인정보가 아니게 되면, 데이터 보호 법률의 적용대상이 되지 아니한다[서문 (26)]. 익명 데이터(anonymous data)는 개인정보(personal data)의 반대개념으로서 식별되었거나 또는 식별될 수 있는 개인과 관련되지 않는 정보 또는 그런 방식으로 처리되어 더는 식별될 수 없는 정보주체에는 적용되지 않는 식별되거나 식별 가능한 자연인과 관련이 없는 정보이다.

익명처리된 데이터는 여러 시나리오에서 사용될 수 있기 때문에 콘트롤러(data controller)는 익명처리된 데이터를 수집하거나 사용하거나 다른 사람들과 공유하기 전에 익명처리를 할 수 있다.

2) 논의 사항

WP 29는 2014년 4월 익명화 기술에 관한 의견서(WP 29, 2014)에서 유럽연합의 데이터보호를 위한 합법적인 프레임워크 신설에 대한 배경을 다음과 같이 제시하였다: 장치, 센서 및 네트워크가 대량의 데이터와 새로운 유형의 데이터를 생성하고 데이터 저장 비용이 무시할 수 없게 되면서 이러한 데이터의 재사용에 대한 대중의 관심과 수요가 증가하고 있다. 익명화는 모든 사람의 권리와 이익을 유지하고 위험을 완화하기 위한 좋은 전략일 수 있다. 데이터가 익명화되고 개인이 더 이상 식별되지 않으면 유럽 데이터 보호법이 더 이상 적용되지 않는다. 그러나 연구 및 분석을 위해 필요한 기본 정보는 그대로 유지한다는 것이 단순한 명제는 아니다. 예를 들어 익명으로 간주되는 데이터 집합은 하나 이상의 개인을 식별 할 수 있는 방식으로 다른 데이터 집합과 결합 될 수 있다. 따라서 기존 익명화 기술의 유효성과 한계를 분석하고 이러한 기술을 신중하고 책임 있게 사용하여 익명화 프로세스를 구축 할 것을 권장한다.

(4) 비식별 정보 활용 현황 및 정책적 추진 방향

정보 주체의 명시적 동의 없는 이차적 활용을 위한 정보처리 가능성과 관련하여 GDPR 제6조(4), 제5조(1)(b)에 따르면, “다른 목적을 위한 처리가 개인정보의 처음 목적과 일치하지 않는 경우” 정보의 이차활용(연구 또는 분석)이 정보주체의 동의 이외의 근거로서 진행될 수 있다. 이러한 정보처리를 결정하는 판단기준 중 하나가 암호화 또는 가명화가 포함된 적절한 안전장치의 유무이다[GDPR 제6조(4)(e)]. "적절한" 안전장치 또는 기술조치는 가명화의 질을 논할 때 항상 사용하는 용어지만, “적절(appropriate)”이라는 용어를 정의하거나 보안 조치의 필요한 질을 측정하는 조항은 없다. 다만 제32조(2)에서 적절한 보안수준의 결정에 관한 지침이 있으나 적절한 안전조치의 기간을 정의하지는 않는다.

콘트롤러가 새로운 목적이 원래의 목적과 양립 할 수 없다고 결론을 내리면, 새로운 목적을 정당화하기 위한 유일한 근거는 동의 또는 법적 의무사항이 구성된 EU 또는 회원국의 국내법이다.

GDPR에서 도입한 가명정보는 제89조에 명시한 대로 ①공익을 위한 유지보존의 목적(archiving purposes in the public interest), ②과학이나 역사적 연구의 목적(scientific or historical research purposes), ③통계목적(statistical purposes)의 개인정보 처리에 활용할 수 있다는 점에 중요한 의의가 있다. 위와 같은 목적을 위한 안전조치로서 가명처리는 데이터 최소화를 포함한 기술적 관리적 조치를 적용한 것이다.

과학적 연구 목적의 개인정보 처리는 기술의 발전과 실증, 기초연구, 응용연구 및 민간 투자 연구(privately funded research) 등을 포괄하는 광범위한 방식으로 해석되어야 한다(Rec. 159). 통계 목적으로 개인정보가 처리되는 경우, 유럽연합 또는 회원국 법률은 본 규정의 한도 내에서 통계 내용, 접근(access) 통제, 통계 목적의 개인정보 처리에 대한 세부사항 및 정보주체의 권리와 자유를 보호하고 통계의 신뢰성을 보장하기 위한 적절한 조치를 결정해야 한다. 통계목적은 통계 조사나 통계 결과를 작성하는데 필요한 개인정보의 수집 및 처리의 작업 등을

의미한다. 통계의 결과는 과학적 연구 목적 등 다른 목적을 위해 추가적으로 활용될 수 있다. 다만, 통계 목적으로의 정보처리 결과는 개인정보가 아닌 집합체 데이터 (aggregate data)이어야 하며 이 결과나 개인정보가 다른 특정 개인에 관한 조치나 결정을 지지하는데 활용되지 않아야 한다(Rec 162).

제89조가 적용될 경우, 정보주체의 정보 열람권(제15조), 수정요구권(제16조), 처리에 대한 제한권(제18조), 반대할 권리(제21조) 등의 권리가 제한될 수 있다.

[표4-3] GDPR 제89조

공익적 기록보존 목적, 과학적 또는 역사적 연구 목적, 또는 통계적 목적을 위한 처리와 관련한 안전조치 및 적용의 일부 제외

1. 공익적 기록보존 목적, 과학적 또는 역사적 연구 목적, 또는 통계적 목적을 위한 처리는 정보주체의 권리 및 자유를 위해 본 규정에 따라 적정한 안전조치가 적용되어야 한다. 그러한 안전조치는 특히 데이터 최소화 원칙이 준수되도록 기술 및 관리적 조치를 이행해야 한다. 그러한 조치에는 가명처리 방식으로 그러한 목적들을 달성할 수 있다면 가명처리가 포함될 수 있다. 정보주체의 식별을 허용하지 않거나 더 이상 허용하지 않는 추가 처리를 통해 그러한 목적들을 달성될 수 있는 경우에는 그러한 방식으로 달성되어야 한다.
2. 개인정보가 과학적 또는 역사적 연구 목적이나 통계적 목적으로 처리되는 경우, 유럽연합 또는 회원국 법률은 본 조 제1항의 조건 및 안전조치에 따라 제15조, 제16조, 제18조 및 제21조에 규정된 권리의 적용을 일부 제외할 수 있다. 단, 그러한 권리가 그러한 특정 목적의 달성을 불가능하게 하거나 중대하게 손상시킬 것으로 예상되고, 그러한 목적을 달성하기 위하여 적용의 일부 제외가 필요한 것이어야 한다.
3. 공익을 위한 기록보존의 목적으로 개인정보가 처리되는 경우, 유럽연합 또는 회원국 법률은 제15조, 제16조, 제18조, 제19조, 제20조 및 제21조에 명시되고 본 조 제1항의 조건 및 안전조치에 따른 권리로 인해 특정 목적의 달성을 불가능하게 하거나 중대하게 손상시킬 것으로 예상되고, 적용의 일부 제외가 해당 목적을 달성하기 위해 요구되는 한, 해당 권리의 적용을 일부 제외하도록 규정할 수 있다.

4. 제2항 및 제3항에 명시된 정보처리가 동시에 다른 목적으로 이루어지는 경우, 적용의 일부 제외는 해당 호에 명시된 목적을 가진 데이터 처리에만 적용되어야 한다.

출처 : GDPR, 개인정보보호위원회 번역

2. 독일

(1) 제도적 개괄

1) 개인정보보호 규제 일반

독일 개인정보법제의 중심이 되는 법률은 1977년 제정된 연방데이터 보호법(Bundesdatenschutzgesetz; BDSG로 표기한다)이다. 같은 법률은 개인정보처리에 관하여 공공부문과 민간부분을 나누어 규율한다. 제1장은 공공부문과 민간부문의 공통사항, 제2장은 공공기관에 의한 개인정보처리, 제3장은 민간기관에 의한 개인정보처리에 관한 규정이다. 그 밖에 정보통신망에 관한 특별법인 텔레커뮤니케이션법(Telekommunikationsgesetz)과 텔레미디어법(Telemediengesetz)에도 개인정보 보호에 관한 규정이 있다.

BDSG에 대하여는 종래부터 매우 제한된 정보처리의 구상에만 얽매여 최신정보처리 기술과 위험을 반영하지 못하고 있으며, 민간부문의 규제가 소홀하다면서 정보주체의 권리를 강화하고 콘트롤러의 자율규제를 고무시키는 방향으로 개정하여야 한다는 요구가 있었다. 이에 독일은 유럽에서 가장 먼저 GDPR을 반영하여 2017년 5월에 BDSG을 개정하였다. BDSG는 연방의 다른 법률이 이 법률이 적용되는 경우를 규정하지 아니하거나 배제하지 아니한다고 규정한 경우에도 적용되며, 개인정보가 처리되는 한 행정절차법보다도 우선한다(제1조 제3항).

2017년 5월 전면 개정된 BDSG의 주요 내용은 다음과 같다.

2017년 개정 BDSG는 전체 4장 19절 85조문으로 구성되어 있다. 그 중 제1장은 일반규정, 제2장은 GDPR의 이행 규정이다. 개정 전과 달리 개인정보 보호 일반에 관한 부분은 거의 그대로 GDPR에 의하는 것으로 하였고, 민감정보에 대하여 제22조부터 제28조에 이르기까지 규정하고 있다. 제3장은 형사사법지침 2016/680/EU의 이행 규정, 제4장은

GDPR과 형사사법지침 2016/680/EU가 적용되지 않는 영역에서의 처리를 위한 특별 규정들이다. 즉 GDPR과 형사사법지침을 철저히 반영하여 개정되어 개인정보처리의 국제적 기준에 최대한 부합한 국내법을 마련함으로써 자국민과 자국기업이 국제규범의 위반으로 발생할 수 있는 불이익을 최대한 방지하고자 하였다. 그중 특히 주목할 만한 조항은 다음과 같다.

첫째, 개정 법률은 고용 관계에서 근로자의 데이터 보호에 대한 포괄적 규칙을 수립하였다. 이러한 규칙은 BDSG의 현행 규칙과 독일 법률, 법원 및 DPA가 제정한 규칙 및 의견을 바탕으로 한 것이다. 개정 법률은 자발적 동의의 필요성을 명시하고 고용관계, 사회적 법률 또는 사회적 보호 하에 권리를 행사하거나 의무를 준수하기 위해 그러한 처리가 요구되는 경우, 고용관계의 목적으로 고용인의 민감한 개인정보를 처리하도록 허용한다.

둘째, 개정 법률은 DPO를 임명하는 요건에 대하여 기존의 기준을 유지하여 회사는 최소한 10명의 직원을 정규로 고용하는 경우, 데이터의 자동처리와 관련된 DPO를 임명해야 한다고 규정한다. 이는 기업의 핵심활동의 성격상 정보주체를 정기적으로 체계적으로 모니터링하거나 특수 데이터 범주를 대규모로 처리를 위한 것이다. DPO에 대한 GDPR의 의무사항과 관련하여 독일만이 DPO 임명에 관한 이전 법에서 유지되는 기준과 새로운 기준인 GDPR을 적용하는 법안을 통과시켰다.

셋째, 개정 법률은 GDPR에 의해 부여된 정보주체의 광범위한 권리를 어느 정도 제한하고 있다. 정보주체의 권한으로서 정보 액세스 권한 및 잊힐 권리 등이 제한된다. 예컨대 개인정보가 법적 또는 계약상 보존의무를 준수하기 위해 저장되는 경우나 데이터가 데이터 보안 및 보호 제어 목적에만 사용되는 경우에는 액세스 권한이 제한될 수 있다. 또한 비자동화된 개인정보가 법적인 보존기간에만 보관되고, 해당 데이터를 제공해야할 때 지나치게 부담이 되는 경우에는 정보주체의 액세스 권한이 제한될 수 있다.

넷째, 개정 법률은 민감정보[GDPR 제9조, 제2조(b),(g),(h),(i),(j)] 처리

를 허용한다. 민감정보 처리는 다음과 같은 경우, 다른 사람의 요구사항과 추가요구 사항에 따라 허용된다. 즉 ① 권리를 행사하고 사회 보장 또는 사회 보장법 분야의 의무를 준수하기 위해 처리가 필요한 경우, ② 예방적 보건의료, 종업원의 근무 능력 평가, 의료, 보건 또는 치료 또는 사회치료, 건강 또는 사회치료 시스템 및 서비스의 관리는 물론 치료 계약을 위한 경우, ③ 심각한 국경을 넘는 건강 위험으로부터의 보호와 같은 보건의료분야의 이익을 위한 경우, ④ 공공의 이익, 과학적 또는 역사적 연구 목적으로 보관하는 경우가 그것이다.

2) 비식별 조치

BDSG는 2017년 개정 전까지는 익명화(Anonymisierung)와 가명화(Pseudonymisierung) 개념을 정의하였고(§ 3 Abs. 6, 6a BDSG), 그 중 가명화를 일정한 경우 정보보안을 위한 의무로 규정하는 개별 규정도 두고 있었다. 다만 익명화의 의미와 정도에 관하여는 학설 대립이 있었다(절대설-상대설). 그러나 이들 규정은 GDPR을 지시하는 규정만을 둔 2018년 개정으로 모두 삭제되어, 지금은 존재하지 아니한다.

(2) 비식별 정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

앞서 본 바와 같이 독일에서 개인정보 보호에 관한 연방의 기본법은 연방데이터보호법(Bundesdatenschutzgesetz)인데, 같은 법은 익명화와 가명화에 대하여 규정하고 있었으나, 이들은 GDPR 입법에 따른 2017년 개정에 의하여 모두 삭제되었다. 다만 연방통계법은 여전히 익명화에 대한 규정을 두고 있다. 그밖에 비식별 조치 데이터의 활용을 위해 BDSG 제28조 제2항 제3호에 따라 가명화된 데이터는 적절한 보호조치

가 된 데이터로서 '연구소(Forschungseinrichtung)에서 수행하는 학술연구(wissenschaftlicher Forschung)'에 한하여 활용 가능하다. 정보처리의 예외를 규정하고 있는 GDPR 제83조를 반영하였다. BDSG 제28조 제6항은 민감정보라 하더라도 위 학술연구를 위한 목적 외 이용의 요건을 충족하면, 제3자 제공은 허용되지 아니하나, 목적 외 이용은 할 수 있다고 규정한다. 그러나 이때 정보처리자의 이익이 정보주체의 이익을 심각하게 침해하지 아니하여야 하고, 민감정보는 가명화되어야 한다.

그 이외에 주(Land)정부는 특정 영역을 관리하는 주 데이터보호 법률이 있다. 16개 주의 데이터보호법은 2016년 9월 GDPR의 동의 유효성에 대해 의견을 발표하였다. 그중에서 바이에른주 DPA는 정기적으로 GDPR의 주제(예 : 데이터 보호 임원, 원 스톱 샵, GDPR 하에서의 CCTV, 잊혀 질 권리, 보안 위반, 데이터 보호 영향 평가, 행동 강령, 접근권, 아동의 동의 및 국제 데이터 전송 등)에 대해 GDPR을 보완할 수 있는 지침을 발표하고 있다.

2) 법령 이외의 형태

독일에는 비식별화나 가명화와 관련하여 공식적인 가이드라인 형식의 문서는 존재하지 아니한다. 그 대신 독일 연방내무부에서 2017년 공표한 가명처리에 관한 지침 백서(Whitepaper zur Pseudonymisierung der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2017)가 있다. 독일에서 가명처리에 관하여는 일반적으로 동 백서의 내용을 인용한다. 이 문서는 익명화된 데이터(anonymisierten Daten)와 가명화된 데이터(pseudonymisierten Daten)의 차이점을 설명하면서, 가명화된 데이터는 식별자와 별도로 저장하여 사용하거나 공개적으로 사용가능한 정보에 의하여 재식별될 수 있는 데이터의 생성인 반면 익명화된 데이터는 데이터 처리 후 재식별할 수 없거나 식별되기 어려운 데이터를 일컫는다고 한다(그림 4-1).

그밖에 익명화기술에 관한 문서로는 2006년 독일연방통계청에서 제시한 워킹페이퍼 개인정보 익명화 방법(Verfaren zur Anonymisierung von Einzeldaten, 2006)을 비롯한 다수의 보고서가 있다. 이 문서는 익명화 방법을 변수의 배제와 같이 정보를 감소시키는 방법과 데이터 값을 수정하는 두 가지 방법으로 설명하고, 두 가지 방법을 혼용하는 것이 더 나은 익명화라고 제시한다. 이 문서를 포함한 익명화에 관한 다수의 보고서가 통계청에서 발간되고 있고, 민간영역의 익명화와 관련하여서도 자주 인용되고 있는데, 이는 독일 연방통계법이 통계자료를 익명화하여 민간 등에 연구목적 등으로 제공하도록 하고 있고, 실제로 연방통계청과 주(Land)통계청이 다수의 통계자료를 가공하여 제공하고 있기 때문이다.

(그림4-1) 익명화된 데이터와 가명화된 데이터의 차이
출처 : Whitepaper zur Pseudonymisierung der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2017

3. 프랑스

(1) 제도적 개괄

1) 개인정보보호 규제 일반

프랑스법상 개인정보 보호는 무엇보다도 공공과 민간부분을 아울러 개인의 사생활과 자유 보호를 위해 1978년 1월 6일 시행된 '정보처리, 파일 및 자유에 관한 법률(Loi Informatique et Libertés, Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés)'에 의하여 규율된다. 같은 법률은 1996년 유럽연합 정보보호지침의 요구사항을 반영하여 2004년 8월까지 여섯 차례 개정되었다. 특히 2004년 8월 개정은 개인정보의 개념을 기명정보(information nominative)에서 인적특성정보(donnée à caractère personnel)로 변경하였다.

같은 법은 2018년 6월 20일 법률 n° 2018-493로 다시 한 차례 개정되었다. 개정의 주요 사항은, 개인정보 규제기관인 정보처리와 자유 국가위원회(Commission nationale de l'informatique et des libertés; 이하 'CNIL'이라 한다)의 권한을 강화하여 그 효율성을 향상시키는 것이다. CNIL은 GDPR 이행의 주관행정청으로서 권한 강화를 추진 중이다. CNIL은 GDPR에 포함된 책임성의 원칙을 구현하기 위한 의견 및 권고 사항을 적용하고 지침 및 인증체계를 승인하는 등의 업무를 한다. 또한 지방 당국, 그룹 및 중소기업에 데이터 보호에 관하여 적절한 정보를 제공하여 기업이 GDPR에 따라 적절한 개인정보보호조치를 하도록 지원한다.

CNIL은 2018. 1. 23. 콘트롤러가 준수하여야 할 "개인정보 보호를 보장하는 안전조치, 보안조치 및 체계를 포함하여 위험을 다루기 위해 고안된 조치[GDPR 제35조(7)]"를 체계적으로 구현할 방법으로 구성된 가

이드라인(Un nouveau guide de la securite des donnees personnelles, 2018)을 발표하였다.

[표4-4] CNIL 개인정보 보안가이드라인의 구성

1. Raising user awareness
2. Authenticating users
3. Access Management
4. Logging access and managing incidents
5. Securing workstations
6. Securing mobile data processing
7. Protecting the internal network
8. Securing servers
9. Securing websites
10. Ensuring continuity
11. Archiving securely
12. Supervising maintenance and data destruction
13. Managing data processors
14. Securing exchanges with other organisations
15. Physical security
16. Supervising software development
17. Encrypting, guaranteeing integrity and signing

출처 : CNIL, Un nouveau guide de la securite des donnees personnelles, 2018

CNIL의 개인정보 보안가이드라인은 콘트롤러의 의무사항인(GDPR 제32조) 위험에 적절한 수준의 보안을 보장하기 위한 적절한 기술 및 조직이 수행하는 위험관리법뿐 아니라 기업이 수행할 수 있는 개인정보의 익명화 및 가명화, 처리 시스템 및 서비스의 지속적 기밀성, 무결성, 가용성 및 복원성을 보장하는 능력, 물리적 또는 기술적인 침해가 발생했을 때 적절한 시기에 개인정보에 대한 가용성 및 접근성을 복원하고 데이터 보안을 보장하기 위한 기술적·조직적 조치로서 정기적 테스트, 평가 프로세스에 관한 전략적 단계를 제시하고 있다(표 4-5).

[표4-5] CNIL 가이드라인이 권고하는 기업의 리스크 관리 단계

단계	내용
1단계	데이터와 처리된 데이터(클라이언트 파일, 계약서) 및 데이터 처리 수단을 복구한다.
2단계	각각의 데이터 처리에 따른 위험을 이해한다. - 주요 대상의 권리와 자유에 대한 잠재적인 위험의 원인(데이터의 무단 액세스, 변경 또는 손실, 내부/외부의 인적 또는 비 인적 요인) - 실현 가능한 위협(발생할 수 있는 사건) - 각 위험에 대처하는 기존 또는 계획된 조치결정(접근통제, 백업, 추적, 보안, 암호화, 익명화) - 위 요소와 관련된 위험평가(추정 가능한 크기의 예: 무시할 수 있음/ 보통/ 중요/ 최대)
3단계	계획된 조치를 실행하고 검증한다.
4단계	정기적인 보안감사 실시한다.

자료 : 연구진 작성(CNIL, Security of personal data Guideline참조 https://www.cnil.fr/fr/un-nouveau-guide-de-la-securite-des-donnees-personnelles)

또한 개인정보를 처리 할 때 체계적으로 취할 수 있는 기본적 주의사항 및 특정 상황에서 요구되는 개인정보 영향평가를 위한 실질적인 도구로서 활용할 수 있도록 체크박스로 된 평가양식을 제공한다(표4-6).

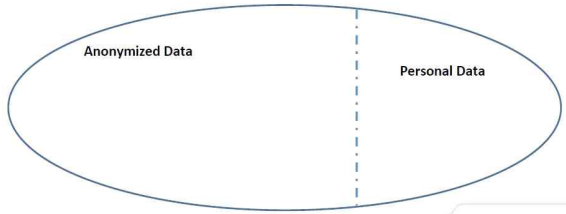
[표4-6] 조직의 개인정보 보안수준을 판단하기 위한 고려사항

	FACTSHEET	MEASURE	
1	Raising user awareness	Inform and raise awareness among individuals handling data	☐
		Write an IT charter and enforce its application	☐
2	Authenticating	Define a unique identifier (login) for each user	☐
		Adopt a user password policy conform to our recommendations	☐
		Require each user to change his or her password whenever it has been resetted	☐
		Limit the number of access attempts to an account	☐
3	Access Management	Define authorisation profiles	☐
		Remove obsolete access permissions	☐
		Carry out an annual review of authorisations	☐
4	Logging access and managing incidents	Implement a logging system	☐
		Inform users of the implementation of the logging system	☐
		Protect logging equipment and the information logged	☐
		Organise the procedures for personal data breach notifications	☐
5	Securing workstations	Organise an automatic session locking procedure	☐
		Use regularly updated antivirus software	☐
		Install firewall software	☐
		Collect the user's consent before any intervention on his or her workstation	☐
6	Securing mobile data processing	Organise encryption measures for mobile equipment	☐
		Undertake regular data backups and synchronisations	☐
		Require a confidential piece of information to unlock smartphones	☐
7	Protecting the internal network	Limit the network traffic to the bare essentials	☐
		Secure remote access to mobile computing devices via VPN	☐
		Implement the WPA2 or WPA2-PSK protocol for Wi-Fi networks	☐
8	Securing servers	Allow access to tools and administration interface only to qualified individuals	☐
		Install critical updates without delay	☐
		Ensure availability of data	☐
9	Securing websites	Use the TLS protocol and check its implementation	☐
		Check that no password or identifier are transferred via URLs	☐
		Check that the user inputs correspond to what is expected	☐
		Place a consent banner for cookies not required by the service	☐
10	Ensuring continuity	Carry out regular backups	☐
		Store the backup media in a secure place	☐
		Organise security measures for the transport of backups	☐
		Organise and regularly test the business continuity	☐
11	Archiving securely	Implement specific access methods to archived data	☐
		Destroy obsolete archives securely	☐
12	Supervising maintenance and data destruction	Record maintenance in a register	☐
		Have a responsible person from the organisation supervise work by third parties	☐
		Delete the data from all hardware before it is discarded	☐
13	Managing dataprocessors	Add a specific clause in the contracts of subcontractors	☐
		Organise the restitution and destruction conditions of data	☐
		Ensure the effectiveness of provided guarantees (security audits, visits, etc.)	☐
14	Securing exchanges with other organisations	Encrypt data before sending it	☐
		Ensure that it is the right recipient	☐
		Send the secret information separately and via a different channel	☐
15	Physical security	Restrict access to the premises via locked doors	☐
		Install anti-intrusion alarms and check them periodically	☐
16	Supervising software development	Offer parameters that respect the privacy of end users	☐
		Avoid comment zones or supervise them strictly	☐
		Carry out tests on fictional or anonymised data	☐
17	Using cryptographic functions	Use recognised algorithms, software and libraries	☐
		Keep the secret information and cryptographic keys in a secure way	☐

출처 : Security of Personal Data, the CNIL's Guides, 2018

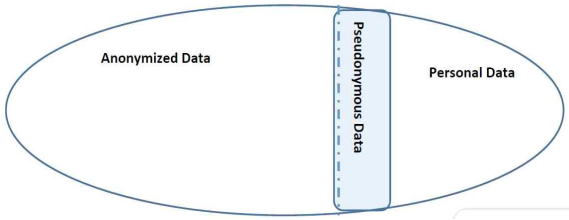
2) 비식별 조치

CNIL은 2015년 WP 29에 익명화 기술에 대한 의견서를 제출하였다(Clinical Trial Data Sharing: Methods and Experiences with De-Identification). 같은 의견서는 개인정보(Personal Data)와 익명화된 데이터(Anonymized Data)를 구분하여 파악할 경우, 익명화된 데이터는 개인정보보호 대상이 아니라고 하면서도, 이러한 인식은 가명화된(pseudonymized) 정보도 익명정보로 인식하는 오류를 범할 가능성이 있음을 지적하고 있다(그림4-2).



(그림4-2) Dichotomy between "Personal Data" and "Anonymized Data"
출처: Clinical Trial Data Sharing: Methods and Experiences with De-Identification

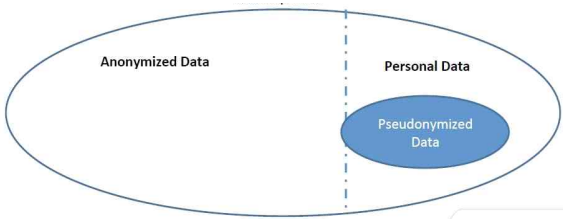
GDPR은 개인정보의 특별한 형태로 가명데이터(pseudonymized data)를 도입하였다. 그러나 가명데이터, 익명화된 데이터와 개인정보 간의 명확한 경계는 없다. 가명처리(Pseudonymisation)는 개인 데이터셋 내에 있는 일반적으로 고유한 속성을 다른 속성으로 교체하는 것을 의미한다. 따라서 자연인은 여전히 간접적으로 식별될 수 있다. 추가적인 조치 없이 가명처리만 단독으로 사용되었을 때는 데이터셋의 익명화를 달성하기 어려울 가능성이 높다.



(그림4-3) Blurred line between anonymized data and personal data
출처: Clinical Trial Data Sharing: Methods and Experiences with De-Identification

콘트롤러가 개인정보(Personally Identifiable Information)를 제거하여 익명화를 할 수는 있지만, 익명화는 가명화(Pseudonymization)와 다르다(그림 4-4). 가명처리된 정보를 익명정보와 동일시하는 것은 위험하다. 가명정보는 여전히 정보주체 식별이 가능하고 다양한 데이터셋에 걸쳐 연결을 허용하므로 결코 익명정보와 동등할 수 없다. 가명처리는 정보주체를 식별할 가능성이 크므로 개인정보보호법 체계의 적용 범위 내에 있다.

데이터의 식별성, 연결 가능성, 추론가능성을 제거하는 것만으로는 익명화 방법으로 사용하기 어렵고, 개인에 대한 환경이나 대상 및 소비자들에 대한 추론가능성은 완벽히 제거할 수 있는 대상이 아니라고 할 수 있다.



(그림4-4) Anonymized data ≠ Pseudonymized data
출처: Clinical Trial Data Sharing: Methods and Experiences with De-Identification

(2) 비식별 정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

정보보호에 관한 일반 법률은 1978년 처음 제정된 '정보기술, 데이터 파일 및 시민의 자유법(Loi Informatique et Libertés)'이다. 같은 법률에서는 익명화(anonymisation)란 단어를 직접적으로 정의하지는 아니한다. 그러나 정보처리자에 의한 데이터 처리 후에는 정보주체를 직접 또는 간접적으로 식별이 불가능해야 한다고 규정하고 있었다. 즉 개인정보는 정보처리자 또는 제3자가 '합리적으로 예상하는 모든 수단'을 동원해도 더 이상 자연인이 식별될 수 없도록 처리되어야 한다. 이러한 데이터 처리의 구체적인 개념은 CNIL의 2010년 개인정보 보안 가이드라인(Sécurité des données personnelles, 2010)에 명시되어 있다. 가이드라인은 비가역적 익명화(irreversible anonymisation)와 가역적 익명화(reversible anonymisation)를 구분한다. 비가역적인 익명화는 직접 및 간접적으로 식별되는 모든 정보를 제거하여 재식별을 불가능하게 하는 것을, 가역적인 익명화는 가명(pseudonym)으로 대체하는 처리를 말한다.

1978년 제정된 '정보처리, 파일 및 자유에 관한 법률'은 2018년 GDPR 및 경찰법(directive police-justice)을 적용하여 개정되었다(La loi n ° 2018-493, l'informatique, aux fichiers et aux libertés). 그에 따라 가이드라인도 2018년 개정되었다. 2018년 개정 가이드라인(Les Guides de la CNIL - Édition 2018)은 데이터 보안 요구를 포함하는 개인정보 보호 응용 프로그램 또는 서비스의 설계와 통합된 데이터 처리를 요구한다. 아키텍처 선택(분산 또는 중앙 집중식), 기능(익명화, 데이터 최소화), 기술(암호화) 등을 활용하여 처리 할 수 있다.

2) 법령 이외의 형태

프랑스에서는 CNIL이 DPA(Data Protection Authority)의 역할을 수행한다. CNIL의 사전승인을 요구하는 데이터 처리는 다음과 같다.

① 공공의 이익이 있거나 짧은 시간 내에 CNIL이 승인 한 익명화 절차가 적용되는 경우 민감정보 처리
② 생체인식 데이터 처리
③ 유전정보의 처리 (예방의학, 의학진단 또는 치료/ 치료에 대한 의사 또는 생물학자에 의한 처리는 제외)
④ 국민사회번호(social security number)가 포함된 데이터 처리 (인사관리를 위해 공공정부기관 및 고용주와 같이 이 번호를 처리할 권한이 있는 조직은 제외)
⑤ 개인의 사회적 어려움에 대한 평가를 포함한 데이터의 처리.
⑥ 범죄, 신념 또는 보안조치와 관련된 데이터 처리(법률을 대표하는 경우는 제외)
⑦ 입법 또는 규정 조항이 없는 사람의 권리, 서비스 또는 계약의 이익을 배제 할 수 있는 데이터 처리
⑧ 서로 다른 목적으로 만들어진 데이터베이스의 조합
⑨ 의학연구의 목적뿐만 아니라 진료 및 예방 관행 또는 활동의 평가 또는 분석을 위한 데이터 처리
⑩ GDPR 상의 유럽위원회의 표준계약사항(Standard Contractual Clauses) 또는 Binding Corporate Rules에 근거한 충분한 수준의 데이터 보호를 제공하지 않는 EU 외부 국가로의 데이터 전송

위 제1항이 명시하듯 일정한 익명화는 CNIL의 승인대상이다. 이 점은 정보기술, 데이터파일 및 시민의 자유법도 명시하고 있다.

(3) 비식별 정보, 익명정보, 가명정보 개념의 도입 과정

1) 배경 및 구체적 절차

CNIL의 WP 29 익명화에 대한 의견은 어떤 익명화 기술로도 모든 재식별 위험을 완벽히 제거하지는 못한다는 것을 인정한다. 데이터를 '익명처리' 하기 위한 각종의 기술 및 관리적 조치와 결부된 데이터 처리는 내재적으로 '잔존 위험'이 있다. 이를 적시하기 위해 '익명성' 또는 ' 익명정보' 대신 '익명처리 기법'이라는 표현을 쓰고, 데이터 처리 프로세스에 집중하였다. 익명화하기 전, 익명화 방법, 익명화 후 내용은 다음과 같다.

익명화하기 전	맞춤형 익명화방법 찾기	익명화 한 후
(a)익명화된 데이터세트의 사용을 지정할 것 (b)중요하지 않은 정보의 주요 속성을 구별할 것 (c)식별자 및 자주 사용하지 않는 값을 제거할 것	(a)여러 익명화 기술의 조합 (b)대상 선택 (c)재발견 위험을 평가하고 그 평가가 유효한지 확인 (d)가능한 모든 데이터 소스(공개 데이터, 익명화된 데이터세트)고려 (e)데이터의 유용성, 인센티브, 공격자, 내부자 등을 고려하여 익명화기술 적용	(a)익명화 기술을 게시하고 검토하기 (b)익명화 및 재식별 기술 개발

(그림4-5) 익명화 프로세스
자료 : 연구진 작성(CNIL, Security of personal data Guideline 참조 https://www.cnil.fr/fr/un-nouveau-guide-de-la-securite-des-donnees-personnelles)

충분한 수준의 익명성을 결정하는 기준은 WP 29 의견(2014년 5월)이 제시한 다음 네 질문에 답을 하고 추측할 확률이 매우 낮은지에 달려 있다(①~③ 중의 하나 혹은 ④). 어떤 제안이 다음 기준 중에서 한 가지를 충족하지 못한다면 그때마다 식별위험을 평가하여야 한다.

① 싱글아웃(Singling out) : 익명화 후에도 데이터세트 내에서 한 개인을 싱글아웃 할 수 있는가?
② 연결하기(Linkability) : 개인과 관련된 레코드를 다른 데이터 세트(동일데이터베이스) 또는 외부의 다른 두 개의 다른 데이터베이스 정보와 연결하여 개인을 식별 할 수 있는가?
③ 추론하기(Inference) : 상당한 확률로 변경되거나 제거된 값에서 원래 속성의 값을 유추할 수 있는가?
④ 위험이 허용가능한 정도로 작다는 것을 입증하기 위한 재식별화 위험(re-identification risk) 분석이 가능한가?

다음 그림은 위 ①~③ 세 가지 기본요건의 관점에 따라 데이터 처리 기법의 강점과 약점을 정리한 것이다. 세 가지 기준을 모두 충족하는 솔루션은 제대로 된 익명처리 과정이라고 할 수 있다. ④번의 의미는 콘트롤러와 제3자가 합리적으로 가장 가능성이 큰 재식별 수단을 가지고 재식별을 시도할 때 이를 막는 솔루션이 되어야 한다는 뜻이다.

	Is Singling out a risk	Is Linkability a risk	Is Inference a risk
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation or K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not
Hashing/Tokenization	Yes	Yes	May not

(그림4-6) Evaluation of the anonymization techniques
출처: Clinical Trial Data Sharing: Methods and Experiences with De-Identification

2) 논의 사항

특수한 범주의 데이터처리(Processing of special category data)는 금지된다(제9조). 하지만 다음 면제조항 중 하나가 적용되는 경우에는 사실상 합법적인 처리를 할 수 있도록 설정되어야 하는 이차활용 자료이다.

① 데이터 주체의 명시적인 동의가 있는 경우
② 의무 이행 및 고용, 사회 보장 및 사회 보장 제도상의 권리 행사를 위해 필요한 사회 보장법 또는 단체 협약 경우

③ 데이터 주체 또는 신체적 또는 법적으로 동의 할 수없는 다른 자연인의 중요한 이익을 보호하기 위해 필요한 경우
④ 제한된 상황에서 특정 비영리 단체와 관련이 있는 경우
⑤ 처리가 데이터 주체에 의해 명시적으로 공개 된 개인정보와 관련이 있는 경우
⑥ 법적 청구의 설정, 행사 또는 방어 또는 법원이 법적 능력을 행사하는 경우 처리가 필요한 경우
⑦ 추구하는 목표와 적절한 안전 조치에 비례하여 연방 또는 회원국 법에 근거하여 실질적인 공공 이익을 이유로 필요한 경우
⑧ 예방 또는 직업 의학에 필요한 경우 직원의 근무 능력 평가, 의료 진단, 건강 또는 사회 복지의 제공 또는 보건 또는 사회 케어 시스템 및 서비스의 관리에 대한 처리
⑨ 건강에 대한 심각한 국경 위협으로부터 보호하거나 보건 의료 및 의료 제품 및 장치에 대한 높은 기준을 보장하는 것과 같이 공중 보건 영역에 대한 공공의 이익을 이유로 필요한 경우
⑩ 공공의 이익, 과학적 또는 역사적 연구목적 또는 제89조(1)에 규정된 제한에 따른 통계적 목적으로 보관이 필요한 경우

이차적 목적을 위한 처리(Processing for a Secondary Purpose)는 데이터를 처음 수집할 때 정보주체에게 동의 받지 않은 새로운 용도와 목적으로 데이터를 사용하기 위한 것이다. 이차적 목적을 위한 처리는 증가하고 있지만, 이는 잠재적으로 목적제한 원칙과 상충된다.

통계자료와 같은 익명처리정보(anonymised data)는 연구 분야에서 꾸준히 사용될 수도 있어 새로운 개인정보보호 문제를 제기하고 있다. 따라서 콘트롤러는 익명처리를 일회성 활동으로 생각해서는 안 되며 수반되는 주요 위험을 정기적으로 재평가해야 한다. 정보주체의 권리를 보호하기 위하여 비식별 조치 프로세스를 통해 개인정보가 처음 수집된 목적과 호환되는지 GDPR 제6조(4)를 적용하여 콘트롤러가 확인한다. 이때 확인하여야 하는 사항은 다음과 같다.

① 원래 목적과 새로운 목적사이의 모든 연관성(link)
② 데이터가 수집 된 맥락(context)
③ 데이터의 특성, 특히 범죄 유죄 판결과 관련된 데이터 또는 특수 범주의 데이터가 처리되는지 여부(추론은 새로운 목적이 호환 가능하다는 견해를 형성하는 것이 훨씬 더 어렵다는 것을 의미함)
④ 새로운 데이터 처리를 통해 데이터 주체를 위해 발생 가능한 결과
⑤ 암호 또는 가명을 포함하는 적절한 안전장치의 존재여부

익명처리 기법의 타당성과 관련하여서는 해당 기법에 의해 익명처리된 데이터의 위험요인을 고려하여 위험의 심각성과 발생 가능성을 평가하여야 한다. CNIL은 GDPR이 강조하는 데이터보호 영향평가[제35조(7), 서문 (84) 및 (90)]를 위하여 GDPR이 제시하는 개인정보 영향평가의 최소특성을 적용하여 실제적인 데이터 보호 영향평가를 실현하고자 개인정보 영향평가 지침서를 제공하였다(CNIL, PIA privacy impact assessment). 구체적으로 말하자면, 아래 핵심 요인에 대해 모든 관련 상황에 비추어 실질적인 평가가 수행되어야 한다.

① 개인정보 수집 목적과 추가 처리 목적 사이의 관계
② 개인정보가 수집된 배경과 그 추가 사용에 대한 정보주체의 합리적인 기대
③ 개인정보의 성격 및 정보주체에 대한 추가 처리의 영향
④ 개인정보의 공정한 처리를 보장하고 정보주체에게 미치는 부당한 영향을 방지하기 위해 개인콘트롤러가 취할 안전조치

(그림4-7) PIA 수행을 위한 반복프로세스
출처 : Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679

(4) 비식별 정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

1978년 정보처리, 파일 및 자유에 관한 법률은 수차례 개정을 거치다가 2016년 디지털공화국법(Loi n° 2016-1547 du 18 novembre 2016 de modernisation de la justice du XXIe siècle)으로 개정되었다. 2016년 법률은 공공의 이익과 관련된 DB의 자동개방, 공공부문 연구자를 위한 암호화된 접근 허용, 공적 연구결과물에 대한 자유로운 접근과 데이터 마이닝을 허용한다(https://www.legifrance.gouv.fr/eli/loi/2016/11/18/JUSX1515639L/jo). 정보 공개정책의 범위는 공역무를 위탁받은

민간기관이 수집한 정보에까지 확대되었다. 다른 한편으로 시민, 사용자, 기업가, 공무원, 소비자 등 전체가 지지하는 현대적인 디지털 정책으로 전환하였다.

GDPR을 국내법으로 적용하게 위해 CNIL은 2016년 6월부터 데이터 이동성(data portability), 데이터 보호 담당관, 데이터 보호 영향평가 및 인증에 대한 공개 상담을 시작하였다. 그리고 GDPR의 결과를 분석하고 기존의 프랑스 데이터 보호법을 재구성하기 위해 법무부가 주도하는 태스크 포스를 만들었다(https://www.dataguidance.com/france-cnil-takes-pragmatic-approach-gdpr-enforcement/). 특히 제재와 관련하여 유럽 데이터 보호당국 간 협력절차에 관한 우려와 벌칙의 구체적 수치에 대한 논의를 했다. 특별히 WP 29에 의해 유럽연합 회원국들 간에 규칙을 균일하게 적용 할 수 있도록 규제에서 언급된 개념을 명확하게 하였다. 2017년 2월 23일부터 GDPR 이행을 위한 실천계획에서 동의, 프로파일링 및 데이터 유출 통지에 관한 온라인 공개상담을 두 차례에 걸쳐 실시했고, 세 번째 주제인 데이터 국제이전과 투명성에 관한 제3차 온라인 공개상담을 실시하였다.

마침내 2018년 GDPR 원칙의 타당성과 관련성(비례성, 적법성, 인격권)을 적용하여 유럽의 데이터 보호원칙을 신실하게 유지하면서 데이터를 처리하는 공공기관 및 민간기관의 의무를 근본적으로 수정한 법률을 공표하였다(Loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles). 그중에서 개인정보보호에 관한 주요 내용은 다음과 같다.

① 개인정보보호법의 적용대상은 프랑스 내 컨트롤러의 설립 여부와는 상관없이, 프랑스에 거주하는 모든 정보주체이다.

② CNIL은 기관의 의견 및 권고안 제시, 인증제도 승인 등 GDPR이 규정하는 책임성의 원칙을 이행할 의무를 가지며, 집행 권한 및 과징금 부과 권한을 행사할 수 있다. CNIL이 GDPR 하에서 제공되는 '협력 메커니즘'을 통해 EU전역의 다른 국가 데이터보호기관과

협력 할 수 있다. 또한 CNIL은 경우에 따라 기업의 연간 글로벌 매출액의 4%까지 또는 2천만 유로 중 가장 높은 금액으로 집행조치를 취하고 재정적 벌칙을 발행 할 수 있다.

③ 개인정보처리 침해사고 시, 정보주체는 기존규정에 따른 즉각적인 처리 중단요청에 물질적·정신적 보상청구를 목적으로 하는 집단소송을 제기할 수 있다. 집단소송은 물질적 처리 및 도덕적 피해를 포함하여 침해된 데이터에 대한 손해배상을 목적으로 한다.

④ 프랑스의 정보주체는 데이터 보호 또는 소비자 보호 분야에서 활동하는 직원대표 또는 비영리조직에게 자신을 대신하여 불만사항을 접수하도록 명령 할 수 있다. 데이터 침해에 대해서는 컨트롤러 또는 프로세서에 대해 조치를 취할 수 있다.

⑤ 웹사이트 및 앱에 대한 접근과 기타 '정보사회서비스' 및 건강정보 처리를 위해 동의를 받을 수 있는 아동의 법적 나이는 15세이다.

4. 핀란드

(1) 제도적 개괄

1) 개인정보보호 규제 일반

2016. 2. 17. 핀란드 법무부는 모든 정부부처의 대표와 민간부문의 대표들로 구성된 실무그룹을 구성하여 핀란드 내의 GDPR 이행을 담당하게 하였다. 그 결과 1999년의 핀란드 데이터 보호법(Henkilötietolaki, 523/1999)은 폐지되고, 2018년 3월 핀란드 정부가 제안한 데이터 보호법 및 관련 법안(HE 9/2018 vp)이 GDPR과 병행하여 적용된다. 동 법을 통해 EU 내에 위치한 콘트롤러나 프로세서가 수행하는 데이터 처리가 규율된다. 나아가 제안된 법안은 적용 가능한 외국법이 공공의 이익을 위한 유지보존, 과학적 또는 역사적 연구 또는 통계적 목적을 위해 GDPR에 따라 데이터 처리를 위축시키는 경우 잠재적으로 적용 가능한 외국 법령에도 불구하고 우선 적용된다. GDPR을 반영한 것이다. 핀란드의 국가 재량에 관한 주요 조항은 다음과 같다.

① 고용 환경에서의 개인정보처리에 대한 엄격한 제한 ; 2009년 3월 개정된 전자통신프라이버시보호법(516/2004; amendments up to 125/2009 included)을 근거로 영업비밀 유출이나 인터넷설비의 오남용을 의심할 만한 사유가 있는 상황에서 고용주가 직원의 이메일 로그기록을 검열할 수 있다. 다만 송수신인의 이름, 송신시간, 첨부파일의 이름 및 크기 등 로그기록만 검열할 수 있을 뿐 이메일 내용은 열람하지 못한다. 즉 직원 모니터링, 기술감시 및 직원 전자메일에 대한 액세스는 엄격하게 규제된다(http://www.finlex.fi/fi/laki/kaannokset/2004/en20040516.pdf).

② GDPR이 시행 된 후에도 직원의 개인정보처리에 대한 현행 상태

를 유지한다. 즉 기업은 식별정보를 처리하기 전에 식별정보처리담당자의 이름과 직무를 정의하여야 한다. 식별정보는 기업의 정보보안 책임자만 처리할 수 있다.

③ 기업은 식별정보처리의 이유와 절차를 노조와 협의하고 근로자, 근로자대표 또는 이용자에게 식별정보처리의 결과를 고지해야 한다. 또 식별정보 수동처리에 관한 보고서를 근로자 대표에게 제공하여야 한다(516/2004; amendments up to 125/2009 included, 제13조). 몇 가지 변경 사항으로서 범죄기록 데이터, 카메라 감시 및 고용주를 대상으로 한 제재범위의 변경이 있다.

④ 개인신원번호를 포함하는 특수 데이터 관련 ; 핀란드는 개인신원번호 처리에 엄격한 제한을 설정하고 있는데, 제안된 법에 따르면, 개인신원번호는 일반적으로 정보주체의 동의를 얻거나 처리가 법률상 요구되는 경우에만 처리할 수 있다. 처리가 요구되는 경우란 정보주체 또는 컨트롤러의 권리 또는 의무를 실현하기 위하여 경우, 법률에 의해 설정된 작업을 수행하기 위하여 정보주체의 식별이 필수적인 경우, 역사적 · 과학적 또는 통계적 연구의 목적으로 필요한 경우, 금융서비스 및 의료 분야의 특정 활동인 경우이다.

⑤ 연구에 사용되는 개인정보는 연구참여자에 관한 정보에 국한되지 않는다. 연구 데이터에는 연구대상의 가족 및 친구 또는 기타 제3자와 관련된 식별자가 포함될 수도 있기 때문에 이러한 사람들과 관련된 정보도 개인정보로 간주된다. GDPR의 정의(제4조 제1항)를 적용한 것으로서 개인정보는 식별되거나 식별 가능한 자연인과 관련된 모든 정보를 의미한다. 자연인은 특히 이름, 식별 번호, 위치 데이터, 온라인 식별자 또는 신체적, 생리학적 기능과 관련된 하나 이상의 요소와 같은 식별자'를 참조하여 직접 또는 간접적으로 식별할 수 있는 경우에 식별가능하다고 간주한다. 또한 유전적, 정신적, 경제적, 문화적 또는 사회적 정체성도 자연인과 관련된 정보로 인정한다.

⑥ 건강관련 데이터 처리를 위해서는 GDPR에 명시적으로 나열되지 않은 특정 목적을 위해 건강관련 데이터 처리를 승인하고 있다. 예를

들어, 보험회사는 보험회사의 책임을 결정할 때 이러한 특수범주의 데이터를 처리 할 수 있다. 의료 서비스 제공자도 이와 유사한 허가를 받아야 한다. 그 이외 아동 데이터 처리를 위한 13세 연령 제한과 벌금조항, 형사상의 제재, 바이오 뱅킹의 지속, 임금통계 및 족보 작성에 관한 조항이 있다.

2) 비식별 조치 가이드라인

유럽연합 WP 29 의견과 연구 분야에 정보를 제공하는 것과 관련된 부문별 법률(http://tukija.fi/en/publications1, 목록은 표 4-7 참조)을 적용하여 2018년 가을에 연구자를 위한 가이드라인(Finnish Social Science Data Archive, Data management guidelines, 2018)이 발표되었다. 동 가이드라인에서는 식별자 유형을 구분하고, 각 유형에 적절한 익명화 방법을 제시하고 있다.

[표4-7] 관련 부문별 법률 목록

▻ Medical Research Act (488/1999, amended 295/2004, 780/2009, 794/2010 and 143/2015)
▻ Government Decree on Medical Research (986/1999, amended 313/2004)
▻ Government Decree on the Amendment of the Decree on Medical Research (65/2016, only in Finnish)
▻ Government Decree on the National Committee on Medical ▻ Research Ethics (820/2010)
▻ Decree of the Ministry of Social Affairs and Health on the Fees Charged for Opinions of the National Committee on Medical ▻ Research Ethics and Regional Ethics Committees (1168/2014)
▻ Decree of the Ministry of Social Affairs and Health on Clinical

Trials on Medicinal Product (841/2010)

▻ Decree of the Ministry of Social Affairs and Health on the Compensation for Research Participation (82/2011)

▻ Biobank Act (688/2012)

▻ Government Decree on Consent for Biobank (643/2013)

▻ Decree of Ministry of Social Affairs and Health on Notification of Biobank (649/2013)

▻ Act of the Medical Use of Human Organs and Tissues (101/2001, amended 547/2007, 778/2009, 653/2010, 336/2011, 689/2012, 277/2013)

자료 : 연구진 작성(http://tukija.fi/en/publications1 참조)

자연인과 관련된 모든 정보는 개인정보가 될 수 있다. 즉 개인정보는 객관적 또는 주관적 일 수 있고, 정보가 사실인지 또는 검증 가능한지 여부는 중요하지 않다. 그리고 개인 식별성은 개인 또는 개인의 신체적, 정신적, 정신적, 경제적, 문화적 또는 사회적 정체성에 특정한 하나 이상의 요인에 기초하여 발생할 수 있다. 또한 진술, 의견, 태도 및 가치 판단이 포함된다. 위 가이드라인은 아래와 같이 다양한 유형의 식별자를 나열하면서, 그 종류를 직접식별자, 강한 간접식별자 및 간접식별자로 구분하고, 처리하는 가장 쉬운 방법으로 식별자 제거(remove), 가명으로 변경(change), 범주화(categorise)를 제시한다.

① 직접 식별자(direct identifier) : 개인을 식별하기에 충분한 정보에는 개인의 성명, 사회 보장 번호, 이름이 포함 된 이메일 주소, 생체 인식정보(지문, 얼굴 이미지, 음성 패턴, 홍채 스캔, 손 모양 또는 수동 서명).

② 강한 간접적인 식별자(strong indirect identifier) : 비정상적인 직종이나 직위, 매우 드문 질병, 한 번에 한 사람만 차지하는 직위, 희소한 사건, 학생의 ID 번호, 보험 또는 은행 계좌번호, 컴퓨터 IP

주소

③ 간접적인 식별자(indirect identifier) : 나이, 성별, 교육, 취업 상태, 경제활동 및 직업상 지위, 사회 경제적 지위, 가구구성, 소득, 결혼상태, 모국어, 민족적 배경, 직장 또는 지역 변수, 우편 번호, 지역, 자치 단체, 공통의 직종, 우편주소, 전화번호, 차량등록번호, 개인출판물에 대한 서지 인용, 웹 주소

④ 민감정보: * 로 표시한다.

[표4-8] 식별자 유형과 익명화 방법

Identifier type	Direct identifier	Strong indirect identifier	Indirect identifier	Anonymisation method
Personal identification number	x			Remove
Full name	x			Remove/Change
Email address	x	x		Remove
Phone number		x		Remove
Postal code			x	Remove/Categorise
District/part of town			x	Categorise
Municipality of residence			x	Categorise
Region			x	(Categorise)
Major region			x	
Municipality type			x	
Audio file	x			Remove
Video file displaying person(s)	x			Remove
Photograph of person(s)	x			Remove
Year of birth		x		Categorise
Age			x	Categorise
Gender			x	
Marital status			x	
Household composition			x	(Categorise)

Occupation		(x)	x	Categorise
Industry of employment			x	
Employment status			x	
Education			x	Categorise
Field of education			x	
Mother tongue			x	Categorise
Nationality			x	(Categorise)
Workplace/Employer		(x)	x	Categorise
Vehicle registration number		x		Remove
Title of publication		x		Categorise
Web page address		(x)	x	Remove
Student ID number		x		Remove
Insurance number		x		Remove
Bank account number		x		Remove
IP address		x		Remove
Health-related information *		(x)	x	Categorise/Remove
Ethnic group *		(x)	x	Categorise/Remove
Crime or punishment *			x	Categorise/Remove
Membership in a trade union *			x	Categorise
Political or religious allegiance *			x	Categorise
Other position of trust or membership		(x)	x	Categorise/Remove
Need for social welfare *			x	Categorise/Remove
Social welfare services and benefits received *			x	Categorise/Remove
Sexual orientation *			x	Remove

출처 :
http://www.fsd.uta.fi/aineistonhallinta/en/anonymisation-and-identifiers.html

위 가이드라인은 연구 데이터에 대한 구체적인 익명화 기술에 중점을 두고 있다. 익명화의 첫 번째 단계는 대개 직접식별자 및 강력한 간접 식별자를 데이터에서 제거하는 것이다. 모든 유형의 데이터에 적합한 단일 익명화 기술은 없다. 익명화는 데이터 기능, 환경 및 유틸리티를 고려하여 항상 사례별로 계획되어야 한다. 데이터는 데이터의 '나이', 민감도, 데이터 주체의 수 및 데이터가 사용되는 맥락에 따라 익명화 한 후 데이터를 사용할 수 있는 방법을 고려해야 한다. 데이터를 누가 언제, 어디에서 사용하는지, 외부의 어떤 출처 데이터를 사용할 수 있는지, 데이터의 실제 저장범위는 어떠한지를 포함하여 익명성을 신중하게 계획하고 관련 익명화 기술 및 프로세스를 문서화해야한다. 모든 익명화 기술에는 장점과 한계가 있으므로 데이터 품질과 유용성에 미치는 영향을 숙지해야 한다. 가령 일반화 후에도 특정 범주에 여전히 변수를 연결할 수 있다면 문제가 되므로 변수의 모든 값을 범주화하여 변수 간의 관계를 판별하기 어렵게 하는 방식이 있다. 무작위화 또한 변수의 분포와 상관관계에 영향을 미칠 수 있기 때문에 관측치가 비교적 적은 경우(1% 미만)에 유용하다.

위 가이드라인은 직접식별자 및 강력한 간접식별자를 제거하는 것만으로 익명화하는 것은 거의 불가능함을 지적한다. 간접식별자의 수와 세부 정보 수준은 익명화 방법의 선택을 하는데 영향을 준다. 숫자가 많을수록 정확도가 높을수록 더욱 익명성을 고려해야 한다. 예를 들어 데이터에 포함 된 정보는 다른 출처에서 사용할 수 있는 데이터도 함께 고려해야한다. 다른 출처의 정보를 사용하는 경우에도 개인을 식별할 수 없도록 데이터를 처리해야 한다.

배경정보 변수는 항상 함께 고려해야 한다. 거주지역의 정보를 데이터에 남기고자하는 경우 식별을 방지하기 위해 해당 개인과 관련된 기타 배경정보(직종, 작업장, 교육, 연령 등)를 세밀하게 조정해야 한다. 반면 참가자의 직업과 연령에 대한 정보가 연구에 중요하다면 참여자와 관련된 지리 정보를 분류해야 한다. 온라인(공개파일, 조직의 웹 사이트 등)에서 볼 수 있는 정보에서 어떤 종류의 간접 식별자가 발견되는지 또한

고려해야 한다.

성공적인 익명화를 위해 직접 식별자를 제거한 후에는 데이터의 간접 식별자여부를 검사하고 이를 근거로 개인을 식별할 수 있는지 평가해야 한다. 모든 종류의 정보에 대한 오픈 액세스가 빠르게 증가하는 오늘날 잔여 위험평가(residual risk assessment)를 정기적으로 수행하는 것이 중요하고, 이전에 익명화된 데이터가 지금도 익명으로 남아 있는지를 확인하여야 한다. 양적 및 질적 데이터의 익명처리 프로세스를 위하여 다음과 같은 질문을 할 수 있다.

① 어떤 종류의 직접식별자 또는 간접식별자가 데이터에 포함되어 있는가?
② 데이터에 희귀하거나 독특한 것이 관찰되는가?
③ 데이터의 어떤 정보가 개인을 식별하기 위해 연결될 수 있는가?
④ 다른 출처의 정보를 데이터에 연결하여 식별이 가능한가?
⑤ 익명화처리에서 데이터의 어떤 특성이 희생되는가?

2014년 5월 발표된 유럽연합의 WP 29 의견서에 따라 여러 익명화 기술을 적용한다. 양적 데이터에 대한 익명화(Anonymisation of quantitative data) 기술은 주로 일반화(generalisation)와 무작위화(randomisation)의 두 가지를 활용한다. 일반화는 정보를 비가역적으로 제거하거나 데이터 주체의 속성값을 범주화하거나 조잡하게 함으로써 규모 또는 순서가 희석되는 것이다. 무작위화는 데이터에 "잡음"을 추가하여 불확실성을 증가시킨다.

질적 데이터의 익명화(Anonymisation of qualitative data)의 출발은 텍스트 데이터셋을 익명으로 만드는 것으로서 연구 참가자의 연락처 세부 정보 및 배경정보 등 식별자가 포함된 배경 데이터를 삭제하는 것이다. 식별자를 제거하거나 편집할 때 데이터에 대한 모든 변경사항을 명확히 표시한다. 변경된 내용은 [변경된 텍스트] 또는 [[변경된 텍스트]]와 같이 대괄호 또는 이중대괄호로 표시 할 수 있다.

가명화는 가명으로 이름을 바꾸는 처리(Replacing personal names with pseudonyms)이다. 이는 식별자가 완전히 삭제될 때까지 데이터 처리를 하는 것이 아니다. 적절한 명사를 가명으로 바꾸는 것은 정성적 데이터에서 가장 일반적으로 사용되는 기술이다. 연구팀은 연구프로젝트 전체에서 가명을 선택하고 사용하는 데 있어 일관성을 유지해야 한다. 모든 팀 구성원이 사용할 수 있는 스프레드시트 파일로 이름 및 가명 목록을 유지·관리할 수 있다. 동일한 가명은 발췌된 데이터에서도 사용되어야 한다. 고유한 이름을 익명으로 지정하는 경우 이름을 모두 삭제하거나 [x] 또는 [--]와 같은 문자 또는 문자열로 바꾸는 대신 가명을 사용하는 것이 내부 데이터의 일관성을 유지하는 차원에서 더 유리하다.

데이터셋은 정치, 사업 또는 기타 업무 관련 활동으로 공적으로 알려진 사람들에 대한 언급을 포함할 수 있다. 그들은 가명으로 변경되지 않는다. 또한 정치, 비즈니스 또는 기타 업무관련 분야의 활동으로 공개적으로 알려진 사람들의 이름은 가명으로 변경되지 않는다. 이 경우에는 가명과 분류를 함께 사용해야 한다(예 : [지역 정치인]). 분류의 종류는 다음과 같다.

① 고유명사 분류(Categorizing proper nouns)

가명의 사용이 모든 고유명사에 항상 필요한 것은 아니다. 한두 번만 데이터에서 언급되고 내용을 이해하는 데 필수적인 중요성이 없는 사람들을 위한 가명을 만들 필요가 없다. 이름을 역할(예 : [여성], [남성], [자매], [아버지], [동료, 여성], [이웃, 남성])로 대체함. 또는 이름을 카테고리(예 : [주거 지역], [고등학교], [고향])으로 대체할 수 있다. 사업장이 언급되거나 데이터에서 간접 식별자를 구성 할 수 있는 사업장에 관한 다른 정보가 있는 경우, 핀란드의 산업분류표를 사용하여 해당 이름을 범주로 대체하거나 [법률 사무소], [축구 클럽], [레스토랑] 등으로 일반화할 수 있다.

② 국가통계의 분류(Classification of Statistics Finland)

장소 명을 [인구 센터], [지역], [마을], [레스토랑] 등과 같은 좀 더 일반적인 표현으로 대체한다. 거주지가 밝혀지지 않을 것으로 결정되면 배경정보와 텍스트 및 데이터 모두에서 거주지와 관련된 지리적 정보를 삭제하고 일반표현으로 대체한다.

③ 민감정보의 변경 또는 제거(Changing or removing sensitive information)

민감정보는 제거 또는 변경해야 한다. 예를 들어 '에이즈(AIDS)'는 [중증 장기 질병]으로 변경 될 수 있고, [질병]으로 언급 될 수 있다. 예를 들어, 만약 연구가 중증환자의 삶에 관한 것이라면, 정보를 변경하는 것보다 다른 익명화 방법을 사용하여 노출 위험을 낮추어야 한다. 만약 '질병'이 '심각한 장기간의 질병'임을 처음부터 추론할 수 있다면(예 : a)응답자가 우연히 언급한 경우 b)정보가 주제와 관련이 없는 경우 c)민감정보의 공개위험이 예상되는 경우) 데이터를 제거하거나 일반화해야 한다.

④ 식별자 값 변경(Changing values of identifiers)

때로는 정보를 왜곡하여 질적인 데이터를 익명화 할 수 있다(예 : 출생일이 내용을 이해하는 데 중요 한 경우).

⑤ 메타데이터 제거(Removing hidden metadata from files)

익명화 과정에서 아카이브 파일에 식별자가 될 수 있는 숨겨진 메타데이터가 포함되어 있는지 확인하는 것이 중요하다. 숨겨진 메타 데이터는 위치정보 및 장치소유자 또는 사용자 프로필에 대

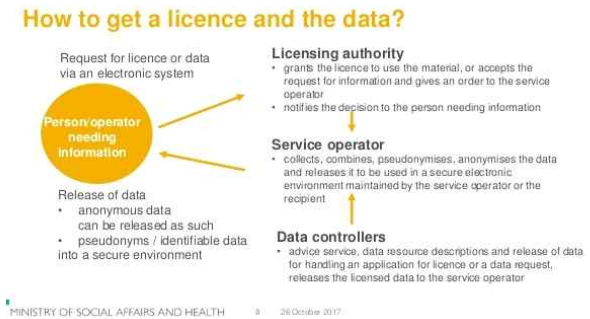
한 정보로 구성된다. 메타데이터는 파일을 만들 때 저장 될 수 있지만 편집할 때도 저장되고 텍스트 또는 이미지 형식의 데이터는 정보 주체 자신이 만든 파일로 구성되기도 한다. 이러한 경우 메타데이터에서 식별위험성이 높다.

(2) 비식별 정보, 익명정보, 가명정보 개념의 도입 과정

1) 비식별 조치의 도입 배경 및 절차

2017년 10월 핀란드 정부는 '건강과 사회적 데이터의 2차 활용을 위한 법률(FI: laki sosiaali- ja terveystietojen toissijaisesta käytöstä, HE 159/2017)을 입법 제안하였다. 법안의 목적은 통계, 연구, 혁신 및 개발, 교육 및 지식관리 목적으로 건강 및 사회 데이터의 이차 활용의 조건과 요청에 대한 데이터 처리를 합리적으로 변화시키는 데 있었다.

구체적인 내용으로 공중보건 데이터와 같은 특수한 범주의 데이터는 공익을 위한 유지보존, 과학적 · 역사적 연구목적 또는 통계목적으로 처리되거나 보관될 수 있다고 규정한다. 하지만 GDPR 제5조의 원칙에 따라 건강 및 사회데이터의 모든 처리과정을 주의 깊게 평가해야 한다. 가령 허가당국(licensing authority)이 건강 및 사회 데이터의 이차적 활용과 관련된 데이터 요청을 처리하고 허가를 부여한다. 이때 사용 목적에 맞추어 데이터 요청을 받고, 다른 레지스터에서 관련 데이터를 가져와서 데이터를 직접 편집, 결합 및 사전처리 및 전송하여야 한다(그림 4-8).



(그림4-8) 라이선스 및 데이터를 얻는 방법
출처: Government proposal of 26 October 2017, Ministry of Social affairs and Health, Finland
(https://stm.fi/en/article/-/asset_publisher/sosiaali-ja-terveystietojen-tietoturvallinen-hyodyntaminen)

2) 논의 사항

가명정보는 개인정보로 간주된다. 사용이 적절하고, 계획이 정당하며, 데이터 처리에 법적 근거가 있는 경우 식별 가능한 데이터라도 과학적 연구에 사용될 수 있다. 연구에 참여하는 사람의 개인 식별번호 대신 케이스 ID를 가지고 종단 연구에 활용하지만, 연구팀은 연구 참여자의 데이터를 연결하는 데 사용할 수 있는 키(key)를 가지고 있다. 암호해독키가 두 번 코딩 되더라도 해당 데이터는 익명화된 것이 아니다. 가명화된 것이며, 코딩 및 이중 코딩(double coding)은 분석에서 식별자 사용을 막는 데 유용하다.

가명을 생성하는 데 사용된 원래 의미와 기술에 대한 정보는 가명화된 데이터와 조직적·기술적으로 분리되어 있어야 한다. 조직적 조치는 물리적 환경 보안 및 액세스 제한에 관한 문서를 갖추는 것이며, 기술적 조치는 보안 데이터 저장 및 암호화 같은 것이다. 만약 별도로 보관된 정보(암호 키, 데이터 및 데이터의 가명부여에 사용된 기술에 대한

정보)가 파괴되면 가명 데이터가 익명화되는 방식이다. 합리적 노력으로 원본 데이터에 연결할 수없는 경우 익명으로 간주한다.

연구 참여자의 관점에서 볼 때, 개인정보를 처리하는 것은 외부인(예 : 가까운 지인들, 고용주 또는 당국)에게 공개되는 것과 같은 기밀 정보의 노출을 의미한다. 따라서 데이터 처리는 철저히 계획하고 신중하게 실행해야한다. 데이터를 처리 할 때 데이터 최소화, 익명화 및 데이터 관리 지침에 제시된 다양한 익명화 방식을 활용목적에 맞게 조정할 수 있다. 익명화는 데이터를 공유하고 재사용할 수 있게 만드는 한 가지 방법이고, 필요한 경우 데이터 보안 솔루션을 통해 데이터를 보다 안전하게 보호할 수 있다.

더 이상 연구를 수행 할 필요가 없는 데이터는 가능한 한 빨리 삭제해야 한다. 가령 이름, 주소 및 기타 유사한 식별자는 더 이상 연구를 수행 할 필요가 없는 경우 즉시 제거해야 한다. 개인 식별번호를 사용하여 데이터를 연결한 경우 더 이상 필요하지 아니하면 삭제해야 한다.

(3) 비식별 정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

핀란드 아카데미(Academy of Finland)는 연구보조금 신청자가 연구계획서에 데이터 수집 방법 및 데이터 관리계획을 설명하게 하고 있다. 데이터 관리계획은 연구 프로젝트가 데이터를 수집하고 사용하는 방법과 연구 완료 후 데이터가 저장되어 다른 연구원에게 제공되는 방법을 말한다. 데이터 보호 및 보안은 전체 데이터 수명주기 동안 유지되어야 한다. 데이터의 재사용이란 원래 수집된 목적 이외의 목적으로 데이터를 사용하는 것이다. 즉 데이터 이용자는 원래 데이터 작성자와는 다른 조사 및 분석방법을 사용한다. 데이터를 재사용하는 경우 연구 데이터는 데이터 및 관련 메타데이터에 접근 가능하고 신뢰할 수 있는 상태로 유지되어야 한다. 연구 데이터를 위한 비식별 조치 계획은 다음과 같다.

① 서면으로 익명화 계획을 세운다.
② 실제 데이터 파일에 포함되지 않지만 익명화 또는 삭제해야 할 식별자가 포함될 수 있는 배경정보를 고려한다(참가자의 연락처, 설문용 서식 등).
③ 제3자와 관련된 식별정보가 데이터에 보관되지 않았는지 확인한다.
④ 통계 패키지를 사용하여 정량 데이터의 익명화를 수행한다.
⑤ 익명 텍스트 데이터의 경우 찾기 및 바꾸기 명령을 사용하여 한 번에 하나씩 변경한다.
⑥ 타인의 성명(성/이름, 닉네임)이 한 사람을 가리키는 데 사용되었는지 확인한다(예: Elizabeth가 Eliza, Beth, Bess, Libby이라고도 함).
⑦ 계획단계에서 각 고유명사 앞에 기호 또는 특수문자를 일관되게 입력하여 향후 텍스트파일의 익명성을 유지하는데 시간을 절약한다.
⑧ 철저한 계획을 하고 보상을 받는다.

혁신을 위한 핀란드 기금운용처(Finnish Funding Agency)는 미래에 데이터를 효율적으로 사용할 수 있는 방식으로 계획·관리되는 연구를 지원한다. 데이터 및 데이터 공유에 대한 오픈 액세스는 연구자 간의 연구 및 협력 혁신을 장려하고, 다른 과학 분야에서 생성 된 데이터를 효율적 사용을 촉진한다. 대규모 설문조사의 데이터에는 원래 연구에서 분석되지 않은 자료가 포함되는 경우가 많다. 데이터 재사용은 데이터의 중복 수집을 피하고, 만나기 힘든(hard-to-reach) 사람들이나 취약한(vulnerable) 집단에 대한 데이터 수집을 최소화하게 해준다. 핀란드 아카데미, 연구 무결성에 관한 핀란드 자문위원회, 핀란드 대학(UNIFI)은 공동으로 연구출판물의 필수조건으로서 연구 데이터의 생산 및 배포를 위한 정책을 마련하였다(http://web.abo.fi/personal/kurskatalog/material/Varantola07042017.pdf). 출판물을 인용할 때 연구자는 새로운 연구를 위해 원래 연구자가 생성한 데이터를 함께 인용하고, 보관된 데이터를 사용하도록 하는 것이 그 핵심이다. 유럽연합과 핀란드 학술

원 및 총리실이 자금을 지원하는 연구 프로젝트는 약 100건이며, 매년 약 800개의 과학논문이 발표되고 있다. OECD 가이드라인은 공공기금으로 조성된 연구 데이터에 대한 오픈 액세스는 국제 연구협력을 강화하는 중요한 조건임을 명시하고 있다http://www.oecd.org/sti/sci-tech/38500813.pdf).

5. 영국

(1) 제도적 개괄

1) 개인정보보호 규제 일반

영국은 1995년 유럽연합 개인정보보호지침이 발효하면서 그에 따라 개인정보보호지침에 제시된 개인정보보호의 원칙들을 반영하여 1997년 데이터보호법(UK Data Protection Act of 1997)을 제정하였다. 그 후 2016년 1995년 개인정보보호지침을 대체하는 GDPR이 확정된 후 2017년 영국은 유럽연합 회원국에서 탈퇴한다는 국민투표 결과가 나오면서 유럽연합의 규범인 GDPR에 직접적인 구속력을 받을 필요가 없어졌다. 하지만 영국 정부는 이른바 브렉시트(Brexit) 이후에도 유럽연합과 동일한 수준의 개인정보 규제를 유지하겠다는 의지를 지속적으로 표명하고 있고, 2018년에 GDPR의 내용을 반영하기 위하여 데이터보호법을 개정하기까지 하였다(UK Data Protection Act of 2018, 이하 "개정법률"이라 한다). 개정법률에는 GDPR의 규정에 따른다는 조항이 다수 포함되어 있다. 가령 대부분의 개인정보의 처리는 GDPR에 기속된다는 규정이 있다[UK Data Protection Act of 2018, Art.1(2)]. 하지만 GDPR은 일부 항목들에 대해서는 GDPR의 내용과 다르게 규제할 수 있게 하고(이를 derogation이라 한다), 개정법률도 개별 조항들에 따라 GDPR을 반영해서 해당 조항들을 개정한다는 내용 또는 GDPR과는 다르게 규율한다는 내용을 포함하고 있다.

2) 비식별 조치

가. 법령

비식별 조치와 관련해서는 개정법률에 별도의 수정조항이 없고, 전적으로 GDPR에 따른다. GDPR에서 규정한 비식별 조치 또는 익명성에 대한 개념과 기준, 가명화의 개념을 특별한 수정 없이 그대로 수용하는 취지이다. 이에 따라 개정법률은 이런 개념에 대한 개념정의를 별도로 두지 아니하였다. 단지 재식별 금지조항이 '비식별 조치된(de-identified)'이란 표현을 사용하고 있을 뿐이다[UK Data Protection Act of 2018, Art.171(1)]. 하지만 개정법률이 재식별 금지조항을 명시적으로 도입한 점은 GDPR과 다르다.

개정법률에 따르면 데이터를 제공받은 주체가 '의식적으로(knowingly)' 또는 '경솔하게(recklessly)' 해당 데이터를 재식별하려는 행위를 하면 벌금 등 금전적 제재의 대상이 된다[UK Data Protection Act of 2018, Art.171(1)]. 이 두 가지 표현이 형법상의 고의 또는 과실 개념에 해당하는지에 대한 논의는 별론으로 하고 단순하게 문언적으로 해석을 하면, 재식별 행위를 한다는 의도 또는 인식이 있는 경우뿐만 아니라 명확한 재식별 행위의 인식이 없는 경우도 제재의 대상이 될 가능성이 있다. 하지만 동시에 비록 금지하는 재식별행위에 해당한다고 해도, 일정한 조건을 만족하면 재식별행위의 처벌에서 면제될 수 있다는 구체적인 면책조항들도 포함하고 있다. 개정법률은 재식별행위의 처벌 조항에 의해 데이터의 활용이 과도하게 위축되는 상황을 피하려는 시도를 하고 있는 것으로 판단된다.

GDPR과 개정법률 시행 이후 비식별 조치 맥락에서 이 법령들에 근거해 구체적으로 발표된 공식 가이드라인은 아직까지 없다. 본 보고서에서는 영국 개인정보보호기구인 ICO(Information Commissioner's Office)가 2012년에 발표한 비식별 조치에 대한 실행규칙(Anonymisation: Managing Data Protection Risk Code of

Practice, 이하 "ICO가이드라인")과 민간네트워크인 UKAN(UK Anonymisation Network)에서 2016년에 발표한 비식별 조치 관련 보고서(The Anonymisation Decision-Making Framework, 이하 "UKAN보고서")의 내용을 소개하기로 한다.

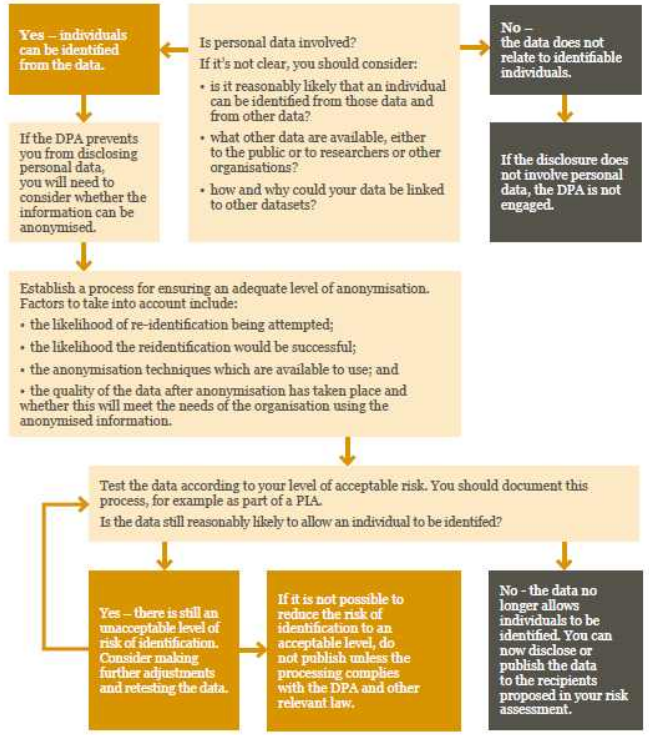
나. 법령 이외의 형태

a. ICO가이드라인

ICO가이드라인은 공식적으로는 법적 구속력이 없는 안내서의 기능을 한다. 아래 그림 4-9가 익명화된 데이터에 해당하는지에 대한 전체적인 메커니즘을 보여준다.

ICO가이드라인 자체에는 법적 구속력이 없지만 ICO은 강력한 집행력을 가지고 있다. 그러므로 ICO가이드라인이 제시하고 있는 사항들을 준수했다는 사실이 입증되면, ICO에서 개인정보 보호 관련 법위반 사실을 조사하거나 법집행을 함에 있어 적극적으로 고려될 수 있을 것이다. 규제대상이 되는 기업이 ICO가이드라인에서 제시하고 있는 익명화에 대한 원칙과 기준, 방법들을 준수하였다면, 실제 데이터를 재식별하려는 제3자에 의해 해당 데이터가 재식별되는 상황이 발생하더라도 행정적 제재를 하지 않을 가능성이 있다.

ICO가이드라인은 허용하는 익명화의 정도를 판단할 때 재식별의 위험성을 기준으로 본다. 이 경우 재식별 위험성이 존재하지 않는 영(zero)인 수준을 요구하지는 않는다. 단지 개인정보를 처리하려는 행위주체는 식별 위험성이 매우 낮은(remote) 수준까지 식별 위험성을 완화하여야 한다고 한다. ICO가이드라인은 합리적 가능성("reasonably likely")이란 표현을 사용해서 식별가능성의 기준을 설명한다. 이 기준이 재식별의 위험성을 판단하는 기준이 된다.



(그림4-9) 익명화된 데이터 해당여부 판단 메커니즘
출처: UK Information Commissioner's Office, Anonymisation: Managing Data Protection Risk Code of Practice, p.17(2012) 그림 인용)

재식별에 대한 위험성 평가를 위해서는 데이터 자체는 물론 그 이외에도 데이터 환경에 대한 파악이 전제되어야 한다. 결국 데이터 환경적 요소에 따라 재식별의 위험성이 달라지기 때문이다. ICO가이드라인도 데이터 환경을 맥락에 따라 세분화하여 위험성 평가 요소를 제시한다. 즉, 위험성 평가를 (1) 데이터의 관리주체, (2)정보가 공유되는 유형, 그

리고 (3) 재식별하려는 데이터공격자의 측면으로 구분해서 프라이버시 침해의 위험성들을 세밀하게 분석하고 있다.

첫째, 데이터를 보유하고 활용하려는 관리주체의 측면이다. ICO가이드라인은 익명화된 데이터가 공개되는 시간과 방법에 대해 데이터를 공개하려는 목적과 이유가 중요하다고 설명한다. 데이터를 공개하는 목적과 이유에 따라서 재식별이 할 수 있는 주체의 동기가 달라질 수 있기 때문이다. 가령 정부 기관이 일반에게 정보를 공개하는 경우에는 다양한 동기를 가진 모든 일반인들이 해당 정보에 접근할 수 있어서 그 정보를 통해 정보주체가 식별될 위험성이 높아진다. 대조적으로 연구적 목적이나 상업적 목적으로 정보를 특정인이나 일부 단체들에만 제한적으로 공개한다면, 해당 정보를 통해 정보주체가 재식별되는 위험성이 일반 대중에게 완정 공개되는 경우에 비교해서 낮다.

둘째, 정보공유의 형태이다. 주어진 데이터 환경 안에서 어떤 종류의 정보가 익명화되고, 외부로 공개되는지에 대한 문제는 재식별의 위험성의 정도와 연결된 중요한 고려 사항이다. 민감정보와 같은 종류의 정보는 다른 정보보다 잠재적 데이터 공격자에게 상대적으로 더 큰 관심을 받는다. 따라서 이러한 정보는 데이터 공격자들의 집중적 공격의 대상이 될 수 있고 더 큰 손실을 가져올 수 있다. 예를 들어, 암이나 HIV와 같은 병력 정보가 외부에 공개가 되면 그 정보주체에게는 높은 수준의 심리적인 피해를 주게 된다. 데이터 공격자도 이와 같은 피해의 규모도 공격의 유인 중 하나가 되기 때문에 다른 유형의 정보에 비해 민감한 정보는 재식별 공격의 대상이 될 위험성이 높다. ICO가이드라인은 구체화되고 세분화된 정보의 유형들을 파악하고 개별 유형 별로 재식별될 경우 피해의 규모도 같이 파악하여야 한다고 설명한다. 정보공유의 형태가 재식별위험을 평가하는 중요한 결정 요소들 중 하나가 된다는 것이다.

셋째, 재식별을 하려는 데이터 공격자가 있다. ICO가이드라인은 데이터 공격자가 중요한 고려요소라는 설명에 그치지 아니하고, 데이터 공격자의 기준을 구체적으로 설명하였다. 그 기준은 '의도적 공격자

(motivated intruder)'이다. 이 표현에서 '의도적'이란 요소는 데이터 공격자가 데이터를 재식별하려는 목적과 관련성이 있다. 공격자의 공격목적에 따라 재식별의 위험성이 달라지기 때문에 데이터 공격자의 의도 또는 목적을 의미하는 '의도적'이란 표현을 포함한 것이다. 구체적 상황을 생각해보면, 데이터 공격자가 완전 공개된 정보가 아니라 이미 다른 방법을 통해 별도의 지식 또는 정보를 가지고 있는 경우에는 외부정보와 해당 데이터셋을 연계하여 재식별을 시도할 가능성이 높다. 데이터 공격자가 별도의 외부정보를 가지고 있다는 것은 재식별의 의도가 강함을 뜻한다.

이 '의도적 공격자' 기준은 재식별의 위험성이 지나치게 확대되어서 실제는 재식별의 위험성이 0(zero)이 되어야만 ICO 가이드라인에 따른 익명화가 되는 현상에 이르지 않도록 조정하는 기능을 한다. ICO가이드라인에서 제시한 기준은 의도적 공격자가 통상적인(ordinary) 재식별의 능력을 갖춘 일반인을 상정한다. 통상적인 일반인이 재식별할 수 없는 수준으로 익명화가 된 데이터라면 ICO가이드라인에 따른 익명화가 된 것으로 본다. 통상적인 일반인을 상정하므로 재식별이 될 수 있는 모든 경우의 수를 고려해야 하는 극단적 상황이 배제된다. 이러한 관점에서 '의도적 공격자'는 사전적인 예비지식이 없고, 일반인의 관점에서 합리적인 수준의 재식별 능력을 갖춘 대상이다. 예를 들어, 일반인들이 대체적으로 가지고 있는 검색 및 조사 능력으로 인터넷, 도서관, 다른 공공문서들에 접근해서 이용할 수 있으면 이 기준에 해당한다. 특정한 고등의 전문가적인 능력 및 지식이나 해킹 기술을 보유하는 것을 조건으로 하지 않는다.

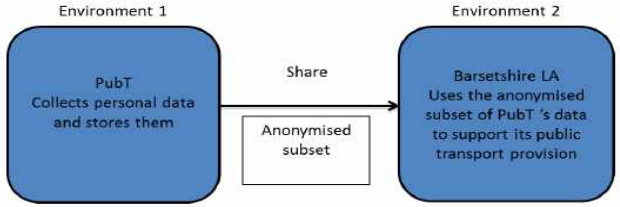
b. UKAN보고서

UKAN은 ICO가이드라인에서 제시한 익명화에 대한 원칙과 기준을 좀 더 구체적으로 설명하고 보충하는 기능을 하고, 나아가 실무자가 현장에서 참조할 수 있는 문헌을 제공하거나 익명화에 관한 자문을 위하여

2012년에 비영리기관으로 설립되었다. UKAN은 여러 분야 전문가들을 통하여 익명화 사례를 연구하고 분석하는 활동을 수행하고 있다. UKAN 내 협업적 작업구조를 기반으로 익명화의 기술적 영역에서부터 제도적 영역까지 광범위한 연구를 한다. 이런 연구 활동과 자문(advice) 등 커뮤니케이션을 지속적으로 유지함으로써 익명화된 데이터의 공유와 사용에 대한 사회적 신뢰를 구축하는 것을 추구하고 있다. 가령 UKAN은 주기적으로 워크숍을 열어 관련 분야의 전문가들의 지식과 경험을 공유한다. 2012년에 설립되어 역사가 길지는 않지만, 설립된 지 4년이 지난 2016년 그간 축적된 논의 및 연구를 바탕으로 170여 페이지의 UKAN 보고서를 발표했다.

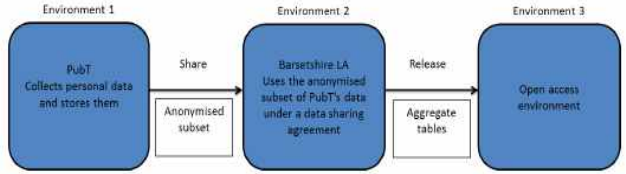
UKAN보고서는 익명화 체제의 형성에 있어서 맥락(context)을 중요한 요소로 평가한다. 효과적인 익명화 기법들을 제시하기 위해서 데이터 환경이란 맥락적 요소를 익명화 체제 형성을 위한 한 가지 변수로 간주한다. UKAN보고서 서문은 특정한 데이터 환경에 위치한 데이터가 식별될 수 있는 위험성을 평가할 때 식별자들의 유형에 따라서도 달라진다는 점을 분명히 한다. 직접식별자 또는 준식별자의 범주에 해당하는지 여부에 따라 재식별의 위험성이 다르다. 식별자의 유형은 데이터 환경과 결합에서 다양한 프라이버시 침해 위험을 발생시킨다. UKAN보고서의 익명화에 대한 접근법은 데이터 환경이란 외부적 요인과의 상호작용의 중요성을 강조하므로 익명화 수준에 대한 판단은 개별적인 데이터 환경이란 맥락에 의존하게 된다. 이처럼 데이터와 데이터 환경을 동등한 요소로서 고려하고 이들 사이의 상호작용의 다양한 형태를 중시하는 접근을 UKAN보고서는 “데이터 상황적 접근법(data situation approach)”이라고 표현한다.

UKAN보고서는 데이터 상황적 접근법을 주장하고 있으므로 데이터가 공유되는 세부 유형에 대한 분석을 중요시한다. 아래의 두 그림이 보여주듯 다양한 데이터 환경 유형을 인지하고 그에 적합한 익명화 방법을 적용하게 된다. 아래의 첫 번째 그림은 데이터 보관 주체가 2인인 경우이다.



(그림4-10) 데이터 보관 주체가 2인 경우
출처: UK Anonymisation Network, The Anonymisation Decision-Making Framework, 그림 3.2(2016) 인용

아래의 그림은 데이터 보관 주체가 2인을 넘어 외부로 완전 공개되는 형태이다.



(그림4-11) 외부로 완전 공개된 형태
출처: UK Anonymisation Network, The Anonymisation Decision-Making Framework, 그림 3.3(2016) 인용

이러한 메커니즘을 기초로 UKAN보고서는 익명화 체계를 단계별로 설명한다. UKAN보고서는 실무에서 참조하기 용이한 실용적인 안내서 방식을 취한다. 보고서는 익명화를 아래 표와 같이 10단계로 설명한다. 다만 이 10단계 과정을 무조건 기계적으로 적용하면 익명화가 된다는 식으로 판단하여서는 안 된다는 점을 경고하였다. 데이터 환경에 따라서 고려해야 할 요소들이 증가할 수도 있고 더 복잡한 판단과정을 적용해야 하는 상황도 있으므로 이는 하나의 참고용 표준에 불과하다고 한다.

[표4-9] UKAN 보고서의 익명화 체계

데이터와 데이터 환경의 확인	
제1단계	데이터 상황의 인식
제2단계	법적 책임의 인식
제3단계	데이터에 대한 인지
제4단계	데이터 활용 형태의 인지
제5단계	윤리 의문의 준수
재식별 위험성의 평가 와 익명화 적용	
제6단계	공개될 위험성의 평가 관련 절차 확인
제7단계	구체적 통제 과정의 인식
사후 관리	
제8단계	이해관계인의 확인 및 논의 계획 수립
제9단계	데이터 공유 이후의 경우에 대한 계획 확립
제10단계	문제 발생 시의 대책 수립

자료 : 연구진 작성(UK Anonymisation Network, The Anonymisation Decision-Making Framework, pp.68-115(2016) 참조)

개별 상황에 가장 적합하게 익명화하려면 데이터와 데이터 환경 모두를 정확히 파악하여야 한다. 데이터와 데이터 환경의 확인과정을 구성하는 1~5단계들의 과정을 차례로 거치게 되면 특정한 데이터 상황과 연관성이 높은 문제점들을 사전에 확인할 수 있다. 본격적으로 재식별의 위험성을 평가하기 전 5단계를 제대로 적용하지 아니하면 전체적으로 잘못된 익명화로 이어지게 되어서 재식별의 위험성 평가가 과대 또는 과소평가되는 잘못된 오류가 발생할 가능성이 높다.

1~5단계 전체를 포괄하는 '데이터 상황'은 데이터와 데이터 환경 요소를 결합한 전체적 개념이다. 데이터 상황은 두 개념 모두를 포괄한다. 그 구체적인 요소를 살펴보면, 사람, 데이터, 기반시설(infrastructure),

거버넌스(governance)가 포함된다. 결국, 주어진 데이터 상황에 대한 판단은 사람, 데이터, 기반시설, 거버넌스란 4가지 요소가 어떠한 형태로 결합하는지에 달려있다. 예를 들어, 특정 데이터를 제3자와 공유하는 동적(dynamic) 구조에 기반을 두는지, 아니면 일반 대중 기타 제3자와 데이터를 공유하지 않고 양 당사자들 사이에서만 데이터 공유 계약 등의 형태로 제공되는 정적(static) 구조에 기초하는지에 의해서 적용될 방법이 달라진다. 좀 더 다양한 동적인 구조들이 등장할수록, 데이터를 공유하게 되는 주체들의 숫자가 증가할수록, 그만큼 재식별의 위험성은 높아진다.

6~7단계에서는 데이터와 데이터 환경에 대한 정확한 이해를 바탕으로 재식별 위험성에 대한 평가를 수행하고 그 결과에 따라 어떤 익명화 기법을 어느 정도로 적용할지를 결정한다. 해당 데이터의 내용 자체가 갖고 있는 재식별의 위험성을 인지하고, 이를 기반으로 이전에 계획된 데이터를 공유하고 제공하는 형태를 반영하여 데이터 공격자가 어떻게 재식별 시도 행위를 할지 구체적인 시나리오를 작성한다. 각각의 시나리오에 따라 데이터 공격자가 공격하는 상태를 데이터 처리에 대한 알고리즘(algorithm) 측면에서 상세하게 설명하고 있다.

8~10 단계인 사후관리 부분은 이전까지의 단계들을 순차적으로 거치면서 확정된 비식별 조치 기법을 적용한 데이터셋이 이후의 상황에 따라 재식별될 수 있는 잠재적 위험성을 관리한다. 사후적 위험성 관리는 특정 데이터를 제공하고 공유하는 과정에 관련된 모든 이해관계자들과 지속적으로 커뮤니케이션을 유지하는 관계를 형성하는 단계이다. UKAN 보고서가 정의하는 이해관계인이란 매우 포괄적인 용어로서 정보주체, 일반대중, 협력적 관계에 있는 단체, 언론매체, 후원자(funder), 그리고 시민단체를 포함한 다양한 이해관계인이 포함된다. 익명처리된 데이터를 제공하는 주체는 이런 이해관계자들과 이메일 교환 등의 다양한 형태로 관계를 지속적으로 유지해서 해당 정보를 제공하고 피드백도 받는 위치에 있다.

외부 주체들과 공유되어 있는 데이터의 경우에는 기술적 또는 계약적

수단을 적용해서 지속적인 관리를 할 수 있다. 예를 들어, 기술적인 방법으로는 '데이터 추적 심사(audit trail)'의 방법을 적용해서 이미 공유된 데이터에 접근하려는 존재가 누구인지를 인지할 수 있다. 이와 별도로 일종의 '위기관리의 정책(crisis management policy)'을 적용해서 데이터의 재식별 행위란 사건을 신속하게 해결하기 위해 일단 재식별 행위가 확인된 이후에는 신속하게 사후적인 해결책을 모색하고 적용할 수 있다. 이러한 위기관리의 정책이 실행되면, 피해가 확산될 시간을 줄일 수 있기 때문에 그 만큼 프라이버시의 침해에 의한 손실의 규모도 작아진다.

(2) 비식별정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

2018년 개정 데이터보호법은 GDPR에서 규정한 익명 개념을 그대로 받아들이고 또 GDPR이 새롭게 도입한 가명화 개념도 그대로 수용하는 전제 하에 이들에 대하여 별도의 정의 규정을 두지 아니한다.

2) 법령 이외의 형태

GDPR이나 2018년 개정 데이터보호법에 근거한 가이드라인, 공식 해설서는 아직 없다. 공식적인 가이드라인이 발표될 때까지는 앞에서 설명한 ICO가이드라인의 내용을 참조할 필요가 있을 것이다.

ICO가이드라인은 익명화(anonymisation)를 포괄적으로 규정하면서 개인정보를 식별이 가능하지 않은 정보로 전환하는 다양한 기술적 처리를 포함한다고 설명한다. 가이드라인에 따르면 익명정보(anonymized information)란 개인을 식별하지 않으면서 외부 데이터와의 결합을 통해서도 개인을 식별할 수 없게 하려는 형태의 데이터를 의미한다. ICO

가이드라인은 식별의 위험성이 0(zero)이 되지 않아도 익명화되었다고 볼 수 있다는 실용적인 접근법을 채택한다. 익명화를 이와 같이 포괄적 의미로 사용하면서 비식별 조치(de-identification)의 개념이나 익명화와 비식별 조치와의 개념적 차이점을 별도로 제시하지는 않는다. 익명화의 개념이 식별가능성을 개념 표지로 정하고 있는 개인정보의 개념 정의에 의존한다고 설명하고 있다.

ICO가이드라인은 가명화의 개념정의도 포함하고 있다. ICO가이드라인에 따르면 가명화란 '실제의(real world)' 정체성을 드러내지 않는 특정 식별자를 사용해서 데이터셋에 있는 개인을 구분하는 방법이다. 익명화의 다양한 기법들 중의 한 가지 방법이란 관점에서 가명화에 대한 개념 정의를 하고 있다.

반면 UKAN보고서는 익명화와 비식별 조치 개념을 각각 정의하면서 두 용어 사이의 관계를 정리하려는 시도를 하고 있다. UKAN보고서는 비식별 조치를 개인의 이름, 주소와 같은 직접식별자(direct identifier)를 제거하는 절차라고 설명한다. 반면 익명화는 특정 데이터에서 어떤 개인이 식별될 위험성을 '무시할 수 있는 수준(negligible)'으로 낮추는 과정이라고 정의된다. 이에 따라 비식별처리는 직접식별자를 제거하는 절차로 좁게 해석하고, 익명화는 직접식별자의 제거를 포함해서 재식별의 위험성을 통제하는 접근법이란 좀 더 넓은 개념으로 보고 있다. 그리하여 UKAN보고서는 비식별 조치를 익명화의 방법 중 직접식별자를 제거하는 한 종류로 설명한다. 재식별의 유형 측면에서 설명하면, 비식별 조치는 해당 데이터 자체에서 직접적으로 재식별될 위험성에 대한 통계적 대체 방안이고 익명화는 해당 데이터 자체뿐만 아니라 외부 데이터와의 결합을 통해서도 간접적으로 재식별될 수 있는 위험성까지 고려하는 포괄적인 통계적 해결책이 된다.

UKAN보고서는 위와 같은 개념 구분에 따라 가명화를 비식별 조치의 한 가지 기법으로 간주한다. 이런 개념 구분을 기반으로 UKAN보고서는 가명화를 직접식별자를 가상의 코드로 대체하는 기법으로 정의한다. 이 경우 대체해서 적용되는 가상의 코드는 해당 코드 그 자체가 직접 개인

을 식별할 수 없도록 선별되어야 한다. UKAN보고서는 이와 같은 개념 정의가 ICO가이드라인의 가명화와 기본적으로 일치한다고 설명한다.

(3) 비식별정보, 익명정보, 가명정보 개념의 도입 과정

1) 배경 및 구체적 절차

ICO가이드라인은 1995년 정보보호지침 체제를 반영한 것이다. 즉, 1995년 정보보호지침이 직접 규정하지 아니한 구체적인 사항들에 대하여 각국이 행동강령(code of conduct)을 정할 수 있는데, EU회원국들 중에서 익명화에 대하여 최초로 발표된 강령이 영국 ICO의 위 가이드라인이다.

ICO는 2012년 ICO가이드라인을 발표하면서 UKAN과의 협력을 통해 더 구체적인 익명화 방법에 대한 정책적 연구를 하겠다는 의지를 표명하였고, 실제로 2012년 UKAN이 설립될 당시부터 2년간 ICO가 UKAN에 재정적 후원을 하였다. UKAN은 익명화 방법을 통하여 개인정보를 처리하고 공유하려는 모든 주체들에게 실용적 자문과 정보를 제공하는 역할을 하고 있다. 2016년에 UKAN이 UKAN보고서를 발표한 것도 이 보고서가 개인정보를 활용하는 기업에 실질적인 안내서가 되게 하려는 것이었다. 특히 UKAN보고서는 GDPR의 최종안이 확정된 이후인 2016년에 발표되어, 그 서문에서 GDPR을 간략하게 언급하고 있는데, 보고서에서 설명하는 익명화와 가명화에 대한 전반적인 접근법이 GDPR의 규정과 일치한다고 한다.

2) 논의 사항

UKAN보고서는 기본적으로 ICO가이드라인과 영국의 데이터보호법의 원칙을 반영하면서 동시에 익명화와 가명화에 대한 GDPR의 접근법을

따르고 있다. UKAN보고서의 기본 원칙과 핵심적인 내용은 영국에서의 익명화, 비식별처리, 가명화에 대한 최근의 논의 결과를 대변한다.

UKAN보고서에 따르면 영국 데이터보호법은 위험성이 0 (zero)인 상태를 익명화의 조건으로 하고 있지 않다. 위험성 수준이 미미한 (remote) 상태 정도로 완화할 수 있다면 익명화 조건을 충족한다. 결국 데이터를 공유하거나 활용하려는 주체는 재식별의 위험성을 매우 낮은 수준으로 하기 위해 데이터의 잠재적 공격자(intruder)의 재식별 의지를 낮추도록 재식별의 위험성을 관리해야 할 의무가 있다. 이를 위해 UKAN보고서는 데이터를 공유하거나 활용하려는 주체에게 다음과 같은 세 가지의 방법을 실천할 것을 강하게 권고하고 있다.

- 통계학자와 컴퓨터분야 전문가(computer scientist)들이 최고의 방법을 상용해서 적절한 수준의 기술적, 물리적, 관리적 보안 조치들을 평가할 것
- 데이터의 유용성을 극대화하면서 재식별의 위험성을 최소화할 것
- 개별 재식별 문제에 대응할 구체적 전략을 세울 것

(4) 비식별정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

ICO가이드라인은 익명화의 기준과 방법들에 대한 설명 이외에 익명화된 데이터가 어떻게 활용될 수 있는지에 대한 11가지 사례를 제시하고 있다. 이 사례들의 내용을 다음의 표에 간략하게 요약한다.

[표4-10] 익명화 된 데이터의 활용 사례

사례 1	-제약(pharmaceutical) 데이터에 대한 접근 사례 -의료전문가와 제약회사 사이의 데이터를 코드화(coded)
사례 2	-핸드폰 데이터를 이용해서 교통 속도를 측정하는 사례 -핸드폰 번호, 대략적 위치, 일자와 시간에 대한 정보의 암호화(encryption)
사례 3	-버스 탑승객의 탑승 시간의 분석 사례 -일종의 교통카드의 사용을 통해 수집한 정보(카드 숫자, 탑승객 생년월일, 출발지, 도착지, 탑승 시간 데이터)의 암호화
사례 4	-선거인등록부(electoral register)와 같은 공공데이터 사례 -공공데이터는 개인정보를 식별할 수 있는 재식별의 위험성이 있으므로 익명화를 적용
사례 5	-공공기관의 통계치와 같은 정보 공개 문제에 대한 사례 -통계치도 경우에 따라 개인정보에 해당할 수 있음
사례 6	-질적(qualitative) 데이터의 익명화 사례 -텍스트와 같은 질적 데이터에서 익명화하는 방법 제시
사례 7	-제 3자의 사전 지식이 재식별 위험성에 미치는 영향에 대한 사례 -제3자가 정보주체의 친구, 직장동료, 가족 구성원이 될 가능성이 있는 경우의 익명화 문제
사례 8	-고객의 구매 습관 데이터에 대한 사례 -고객의 구매 습관 데이터가 익명화된 다른 정보화 결합해서 특정 고객의 식별 가능한 경우 추가적인 암호화 적용
사례 9	-고객에 대한 분석 사례 -익명화된 고객의 거래 데이터의 분석을 통해 소매업자들의 매출 증대 가능
사례 10	-설문조사 등을 통해 형성된 종단면(longitudinal) 데이터의 활용 사례

	-종단면 데이터에 일반화(suppression)과 같은 익명화 기법 적용
사례 11	-해시(hash) 기법의 적용 메커니즘에 대한 간략한 설명 사례

자료 : 연구진작성(UK Information Commissioner's Office, Anonymisation: Managing Data Protection Risk Code of Practice, pp.66-79(2012) 참조)

영국은 국가가 의료보장을 하고 있어 보건의료정보가 NHS(National Health Service) Digital라는 기관에 집적된다. 막대한 양의 데이터가 집적되어 있어 그 외부 공유와 활용을 통해 보건의료서비스의 개선 등 사회적 가치를 창출하려는 시도를 하고 있다. 하지만 2018년 5월 NHS Digital과 국토관리국(Home Office) 사이의 데이터 공유 약정의 실행을 영국 정부가 정지하는 일이 있었다. NHS Digital이 보유한 민감한 보건의료정보가 외부로 제공될 경우 환자들인 자신의 프라이버시가 침해될 수 있다는 우려로 의료서비스 이용을 기피할 것이라는 우려 때문이다. 이처럼 보건의료데이터 구조의 측면에서 영국은 데이터의 활용을 효율적으로 할 수 있는 제도를 갖추고 있지만, 프라이버시 침해의 위험성에 대한 사회적 우려도 낮지 않은 상태이다.

제 2 절 북미

1. 미국

(1) 제도적 개괄

1) 개인정보보호 규제 일반

미국에서는 개인정보보호를 포괄적으로 규제하는 연방 차원의 법률이 마련되어 있지 않다. 개인정보의 보호에 관련한 규제 체제는 현재 분야별로 규율되는 형태(sectoral regulation)로 되어 있다. 이런 접근법 하에서 비식별에 관한 논의의 맥락에서 특히 중요한 것이 보건의료영역을 규제하는 연방법으로 제정된 「보건의료정보의 이동과 책임에 관한 법률」(Health Information Portability and Accountability Act, "HIPAA")이다. 이 법률은 보건의료정보의 보호와 활용에 관한 규제를 포함하고 있고, 그 하위법령인 'HIPAA 프라이버시규칙(HIPAA Privacy Rule)'이 구체적인 지침을 제공하고 있다.

이 HIPAA 프라이버시규칙은 비식별 조치를 실제 적용하는 기준과 방법들을 규정하고 있다. 그래서 미국의 비식별 조치에 대한 개인정보보호의 법제도 측면을 논의할 때 HIPAA 프라이버시의 내용을 살펴보는 것은 필수적인 연구가 된다. 더 나아가 HIPAA 프라이버시규칙의 비식별 조치에 대한 규정에 대해서 더 상세하게 Q&A 형식으로 설명해주는 가이드라인도 있다(Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule). 이 가이드라인은 미국보건부(Health and Human Service, 이하 "HHS")의 인권국(Office of Civil Rights, 이하 "OCR")에서 2012년에 발표했다

HIPAA 이외에 주목할 만한 최근의 주법으로 2018년 6월 제정된 '캘리포니아 소비자 프라이버시법(The California Consumer Privacy Act of 2018(AB 375), 이하 "CCPA")이 있다(발효예정일은 2020. 1. 1.). CCPA는 개인정보의 보호 측면에서 미국에서 가장 엄격한 법률로 유럽연합의 GDPR에 버금가는 개인정보보호 원칙들을 일부 포함한다. 캘리포니아에서 이와 같이 강력한 데이터보호 법률 체제가 도입될 수 있었던 배경에는 1974년 주민투표로 프라이버시권을 주 헌법의 기본권에 도입하였다는 점이 꼽힌다(California Constitution Art.1, Sec.1). CCPA는 단 하나의 주에만 적용되는 주법이지만 특히 데이터가 핵심적 요소를 차지하는 IT산업 관점에서 캘리포니아 주가 차지하는 비중 때문에 매우 중요한 주법이기도 하다. 나아가 GDPR이 확립되고 CCPA가 제정되면서 그 영향으로 다른 주에서도 CCPA처럼 강력하면서 포괄적인 규제 체제를 채택하려는 움직임을 보일 가능성이 상당하다는 예측도 이루어지고 있다[21].

CCPA는 정보주체가 본인의 개인정보를 통제할 권리가 있다는 원칙을 전제한다. 즉 CCPA는 정보주체이자 소비자는 자신의 개인정보에 대한 통제권을 행사할 수 있어야 하고 이런 통제권에 기초해서 소비자 자신의 개인정보가 오용되지 않게 하는 안전장치가 있음을 확인할 권리가 있다고 선언한다[US California Assembly Bill No.374, Chap. 55, Sec. 2(h)]. 이러한 기본 원칙하에 CCPA는 소비자들에게 4개의 권리를 부여하고 있다[US California Assembly Bill No.374, Chap. 55, Sec. 2(i)].

첫 번째 권리는 소비자들이 자신의 개인정보가 어떻게 이용되는지에 대한 사실을 확인할 수 있는 권리이다. 이런 사실을 대체로 프라이버시 약관을 통한 사실 확인이 가능하고 좀 더 상세한 사실들에 대해서는 개별 소비자의 요청에 따라 기업들이 제공하는 형태가 된다. 이런 방법을

21) https://www.forbes.com/sites/forbestechcouncil/2018/08/20/how-will-californias-consumer-privacy-law-impact-the-data-privacy-landscape/#5ca73677e922.

통해 기업이 수집하는 개인정보의 종류, 개인정보를 수집한 장소, 개인정보의 사용 목적, 공개 또는 판매 여부, 그리고 공개 또는 판매의 대상에 대한 사실들을 요청할 수 있다. 두 번째 권리는 철회(opt-out)할 수 있는 권리이다. 기업이 제3자에게 개인정보를 판매할 경우 그 철회를 요청할 수 있는 권리를 의미한다. 나아가 해당 개인정보의 주체가 16세 미만인 경우에는 사후적 철회가 아니라 사전동의(opt-in) 원칙을 적용하여 판매 전에 소비자의 동의를 획득해야 한다. 세 번째 권리는 삭제권이다. 일정한 예외적인 상황을 제외하고는 소비자는 기업에게 자신의 개인정보를 삭제할 것을 요청할 수 있는 권리가 있다. 네 번째 권리는 평등한 서비스와 가격을 받을 권리이다. 프라이버시권을 적극 행사하는 소비자와 그렇지 않은 소비자 사이의 차별적 대우를 인정하지 아니하여 프라이버시권의 행사가 소비자에게 불이익이 되지 않도록 하고 있다.

2) 비식별 조치

가. 법령 및 가이드라인

a. HIPAA 프라이버시규칙

앞서 언급한 바와 같이 미국에서 비식별화에 관하여 논의함에 있어서는 HIPAA 프라이버시규칙을 빼놓을 수 없다. HIPAA 프라이버시규칙이 규정한 구체적인 비식별 조치의 방법들을 논의하기에 앞서서 HIPAA 및 HIPAA 프라이버시규칙의 적용을 받는 보건의료정보의 개념을 간략하게 살펴본다. HIPAA는 '보건의료정보의 이동과 책임에 관한 법률'이다. 이 법률의 명칭에 표현된 보건의료정보에 해당해야 HIPAA의 규제 대상이 된다. 그렇다면 HIPAA의 보건의료정보에 해당하지 않는 정보 유형은 처음부터 HIPAA의 적용도 받지 아니할 것이다. 그런데 HIPAA 프라이버시규칙은 규제의 대상인 보건의료정보에 대해서, "개인을 식별하지 않고, 개인을 식별할 수 있게 한다는 합리적(reasonable) 근거가 없는 보

건의료정보는 개인적으로 식별가능한 보건의료정보가 아니다"라고 규정하고 있다(45 CFR 164.514). 즉 적용대상이 되는 보건의료정보의 판단기준을 개인의 합리적 식별가능성에 두고 있다. 결국 합리적으로 식별가능하지 않은 정보는 HIPAA의 적용 대상이 되지 않기 때문에, HIPAA 프라이버시규칙이 부과하는 의무들을 부담할 필요가 없다. 비식별 조치도 이러한 개념 규정을 근거로 가능하게 된 것이다.

이런 전제 하에 HIPAA 프라이버시규칙은 비식별 조치의 두 가지 방법을 제시하고 있다. 이 두 방법은 선택 관계에 있어서 HIPAA 프라이버시규칙의 수범기관은 그중 원하는 방법을 선택하고 그 방법을 적용해서 보건의료정보의 비식별 조치를 하면 된다.

첫 번째 방법은 비식별 조치를 하게 되는 구체적인 상황이 발생하면 비식별 조치 방법을 사용하려는 수범기관이 관련 전문가를 선임해서, 이 전문가가 재식별의 측면에서 해당 데이터에 적용된 비식별 조치가 제대로 이행되었는지를 평가하는 것이다. 이를 전문가판단(expert determination) 방법이라고 한다(45 CFR 164.514). 전문가판단 방법은 해당 데이터셋의 식별자를 일률적으로 제거하는 방법이 아니라 데이터의 활용 목적에 비추어 데이터의 비식별처리를 이행하고 이렇게 처리된 비식별 조치가 재식별의 위험성이 없는지를 전문가가 사안별로 판단하는 것이다. OCR가이드라인은 전문가판단 방법이 실제로 진행되는 과정을 다음의 그림을 통해 보여준다.

전문가 판단 절차는 수범기관이 재식별의 위험성을 판별할 전문가를 선임함으로써 시작된다. HIPAA 프라이버시규칙은 특정한 자격증과 같이 전문가에 해당하기 위한 구체적인 자격 요건을 제시하고 있지 않다(45 CFR 164.514). HIPAA 프라이버시규칙의 전문가에 해당하려면 "비식별 조치와 관련해 일반적으로 활용되는 통계적 그리고 과학적 이론과 방법론에 대해 적정한 수준의 지식과 경험이 있는 주체"여야 한다(45 CFR 164.514).

(그림4-12) 전문가판단 방법이 실제로 진행되는 과정
출처: US Office of Human Rights, 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(2012)

이러한 HIPAA 프라이버시규칙의 규정에 기반해서 OCR가이드라인은 좀 더 자세하게 전문가의 자격에 대한 설명을 하고 있다. OCR가이드라인에 따르면 해당 전문가가 되기 위해 특정의 학위나 자격증은 필수 요건이 아니다. 전문가에 대한 HIPAA 프라이버시규칙의 해당 조항은 일정 수준의 지식과 관련 경험이 필요하다는 정도로 규정하고 있는데, 여기에서 관련 경험은 다양한 방법론이 적용된 교육이나 경험을 통해서 획득가능하다고 한다. 전문가로서의 경험을 얻기 위한 특정한 방법이 있는 것이 아니라 다양한 방법으로 경험을 축적할 수 있다는 것이다.

게다가 해당 경험의 분야가 반드시 통계, 수학, 또는 기타 과학 영역일 필요도 없다.

이렇게 선임된 전문가가 대상이 되는 보건의료정보가 식별될 수 있는 위험성이 '매우 작다(very small)'고 결정하면, 그 보건의료정보는 식별할 수 없는 데이터로 판단되어서 HIPAA의 적용대상에서 제외된다(45 CFR 164.514). OCR가이드라인은 재식별 위험성이 '매우 작다'는 기준을 일률적으로 충족할 수 있는 특정의 수치나 기준을 제시하지 아니한다. 구체적인 데이터 상황에 적합한 방법을 적용해서 재식별의 위험성이 어느 정도인지를 판단하여야 한다. 위험도를 판단할 때 고려하는 주요요소 하나는 데이터에 대한 잠재적 공격자이다. HIPAA 프라이버시규칙은 이 잠재적 공격자에 대해 해당 데이터를 식별할 의도를 가지고 있는 '예상 주체(anticipated recipient)'라고 표현하면서, 예상 주체의 개별적 능력을 고려하여야 함을 명시한다(45 CFR 164.514). 이러한 기준을 적용해서 재식별의 위험성을 평가한 이후 전문가는 행정기관이 사후적으로 판단의 근거와 결과를 확인할 수 있도록 비식별 조치에 적용된 기술적(technological) 방법들과 재식별의 위험성에 대한 판단의 기준과 결과를 문서화해야 한다.

두 번째는 세이프하버(safe-harbor) 방법이다. 기본적으로 HIPAA 프라이버시규칙에서 나열한 18가지의 식별자들을 제거한 데이터는 비식별 조치가 적용된 것으로 간주한다(45 CFR 164.514). 18가지의 식별자의 제거 이외에도 18가지 식별자를 제거하고 남은 데이터가 재식별되지 아니한다고 '실제 인식(actual knowledge)'하여야 한다(45 CFR 164.514). 이러한 두 요건을 충족하면 세이프하버 규정이 적용되어 HIPAA의 적용대상이 되지 않는다. 18가지 식별자에는 이름, 주소 등과 같이 개인을 직접적 또는 간접적으로 식별할 수 있는 정보들이 포함되어 있다. HIPAA 프라이버시 규칙이 나열한 18가지 식별자는 다음의 표와 같다(45 CFR 164.514).

[표4-11] HIPPA 프라이버시 규칙이 나열한 18가지 식별자

1. Names; 2. All geographical subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a 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. 3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; 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; 4. Phone numbers; 5. Fax numbers; 6. Electronic mail addresses; 7. Social Security numbers; 8. Medical record numbers; 9. Health plan beneficiary numbers; 10. Account numbers; 11. Certificate/license numbers; 12. Vehicle identifiers and serial numbers, including license plate numbers; 13. Device identifiers and serial numbers; 14. Web Universal Resource Locators (URLs); 15. Internet Protocol (IP) address numbers; 16. Biometric identifiers, including finger and voice prints; 17. Full face photographic images and any comparable images; 18. Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data)

출처: Health Information Portability and Accountability Act(HIPAA) Administrative Simplification(including Privacy Rule), Art.164.514

b. CCPA

CCPA에도 비식별 조치에 관한 규정이 있다. 즉 CCPA는 비식별 조치된(deidentified) 소비자 정보의 수집, 사용, 보유, 또는 공개의 경우에는 사업자들이 CCPA에서 규정한 의무를 부담할 필요가 없다고 규정한다[California Consumer Privacy Act of 2018, Art.1798.145(a)(5)]. 즉, CCPA도 비식별 조치된 정보는 해당 법률의 적용 대상이 되지 않음을 명확하게 밝히고 있는 것이다. CCPA는 HIPAA프라이버시규칙과 같은 구체적인 비식별 조치의 방법을 제시하고 있지는 아니하나, 비식별 조치 개념의 규정을 통하여 어느 정도의 지침을 제공한다. CCPA에 따르면 비식별 조치된 정보는 특정 소비자를 합리적으로(reasonably) 식별할 수 없는 정보이다[California Consumer Privacy Act of 2018, Art.1798.140(h)]. 이런 정보에는 특정 소비자와 직·간접적으로 연결될 수 없는 정보도 해당된다. 그 이외에 비식별 조치된 정보에 해당하기 위해 추가적으로 요구되는 요건들은 다음과 같다.

- 재식별을 방지하는 기술적 안전장치 적용
- 재식별을 특정적으로 금지하는 절차 마련
- 비식별 조치된 정보의 의도치 않은 공개를 예방하는 절차 마련
- 재식별 행위의 시도를 하지 않을 것

이와 같이 재식별 행위를 통제할 수 있는 기술적 및 관리적 방법들의 적용을 의무화함으로써 CCPA는 비식별 조치된 정보의 활용으로 생기는 프라이버시 침해의 위험성을 최소화하고 소비자의 신뢰를 높이기 위한 제도적 장치들을 마련하고 있다.

나. 기타 보고서

그 밖에 미국 상무부(US Department of Commerce) 산하 기술의 표준화를 담당하는 기구인 국가기술표준원(National Institute of Standards and Technology, “NIST”)이 발간한 보고서도 중요하다. NIST는 비식별 조치에 대한 최근 논의를 정리하는 두 가지 보고서들을 1년 간격으로 연달아 발표하였는데, 2015년에 발표한 ‘개인데이터의 비식별 조치(De-Identification of Personal Data, 이하 “NIST2015”)’와 2016년에 발표한 ‘정부데이터셋의 비식별 조치(De-Identifying Government Datasets, 이하 “NIST2016)”가 그것이다. 이 두 보고서는 비식별 조치에 대한 최근의 논의의 흐름을 기술적 측면에서 설명하고 있다. 다만 NIST2016은 아직 초안이고, 공식적인 최종본은 아니다.

a. NIST2015

NIST2015는 비식별 조치의 전체적 체제를 구성할 때 필수적으로 반영해야하는 두 요소를 제시한다. 특정 데이터기 위치하고 있는 데이터 환경을 가리키는 맥락(context)과 재식별의 위험성이 그것이다. 이 요소들을 결합해서 특정한 맥락에 있는 데이터가 재식별될 수 있는 위험성을 구체적으로 판단함으로써 재식별 위험성 문제를 해결하는 방법을 제시하고 있다. 기본적으로 NIST2015는 어떤 비식별 조치 기법들을 적용해도 재식별의 위험성이 완전히 제거되지 않는다고 전제한다. 그러한 관점에서 주어진 데이터 환경하에서 특정 종류의 비식별 조치가 어떻게 적용되는지에 따라 재식별 위험이 달라진다는 점을 강조한다.

NIST2015에 따르면 급속하게 진보하는 기술의 변화 속도를 고려할 때 특정한 데이터 환경에서 어떤 비식별 조치 방식이 더 적합한지에 대한 판단 기준을 모든 경우로 일반화하여 말할 수는 없다. 다른 정보와의 연결을 통해 식별가능성을 높이는 직접식별자(direct identifier)가 해당 데이터베이스에서 없도록 하는 방향으로 처리를 하는 것이 재식별

의 위험성을 크게 줄일 수 있다는 말을 할 수 있을 뿐이다. 직접식별자 이외의 간접식별자(indirect identifier 또는 quasi-identifier)는 재식별 기술의 진보함에 따라 확대될 가능성이 높기 때문에 데이터를 보유하고 활용하려는 기업은 이런 간접식별자의 처리에도 더 주의를 기울일 필요가 있다. 그래서 간접식별자를 확인하고 제거하는 기술적 방법 이외에도 사후적으로 통제할 수 있는 다양한 관리적 방법도 같이 고려하여야 한다. 관리적 방법의 예로는 ‘데이터 활용에 관한 합의서(data use agreement)'와 같은 계약이 있다. 이런 계약의 내용에 데이터의 이용자는 재식별 행위를 해서는 안 된다는 금지 조항이 포함될 수 있다.

NIST2015는 비식별 조치, 재식별, 데이터공유모형(data sharing model) 등의 다양한 기술적 개념들을 설명하고 있다. 비식별 조치는 그 자체로 완벽한 방법은 아니지만 정보주체의 프라이버시를 보호하기 위한 유용한 통제방법이라는 결론을 내린다. 비식별 조치 기법은 데이터의 유용성을 확보하면서 개인을 식별하는 데이터를 제거하기 위한 것이다. 연구목적에 의하여 개별적으로 결정되는 데이터의 유용성 정도가 다양해지는 환경에서 특정한 일률적인 기준을 적용해서 여러 데이터의 통계적 처리 기법들을 비교하는 것은 의미가 없다고 한다. 즉 비식별 조치는 특정한 단일한 기법의 적용으로 충분하지 않고, 개별적인 연구의 목적이나 상황에 맞는 다양한 방법들과 알고리즘들을 결합하여야 한다는 것이다.

NIST2015는 재식별의 위험성에 대한 문제에 대처하기 위해 대상이 되는 데이터셋에서 특정한 개인을 식별하려는 주체인 데이터공격자에 대한 분석이 우선적으로 요구된다는 점을 강조한다. 그리하여 데이터공격의 개념정의와 그 유형에 대한 설명에서부터 재식별을 하려는 공격행위로 실제 이어지는 동기까지를 포괄적으로 설명하고 있다. 재식별의 위험에 대한 평가는 가상의 데이터공격자들의 존재와 재식별 공격 행위 자체가 실제 재식별의 성공으로 이어지는 예상 확률에 달려있다. 결국 데이터공격자에 대한 정의에 따라 비식별 조치의 실제 적용방법도 달라진다고 밝히고 있다.

b. NIST2016

NIST2016은 정부 기관이 보유하고 있는 정부데이터의 비식별 조치를 집중적으로 다룬 보고서이다. 따라서 NIST2015에서는 설명하지 않은 공공데이터의 공유와 그에 따른 보호의 문제에 대해 추가적으로 생각할 정책적 가치들을 제시하고 있다. 공공데이터로서의 정부데이터 맥락에서 발생하는 정책적 가치들이 생기는 원인에 프라이버시와 공공이익 사이의 관계라는 새로운 측면을 제시한다. 일반적으로 개인정보의 보호와 관련된 문제는 프라이버시 침해의 위험성과 데이터 유용성이란 두 가지 가치 사이의 상충관계(trade-off)에 놓인다. 그러나 공공데이터의 경우에는 일반인 모두가 사회적 편익을 누리게 되는 공공재(public goods)로서의 특성이 포함되기 때문에 여기에 공공의 이익이란 가치가 추가적으로 등장한다고 한다.

NIST2016에 따르면 정부데이터에 대한 비식별 조치 및 공개에 대한 결정 및 관행은 정부 기관의 사명과 적절한 기능 역할에 필수적일 수 있다. NIST2016은 이러한 활동이 기관의 사명과 법적 권한에 부합하는 성과와 결과를 보장하는 방식으로 기관의 리더쉽에 의하여 관리되어야 한다고 밝힌다. 이러한 관리를 통해 비식별 조치를 실제로 수행하기 전에 기관들은 비식별 조치를 수행하는 데 있어서의 목표, 비식별 조치할 데이터의 유형, 비식별 조치된 데이터에 대한 예상 용도를 분명히 해야 한다. 기관들은 대개 이전에 개인의 민감정보를 포함하고 있던 데이터에 대한 더 넓은 접근을 허용하기 위해 비식별 조치를 수행한다. 또한, 민감정보를 수집, 저장, 처리하는데 있어서 위험을 줄이기 위해 비식별 조치를 수행할 수도 있다.

NIST2016은 정부기관들은 데이터에 비식별 조치를 적용한 이후 기관 외부로 공개하기 위해 사용될 데이터 공개 모델(data sharing model)을 정해야 하며, 데이터 공개 모델은 다음을 포함한다고 설명한다.

- Release and Forget Model: 비식별 조치된 데이터는 보통

인터넷에 업로드 되면서 대중에게 공개될 것이다. 그렇기 때문에 이러한 방식으로 일단 공개된 데이터는 기관입장에서 다시 확인하기가 어렵거나 불가능할 수 있으며, 추후 데이터 공개 시 드러나는 정보를 제한할 수 있다.
- Data Use Agreement (DUA) Model: 데이터를 사용하여 수행 할 수 있는 작업과 수행 할 수 없는 작업에 대해 자세히 설명하는 법적 구속력이 있는 데이터 사용 계약에 따라 비식별 데이터를 사용할 수 있다. 일반적으로 데이터 사용 계약은 재식별 시도 또는 데이터의 링크 및 유사하게 구속력 있는 DUA가 없는 데이터의 재배포를 금지 할 수 있다.
- Simulated Data with Verification Model: 원본 데이터셋은 원본 데이터셋의 여러 측면을 포함하는 시뮬레이트 된 데이터셋을 만드는 데 사용된다. 이러한 데이터셋은 공개적으로 또는 조사된 연구원에게 공개된다. 시뮬레이션 된 데이터는 쿼리 또는 분석 소프트웨어 개발에 사용될 수 있다.
- Enclave Model: 비식별 데이터는 원본 데이터의 외부제공을 제한하고, 연구자의 쿼리를 대신 받아서 비식별 데이터에 대한 쿼리를 실행하고 결과로 응답하는 분리된 개체로 유지 될 수 있다

정부데이터 맥락에서의 비식별 조치의 한 가지 체제로서 NIST2016은 '다섯 가지 안전성(The Five Safes)'을 제시한다. 데이터 접근 시스템을 디자인, 구성, 평가하기 위해 만들어진 프레임워크(framework)이다. 특히, 이 시스템은 미국 통계국 또는 영국 통계국과 같은 국가적 통계기관들이 외부 주체들과 정보를 공유하기 위해 만들어진 접근 시스템에 주로 사용된다. 이 프레임워크는 다음과 같은 다섯 가지 확인 사항들로 구성된다.

(1) 안전한 프로젝트(safe projects): 데이터의 이러한 사용이

적절한가?

(2) 안전한 사람(safe people): 연구자들이 적절한 방법으로 데이터를 사용할 것을 믿을 수 있는가?

(3) 안전한 데이터(safe data): 데이터 자체가 노출 위험이 있는가?

(4) 안전한 세팅(safe settings): 데이터 접근 시설이 비허가된 사용을 제대로 제한하고 있는가?

(5) 안전한 결과물(safe outputs): 통계 결과가 비공개인가?

NIST2016은 비식별 조치의 전체적인 체계를 제시하고 있다. 이 체계는 비식별 기법의 적용에서부터 사후적인 관리에 이르기까지의 과정을 포괄적으로 담고 있다. 이는 다음의 10단계로 구성되어 있다.

[1 단계] 공개된 비식별 조치된 데이터의 용도를 파악한다.

[2 단계] 비식별을 수행하기 전에 식별된 데이터를 공개함으로써 발생할 수 있는 위험요소들을 파악한다.

[3 단계] 비식별된 데이터에 존재하는 데이터 유형(형식)들을 파악한다.

[4 단계] 비식별을 수행하는데 사용될 방법들을 파악한다.

[5 단계] 외부 파일과의 연결 가능성을 검토하고 제거한다.

[6 단계] 승인된 방법을 사용하여 재식별을 수행한다. 예를 들어, 식별자를 제거하고 준식별자를 변환, 합성 데이터를 생성하거나 대화식 쿼리 인터페이스를 개발하여 비식별을 수행 할 수 있다.

[7 단계] 테스트 및 유효성 확인을 위해 변형 된 데이터를 다른 시스템으로 내보낸다.

[8 단계] 비식별 데이터 품질을 테스트한다. 비식별 데이터에 대한 분석을 수행하여 충분한 유용성과 데이터 품질을 갖고 있는지 확인해야 하다.

[9 단계] 재식별 시도. 비식별 데이터를 검사하여 데이터를 다시 식별 할 수 있는지 확인한다. 이 단계에서는 외부 팀이 참여할 수 있다.

[10 단계] 비식별 기술과 결과를 보고서로 문서화해야 한다.

비식별 조치가 적용된 데이터의 사후 검증에 대한 사항들에 대하여는 간략하게 설명하고 있다. 정부 기관은 결과 데이터 집합이 개인 정보 보호와 데이터 유용성 측면에서 기관의 목표를 충족하는지를 확인하기 위해 데이터 집합을 비식별한 후에 데이터 집합의 유효성을 검사해야 한다. 개인 식별 정보가 제공하는 개인 정보 보호를 검증하기 위한 몇 가지 접근 방법이 있다. 우선 데이터 파일을 검사하여 식별정보가 파일 데이터 또는 메타 데이터에 포함되어 있지 않은지 확인한다. 다음으로 특정한 거증 분석을 수행하여 외부의 개인이 공개적으로 사용 가능한 데이터셋 또는 기밀처리된 기관 데이터를 사용하여 재식별할 수 있는지를 확인한다. 동시에 데이터 유용성을 검증하기 위한 몇 가지 방법이 존재한다. 가령 원본 데이터셋과 비식별 데이터셋 모두에 대해 동일한 통계 계산을 수행하여 그 결과를 비교함으로써 받아들일 수 없을 만큼 중대한 변경이 이루어졌는지를 확인할 수 있다.

(2) 비식별정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령 및 가이드라인

가. HIPAA 프라이버시규칙

HIPAA의 적용대상이 되는 보건의료정보는 '개별적으로 식별가능한 보건의료정보(individually identifiable health information)'이다(45 CFR 164.103). 따라서 HIPAA 프라이버시규칙은 개별적으로 식별할 수

없는 보건의료정보는 HIPAA의 보호 대상이 아니라고 밝히고 있다(45 CFR 164.514). HIPAA 프라이버시규칙은 비식별 조치된 정보의 개념을 정의하지는 않지만 HIPAA의 적용 대상이 되지 않기 위해 적용되는 비식별 조치의 기준과 구체적 방법들을 제시하고 있다. 앞서 설명한 바와 같이 비식별 조치의 기준으로 합리적(reasonably) 기대가능성을 제시하고, 비식별 조치의 방법으로 18가지 식별자를 제거하는 방법과 전문가 판단 방식을 정하고 있다. 하지만 익명화 또는 가명화에 대한 개념정의는 HIPAA 프라이버시규칙과 OCR가이드라인에 포함되어 있지 않다. 이 두 문헌은 익명화라 표현 자체를 사용하지 않고 있고, OCR가이드라인은 재식별의 맥락에서 가명(pseudonym)이란 표현을 쓰고 있을 뿐이다.

나. CCPA

CCPA는 비식별 조치된 정보의 개념 정의를 하고 있다. 하지만 익명화 개념은 CCPA에 없다. 익명화와는 대조적으로 CCPA는 가명화의 개념 정의 조항을 포함하고 있다. 가명화란 추가정보의 활용 없이는 개인정보가 특정 소비자의 속성을 드러내지 않도록 적용하는 조치를 뜻한다(California Consumer Privacy Act of 2018, Art.1798.140(r)). 이 경우 추가정보는 분리해서 보관되어야 하고 식별 가능한 소비자의 속성이 드러나지 않도록 기술적 및 관리적 방법을 적용해야 한다. 이 가명화에 대한 정의 규정은 GDPR의 내용과 동일하다.

2) 법령 이외의 형태

NIST2015는 비식별 조치와 익명화의 개념이 정리되어 있지 않다는 점을 밝히고 있다. 비식별 조치와 익명화의 다양한 개념 정의를 제시한 뒤 일관된 기준이 없다고 한다. 비식별 조치, 익명화, 가명화에 대한 개념을 다음과 같이 정의하고 있다.

- 비식별 조치: 일련의 식별데이터와 정보주체 사이의 연관성을 제거하는 모든 절차에 대한 포괄적인 개념
- 익명화: 식별데이터셋과 정보주체 사이의 연관성을 제거하는 절차
- 가명화: 정보주체와의 연관성을 제거하면서 해당 정보주체의 특징들과 가명 사이의 연관성을 추가하는 특정 유형의 익명화

그럼에도 NIST2015는 비식별 조치와 익명화에 대한 위 개념 정의의 불명확성과 익명화의 비일관적 사용을 고려하여, 불필요한 혼란을 피하기 위하여 비식별 조치라는 하나의 용어로 통일하겠다고 한다. NIST2016은 NIST2015에서 설명한 위 세 가지 개념 정의와 접근 방식을 기본적으로 받아들였다. NIST 2016은 익명화 개념과 관련해서 일부 연구자들은 “비식별 조치” 와 “익명화”를 번갈아가면서 사용하기도 하며, 다른 이들은 “비식별 조치”를 하나의 프로세스로, “익명화”는 되돌릴 수 없는 비식별 조치의 한 종류로 표현하기도 한다고 설명한다. 즉, ‘비식별’, ‘익명’ 등의 용어사용에 관해 아직까지 일반적으로 인정되는 합의가 존재하지 않음을 강조한다. 또한 가명화된 데이터는 잠재적으로 재식별 가능한 데이터로 간주할 것을 기관들에게 권고하고 있다.

(3) 비식별정보, 익명정보, 가명정보 개념의 도입 과정

1) 배경 및 구체적 절차

2003년에 제정된 HIPAA 프라이버시규칙은 보건의료정보 관련 기술의 발전에 따라 보건의료정보에 대한 활용가치가 높아지는 사회적 배경을 반영하고 있다. OCR가이드라인에 따르면 HIPAA 프라이버시규칙이 도입한 비식별 조치 방법은 프라이버시 침해의 위험을 완화하면서 보건

의료정보의 '2차적 이용(secondary use)'을 용이하게 하는 방식이라는 제도적 가치를 가지고 있다. 예를 들어, 다양한 비교효과 연구, 정책 평가, 생명과학 연구에 비식별 조치에 따른 데이터 활용을 적용할 수 있다. 특히 비식별 조치가 적용되어 식별자들이 제거된 잔존 데이터의 잠재적 유용성의 가치가 인지된 상태이기 때문에 HIPAA 프라이버시규칙에서 규정한 비식별 조치의 방법은 실질적 효용성도 높다고 한다.

CCPA의 경우 비식별 조치된 정보와 가명화에 대한 개념 정의 조항을 도입한 입법의도를 명시적으로 밝히고 있지 않다. 하지만 CCPA가 강력한 소비자의 권리보호 조항들을 도입하고 있기 때문에 기업들이 사업적 가치 창출이 지나치게 위축될 수 있다는 우려를 반영해 비식별 조치된 정보는 CCPA의 적용 대상이 아니라는 명시적 예외 규정을 도입한 것이다. 특히 비식별 조치된 정보의 개념정의 조항에 재식별 금지와 관련된 추가 요건을 구체적으로 제시한 사실을 통해 볼 때 프라이버시 침해의 위험성을 최소화하면서 데이터의 활용성을 유지하려는 방법으로 비식별 조치된 정보의 개념정의를 규정하였다고 이해된다.

2) 논의 사항

HIPAA 프라이버시규칙은 비식별화의 방법으로 18가지 식별자의 제거와 전문가 판단의 두 가지 선택지를 제시하고 있다. 18가지 식별자의 제거의 경우 어떤 식별자를 어느 정도의 범위까지 삭제해야하는 것인지에 대한 다양한 논의를 배경으로 한다. 예를 들어, 18가지 식별자들 중의 한 가지로 나열된 우편번호(zip code)는 5자리로 이루어진 숫자 체계인데 HIPAA 프라이버시규칙은 인구 2만 명을 기준을 최초 3자리 숫자의 제거의 필요성을 판단하고 있다. 이 경우 최초 3자리 숫자를 삭제한다는 것은 실질적으로 000과 같은 표시로 대체한다는 의미이다. 우편번호와 대한 이런 기준은 1997년에 발표된 Latanya Sweeney의 연구결과에 영향을 받은 것이다(Sweeny, 1997). Sweeney는 생년월일과 전체 우편번호 정보를 활용하면 미국 메사추세츠(Massachusetts) 주의

캠브리지(Cambridge) 에 등록된 투표자명부 등록자들의 97%의 이름과 주소를 식별할 수 있음을 보였다. 이 연구의 영향으로 다른 외부 정보와의 결합으로 재식별될 위험성이 나타날 수 있다는 점을 인지하면서 삭제해야할 우편번호의 기준을 마련하였다.

한편 CCPA에서 비식별 조치된 정보와 가명화 개념정의 조항을 위한 어떤 구체적인 논의가 있었는지에 대해 명확하게 밝혀진 사실은 없다. 하지만 개념정의 조항을 보면 HIPAA 프라이버시규칙과 GDPR의 영향력을 짐작할 수 있다. HIPAA 프라이버시규칙은 비식별 조치된 정보의 판단 기준으로 합리성(reasonable)을 제시하고 있다(45 CFR 164.514). CCPA 또한 합리적 식별가능성 기준을 제시하고 있다. 비식별정보는 CCPA의 적용 대상이 되지 않는다는 조항 또한 이미 HIPAA 프라이버시규칙에 있다. CCPA의 가명화 개념정의는 GDPR의 가명화 개념 정의와 그 구조와 내용이 동일하다.

(4) 비식별정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

오바마 행정부는 정부 기관이 보유한 데이터를 적극적으로 공유하려는 "정보공개(Open Data)" 정책을 추진한 바 있다. 그에 따라 2009년부터 2013년까지 정보공개 정책에 기반으로 아래와 같은 4개의 주요 정책 문서들을 발표했다[22].

- 투명성과 공개 정부에 대한 각서(Memorandum on Transparency and Open Government, 2009)
- 공개정부 지침(The Open Government Directive, 2009)
- "디지털정부: 미국 국민들에게 개선된 서비스를 제공하기 위한 21세기 플랫폼 건설"(Digital Government: Building a 21st Century Platform to Better Serve the American

22) https://opengovdata.io/2014/us-federal-open-data-policy/ 및 https://www.ncbi.nlm.nih.gov/books/NBK9573/.

People, 2012) 보고서

- 공개데이터 정책에 대한 각서 - 자산으로서 정보관리 (Memorandum on Open Data Policy—Managing Information as an Asset, 2013)

2017년에 트럼프 행정부가 등장하면서 최근에는 연방정부 차원에서 위와 같은 적극적 데이터 정책을 추구하지 않는 것으로 판단된다. 트럼프 행정부는 2018년 페이스북 대규모 정보유출 사건이 발생하면서 소비자 데이터 측면에서 소비자 프라이버시 정책을 제시하기 위한 작업을 진행 중인 것으로 알려져 있다[23].

23) https://www.reuters.com/article/us-usa-internet-privacy/trump-administration-working-on-consumer-data-privacy-policy-idUSKBN1KH2MK.

2. 캐나다

(1) 제도적 개괄

1) 개인정보보호 규제 일반

캐나다는 연방법과 주(state)법으로 구성되는 이원적 법체계를 가지고 있다. 연방법 차원에서 개인정보를 일반적으로 규율하는 일반법으로는 연방정부 산하 기관에 적용되는 Privacy Act와 민간 기업에 적용되는 법률인 Personal Information Protection and Electronic Documents Act(이하 "PIPEDA")가 있다. 그 이외에 개별 영역을 규율하는 법들이 있다. 가령 Federal Bank Act는 개인금융정보의 보호와 활용에 대한 규정들을 포함하고 있다.

2) 비식별 조치

위에서 언급한 두 일반적 연방법 모두 비식별 조치의 기준 등 구체적인 내용을 제시하는 조항은 포함하지 않고 있다. 단지 PIPEDA는 개인정보의 개념을 '식별가능한(identifiable)' 개인에 대한 정보로 정의하고 있으므로, 식별가능성 개념을 근거로 한 비식별 조치 방법의 가능성을 인정하고 있다는 해석이 가능하다(PIPEDA, Part I Section 2-4). 한편 2016년 온타리오주(Ontario)의 개인정보보호기구(Information and Privacy Commissioner of Ontario)는 구조적 데이터(structured data)에 대한 비식별 조치 가이드라인(Canada Information and Privacy Commissioner of Ontario, De-identification Guidelines for Structured Data, 2016, 이하 "IPCO가이드라인"이라 한다)을 발표했다. 이 가이드라인은 현행법 하에서 데이터를 비식별 조치하기 위한 절차를 구체적으로 알려준다.

이 가이드라인은 비식별 조치가 데이터가 재식별될 수 있는 위험성을 완전히 제거하는 절차가 아니라하는 전제를 명확하게 하고 있다. 단지, 비식별 조치는 재식별의 위험성을 매우 적은 수준으로 만드는 절차라고 함으로써 비식별 조치는 위험성의 제거가 아니라 위험성의 관리를 목적으로 하는 도구임을 밝힌다. 이런 전제 하에 이 가이드라인은 비식별 조치의 기본 개념과 여러 기술적인 비식별 조치 방법을 소개한다. 특히 데이터의 유형 중 가장 전형적인 구조적 데이터 형태의 개인정보를 비식별 조치할 때 고려할 문제점들을 제시하고, 데이터셋(dataset)에 포함된 개인정보를 제거하는 단계적인 절차들을 설명한다.

위험성을 매우 작게 만드는 절차가 중요하기 때문에 이 가이드라인은 위험성에 기반한 접근법(risk-based approach)을 제시하고 있다. 따라서 데이터를 공유하거나 공개할 때 허용되는 위험성 수준을 산정하는 과정이 필요하다. 위험성 수준을 정확하게 산정하게 위해서는 주어진 데이터의 잠재적 공격자과 같은 데이터 환경에 대한 정확한 파악이 필요하다. 일반적으로 잠재적 공격자의 주체로 특정 개인정보가 데이터셋에 포함되어 있는 것을 알고 있는 경우와 알지 못하는 경우를 구분, 개별적으로 위험성을 계산한다. 가이드라인에서는 전자의 경우를 상정한 데이터 환경 하에서의 재식별 위험성을 산출하는 것을 목표로 한다. 후자의 경우보다 전자의 경우가 재식별의 위험성이 높기 때문에 전자의 경우로 논의의 범위를 한정해서 적용한 논리 구조가 후자의 경우에도 적용될 수 있다.

이런 이론적 전제 하에서 이 가이드라인은 9단계의 비식별 조치 모형을 제시한다. 전체 모형을 다음의 표로 정리할 수 있다.

[표4-12] PIPEDA의 9단계 비식별 조치 모형

1단계	공개 모형 선택
2단계	변수 분류
3단계	재식별 위험성 기준치(threshold) 결정
4단계	데이터 자체의 위험성 측정
5단계	데이터 맥락(context)의 위험성 측정
6단계	총 위험성 산정(4단계 위험성 x 5단계 위험성)
7단계	데이터의 비식별 조치 처리
8단계	데이터의 유용성(utility) 평가
9단계	전과정 기록 및 문서화

자료 : 연구진 작성
(Canada Information and Privacy Commissioner of Ontario, De-identification Guidelines for Structured Data, pp.6-20(2016) 참조)

1단계는 공개 모형의 선택으로 데이터를 공개하는 유형을 우선적으로 결정하는 단계이다. 비식별 조치된 정보는 완전공개(public release) 또는 부분공개의 형태로 공개가능하다. 완전공개의 예로 인터넷 환경에 누구라도 접근할 수 있는 방식으로 공개하는 방식이 있고, 부분공개의 예로 특별한 계약 관계 하에서만 공개하는 방식이 있다. 공개 모형을 선택하는 것을 가장 먼저 고려해야 하는 이유는 완전공개인지 아니면 부분공개인지에 따라 공개되는 데이터의 양이 달라지기 때문이다. 완전공개의 경우에는 공개된 데이터에 대한 사후통제가 어렵기 때문에 높은 수준의 비식별 조치가 필요하다. 반면 부분공개의 경우에는 공개되는 데이터에 대한 통제가 가능하기 때문에 이미 데이터보호의 정도가 상대적으로 높은 수준에 있어서, 비식별 조치의 정도가 상대적으로 낮아지게 된다.

2단계는 변수의 분류이다. 만일 데이터셋이 개인에 관한 것이라면 한 파일에서 흔히 각 행은 한 개인을 나타내고 각 열은 개인에게서 수집된

정보의 변수들을 나타낸다. 정보의 종류에 따라 어떤 변수는 직접적으로나 간접적으로 개인을 식별하는데 사용될 수 있지만 어떤 것들은 그렇지 않을 수 있다. 비식별 조치는 개인을 식별하는 데 사용되는 변수만 고려한다. 위에서 언급했듯이 두 종류의 변수들이 있는데 이는 직접식별자와 간접식별자 또는 준식별자이다.

직접식별자는 단독 또는 다른 이미 공개된 정보의 소스와 통합하여 한 개인을 식별하는데 사용되는 하나 또는 그 이상의 변수를 포함한다. 예로는 이름, 주소, 이메일 주소, 전화번호, 팩스 번호, 신용카드 번호, 자동차 번호판 번호, 자동차 등록 번호, 사회보험 번호, 건강카드 번호, 의료기록 번호, 장치 식별명, 생체 측정 식별자, 인터넷 규약 (IP) 주소 번호 및 인터넷 콘텐츠 식별 체계(URL)가 있다. 준식별자는 두 가지의 중요한 특성을 가진 변수로 이들은 (1) 애드버서리(adversary, 공격자)는 그것에 대한 배경지식이 있다는 것으로 추정하고 (2) 데이터셋 내에 있는 개인을 재식별하기 위해 개별적으로나 통합적으로 사용될 수도 있다는 것이다. 변수는 애드버서리(공격자)가 그것에 관한 배경지식이 있는 경우에만 준식별자가 될 수 있다. 애드버서리(공격자)는 데이터셋에 있는 하나 또는 여러 개인에 대한 배경 지식을 다음과 같은 여러 방법으로 얻을 수 있는 가능성이 있는데, 실제로는 개별상황에 따라 각기 다르게 나타난다:

- 개인에 대한 정보는 공용 레지스트리(유권자 목록 또는 법적 기록), 미디어에서(예, 사망 기사), 전문적인 기관에서(예, 멤버 목록) 또는 직원(예, 직원 주소록)
- 애드버서리(공격자)는 하나 또는 그 이상의 개인을 알고 있을 수도 있다 (예, 이웃, 동료 또는 전 배우자)
- 하나 또는 그 이상의 개인이 유명 인사여서 그에 대한 공개된 정보가 있다.
- 애드버서리(공격자)가 개인에 대한 추가적인 정보소스에 접속권을 가지고 있을 수도 있다 (예, 다른 연구 프로젝트를 통한 데이터셋)

- 개인들이 자신에 대한 정보를 인터넷에 게시했을 수도 있다 (예, 소셜네트워크 사이트 또는 개인 블로그)

제3단계는 재식별 위험의 기준치를 결정하는 단계이다. 비식별 조치는 개인을 식별하는 정보 또는 개인을 식별하기 위해 하나 단독으로 또는 다른 정보와 같이 사용될 수 있다고 볼만 한 타당한 근거를 가진 정보를 제거하여 개인의 프라이버시를 보호한다. 프라이버시를 보호하기 위해서 적용되어야 하는 비식별 조치의 정도는 데이터셋을 공개하는 경우 재식별 위험의 정도에 비례한다. 데이터 공개의 재식별 위험도가 높을수록 더 높은 수준의 비식별 조치가 요구된다. 한 데이터셋에 대한 적절한 정도의 재식별 위험을 결정하기 위해서는 데이터셋의 공개가 개인의 프라이버시를 얼마나 침해하는지를 평가해야 한다. 평가의 결과는 질적 가치로 “상(high)”, “중(middle)” 또는 “하(low)”의 범위로 정할 수도 있다. 프라이버시 침해가능성의 평가 결과는 질적 가치이지만, 데이터셋에 적용되어야 하는 비식별 조치의 정도는 계량화가 필요할 수 있다. 이 차이를 줄이기 위해서는 프라이버시 침해 값을 평가한 후 그 결과를 그 위험 정도와 비례하는 비식별 조치 정도를 나타내는 숫자 값으로 변환할 필요가 있다. 이러한 “재식별 위험 한계치”는 일반적으로 비식별 조치 되었다고 고려되는 정도 즉, 더는 개인 정보가 포함되지 않도록 하는 최소한의 비식별 조치 작업을 데이터셋에 적용해야 한다.

제4단계는 데이터 자체의 위험성을 측정한다. 구조적 데이터 세트에 있는 재식별 위험성의 정도를 측정하는 데는 두 단계 절차를 거친다. 우선적으로 각 행(row)별 재식별성의 확률을 산정한다. 개인들에 대한 데이터 세트의 각 행에는 한 개인에 대한 정보가 들어있다. 따라서 각 행은 재식별 확률을 가지고 있다. 한 행의 재식별 확률은 데이터 세트 내에 준식별자인 변수와 동일한 값을 가진 다른 행들이 몇 개인지에 따라 달려있다. 재식별 확률은 1에서 이런 동일한 값을 가지는 레코드들의 개수를 나눈 값이다. 예를 들면, 준식별자의 변수에 해당하는 값들이 동일한 레코드가 5개인 경우에는 각 행의 재식별 확률은 0.2이다. 그

다음 적합한 위험성 측정 방법을 선택한다. 이 경우에는 데이터의 접근이 엄선한 식별된 수령자의 수로 제한되어 있어 재식별 공격에 대해 더 취약한 행이 없음을 가정한다. 여기서는 모든 행을 거쳐 얻은 재식별화의 평균 확률을 사용하여 데이터 세트 내에서의 재식별화 위험 정도를 측정해야 한다.

1단계에서 선택한 공개 모형에 따라 재식별 위험을 최고 위험성 기준으로 설정하거나 평균 위험성 또는 엄격한(strict) 평균 위험성 기준으로 설정한다. 데이터 공개의 경우에는 누군가가 재식별을 위해 시범적 공격을 시도할 것이라고 가정해야 한다. 이와 같은 공격은 데이터 세트 내의 가장 취약한 행들인 가장 작은 등가류 및 재식별화의 확률이 가장 높은 것을 타겟할 것이다. 때문에 재식별 위험의 정도를 측정하기 위해서는 모든 행을 거쳐 재식별의 최대확률인 최고 위험성을 사용해야 한다. 반면 부분 공개의 경우에는 평균 위험성 또는 경우에 따라 최고 위험성 기준을 적용한다.

5단계는 데이터를 둘러싸고 있는 데이터 환경이란 맥락(context)의 위험성을 측정하는 단계이다. 데이터 세트에서의 위험이 데이터 세트의 공개와 관련된 재식별화 위험의 정도를 결정하는 중요한 역할을 하지만 이 요인만 고려되는 것은 아니다. 재식별 위험은 주어진 공개 모형에서 데이터 세트에 가해질 수 있는 재식별화 공격의 여러 종류의 함수이기도 하다. 가능한 공격에 관해서 재식별화 위험을 추가로 분석하여 파악한다. 데이터 위험과 함께 이 값은 데이터 세트의 공개와 관련된 재식별의 총 위험을 산출하는데 사용된다. 문맥적 위험은 데이터셋에 대한 하나 또는 그 이상의 재식별화 공격의 확률이다. 비식별 조치된 데이터셋이 공개되자마자 재식별 공격이 가해질 수 있다. 그러나 데이터 공격자와 공격의 종류는 주어진 공개 모형에 따라 달라진다. 크게 아래와 같이 3가지 공격 유형으로 구분해서 설명한다.

1. 내부자의 고의적 공격
2. 의도없이 개인을 인식한 경우

3. 데이터 유출(breach)

제6단계는 제4단계와 제5단계에서 산출한 위험성에 근거해 총 위험성을 산정하는 단계이다. 데이터 위험 및 문맥적 위험이 측정된 후에는 총 위험을 산출할 수 있다. 총 위험은 데이터 위험에 문맥적 위험을 곱한 조건부 확률이다.

[총 위험= 데이터 위험 x 문맥적 위험]

총 위험은 공격이 가동되었을 경우 하나 또는 그 이상의 행이 재식별될 확률과 같다. 예를 들면, 한 데이터 세트의 재식별 위험성이 0.2이고 문맥적 리스크가 0.5일 경우 그 데이터 세트의 총 위험성은 0.1이다.

제7단계에서는 해당 데이터셋에 비식별 조치 처리를 적용한다. 데이터 세트가 비식별 조치된 것으로 고려되려면 식별가능한 모든 정보가 제거되어야 한다. 개인을 식별하는 모든 정보 또는 개인을 식별하기 위해서 단독으로 또는 다른 정보와 함께 사용될 수 있다고 합당하게 추론되는 정보를 제거하기 위해서는 데이터 세트의 값이 여러 방법으로 변형될 수 있다. 식별자의 종류 및 특성에 따라 다른 기술들을 적용할 수 있다. 식별가능한 정보를 제거하기 위해서는 세 가지 단계를 거친다.

우선적으로 직접식별자를 마스킹(masking) 한다. 직접식별자로 분류되는 변수들은 데이터 분석에 사용되지 않는다. 일반적으로 연구목적을 위한 사용에 적합하지 않기 때문이다. 단순히 프라이버시-보호적으로 다루는 법은 직접적인 식별 변수가 있는 열을 제거하여 데이터 세트에 있는 그 값을 삭제하는 것이다. 하지만 연구 목적에 따라 관여된 개인들에게 연락하고 결과를 통지해야 할 필요가 있을 수도 있다. 이 같은 경우에는 다음의 마스킹 기술을 사용하여 직접적 식별 변수를 변형해야 한다. 해당 값을 가명으로 대체하고 이와 연결된 매칭표(matching table)를 안전한 장소에 보관하거나, 그 값을 암호화하고 키를 안전한 곳에 보관한다. 다음 준식별자 변수에 해당하는 데이터 값들을 변형한

다. 한 데이터셋이 비식별 조치된 것으로 보려면 재식별의 총 위험이 재식별 위험 한계치 이하여야 한다. 이를 행하기 위한 일반적인 기술에는 일반화(generalization) 등이 있다. 마지막으로 이렇게 변형된 데이터에 기초한 재식별의 총 위험성을 계산해 재식별 위험 한계치와의 비교를 다시 해야 한다.

제8단계는 비식별 조치 처리된 데이터의 유용성(utility)을 평가하는 단계이다. 데이터셋에 적용된 비식별 조치의 정도와 발생된 정보의 유용성간에 상충효과가 있을 수도 있다. 준식별자라 여겨지는 변수들이 일반화 또는 억제와 같은 기술들을 사용해 비식별 조치 될수록 이에 해당하는 데이터셋의 유용성에 손실이 커진다. 재식별의 총 위험이 재식별 위험 한계치 이하가 되도록 데이터셋에 일반화와 삭제가 적용되나 이 결과를 얻기 위하여 비식별 조치 기술들이 다른 방법 및 그 조합으로 바뀌어야 할 수도 있다. 예를 들어 한 접근법은 등가류의 크기를 증가시키기 위해 일반화 및 범주의 정확성을 감소시키는 것에 중점을 두고 다른 접근법은 억제 및 등가류가 너무 작은 변수의 셀 또는 열을 제거하는 것에 중점을 두고 있을 수가 있다. 데이터셋의 특성에 따라 다른 적용 또는 일반화 방법 등의 조합이 필요하다.

제9단계에서는 이 모든 비식별 조치 절차를 기록하여 문서화한다. 개인정보가 들어있는 데이터셋을 비식별화하는 각 시도는 동일한 단계를 따르고 동일한 문제들을 평가해야 한다. 하지만, 변수들과 값, 그리고 비식별 조치의 종류 및 양을 결정하는 분석은 데이터 공개마다 다를 것이다. 이 때 개인정보를 비식별 조치 작업을 하는 절차 및 결과를 기록하는 보고서를 작성하는 것을 고려해야 한다. 우수 관행에는 다음을 포함한 여러 이점이 있다. 가령 프라이버시 위반의 경우 중요한 역할을 할 수 있는 준수의 증거 및 성실함을 입증할 수 있다.

이 가이드라인도 재식별의 위험성은 지속적으로 관리되어야 한다는 점을 강조하면서 재식별의 위험성을 정기적으로 평가하고 데이터셋을 관리해야 한다고 한다. 이미 비식별 조치된 정보와 새로운 정보가 조합되어 개인이 재식별될 가능성은 고려되어야 할 중요한 프라이버시 우려

이다. 비식별 조치 과정을 거칠 시기에는 제공되지 않은 예상 밖의 정보들이 제공되고 그것이 개인을 재식별하는 데 사용될 수 있다. 비식별된 데이터셋을 공개하자마자 새로운 정보가 나타나는지 그리고 그 소스들이 데이터 세트 안에 있는 개인들을 재식별하는 데 사용될 수 있는지를 검토하여야 한다. 지속적인 재평가에 따라 재식별의 총 확률이 재식별 위험 한계치 이하라는 것을 확인하여야 한다.

(2) 비식별정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

PIPEDA는 비식별정보, 익명정보, 가명정보의 개념을 정의하지 않고 있다. 단지 어떤 특정 목적을 이루기 위해 개인정보가 더는 필요하지 아니하다면 해당 개인정보는 파기되거나 삭제되거나 익명화되어야 한다는 조항이 있는 정도이다(PIPEDA, Art.4.5.3).

2) 법령 이외의 형태

IPCO가이드라인은 비식별 조치를 “레코드(records) 또는 데이터셋에서 개인정보를 제거하는 절차에 해당하는 포괄적인 용어”라고 정의한다. 이 개념정의를 좀 더 자세하게 본다. 가이드라인에 따르면, 비식별 조치는 더 정확히는 (i) 개인을 식별하는 정보 또는 (ii) 단독으로 또는 다른 정보와 함께 개인을 식별하기 위해 사용할 가능성이 높은 정보를 모두 제거하는 절차이다. 그러나 익명화에 대한 개념정의나 익명화와 비식별 조치의 관계에 대한 별도의 설명은 하지 아니한다. 가명화에 대한 개념정의도 없다. 단지 마스킹(masking)에 대한 개념을 “변수를 제거하거나 익명화되거나 암호화된 정보로 대체하는 절차”로 제시하면서 익명화란 표현을 사용할 뿐이다.

(3) 비식별정보, 익명정보, 가명정보 개념의 도입 과정

1) 배경 및 구체적 절차

IPCO가이드라인은 비식별 조치의 방법이 정부가 보유한 데이터의 안전한 활용을 가능하게 하는 유용한 도구라고 밝히고 있다. 즉 프라이버시 침해의 위험성 문제와 데이터 활용의 경제적 가치 모두를 효과적으로 달성하는 제도적 도구로서 비식별 조치의 개념과 구체적인 방법을 제시한다. 특히 빅데이터 분야에서의 성공을 위한 전제조건이 정보주체가 자신의 프라이버시가 존중받고 있다는 신뢰를 가지는 것인데, 이러한 신뢰성 확보를 위한 가장 효과적인 방법들 중 하나가 비식별 조치라는 점을 지적한다.

2) 논의 사항

IPCO는 2016년 가이드라인을 발표하기 이전인 2014년 “Big Data and Innovation, Setting the Record Straight: De-identification Does Work”라는 보고서를 발표했다. 이 보고서는 비식별 조치에 대한 학술적 논의를 정리하고 그 유용성을 논의하고 있다. IPCO가이드라인을 통해 비식별 조치에 대한 전반적인 체계와 방법론을 제시하기 전에 먼저 전반적인 학술적 논의를 다룬 일종의 사전 연구보고서의 기능을 한다는 인상을 준다. IPCO가이드라인에서 비식별 조치란 개념을 전면적으로 설명하기 위한 기본적 논의 사항들이 이 보고서의 내용을 통해 추측할 수 있다고 판단된다.

이 보고서는 재식별 위험성과 관련하여 비식별 조치의 유용성 자체를 의심하는 연구 자료를 소개하면서 이런 태도에 비판적인 관점을 견지하고 있다. 일부 전문가가 자신들의 연구결과를 잘못 해석해서 비식별 조치가 무용하다는 주장을 하고, 나아가 자신들의 연구결과를 과장해서

이런 부정적인 결론을 제시한다고 한다. 하지만 이 보고서는 높은 수준의 데이터 품질(quality)을 유지하면서 재식별 위험을 최소화하는데 비식별 조치를 쓸 수 있다고 한다. 완벽한(perfect) 장치는 없다는 전제하에 적절한 방법으로 비식별화하면 실제 재식별의 위험이 충분히 낮은 상태를 달성하고 유지할 수 있다고 한다.

(4) 비식별정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

IPCO는 2017년 “Big Data Guideline(이하 ”빅데이터 가이드라인“)을 공개했다. 빅데이터 가이드라인은 주로 정부기관들이 빅데이터의 활용에 초점을 맞추어 설명을 하고 있는데, 빅데이터의 효과적인 활용을 위한 제도적 장치로 비식별 조치를 소개하고 있다. 즉 빅데이터 가이드라인은 비식별 조치의 방법을 적용해서 빅데이터의 효과적인 활용을 추구한다는 정책 방향을 제시한 것이다.

빅데이터 가이드라인에 따르면 정부기관의 빅데이터 활용 문제는 중요하면서 시의적절한 주제로 다양한 정책 목표의 달성을 위한 방법으로 이용 가능한 새로운 정보기술이 관련되어 있다. 이런 인식 하에 IPCO는 정부기관들의 데이터 활용이 효과적인 정책 이행에 도움이 될 안내서를 제공하는 측면에서 빅데이터 가이드라인을 발표했다. 행정적 처리와 정책 분석을 위해서는 다양한 형태의 정보가 필요하므로 개별 기관의 업무 목적에 따라 정보의 처리 형태가 달라진다. 만약 특정 개인들에게 정부 프로그램을 전달할 목적이라면 식별가능한 정보를 제공해야 하는 반면, 정책분석의 목적을 달성하기 위해서는 개인이 아닌 단체 또는 전 국민을 대상으로 한 데이터분석이 필요하다. 후자의 경우 식별성이 제거된 데이터가 필요하다. 결국 식별성이 제거되는 데이터가 필요한 상황을 위해서 적절한 비식별 조치가 적용된 데이터가 있어야 정책 목적을 효과적으로 달성할 수 있다. 빅데이터 가이드라인은 이러한 정책목적의 달성이란 맥락에서 해당 데이터에 적용된 비식별 조치가 빅데이터

와 데이터 최소화(data minimization) 사이의 긴장관계를 해결하기 위한 장치로 작용한다고 한다. 빅데이터 가이드라인은 프라이버시 침해를 완화하는 비식별 조치의 장점으로 도난, 손실(loss), 비합법적 공개와 활용으로부터의 보호를 제시하고 있다.

제 3 절 아시아

1. 일본

(1) 제도적 개괄

1) 개인정보보호 규제 일반

일본은 개인정보를 포괄적으로 규율하는 법으로 개인정보의 보호에 관한 법률(個人情報の保護に関する法律, 이하 “일본 개인정보보호법”)을 2003년 제정했다. 그 후 같은 법은 2015년 전면 개정되어 2017. 5. 30.부터 시행되고 있다. 개정 개인정보보호법의 주안점은 빅데이터 산업의 활성화를 위해 데이터의 활용성을 높이는 것이었다. 이러한 개정 의도에 따라 개정 개인정보보호법은 ‘익명가공정보(匿名加工情報)’란 새로운 개념을 도입해서, 개인정보를 익명가공처리를 적용해서 익명가공정보로 변환할 경우에는 정보주체의 사전동의 없이도 수집 목적 이외의 활용이 가능하도록 하였다. 이후 시행령과 시행규칙이 순차로 마련되었다.

2) 비식별 조치

가. 법령

개정 개인정보보호법은 익명가공정보란 새로운 법적 개념을 도입하고 있다. 익명가공정보란 “개인정보의 구분에 따라 정해진 조치를 강구해서 특정 개인을 식별할 수 없도록 개인정보를 가공하여 얻어진 정보”로서 “해당 개인정보를 복원할 수 없도록 하는 것”이다(個人情報の保護に関する法律 第2條 第9項).

나아가 같은 법은 부연하여 복원할 수 없도록 조치를 취한다는 것은

개인정보에 포함된 특정 정보나 부호를 삭제하거나 다른 방법으로 대체하는 것을 의미한다고 설명하였다(個人情報の保護に関する法律 第2條 第9項). 이로부터 알 수 있듯이 특정 개인을 알 수 없도록, 즉 식별성을 제거하는 방법을 적용한 정보로서 익명가공정보는 일종의 비식별정보에 해당한다. 익명가공정보에 해당하는 데이터는 정보주체의 사전동의 면제라는 법적 효과를 가지기 때문에 제도적 관점에서도 사전동의 방법과 동일한 위치에 있는 비식별 조치를 통한 개인정보 규제라고 볼 수 있다.

개정법은 나아가 익명가공정보를 활용하는 개인정보취급사업자에게 구체적인 법적 의무를 부과하고 있다. 우선 익명가공정보로 가공하는 경우의 의무이다(個人情報の保護に関する法律 第36條 第1項). 이때에는 개인정보보호위원회규칙에서 정한 기준을 준수해서 가공해야 한다. 다음 익명가공정보 가공의 방법을 외부에 누설해서는 안 된다는 의무가 있다(個人情報の保護に関する法律 第36條 第2項). 누설을 막기 위하여 개인정보취급사업자는 적절한 안전조치를 하여야 한다. 세 번째는 공표의무이다(個人情報の保護に関する法律 第36條 第3項). 익명가공정보를 작성할 때 그에 포함되는 개인에 관한 정보의 내용을 공표해야 한다. 또한 익명가공정보를 제3자에게 제공할 경우에도 미리 익명가공정보에 포함된 개인에 관한 정보의 내용과 제공 방법을 공표하고, 그 제3자에게는 제공되는 정보가 익명가공정보임을 명시해야 한다(個人情報の保護に関する法律 第36條 第4項). 개인정보취급사업자의 재식별행위를 금지하는 조항도 있다. 개인정보취급사업자는 당해 익명가공정보의 작성에 사용된 개인정보와 관계된 본인을 식별하기 위하여 익명가공정보를 다른 정보와 결합해서는 안 된다(個人情報の保護に関する法律, 第36條 第5項)

앞서 언급한 바와 같이 익명가공정보를 가공할 때에는 개인정보보호위원회규칙(個人情報保護委員会規則)을 준수하여야 한다. 이와 관련하여 개인정보보호위원회규칙은 구체적인 가공의 방식으로 다음의 5가지를 제시하고 있다(個人情報保護委員会規則 第19條).

(1) 개인정보에 포함된 특정 개인을 식별할 수 있는 기술(記述) 등의 전부 또는 일부를 삭제 할 것.(해당 전부 또는 일부 기술 등을 복원할 수 있는 규칙성을 갖지 아니한 방법에 따라 다른 기술 등으로 대체하는 것을 포함.)

(2) 개인 정보에 포함된 개인 식별 부호의 전부를 삭제할 것. (해당 개인 식별 부호를 복원할 수 있는 규칙성을 갖지 아니한 방법에 따라 다른 기술 등으로 대체하는 것을 포함.)

(3) 개인 정보와 해당 개인 정보에 조치를 취해 얻은 정보를 연결하는 부호 (실제로 개인 정보 취급 사업자가 취급하는 정보를 서로 연결하는 부호에 한함)를 삭제할 것. (해당 부호를 복원할 수 있는 규칙성을 갖지 아니한 방법에 따라 해당 개인 정보와 해당 개인 정보에 조치를 취해 얻은 정보를 연결할 수 없는 부호로 대체하는 것을 포함한다.)

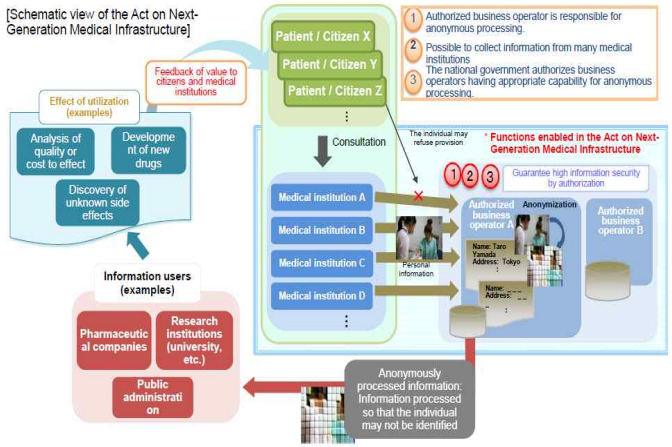
(4) 특이한 기술 등을 삭제할 것.(해당 특이한 기술 등을 복원할 수 있는 규칙성을 갖지 아니한 방법에 따라 다른 기술 등으로 대체하는 것을 포함.)

(5) 이전 각호의 정한 조치 외, 개인 정보에 포함된 기술 등과 해당 개인 정보를 포함하는 개인 정보 데이터베이스 등을 구성하는 다른 개인 정보에 포함된 기술 등과의 차이, 기타 해당 개인 정보 데이터베이스 등의 성질을 고려하고 그 결과에 따라 적절한 조치를 취할 것.

익명가공정보에 대한 내용은 위 법률과 시행령 이외에 가이드라인에도 있다[個人情報の保護に関する法律についてのガイドライン (匿名加工情報編) , 이하 “일본 가이드라인”이라 한다]. 일본 개인정보보호위원회가 마련한 일본 가이드라인은 이 법률과 시행령 규정의 내용을 좀 더 구체적으로 설명하고, 필요한 경우에는 구체적인 예시도 제시하고 있다. 이 가이드라인 이외에 일본의 개인정보보호위원회 사무국이 2017년에 ‘개인정보의 이용 및 활용 촉진과 소비자 신뢰성 확보의 양립을 위하여

(匿名加工情報- パーソナルデータの利活用促進と消費者の 信頼性確保の両立に向けて)'라는 보고서(이하 "일본 보고서")를 발표하였다. 보고서에서는 익명가공정보의 취급 및 생성, 활용 등에 수반되는 주의 사항과 구체적인 활용 사례들이 포함되어 있다. 이 두 문헌은 아래에서 살펴보기로 한다.

위에서 언급한 두 가지 문헌을 살펴보기 이전에 일본에서 2017년에 제정되어 2018. 5. 11.부터 시행되고 있는 '의료분야의 연구개발에 기여하기 위한 익명가공 의료정보에 관한 법률인 "차세대의료기반법(次世代医療基盤法)"을 볼 필요가 있다. 이 법률은 의료 영역에서의 첨단 연구개발 및 신산업 창출을 위해 일본 국민의 권리보호와 동시에 익명 가공된 의료정보의 안전한 활용을 촉진하기 위해 제정되었다. 같은 법은 특히 익명가공 의료정보의 작성 및 취급에 관한 사항들을 구체적으로 규정한다. 차세대의료기반법은 익명가공 의료정보를 "특정의 개인을 식별할 수 없도록 의료정보를 가공한 개인에 관한 정보이며, 해당 의료정보를 복원할 수 없도록 한 것"으로 정의하고 있다(次世代医療基盤法 第2條). 이 개념규정에서도 확인할 수 있듯이 차세대의료기반법은 일본 개인정보보호법의 익명가공정보 개념을 그대로 도입하고 있다. 이러한 개념을 전제로 일종의 opt-out 방식으로 익명가공정보를 활용하는 것을 허용한다. 즉, 개인에 대한 정보를 보유한 의료기관은 정보주체 자신이 거부하지 않는 한 "인정 익명가공 의료정보작성 사업자"에게 익명가공 의료정보를 제공할 수 있다(한국인터넷진흥원, 2018). 익명가공정보의 활용이 이루어지는 전체적인 흐름을 (그림 4-13)을 통해 볼 수 있다.



(그림4-13) 익명가공정보의 활용이 이루어지는 전체적인 흐름
출처: Tadasi Mina, "Trend about Personal Data in Japan", 2018/8/6 한국인터넷진흥원 비식별 조치 세미나 발표 자료

나. 법령 이외의 형태

a. 일본 가이드라인

일본 가이드라인은 일본 개인정보보호법에서 규정하는 익명가공정보의 개념과 관련하여 이 법률에서 표현한 '특정 개인을 식별할 수 있다' 함은 정보 객체 또는 복수의 정보를 조합해서 보관한 것 중 사회 통념상 그렇게 판단할 수 있는 것을 의미한다고 설명한다. 일반인의 판단능력 또는 이해력을 갖고 생존하는 구체적인 인물과 정보 간에 동일성을 인정할 수 있는지 여부에 따른다는 취지이다. 익명가공정보에 요구되는 '특정 개인을 식별할 수 없다'는 요건은 모든 방법을 동원해 특정할 수 없도록 기술적 측면에서 모든 가능성을 배제하는 것이 아닌 적어도 일반인 및 일반 사업자의 능력, 수법 등을 기준으로 할 때 그 정보

를 개인정보취급사업자 또는 익명가공정보취급 사업자가 기존 방법으로 특정할 수 없는 상태가 되도록 하는 것을 의미한다.

또한 '해당 개인정보를 복원할 수 없게 한 것'은, 익명 가공정보에서 익명가공정보의 작성원(作成源)인 개인정보에 포함된 특정 개인을 식별하게 될 기술 등 또는 개인식별부호의 내용을 특정하여 익명가공정보를 개인정보로 되돌릴 수 없는 상태로 만드는 것을 뜻한다. 개인정보취급사업자가 취급하는 개인정보에는 성명, 주소, 생년월일, 성별 외에 다양한 개인에 관한 기술(記述) 등이 포함된다. 이 기술(記述)에는 성명처럼 그 정보 하나로 특정 개인을 식별할 수 있는 것 외에도 주소, 생년월일 등 기술(記述)의 조합으로 특정 개인을 식별할 수 있는 것이 있다. 이와 같이 특정 개인을 식별할 수 있는 기술 전부 또는 그 일부를 삭제 또는 그렇지 아니한 다른 기술(記述)로 대체함으로써 특정 개인을 식별할 수 없도록 가공해야 한다.

위 가이드라인은 익명가공정보의 안전 관리 조치 의무에 대한 구체적인 설명을 제시하고 있다. 개인정보취급사업자는 익명가공정보를 작성할 때 가공 방법 등에 관한 정보의 유출을 방지하기 위해 규칙에서 정하는 기준에 따라 필요한 조치를 마련해야 한다. 해당 조치 내용은 대상인 가공 방법 등 정보가 유출된 경우의 재식별 리스크의 규모를 고려하고, 해당 가공 방법 등 정보의 양, 성질 등에 따른 내용을 포함하여야 한다. 개인정보취급사업자나 익명가공정보취급사업자는 익명가공정보의 안전 관리 조치, 불평 처리 등 익명가공정보의 적절한 취급을 확보하기 위해 필요한 조치를 스스로 강구하고 해당 조치 내용을 공표하도록 노력해야 한다. 해당 안전 관리 등의 조치는 개인정보와 동일하게 취급할 필요는 없지만, 개인데이터 안전 관리, 종업원 감독 및 위탁처 감독과 개인정보 취급에 관한 불평 처리에서 요구되는 조치의 사례를 참고할 수 있다. 구체적으로는 사업의 성질, 익명가공정보의 취급 상황, 취급할 익명가공정보의 성질과 양 등에 따라 합리적이면서도 적절한 조치를 마련하는 것을 권장한다.

이외에 익명가공정보의 공표의무와 관련하여 그 방법을 구체적으로

설명하고 있다. 개인정보취급사업자는 익명가공정보를 작성하였을 때는 익명가공정보의 작성 후 지체없이 인터넷 등을 이용, 익명가공정보에 포함된 개인에 관한 정보의 항목을 공표해야 한다. 여기서 '지체없이'는 정당하면서도 합리적인 기간이라면 익명 가공 정보를 작성한 직후 공표를 하지 않더라도 인정할 수 있음을 뜻한다. 다만 적어도 익명가공 정보의 이용 또는 제3자 제공 전에 익명 가공 정보를 작성했음을 것을 일반인에게 알리는데 충분한 기간을 확보해야 한다. 허용되는 구체적인 기간은 업종 및 비즈니스 업태에 따라 다르므로 개별적으로 구체적인 판단을 할 필요가 있다. 개인에 관한 정보항목이 동일한 익명가공정보를 동일한 방법으로 반복적 계속적으로 작성할 경우에는 처음 익명가공정보를 작성해서 개인에 관한 항목을 공표할 때 작성 기간 또는 계속된 작성이 예정되어 있음을 명기하는 등 계속적으로 작성될 것임을 확실히 해둘 수 있다. 이때에는 그 후 작성될 익명가공정보에 관한 공표는 그 이전의 (포괄적인) 공표에 따라 이미 이루어졌다고 해석한다.

재식별행위의 금지의무와 관련하여서는 금지되는 재식별 행위가 무엇인지에 대한 구체적인 인식을 제공하기 위해 아래의 표의 내용과 같이 금지되는 재식별행위에 해당하는 사례와 그렇지 않은 사례를 제시하고 있다. 해당 익명가공정보의 작성원이 된 개인정보의 본인을 식별하기 위해 다른 정보와 대조하는 것은 금지된다. 개인정보로써 이용목적의 범위 내에서 취급하는 경우에는 대조를 금지하지는 않는다.

[표4-13] 재식별행위에 해당 사례와 그렇지 않은 사례 비교

해당 사례	보유한 개인 정보와 익명 가공 정보에 대해서 공통된 기술 등을 선별해서 이들을 대조하는 것
	직접 작성한 익명 가공 정보를 해당 익명 가공 정보의 작성원이 된 개인 정보와 대조하는 것

해당하지 않는 사례	복수의 익명 가공 정보를 조합해서 통계 정보를 작성하는 것.
	익명 가공 정보를 개인과 관계없는 정보 (예: 기상 정보, 교통 정보, 금융 상품 등의 거래 금액)와 함께 경향을 통계적으로 분석하는 것.

출처: 日本 個人情報保護委員会,
個人情報の保護に関する法律についてのガイドライン (匿名加工情報編) (개인정보보호위원회 번역), p.16(2016)

b. 일본 보고서

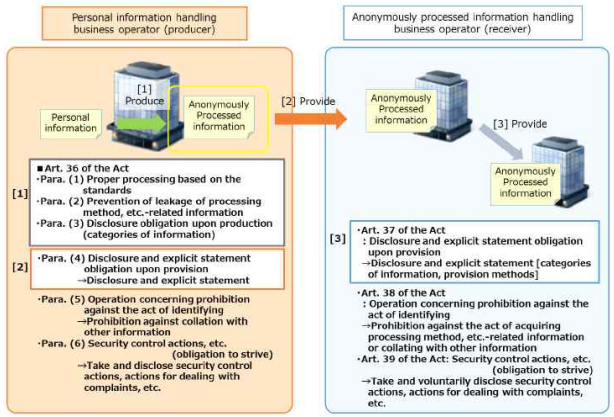
일본 보고서는 위에서 설명한 법령들이 제시한 익명가공정보의 기준들을 기반으로 좀 더 구체적으로 익명가공정보의 수집, 제공 및 공유의 사례와 방법들을 설명하고 있다. 이 보고서는 Q&A 형태로 작성이 되어 있는 실용적인 안내서이다. 보고서는 (1) 익명가공정보 생성의 접근법과 방법, (2) 재식별 행위의 금지, (3) 관련 정보의 처리를 위한 보안 통제 행위의 설명에 주안점을 두고 있다. 나아가 익명가공정보 제도의 구체적인 내부 규칙들을 마련하려는 인증기관(accredited organization)과 실제로 익명가공정보를 생성하려는 기업들에게 유용한 정보를 제공하는 것을 목표로 한다. Q&A 형태지만 익명가공정보의 설명이 체계적으로 되어 있어서, 이 보고서의 대략의 목차를 살펴보는 것만으로도 의미가 있다. 전체 7장으로 구성되어 있는데 익명가공정보의 핵심적 내용으로 볼 수 있는 3장, 4장, 5장의 큰 목차를 아래의 표에 소개한다. 보고서는 익명가공정보에 대한 법령과 가이드라인의 내용을 서술하면서 각각의 개요에 대한 보충 설명을 하는 구조로 작성되어 있다.

[표4-14] 일본 보고서 3장, 4장, 5장의 목차

3장	익명가공정보란 무엇인가?
3.1	익명가공정보 이용의 장점
3.2	익명가공정보의 정의
3.3	익명가공정보 처리의 제한
3.4	익명가공정보에 대한 주의사항
4장	익명가공정보 생성 시에 필요한 처리
4.1	익명가공정보의 처리 기준에 대하여
4.2	익명가공정보 생성 시의 유의사항
4.3	익명가공정보 생성의 참조사항
5장	익명가공정보 등에 대한 안전관리조치
5.1	가공방법 등 관련 정보에 대한 안전관리조치
5.2	익명가공정보에 대한 안전관리조치 등

자료 : 연구진 작성(日本 個人情報保護委員会, (匿名加工情報-パーソナルデータの利活用促進と消費者の 信頼性確保の両立に向けて), 목차(2017) 참조)

일본보고서에 따르면 익명가공정보 체제의 목적은 사업자들 사이의 데이터 거래와 개인정보 유용성의 증진이다. 익명가공정보 처리를 통해 비식별화된 데이터는 정보주체의 사전동의 없이 수집 목적 이외의 활용이 가능해진다. 데이터의 자유로운 이동이 가능해지면서 새로운 사업 모형이 만들어지고 공공 서비스의 품질도 개선될 것이 기대된다. 이와 같이 익명가공정보가 사업자들 사이의 데이터 이동을 쉽게 하는 기능을 제대로 수행하기 위해서는 데이터의 자유로운 이동은 물론 안전한 이동이 가능해야 하기 때문에 사업자들은 익명가공정보의 가공 및 처리 시에 준수해야 할 법적 의무들을 인지할 필요가 있다. 이런 측면에서 일본 보고서는 개별 사업자들이 익명가공정보와 관련해서 준수해야 할 법적인 의무를 다음의 그림과 같이 소개하고 있다.



(그림4-14) 개별사업자들이 익명가공정보와 관련해서 준수해야 할 법적의무
출처: 日本 個人情報保護委員会事務局,
パーソナルデータの利活用促進と消費者の信頼性確保の両立に向けて,
그림3.1(2017)

일본보고서는 익명가공정보과 관련하여 유의사항으로 통계데이터와 식별가능성의 해석 문제를 언급한다. 통계데이터란 다수의 사람들로부터 획득한 정보에서 공통 요소들을 추출하고 이런 요소들을 영역(category) 별로 합쳐서(tallying) 얻은 정보이다. 이 통계데이터는 특정 단체의 특징이나 성향을 정량적으로 판단하는 데 이용된다. 일본보고서는 일본 가이드라인에 근거해서 해당 데이터와 이에 대응하는 특정인 사이의 관련성이 제거되어 있다면 그 통계데이터는 규제의 대상이 되는 개인정보가 아님을 분명히 한다. 개인정보와 익명가공정보 모두 특정 개인에 대한 정보이기 때문에 통계데이터는 처음부터 익명가공정보에 해당하지 않고 동시에 개인정보에도 해당하지 않는 별개 영역의 정보임을 강조하는 것이다.

다음 원본 개인정보를 보유하면서 동시에 익명가공정보를 생성해서 활용하려는 기업 관점에서의 식별가능성의 해석 문제이다. 이런 기업이 원본 개인정보와 이 개인정보를 처리한 익명가공정보를 모두 보유하는 경우 이 두 정보의 대조(collation)를 통해서 익명가공정보로부터 개인정보를 재식별 해낼 위험이 있다. 이런 대조 행위를 금지하는 통제 장치가 마련되어 있지 않으면 해당 익명처리된 정보는 규제의 대상이 되는 개인정보의 범주에 해당될 가능성이 높다. 일본 보고서는 이런 가능성을 제시하면서 익명가공정보가 이러한 대조 행위를 통해 개인정보가 재식별될 수 없다는 신뢰를 주는 장치가 있어야 비로소 개인정보가 아닌 익명가공정보에 해당한다고 한다. 기업들로서는 원본 데이터와 익명가공처리된 데이터 사이의 대조 행위를 금지하거나 예방하는 다양한 방법을 마련할 필요가 있다.

이외에 일본 보고서는 익명가공정보 생성시의 유의사항을 설명하고 있다. 익명가공정보는 일반인 또는 일반 기업이 가용할 수 있는 능력과 수단을 기준으로 특정인이 식별될 수 없거나 그로부터 개인정보가 추출될 수 없도록 처리되는 정보이다. 처리의 강도는 익명가공정보의 목적과 재식별 위험성의 측정에 따라 달라진다. 그러므로 기업은 익명가공정보의 내부적 처리 원칙을 고려할 때, 익명가공정보를 생성하는 목적과 사용 방법을 우선적으로 검토할 필요가 있다. 예컨대 익명가공정보의 사용 목적에 따라 해당 정보의 필요한 범위와 세밀성(granularity)이 정해진다. 이외에도 일본 보고서는 익명가공정보를 생성하는 처리 방법으로 k-익명성(k-anonymity)과 같은 다양한 통계적 기법과 함께 개별 식별자의 재식별 위험의 정도와 유출시의 피해 정도를 제시하고 있다.

(2) 비식별정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

앞서 설명한 바와 같이 개정 일본 개인정보보호법은 익명가공정보라는 새로운 개념을 도입했다. 전체의 개념 정의 조항은 다음과 같다(個人情報の保護に関する法律, 第2條 第9項).

> "익명가공정보"= 특정 개인을 식별할 수 없도록 개인정보를 처리해서 생성될 수 있는 개인에 대한 정보. 특정 개인을 식별할 수 없도록 하는 처리 행위는 다음 두 가지와 같다.
> (i) 해당 개인정보에 포함된 기술(description)의 일부분을 삭제할 것(이 삭제에는 해당 기술의 일부분이 추출될 수 있게 하는 규칙성이 없는 방법을 사용해서 다른 기술(description)로 대체하는 것을 포함)
> (ii) 해당 개인정보에 포함된 모든 개인 식별 코드(code)를 삭제할 것(이 삭제에는 해당 기술의 일부분이 추출될 수 있게 하는 규칙성이 없는 방법을 사용

2) 법령 이외의 형태

일본 가이드라인은 위와 같이 일본 개인정보보호에서의 익명가공정보 개념정의 조항에 대한 해설을 하고 있다. 익명가공정보의 개념 표지로 포함된 "특정 개인을 식별할 수 없도록 하는"이란 표현은 특정 개인이 모든 가능성 있는 수단의 사용을 전제로 기술적으로(technically) 식별될 수 있는 가능성의 완벽한 제거를 의미하지 않는다. 대신 익명가공정보는 적어도 개인정보를 취급하는 사업자 또는 익명가공정보를 취급하는 사업자가 가용한 능력과 수단을 적용한 통상적인(ordinary) 방법을 적용해 식별할 수 없다면 충분하다.

(3) 비식별정보, 익명정보, 가명정보 개념의 도입 과정

1) 배경 및 구체적 절차

개정 개인정보보호법의 입법 의도는 법 제1조에서 확인할 수 있다. 이 조항은 일본 개인정보보호법의 목적을 포괄적으로 서술하고 있는데, 이번 개정으로 기존에 나열된 목적들 이외에 "개인정보의 적정하고 효과적인 활용이 새로운 산업의 창출 및 활력 있는 경제사회와 풍요로운 국민생활의 실현에 이바지하는 것"이란 표현이 추가되었다(個人情報の保護に関する法律, 第1條). 개정 개인정보보호법의 가장 큰 변화 하나는 익명가공정보의 도입이므로, 익명가공정보의 도입 취지가 위 표현과 같다고 간주할 수도 있을 것이다. 즉, 익명가공정보는 개인정보의 효과적인 활용과 그에 따른 산업적 가치의 창출을 염두에 두고 도입된 개념이다.

이렇게 도입된 익명가공정보가 실질적인 활용으로 이어지기 위해서 일본 개인정보보호법이 개정된 이후 익명가공정보의 기준과 방법에 대한 구체적인 사항들을 제시하는 시행령과 가이드라인, 더 나아가 개인정보보호위원회 차원의 공식적인 보고서가 순차적으로 발표되었다. 이와 같은 체계적인 제도화 작업을 통해 익명가공정보의 안전하면서 효과적인 활용을 위한 제도적인 틀은 어느 정도 갖추어 놓은 상태이다.

2) 논의 사항

일본 보고서는 개정 개인정보보호법이 익명가공정보를 도입하기 이전의 과정을 간략하게 설명하고 있다. 이 과정의 확인을 통해 익명가공정보의 도입에서 기본적으로 논의되었던 사항을 추측할 수 있다. 2013년에 '세계 최고의 선진 IT 국가 창설에 대한 선언(Declaration on the Creation of the World's Most Advanced IT Nation)'이 발표되었다. 이 선언에는 데이터 활용을 위한 환경을 조성하고 공개 데이터와 빅데이터의 이용을 증진하기 위해 새로운 조직이 즉시 창설된다는 내용이

있다. 더 나아가 이 조직은 데이터 활용에 대한 규칙들을 개발함으로써 개인정보의 보호의 확보와 함께 데이터 유용성을 증진하는 것을 목적으로 한다는 내용도 포함되어 있다. 이 선언의 내용에 따라 '개인데이터에 대한 연구모임(Study Group of Personal Data)이 신설되었고, 이 모임은 익명화에 대한 연구를 수행했다.

이 모임에서 작성한 연구보고서가 제출되었고 상위 기관은 이 보고서를 바탕으로 개인데이터의 유용성에 대한 시스템 개혁에 대한 전체적인 체계(The Outline of the System Reform Concerning Personal Data Utilization)를 작성했다. 이 문헌은 그 당시의 일본 개인정보보호법(개정 이전)이 개인정보 개념이 해석이 어려운 '중간 영역(grey area)'을 만들어낸다는 점을 지적하면서 이 중간 영역의 문제를 해결할 필요성을 주장하였다. 이러한 중간 영역의 문제가 실제 기업의 데이터 활용의 의지를 약화시킨다는 문제의식을 드러내면서, 이런 문제점을 극복하기 위해 익명가공정보란 새로운 개념을 제안하고 있다. 이 문헌의 내용에 따라 2015년에 일본 개인정보보호법이 개정되었고, 이 개정법에 익명가공정보가 포함된 것이다. 결국 익명가공정보 도입과정에서 핵심적으로 논의되었던 사항은 데이터를 활용하려는 수범자의 입장에서 규제의 불명확성을 해소하는 제도적 방법이 무엇인지에 대한 사항이었다고 볼 수 있다. 비식별정보 또는 익명정보의 기본 개념 규정으로부터 비식별 조치 또는 익명화의 방법을 제시한 것이 아니라 익명가공정보란 새로운 개념을 제시함으로서 법적인 불명확성을 해소하려는 의도가 있었던 것이다.

(4) 비식별정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

데이터의 활용성을 높이기 위하여 익명가공정보 개념이 도입되었고 법률에서부터 가이드라인, 보고서에 이르기까지 다양한 문헌들이 지속적으로 발표된 상황에 비추어볼 때, 익명가공정보의 활용을 적극적으로

장려하고 그에 따라 데이터 산업의 국제적 경쟁력을 확보하려는 일본 정부의 의지를 확인할 수 있다. 특히 일본 보고서에는 익명가공정보의 실제 사용의 예시를 다음의 표와 같이 제시하고 있다.

[표4-15] 익명가공정보의 실제 사용의 예시

사례	내용
탑승 내역의 사례	-철도회사가 보유하는 탑승내역의 정보의 익명가공처리 후 일반사업자에게 제공 -일반사업자는 철도 이용자의 연령이나 성별과 같은 정보 또는 탑승내역에 근거해서 상권 분석 또는 맞춤형(targeting) 광고에 활용 예상
이동 내역의 사례	-자동차회사가 보유하는 이동내역의 정보의 익명가공 처리 후 소매업자와 같은 일반사업자에게 제공 -일반사업자는 자동차의 이동내역, 해당 소유자의 연령대 등의 속성에 근거해 점포에서의 상품라인업의 검토 또는 새로운 점포의 출점 계획에 활용 예상
전력이용 내역의 사례	-2020년대 초까지 모든 가구에 스마트미터 도입을 추진하는 정책을 배경으로 가정의 전력 사용량과 전력 사용 유형 등의 정보가 축적될 것으로 예상 -전력 관리 사업자가 보유하는 전력 이용량 정보의 익명가공처리 후 가전 제조업체 등의 일반사업자에게 제공 -일반사업자는 가족 구성원과 개별 집안의 사용 패턴 등의 생활 형태를 분석해서 기존 제품의 광고 전략이나 신상품 개발에 이용 예상

자료 : 연구진 작성(日本 個人情報保護委員会, (匿名加工情報-パーソナルデータの利活用促進と消費者の 信頼性確保の両立に向けて), pp.99-120(2017) 참조)

2. 중국

(1) 제도적 개괄

1) 개인정보보호 규제 일반

중국에는 개인정보의 보호만을 다루는 일반적인 법률은 없다. 그러나 보안 관점에서 2016년에 제정된 '중화인민공화국 사이버보안법(中华人民共和国网络安全法, Cybersecurity Law of the People' Republic of China, 이하 "사이버보안법")'이 개인정보의 보호에 대하여도 함께 규율하고 있다. 사이버보안법은 아래의 표와 같이 크게 6장으로 구성되어 있다.

[표4-16] 중국 사이버보안법 목차

제1장	일반 규정
제2장	사이버보안의 보완과 증진
제3장	네트워크 운영 보안
제4장	네트워크 정보 보안
제5장	감시, 초기 경고 그리고 비상 대응
제6장	법적 책임

자료 : 연구진 작성(中华人民共和国网络安全法 참조)

위 6장 중 제4장인 '네트워크 정보 보안'의 '개인정보와 주요 데이터 보호 시스템'에 개인정보에 대한 규정이 포함되어 있다. 가령 네트워크 상품과 서비스가 인터넷이용자의 정보를 수집하는 기능을 하는 경우 이 상품 또는 서비스 제공자는 그 이용자에게 명료하게 고지를 하고 동의를 얻어야 한다는 조항이 있다(中华人民共和国网络安全法 第40條).

2) 비식별 조치

사이버보안법은 비식별 조치에 대하여 구체적으로 규정하고 아니한다. 다만 개인정보의 공개에 관한 원칙을 규정한 조항에서 '식별되지 아니한'이라는 표현을 쓰고 있다(中华人民共和国网络安全法 第41條). 같은 조항에 따르면 사이버보안법의 수범자인 네트워크운영자는 정보주체의 개인정보가 식별되지 아니하도록 처리하지 아니하는 한, 그 정보주체의 사전동의 없이 해당 개인정보의 공개 또는 변경을 할 수 없다. 비식별이란 표현은 명시적으로 쓰고 있지 않지만 식별되지 않도록 처리되는 것을 비식별 조치라고 해석할 수도 있다. 사이버보안법상 비식별 조치된 데이터는 같은 법상 의무로부터 면제된다고 해석될 가능성이 있다.

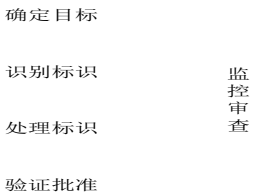
이를 전제로 중국 정부는 2017년 '정보보안기술-개인정보 비식별 조치 가이드라인(信息安全技术-揭秘个人信息指南, 이하 "중국가이드라인"이라 한다)' 초안을 발표하였다. 중국가이드라인은 개인정보 비식별 조치와 관련된 전 세계의 연구 이론들을 반영해서 비식별 조치를 적용하는 방법에 대한 지침을 제시하는 것을 목적으로 한다. 특히 중국가이드라인은 마이크로데이터(microdata)를 위한 구체적 개인정보 비식별 조치 지침의 기능을 한다. 개인정보를 처리하는 기업이나 네트워크 보안 및 관리 기능을 하는 정부의 관련 부처 등의 조직에서 실제로 개인정보 보안 감독 관리, 평가 등의 작업을 진행할 경우에 이 중국가이드라인이 적용된다.

중국가이드라인에 따르면 식별자를 삭제하거나 변환해서 공격자가 정보주체를 직접적으로 식별하거나 외부 정보와의 결합을 통해 간접적으로 식별하는 것을 방지하는 것이 비식별 조치의 목표이다. 데이터의 환경에 적합한 비식별 조치의 모형과 기술을 선택해서 이 재식별의 리스크(risk)를 수용 가능한 범위로 제어한다. 가령 공격자가 외부 정보들을 다양하게 연계함으로써 해당 데이터의 재식별 리스크가 증가되는 데이터 환경과 그렇지 않은 데이터 환경을 구분해서 각각의 데이터 환경에

적합한 비식별 조치의 모형을 설계한다.

중국가이드라인은 나아가 비식별 조치 시에 적용해야할 기본적인 원칙을 제시하고 있다. 우선적으로 수범자는 중국의 관련 법규의 개인정보 보안과 관련된 규정을 지속적으로 준수해야 한다. 또한 개인정보 보호의 중요성을 인지하고 개별 데이터 업무의 목표와 보안 및 보호와 관련된 사항들 모두를 고려해 개인정보에 대한 비식별 조치를 진행해야 한다. 그래서 비식별 조치를 통한 데이터의 활용 가치는 개인정보가 보호된다는 전제 하에 추구되어야 한다. 이런 기본 전제 하에 비식별 조치의 최적 모형은 기술적(technological) 측면과 관리적 측면을 결합하는 방향으로 설정된다. 이와 같이 비식별 조치의 모형이 설정되어서 비식별 조치가 완료된 이후에도 지속적으로 재식별 리스크에 대한 평가가 진행되어야 한다.

일반적으로 비식별 조치 과정은 목표확정, 식별, 식별 처리, 검증 승인의 단계로 나눌 수 있으며, 각 단계 실행 과정 중이나 완료 후 효과적인 모니터링 및 심사를 진행한다. 다음의 그림은 비식별 조치의 과정을 나타낸다.



미, 데이터 유형 등의 내용을 포함한다. 이외에 규칙 판정법은 자동화 프로그램의 적용을 통해 데이터셋 규칙을 분석하여 비식별 조치가 필요한 직접식별자와 준식별자를 자동으로 알아내는 방법이다. 규칙 판정법은 특정한 상황에서 테이블 조사 식별법으로 확인하지 못한 식별자를 알아내는 데 도움을 준다. 그래서 일반적으로 테이블 조사 식별법을 적용하고, 데이터의 양 및 복잡성 정도에 따라 규칙 판정법과 테이블 조사 식별법을 결합해서 사용한다.

식별자를 확인한 이후에는 세 번째 단계로 식별성 관련 처리를 한다. 식별 처리 단계는 전(前)처리, 비식별 조치의 적용 모형의 선택, 비식별 조치의 적용이란 세 가지 과정으로 구성된다. 전처리는 데이터셋에 비식별 조치를 본격적으로 적용하기 이전에 준비하는 예비 과정이다. 비식별 조치를 정식으로 적용하는 단계를 효율적으로 이행하기 위해 일반적으로 전처리의 단계 자체에 데이터셋에 변화를 주는 과정이 포함된다. 전처리에서는 데이터를 특정 형식에 맞추어 재구축하거나, 추출하는 데이터셋 규모를 축소하는 등의 처리를 하게 된다. 이 후 단계인 비식별 조치의 구체적인 적용 모형을 선택할 때에는 다양한 유형의 데이터에 적합한 비식별 조치의 기술을 선택해야 한다는 사실을 염두에 두어야 한다. 그러므로 데이터의 유형과 업무 특성을 파악한 후 이에 적합한 비식별 조치의 모형 및 그에 따른 구체적인 비식별 조치의 기법들을 선택하는 것이 중요하다. 구체적인 모형을 선택하는 경우에 재식별 리스크의 수량화 문제, 데이터 삭제의 가능성 문제, 원본 데이터로서의 형태의 유지 필요성, 비식별 조치가 적용된 데이터의 가역성 문제, 기존 데이터의 특성의 유지 필요성 등을 고려한다. 이러한 단계들을 거친 이후에 선택한 비식별 조치의 모형에 따라 데이터셋을 조작한다. 가령 비식별 조치가 필요한 식별자가 여럿일 경우 데이터 및 업무의 특징에 따라 적용 순서를 결정한다. 데이터와 데이터 환경의 특성에 따라 선택된 비식별 조치의 도구 혹은 프로그램을 이 적용 순서에 따라 실행해서 데이터에 대한 비식별 조치를 적용한다.

마지막 절차는 검증과 승인의 단계이다. 데이터셋을 비식별 조치한 뒤

생성된 데이터셋이 사전에 설정한 재식별 리스크의 한계 및 데이터 유용성 목표에 부합하는지에 대한 검증 절차를 거친다. 비식별 조치가 적용된 이후의 재식별 리스크를 평가해야 하며 실제 리스크를 계산해서 수용할 수 있는 최대치의 리스크 수치와 비교한다. 산정된 리스크가 최대 수치를 초과할 경우에는 이 수치 이하로 통제될 때까지 비식별 조치의 적용 과정을 조정해야 한다. 기술 발전에 따라 재식별의 공격 능력이 빠른 속도로 진보하고 있기 때문에 비식별 조치의 전문가가 재식별 리스크를 정기적으로 검증하고 평가할 필요가 있다. 재식별 리스크를 포함한 전체적인 보안 상황을 검증하는 방법들이 다수 존재한다. 예를 들어, 직접식별자와 준식별자가 포함되지 아니하도록 처리된 데이터 문서 또는 메타데이터를 직접적으로 검사하는 방법이 있다. 이외에 비식별 조치에 사용된 소프트웨어의 이론적 구조를 평가하거나, 침입자 테스트를 실시해서 데이터셋이 공개될 경우의 외부 데이터 공격자를 상정하는 방법들도 있다.

중국가이드라인은 위에서 설명한 비식별 조치의 전 과정에 걸쳐 매 단계마다 지속적으로 모니터링(monitoring)하는 것이 중요하다는 사실을 강조하고 있다. 비식별 조치의 개별 단계마다 설정된 목표를 달성할 수 있도록 세밀한 모니터링이 필요하다. 각각의 단계별 임무를 제대로 이행하기 위해 목표를 확정하는 단계에서부터 비식별 조치를 적용하는 전체적인 작업 계획을 작성한다. 비식별 조치의 단계별 작업을 명확히 적용하기 위해 식별에서 부터 검증의 단계까지의 작업 과정과 결과를 자세하게 기록해서 문서화해야 한다. 모니터링 업무를 관장하는 주체는 각각의 단계가 완료될 때마다 문서화된 해당 기록을 심사하고 누락이나 오류 사항이 없는지 점검한다. 이러한 모니터링 심사 과정 자체도 문서화함으로써 객관적인 심사가 되도록 운용한다. 문서화된 모니터링 심사 내용에는 심사의 대상, 시간, 과정, 결과 등의 내용이 포함된다. 엄격한 문서화를 기반으로 한 지속적인 모니터링을 통해 데이터의 비식별 조치가 적용된 이후의 상황 변화에 따라 재식별 리스크를 산정해서 이 결과값과 수용 가능한 리스크의 최대치를 비교하는 과정을 수시로 적용한

다. 더 나아가 이런 비교의 과정을 정기적으로 적용하는 방법도 고려할 수 있다. 특히 비식별 조치가 적용된 데이터에 예상되는 데이터 공격자가 접근할 수 있는 가능성이 높은 경우에는 데이터에 대한 재식별 리스크 평가를 정기적으로 진행해야 한다.

(2) 비식별정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

사이버보안법 자체에는 비식별정보, 익명정보, 가명정보에 대한 별도의 개념정의 등이 존재하지 아니한다. 이 법률은 전반적으로 정보보안의 측면에 규정되어 있고, 개인정보의 개념과 같은 정보보호의 기본적 내용들에 대한 규정은 여기에 존재하지 아니한다.

2) 법령 이외의 형태

반면 사이버보안법에 터 잡은 중국가이드라인은 비식별 조치의 개념을 규정하고 있다. 중국가이드라인에 따르면 비식별 조치는 개인정보에 대한 기술 처리를 통해서 추가 정보 없이는 정보주체를 식별할 수 없는 과정이다. 하지만 익명화에 대한 별도의 개념 규정은 하지 않고 있다. 가명화는 비식별 조치의 한 가지 기법으로서 제시되어 있다.

(3) 비식별정보, 익명정보, 가명정보 개념의 도입 과정

중국가이드라인은 서론에서 빅데이터, 클라우드 컴퓨팅, 사물인터넷 시대를 맞이하여 데이터 기반 부가가치창출이 활발하게 이루어지면서 개인정보 보안 문제가 크게 대두되고 있다는 사실을 제시한다. 이러한 사회적 배경 하에 보안의 위험성을 방어하면서 동시에 데이터 산업의

발전에 대한 사회적 요구에 부응하기 위한 장치로 비식별 조치의 개념과 구제척인 방법들을 제시하기 위해 중국가이드라인을 마련한 것이다.

(4) 비식별정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

중국은 2016년 사이버보안법 개정 이외에 개정 사이버보안법을 기반으로 한 다수의 정책 문서를 발표했다. 그 중 2018년 1월 발표한 '정보기술-개인정보 보안규범(信息安全技术 个人信息安全规范(GB/T 35273-2017), Information Technology - Personal Information Security Specification, 이하 "보안규범"이라 한다)이 중요하다. 이 보안규범은 2018. 5. 1. 시행되었다. 직접적인 법적 구속력을 가지는 법령이나 규칙은 아니지만, 정부기관들은 이 보안규범의 내용을 기준으로 실제 정책적 판단을 할 가능성이 높다고 보고 있다(https://www.chinalawblog.com/2018/02/chinas-personal-information-security-specification-get-ready-for-may-1.html). 보안규범은 사이버보안법의 내용들 중 개인정보의 보호 영역에 대한 구체적인 규제원칙들을 제시하고 있다. 다른 국가들의 일반적인 개인정보보호법의 구성과 마찬가지로 정보의 수집, 제공, 활용 시 준수해야 할 원칙들을 구체화한다. 물론 개인정보 관련 규제 내용을 다른 국가 수준으로 정비하고 있다는 점만으로 중국 정부가 비식별정보 또는 익명정보의 활용에 대하여 어떠한 정책을 추구할지 예상할 수는 없지만, 규제 정비가 신속하게 이루어지고 있는 점에 비추어보더라도 데이터의 공유와 보호를 위한 제도적 기반을 마련하고 있는 단계라고 할 수 있다.

3. 싱가포르

(1) 제도적 개괄

1) 개인정보보호 규제 일반

싱가포르법상 개인정보를 포괄적으로 규율하는 법률은 개인정보보호법(Personal Data Protection Act 2012, 이하 "PDPA"라 한다)이다. PDPA는 개인정보의 공개, 제공, 활용 등에 광범위하게 적용된다.

2) 비식별 조치

가. 법령

PDPA에는 비식별에 대한 구체적인 기준이나 방법이 포함되어 있지 않다. 단지 연구 목적의 활용인 경우에는 '개별적으로 식별가능한 형태(individually identifiable form)'로 되어 있는 데이터를 제공해야 할 합리적 필요성이 있어야 정보주체의 사전동의를 필요로 하지 않는다고 규정할 뿐이다[Personal Data Protection Act 2012, Sec.17(2) Art.1(i)]. 다만 이 규정을 통하여 PDPA는 비식별정보는 정보주체의 동의 없이 제공, 활용할 수 있을 것으로 해석할 수 있다.

이에 터 잡아 싱가포르의 개인정보보호기관인 개인데이터보호위원회(Personal Data Protection Commission, 이하 "PDPC"라 한다)는 2013년 이 법률의 일부 내용에 대한 가이드라인을 발표했다('Proposed Advisory Guidelines on the Personal Data Protection Act for Selected Topics', 이하 "PDPC 가이드라인"이라 한다). 이 가이드라인은 비식별 조치 이외에도 여러 측면을 다루고 있다. 이하에서는 이 가이드라인의 내용 중 비식별 조치에 대한 부분에 집중하여 간략하게 살

펴보기로 한다. 이외에 PDPC가 2018년 발표한 익명화에 대한 안내서(Guide to Basic Data Anonymisation Techniques, 이하 "PDPC 익명화가이드"라고 한다)도 있다.

나. 법령 이외의 형태

a. PDPC 가이드라인

PDPC가이드라인은 권고적 성격을 가질 뿐이고 법적 구속력은 없다는 점을 명확하게 밝히고 있다. 가이드라인의 내용이 PDPA의 내용이나 법적인 해석을 수정하거나 보충하는 기능을 하지 않는다.

PDPA는 정보주체의 사전동의를 개인정보 활용의 기본원칙으로 설정하고, 사전동의 적용의 예외들 중 한 가지 방법으로 연구 목적의 활용을 제시하고 있다[Personal Data Protection Act 2012, Sec.17(2) Art.1(i)]. 가이드라인은 정보주체의 사전동의 없이 연구목적의 활용을 가능하게 하는 방법으로 위 조항의 적용 이외에도 데이터의 익명화가 있다고 한다. 익명화된 데이터는 개인정보가 아니기 때문에 더 이상 PDPA의 적용대상이 되지 않는다고 한다.

가이드라인에 따르면 익명화란 식별을 가능하게 하는 정보를 제거함으로써 잔존 데이터로는 어떤 특정 개인을 식별할 수 없게 하는 절차를 의미한다. 그러므로 익명화는 개인정보를 개인을 식별하기 위해 활용될 수 없는 데이터로의 변환을 가리킨다. 따라서 익명처리된 정보, 즉 익명정보는 PDPA의 적용대상인 개인정보에 해당하지 않는다. 가이드라인은 익명화에 대하여 이처럼 데이터의 이용 가능성을 창출하는 하나의 방법으로써의 측면 이외에 보안 침해 및 예상치 못한 정보공개를 예방할 수 있는 개인정보의 보호 방법으로서의 의미도 있다고 한다.

가이드라인은 익명화의 기법들에 대한 설명을 포함한다. 익명화의 기법들로 널리 알려진 총계화(aggregation), 대체(replacement), 데이터 삭제(data reduction), 데이터억제(data suppression), 마스킹

(masking) 등을 나열하면서, 이러한 기법을 적용한 이후에도 여전히 재식별의 가능성이 있다고 밝히고 있다. 예컨대, 해당 기관이 접근할 수 있는 다른 정보와 해당 데이터를 결합하던지, 적정한 알고리즘을 활용해서 익명화 과정을 역추적하여 원래의 식별정보를 알아낼 수 있는 경우가 있고, 그러한 가능성이 있는 한 여전히 개인정보에 해당한다고 한다. 이와 같은 재식별의 위험성은 기업들이 인식하지 못하거나 통제 범위 밖에 있는 상황에 있을 수 있기 때문에 기업들은 데이터의 활용에 심리적 제한을 가지게 된다. 위에서 언급한 비식별 조치 기법들의 단순 적용으로는 개인정보가 익명정보로 변환된다는 제도적 보장이 없다면서, 재식별의 위험성 문제의 처리를 어떻게 할 것인지를 논의하고 있다.

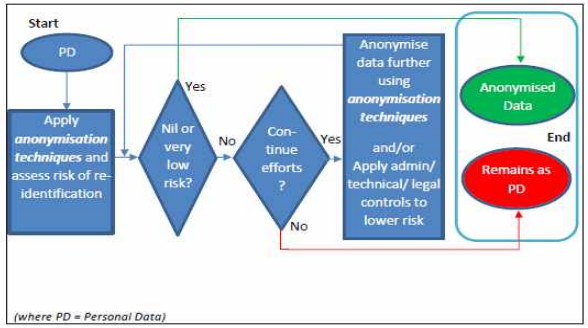
이 가이드라인은 재식별의 위험성 문제를 재식별의 위험성(risk)에 대한 관리(management)의 측면에서 해결하려는 시도를 하고 있다. 재식별의 위험성에 대한 적절한 관리를 통해 익명화된 데이터가 다시 개인정보로 될 가능성을 낮출 수 있다는 것이다. 결국 재식별의 위험성을 가능하게 하는 다른 정보의 가용성에 대한 평가를 효과적으로 적용해야 하는데, 현실적으로는 이런 작업이 매우 어렵다는 한계가 있다. 특정 시점에서는 익명화된 데이터라고 하더라도 시간이 지남에 따라 다른 관련 요인들의 변동에 따라 재식별의 위험성이 달라지기 때문이다. 그래서 데이터의 변형 정도, 다른 정보의 가용성, 그리고 재식별의 유인(motivation) 등의 관련 요소들에 따라 다양한 재식별의 위험성을 가지게 됨을 인식해야 한다.

이런 재식별의 위험성을 인식하면서 이 가이드라인은 익명화에 대해 실용적인 접근법을 적용하고 있다고 밝히고 있다. 그래서 재식별의 위험성이 높으면 개인정보이지만, 재식별의 위험성이 사소한(trivial) 수준이면 익명화된 정보로 본다고 결론을 내리고 있다. 즉, 재식별의 위험성이 낮은 수준으로만 관리가 되면 익명화 방법을 통해 데이터를 활용할 수 있다는 것이다. 이런 실용적 접근법 하에서 재식별의 위험성 판단 기준으로 실제 재식별을 하려는 잠재적 공격자의 기준이 중요하다고 판단해서, 영국의 ICO에서 발표한 익명화 가이드라인의 '유인된 공격자

기준(motivated intruder test)'을 적용할 수 있는 방법으로 제시하고 있다. 그래서 이 기준에 따라 인터넷과 공개된 정보와 같은 자원들에 접근할 수 있는 합리적 수준의 능력자가 이 공격자 기준에 해당한다고 설명한다.

b. PDPC 익명화 가이드

가이드는 익명화의 기술적(technological) 측면에 대하여 일반적으로 소개하는 것을 목적으로 한다. 위에서 설명한 PDPC 가이드라인의 익명화 내용을 기반으로 좀 더 구체적으로 익명화 기법들을 설명한 것이다. PDPC가이드라인에서 제시한 내용과 마찬가지로, 가이드에서도 익명화된 데이터란 재식별의 위험성 평가와 함께 익명화 기법들을 활용해서 변형 과정을 거친 데이터를 의미한다고 한다. 즉, 재식별의 기법적 측면과 재식별의 위험성 관리 측면 모두를 전체 익명화 절차의 구성 요소로 보는 것이다. 이 가이드는 익명화의 전반적인 과정을 다음의 그림으로 표시하고 있다.



(그림4-16) 익명화의 전반적인 과정
출처: Singapore Personal Data Protection Commission, Guide to Basic Data Anonymisation Techniques, p.4 그림(2018) 인용

그리고 가이드에서 제시하는 기법의 적용만으로 PDPA의 적용을 받지 않는 데이터라고 언제나 확언할 수 없다는 사실을 강조한다. 이와 함께 이 가이드의 내용이 기업들에 도움이 될 수는 있지만, 모든 기업들에 일률적으로 적용되는 익명화 해결책은 없다는 사실도 확인하고 있다. 기업은 개별적 데이터의 특징과 데이터 환경에 적합한 익명화 접근방법을 적용해야 한다고 한다.

가이드의 전반부에는 익명화 방법의 개념적 배경과 전제를 설명하고 있다. 익명화 기법은 기본적으로 원래의 데이터셋으로부터 개인이 식별될 수 있는 가능성을 해당 기업의 위험성 배분구조(portfolio)에 따라 수용할 수 있는 수준으로 낮춘다. 위험성 평가를 위해서는 익명화 기법이 적용되기 전과 후 모두에 식별가능성에 대한 평가가 수행될 필요성이 있다. 적용 전 단계에서는 원래 데이터의 위험성을 평가하는 것이고, 적용 이후 단계에서는 익명화 기법이 적용된 데이터의 잔존 위험성을 평가한다. 이 두 가지 위험성의 비교를 통해 위험성을 어느 수준까지 낮추어야하는지 판단하게 된다.

이 가이드는 다양한 익명화 기법들을 자세히 설명하면서 해당 기법들이 활용되는 경우와 적용 방법, 그리고 구체적인 예를 상세히 제시한다. PDPC 가이드라인과 달리 기업들에 좀 더 실질적인 도움이 되도록 개별 기법들의 상세한 특징을 설명한다. 익명화 기법들에 대한 설명 이외에도 익명화가 적용된 이후의 사후적인 관리 방법에 대해서도 설명하고 있다. 이러한 사후적 통제의 적정성 여부는 해당 데이터셋에 대한 접근 방침에 따라 달라진다. 사후적 통제 방식의 예로 데이터에 대한 접근 권한을 철회할 수 있는 방법이 있다. 이외에 데이터 자체에 대한 직접적 접근은 배제하고 쿼리(query)만을 허용하는 방법도 있다. 또한, 별도의 인증(authentication) 절차를 적용하거나, 물리적으로 지정된 공간 내에서만 데이터에 대한 접근을 허용하는 방법도 있다. 사후적 통제뿐만 아니라 데이터 자체에 대한 전반적인 관리(governance) 측면에서도 익명화된 데이터의 이동경로를 추적하거나 가명처리에 활용된 키(key)

와 이 키와 원본 식별자의 대응 관계를 알려주는 표(mapping table)를 분리해서 보관하는 등의 관리 정책들도 제시하고 있다.

(2) 비식별정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

PDPA는 비식별정보, 익명정보, 가명정보의 개념정의 규정을 포함하지 아니할 뿐 아니라, 나아가 PDPA 내에서는 이러한 용어가 사용되지도 아니한다.

2) 법령 이외의 형태

PDPC 익명화 가이드는 익명화의 개념을 정의하고 있다. 즉, 익명화는 다양한 익명화 기법을 적용해서 개인정보를 익명정보로 변환하는 처리를 의미한다고 한다. 하지만 비식별 조치라는 표현은 쓰지 아니한다.

PDPC 가이드라인과 PDPC 익명화 가이드 모두 가명화에 대한 개념정의를 한다. PDPC가이드라인은 가명화를 ”식별자를 다른 참조표시로 대체하는 것“으로 정의하면서, 이름을 참조번호(reference number)로 대체하는 것이 가명화의 예라고 설명한다. PDPC 익명화 가이드는 가명화를 ”식별자를 관련성이 없으면서도 특정성을 갖춘 값으로 대체하는 기법“이라고 정의한다. 가명화에 대한 기본적인 정의는 같지만, PDPC 익명화 가이드가 원래의 개인정보를 대체하는 정보의 특징을 보다 구체적으로 설명하고 있다.

(3) 비식별정보, 익명정보, 가명정보 개념의 도입 과정

1) 배경 및 구체적 절차

PDPC는 익명화의 적용을 통해 생성된 데이터는 더 이상 개인정보가 아니므로 해당 데이터는 연구와 데이터마이닝(data mining)에 활용될 수 있는 적합한 형태가 될 수 있음에 익명화의 제도적 가치가 있다고 설명한다(https://www.pdpc.gov.sg/-/media/Files/PDPC/New_DPO_Connect/nov_15/pdf/Anonymisation.pdf). 데이터분석은 일상생활에 커다란 가치를 제공할 수 있는데, 교통 상황의 개선부터 공공안전의 보장에까지 그 효용성이 높다. 특히 특정 데이터가 아니라 다량의 데이터의 총계처리(aggregating)가 더 적합한 경우 익명화된 데이터의 가치가 높아진다. 데이터 분석의 사회적 가치와 관련하여 익명화된 데이터의 유용성이 높다는 인식이 PDPC가 PDPC 가이드라인이나 PDPC 익명화가이드를 발표한 배경이라고 볼 수 있다.

2) 논의 사항

PDPC는 익명화의 사회적 가치를 인지하고 있지만 동시에 익명화의 도전과 한계점이 있다는 점도 간과하지 않고 있다. 이런 인식들이 PDPC 가이드라인이나 PDPC 익명화가이드를 실제 작성할 때의 주요 고려였을 것으로 예상할 수 있다. PDPC는 익명성과 데이터 완결성(integrity) 사이의 상충관계에 주목한다. 너무나 강한 익명화를 적용하면 데이터의 완결성이 약화되어서 결국 데이터의 효용가치가 상실되는 문제가 발생할 수 있다. 개별적인 데이터 활용 목적에 따라 적합한 익명화 방법을 적용하는 접근법이 위 두 문헌에서 제시된 까닭이다. 또 다른 문제는 재식별의 문제이다. 재식별의 위험성을 관리하기 위해 데이터의 맥락에 따른 종합적인 관리법이 중요하다는 점이 위 두 문헌에

서 반영되어 있다.

(4) 비식별정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

PDPC 익명화가이드가 발표된 이후 2017년 이 가이드에 대해 PDPC가 싱가포르의 한 언론사와의 인터뷰가 있다. 그 내용에서 익명정보에 대한 싱가포르 정부의 정책적 추진 방향을 대략적으로 읽을 수 있다. 이 인터뷰에서 PDPC는 기업이 자신들이 수집한 데이터가 기업 경쟁력 확보를 위한 전략적 자산이라는 사실을 인지했다고 밝히고 있다(https://www.straitstimes.com/singapore/new-guide-to-clear-doubts-on-sharing-customer-data). 기업들이 PDPC 익명화이드에 따라 보유 데이터를 활용하는 것을 활성화하려는 강력한 정책적 의지가 있음을 밝히고 있는 것이다. 일부 기업은 PDPC 익명화가이드가 유용하다는 의견을 내놓으면서, 카드 사용 내역에서의 데이터를 익명화해서 통근자(commuter)들에게 혜택을 줄 것이라고 설명하고 있다.

제 4 절 호주

(1) 제도적 개괄

1) 개인정보보호 규제 일반

호주는 프라이버시를 포괄적으로 규제하는 프라이버시법(Privacy Act 1988)을 두고 있다. 이 프라이버시법은 프라이버시 원칙들을 제시하고 있는데, 이 원칙들이 개인정보의 활용, 공개 및 수집을 규율한다. 특히 이 프라이버시 원칙들 중 6번째 원칙은 개인정보의 활용과 공개를 제한하는 내용을 포함하고 있다.

2) 비식별 조치

가. 법령

프라이버시법은 비식별의 개념 정의를 하고 있다. 프라이버시법에서 규정하는 비식별(de-identified)이란 '특정 정보가 더 이상 식별가능한 개인 또는 합리적으로 식별가능한 개인'에 대한 정보가 아닌 경우이다[Privacy Act 1988, Art.6(1)]. 합리적 식별가능성을 개인정보의 판단 기준으로 설정함과 동시에, 합리적 식별가능성을 배제하도록 비식별처리된 데이터는 프라이버시법의 적용대상이 아님을 밝힌 것이다. 이 이외에도 프라이버서법에서 규정한 프라이버시 원칙들을 살펴보면, 더 이상 필요 없는 정보는 개인정보의 보안 강화를 위해 파기하거나 비식별 조치를 적용해야 한다는 등[Privacy Act 1988, Art.11.2(d)] 비식별 조치를 전제하여야 이해될 수 있는 규정들이 있다. 하지만 비식별 조치의 구체적인 기준이나 방법에 대한 규정은 없다.

호주의 개인정보보호기관(Office of the Australian Information

Commissioner, 이하 "OAIC"라 한다)은 이에 터 잡아 비식별 조치에 대한 두 가지 안내서를 발표했다. 하나는 2018년에 발표한 '비식별 조치와 프라이버시법(De-identification and the Privacy Act, 이하 "OAIC2018"이라 한다)이다. 이는 비식별 조치에 대한 대략적인 방법론과 규제적 의미를 제시하고 있다. 다른 하나는 2016년 발표한 '비식별 조치 의사결정 체제(De-identification Decision-Making Framework, 이하 "OAIC2016"이라 한다)이다. OAIC2016은 UKAN의 익명화 보고서를 거의 그대로 반영한 비식별 조치에 대한 포괄적이고도 상세한 안내서이다.

나. 법령 이외의 형태

a. OAIC2018

OAIC2018은 비식별 조치가 프라이버시를 강화하기 위한 도구라는 점을 명시한다. 비식별 조치가 적절하게 적용되면 프라이버시법에서 나열된 법적 의무들의 준수에 도움이 될 수 있다. 어떤 데이터가 개인정보인지 아니면 비식별정보인지 여부는 맥락, 즉 특정 데이터 환경에 따라 달라진다. 특정 데이터 환경하에서 개인의 재식별 위험이 '매우 낮은(very low)' 경우에는 비식별 조치된 정보라 한다.

비식별 조치의 구체적인 과정은 두 단계로 구성된다. 우선 직접식별자를 제거하고 다음으로 추가 조치를 한다. 이러 추가 조치에는 재식별의 잠재적 위험성을 높게 하는 다른 정보를 제거하거나 다른 기호로 대체하는 방법과 재식별의 예방을 위해 데이터에 접근하는 환경을 안전하게 관리하는 방법들이 포함된다. 정리하면 이 안내서는 재식별의 위험성을 맥락적으로 평가해야 한다는 기본 전제 하에 이 두 가지 요소들은 어떤 비식별 조치의 기법이 데이터 식별에 필요한지도 살펴본다. 이런 맥락적 평가 이외에 비식별 조치는 위험성 관리의 실제이다.

위에서 언급한 '합리적 식별가능성'을 고려할 때는 여러 가지 요소들

을 고려한다. 정보의 특징과 양, 정보를 보유하고 접근할 수 있는 주체, 가용한 다른 정보, 그리고 해당 정보를 적용해서 개인을 식별하려는 방법의 실용성이 있다. 그래서 재식별의 발생이 기술적으로 가능하고 재식별이 발행할 합리적인 개연성이 있다면 이 식별가능성의 기준에 부합한다는 결론에 이르게 된다.

b. OAIC2016

OAIC2016은 2017년 9월 발표되었다. 앞서 언급한 것처럼 OAIC2016은 UKAN보고서의 영향을 많이 받았다. OAIC2016을 직접 작성한 저자 중의 일부가 당초 UKAN보고서를 작성한 저자이고, OAIC2016의 서문(recital)도 그것이 UKAN보고서를 호주의 맥락에 맞추어 재구성하여 작성된 것임을 명확히 밝히고 있다. 그러나 UKAN보고서는 형식적으로 정부 기관에서 발표한 자료가 아니라 UKAN이라는 일종의 민간조직에서 만든 것인 데 비하여 OAIC2016은 호주의 개인정보보호 기관이 발표한 정부문건이라는 차이가 있다. 그 이외에도 UKAN보고서의 제목은 '익명화'란 표현을 사용한 반면, OAIC2016의 제목은 '비식별 조치'란 표현을 사용하는 차이점도 있다. 특히 서문(recital)은 이와 같은 용어의 변화가 UKAN보고서와 OAIC2016의 주요한 차이라고 명백하게 설명하고 있다.

제시하고 있는 전반적 비식별 조치 원칙이나 이를 적용하고 이해하는 관점 그리고 설명하는 방식들 모두 UKAN보고서와 거의 일치한다. OAIC2016 는 UKAN보고서와 마찬가지로 데이터와 데이터 상황 모두를 확인해야 하는 접근법을 선택하고 있고, 재식별의 위험성이 0 (zero)이 되어야 한다는 원칙은 비현실적이며, 비용편익분석의 관점에서 위험성 관리를 위해 투자한 수단들과 재식별의 위험성은 비례적 관계에 있다는 명제를 논의의 전제로 하고 있다. 이와 함께 구체적으로 제시된 비식별 조치 또한 UKAN보고서의 10단계의 익명화 체제를 그대로 인용했다. 익명화 방식의 적용은 늘 데이터 환경이라는 맥락에 의존하며, 맥락에 따라 위험성 관리를 수행한다는 맥락적 접근법도 그대로 원용했다.

호주의 개인정보보호법에 대한 설명을 제외한 대부분의 내용들이 UKAN보고서와 OAIC2016에서 일치하고 있지만, OAIC2016의 경우에는 보안적 관점을 기반으로 형성된 '다섯 가지 안전성(Five Safes)' 모형에 대한 간략한 설명을 하고 있다는 차이점이 있다. 2003년에 개발된 모형으로서 데이터의 접근에 대한 복잡한 논의들을 다섯 가지의 질의 형태로 단순화하는 형태이다. 이 다섯 가지 질의는 다음의 표에서 확인할 수 있다. 이는 미국의 NIST2016에 제시된 것과 마찬가지의 것이다.

[표4-17] OAIC 2016의 다섯 가지 안전성 모형

1. 안전한 프로젝트 측면	데이터의 활용이 적절한가?
2. 안전한 사람 측면	연구자가 적절한 방식으로 데이터 활용을 할 것을 신뢰할 수 있는가?
3. 안전한 데이터 측면	데이터 그 자체에 공개의 위험성이 있는가?
4. 안전한 환경(settings) 측면	데이터 접근이 이루어지는 환경이 허용되지 않는 이용을 적절한 수준에서 통제하는가?
5. 안전한 결과물 측면	통계적 결과물 자체에 외부공개적 요소가 있는가?

자료 : 연구진 작성(Office of the Australian Information Commissioner, De-identification Decision-Making Framework, p.20(2016) 참조)

위 다섯 가지 항목들은 일종의 체크리스트(check list) 기능을 한다. 기업 입장에서는 본인들이 보유하고 활용하려는 데이터가 안전한 상태에 있는지를 다섯 가지 항목만 확인하면 되는 구조이다. 그래서 다섯 가지 안전성 모형을 적용하면 데이터에 접근해서 활용하게 되는 상황이 발생할 때마다 일일이 생각해야 하는 요소들을 미리 유형화해서 데이터의 활용이 가능한지에 대한 여부를 신속하면서 일관적으로 결정할 수

있게 된다.

(2) 비식별정보, 익명정보, 가명정보 개념의 도입 현황 및 법적 효력

1) 법령

앞서 설명한 바와 같이 프라이버시법은 비식별 조치의 개념을 정의하고 있다. 하지만 익명정보와 가명정보는 별도로 다루지 않는다. 단지 프라이버시법은 '호주 프라이버시 원칙(Australian Privacy Principle)'의 하나로서 '익명성(anonymity)과 가명성(pseudonymity)'을 설명하고 있다(Privacy Act 1988, Schedule 1, Art.2). 이 원칙에 따르면 개인은 자신을 식별하지 않거나 가명을 사용할 선택권을 기본적으로 가지고 있다. 이 원칙은 프라이버시의 측면에서 익명과 가명이란 용어를 사용한 것으로 보인다. 개인정보를 보관하는 제3자의 입장이 아니라 정보주체가 스스로 자신의 프라이버시의 권리를 보호하는 수단으로 스스로의 정체성을 드러내지 않아도 된다는 원칙을 밝힌 것이다.

2) 법령 이외의 형태

OAIC2018은 비식별정보의 개념정의를 하고 있다. 개인정보는 합리적으로 식별가능한 개인에 대한 정보이다. 그래서 OAIC에 의하면 비식별정보는 비식별 조치의 과정을 거쳐서 더 이상 위 개인정보의 개념에 해당하지 않는 정보이다. 익명정보와 가명정보에 대한 별도의 개념정의는 하고 있지 않다. 하지만 호주 내에서도 비식별 조치와 유사한 개념을 익명화과 기밀화(confidentialisation)와 같은 다양한 용어로 사용하고 있다고 OAIC2018이 밝히고 있다.

OAIC2016은 OAIC2018과 마찬가지로 익명화와 가명화에 대한 개념정의를 제시하고 있지 않지만, 비식별 조치의 개념정의는 포함하고 있

다. OAIC2016에 따르면 비식별 조치란 데이터셋에서 직접식별자를 제거하거나 대체하는 절차를 의미한다. 비식별 조치를 적용할 때는 추가적인 기술적 통제 조치가 수반되어야 하며 이런 조치의 적용을 통해 개인이 식별되지 않도록 데이터를 제거하거나, 총계처리하거나(aggregate), 대체한다.

(3) 비식별정보, 익명정보, 가명정보 개념의 도입 과정

1) 배경 및 구체적 절차

OAIC가 비식별 조치에 대한 개념과 구체적인 방법들을 도입한 배경은 OAIC2016의 서문에서 확인할 수 있다. OAIC2016에 따르면 데이터를 분석하고 활용할 수 있는 능력을 갖춘 주체가 데이터에 접근해서 이를 공유할 수 있을 때 비로소 데이터의 가치가 실현될 수 있다. 그러므로 프라이버시의 규제 기준과 데이터의 가치에 대한 사회적 기대 모두를 실현할 수 있게 하는 데이터 공유방법을 찾는 것이 중요한 사회적 과제가 된다. OAIC는 비식별 조치가 잠재적인 해결책이 될 수 있다고 강조함으로써 비식별 조치의 개념과 방법이 도입된 배경을 설명하고 있다.

비식별 조치가 제대로 적용되면, 좀 더 많은 사람이 개인의 프라이버시를 보호하면서 데이터를 공유할 수 있게 된다. 비식별 조치가 공공영역 또는 민간 영역을 불문하고 전세계의 모든 기업이나 기관의 많은 관심을 불러일으키는 이유이다. 하지만 최근 10년 사이에 비식별 조치에 대한 문헌들이 많이 발표되었음에도 불구하고 실제 비식별화의 사례들을 살펴보면 일관적이지 못한 방식으로 적용이 되는 문제점이 발생했다고 한다. 이런 문제 때문에 잘못된 비식별 조치를 적용한 데이터들이 공개될 경우에 재식별의 위험성이 발현되면서 심각한 프라이버시 침해의 문제점을 드러내왔다. OAIC2016은 비식별 조치의 일관적 활용이 가

능하도록 그 표준을 제시할 목적으로 작성되었다. OAIC는 오랜 기간 동안 적절한 안전장치가 갖추어진 상태로 데이터의 활용성을 높이는 접근법의 중요성을 인지해왔다.

2) 논의 사항

OAIC2016의 서문은 비식별 조치의 개념의 기본전제는 재식별의 위험성은 맥락적으로(contextually) 평가되어야 한다는 사실이라고 강조하고 있다. '맥락적' 접근법은 OAIC2016에서 비식별 조치의 전반적인 개념과 원칙을 설정할 때에도 핵심적인 논의사항이었던 것이다. 맥락적 접근법과 같은 의미로 OAIC2016은 기능적(functional) 비식별 조치란 개념을 든다. 기능적 비식별조치란 데이터를 보유하는 주체가 데이터 그 자체만이 아니라 데이터가 공개되는 환경도 고려해야 한하는 것을 의미하는 방법이다. 즉, 데이터가 공개되는 환경, 즉 데이터 환경의 중요성을 강조한다. 이 두 요소 모두에 대한 고려를 통해서만 안전하면서 유용한 데이터를 생산할 수 있는 기술과 관리 방법을 선정할 수 있다고 한다.

(4) 비식별정보 또는 익명정보의 활용 현황 및 정책적 추진 방향

OAIC는 비식별 조치를 통하여 데이터의 안전한 공유를 활성화하려는 강력한 의지를 보이고 있다. 2016년 OAIC2016이 발표된 이후 2018년에 Q&A 형식의 실질적인 안내서 형태의 OAIC2018을 공개한 사실도 이를 방증한다. OAIC2016의 발표 이후 현재까지 비식별 조치를 적용한 데이터 공유 사례가 얼마나 증가했는지는 파악하기 어렵다. 하지만 2017년 언론을 통해 공개된 재식별 사례는 최소한 정부기관들이 비식별 조치를 적용한 데이터의 공개를 실행하고 있음을 보여준다(https://www.smh.com.au/technology/australians-health-records-unwittingly-exposed-20171218-p4yxt2.html).

위 사례에서는 호주 보건복지부(Department of Health)가 관리하는 호주 국민 중 약 10%의 건강정보가 재식별의 위험성에 노출되었다고 한다. 이 건강정보에는 개인의 HIV 병력 여부, 임신 중절 여부 등과 같은 민감한 정보까지 잠재적으로 드러날 수 있는 데이터가 포함되어 있었다. 해당 건강정보는 비식별 조치를 적용해서 대중에 공개된 데이터인데, 호주 멜버른(Melbourne) 대학의 연구팀이 출생연도와 수술사실과 같은 공개된 외부 데이터와 비식별 조치가 적용된 해당 건강데이터 사이의 연결(linking) 가능성을 고려하여 위에서 언급한 바와 같은 재식별 위험성이 있음을 밝힌 것이다.

이외에도 호주정부가 운영하는 "My Health Record"라는 프로그램이 있다. 이 프로그램은 환자에게 자신의 보건의료정보에 대한 통제권을 부여함으로써, 환자가 자기 보건의료정보를 누구에게 그리고 어느 범위까지 공유를 허락할지를 결정할 수 있게 하는 것이다(https://www.myhealthrecord.gov.au/for-you-your-family/secondary-uses-data). 이러한 공유는 1차적으로는 환자 본인의 효과적인 치료를 위함이기도 하나 동시에 연구 목적으로의 이차적 활용도 배제하지 아니한다. 특히 연구 목적 이차적 활용의 경우에는 대체적으로 비식별 조치가 적용된 비식별정보 형태로 공유가 이루어진다는 점에서, 이 프로그램은 그 자체 비식별정보의 활용성을 높이려는 호주정부의 의지를 보여주는 사례이다.

제 5 장 비식별 조치에 대한 국내 법제도 분석 및 정책적 제언

제 1 절 국내 법제도에 대한 분석

1. 현황

(1) 법적 기초

현행법상 비식별 조치를 정면에서 규율하는 규정은 없다. 그러나 앞서 살펴본 바와 같이(제2장) 개인정보 보호법 제2조 제1호는 “개인정보”를 “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”고 규정하고 있고, 다른 개인정보 관련 법령도 대체로 이러한 개념에 기초하고 있다. 그러므로 개인에 관한 정보라 하더라도 그 특정 개인, 즉 정보주체를 알아볼 수 없게 하면 더는 개인정보라고 할 수 없다는 해석이 가능하다. 개인정보 보호법은 개인프로세서가 개인정보를 처리하는 경우에 적용되므로 개인정보가 아니면 더는 적용되지 아니한다. 즉, 식별가능성을 제거하면 개인정보 보호법의 규제를 피할 수 있다. 2016년 정부부처 합동으로 발표한 「개인정보 비식별 조치 가이드라인」도 같은 전제하에 비식별 조치에 관하여 여러 제안을 하고 있고, 제2장에서 소개한 서울중앙지방법원 2017. 9. 11. 선고 2014가합508066, 538302 판결도 “개인정보는 해당 정보를 처리하는 자의 입장에서 특정 개인을 식별할 수 있는(identifiable) 정보이므로, 개인정보에 암호화 등 적절한 비식별화(de-identification) 조치를 취함으로써 특정

개인을 식별할 수 없는 상태에 이른다면 이는 식별성을 요건으로 하는 개인정보에 해당한다고 할 수 없고, 따라서 정보주체의 동의 없이 통계작성 등의 용도로 이용되거나 제3자에게 제공되더라도 개인정보 보호법을 위반한 것이라고 볼 수 없다”고 하였다.

다만, 지난 9월 20일 국회를 통과한 「지역특화발전특구에 대한 규제특례법 전부개정법률안(대안)」(이른바 규제프리존법)은 비식별화에 관한 명문 규정을 하나 신설하는 것으로 하고 있다.

[표5-1] 지역특화발전특구에 대한 규제특례법(대안)

지역특화발전특구에 대한 규제특례법
제118조(「정보통신망 이용촉진 및 정보보호 등에 관한 법률」에 관한 특례) ① 규제자유특구 내 혁신사업 또는 전략산업과 관련하여 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 제2조제3호에 따른 정보통신서비스 제공자 중 규제자유특구사업자에 대하여는 규제자유특구 내 설치된 사물인터넷 기반을 통하여 수집한 같은 법 제2조제6호에 따른 개인정보에 대하여 비식별화를 하는 경우에 같은 법 제24조 및 제24조의2를 적용하지 아니한다. ② 제1항에 따른 규제자유특구사업자는 비식별화 정보를 이용하는 과정에서 개인정보가 생성되는 경우 이를 지체없이 파기하거나 추가적인 비식별화 조치를 하여야 한다. ③ 제2항을 위반한 자에게는 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 제24조 및 제24조의2를 적용한다. ④ 제1항부터 제3항까지에서 규정한 사항 외에 규제자유특구사업자의 지정방법, 관리방법 및 기타 절차에 관한 사항과 비식별화의 수준 및 방법 등 필요한 사항은 과학기술정보통신부장관과 방송통신위원회가 협의하여 정하는 바에 따른다.

출처 : 국회 홈페이지

위 규정에 따르면 이른바 규제프리존 내에서 관련 사업을 하는 경우 규제프리존 내에 설치된 사물인터넷(IoT)을 통하여 수집한 개인정보에 대하여 비식별화를 하는 경우에는 일정한 규제를 피하는 취지로서, 그

과정에서 개인정보가 생성되는 경우 즉시 파기 또는 비식별화는 것을 전제로 일체의 규제를 피할 수 있게 하고 있다. 그리고 그와 관련하여 비식별화의 수준 및 방법 등 필요한 사항을 과학기술정보통신부장관과 방송통신위원회가 협의하여 정하는 바에 따르도록 위임하고 있다. 법률은 공포 후 6개월 뒤에 시행되므로 좀 더 기다려보아야 하나, 위 법률이 시행되는 경우 최초로 「비식별화」라는 용어를 명시적으로 채택하고, 그 구체적인 수준과 방법 등에 관하여 법적 구속력이 있는 하위법령을 갖게 될 것으로 전망된다. 규제프리존법의 적용범위는 규제프리존 내 IoT에서 수집하는 정보로 제한되어 있지만, 위와 같은 규정이 마련되어 실제 시행되는 경우 그 밖의 비식별 조치에 대하여도 영향을 미칠 가능성이 있다.

한편, 오히려 현행법에는 가명처리를 규율하는 규정이 있다. 개인정보보호법 제3조 제7항, 제18조 제2항 제4호 등은 '익명' 등의 표현을 써 일정한 경우 가명처리를 요구하고 있고, 생명윤리 및 안전에 관한 법률도 익명화라는 이름하에 가명처리를 요구한다.

(2) 가이드라인

개인정보 보호법 등의 개인정보 개념에 근거한 비식별 조치가 하나의 대안으로 대두하면서, 비식별화의 방법에 관한 논의도 활발해졌다. 관련 논의는 정부와 민간 모두에서 이루어졌으나, 특히 데이터 기반 경제의 육성 등을 위하여 정부에서 일련의 가이드, 가이드라인 등을 발표해온 점이 주목된다. 행정자치부가 2013년 9월 발표한 「공공정보 개방·공유에 따른 개인정보 보호지침」(개인정보 처리에 관한 주의사항 이외에 비식별화 조치방법에 관한 소개를 포함), 미래창조과학부가 2014년 5월 발표한 「빅데이터 활용을 위한 개인정보 비식별화 사례집」, 행정자치부가 2014년 12월 발표한 「개인정보 비식별화에 대한 적정성 자율평가 안내서」(개인정보 재식별 및 위협요소에 관한 설명과 비식별화에 대한

적정성 평가, 재식별 위험관리 방안 등을 다루었다), 미래창조과학부가 2015년 6월 발표한 「빅데이터 활용을 위한 개인정보 비식별화 안내서」(분야별 관련 법령 및 비식별화 기법의 실무상 활용방법에 대한 안내를 포함하였다) 등이 있다. 그중 첫 번째 것을 제외한 나머지는 공공부문과 민간부문 모두에 적용할 것을 예정하였고, 세 번째 것은 비식별 조치의 절차적 및 관리적 측면에 주목하였다는 점에서 의의가 있다.

이후 방송통신위원회가 2014. 1. 23.「빅데이터 개인정보보호 가이드라인」을 발표하였고, 2015. 2. 23. 그 해설서도 발표하였다. 이들은 "비식별화"를 "데이터값 삭제, 가명처리, 총계처리, 범주화, 데이터 마스킹 등을 통해 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 쉽게 결합하여도 특정 개인을 식별할 수 없도록 하는 조치"로 규정하고, 이러한 조치를 취하면 정보주체의 동의 없이 처리할 수 있음을 분명히 하였다. 이를 참조하여 2016. 6. 30. 국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부 6개 부처 합동으로 「개인정보 비식별 조치 가이드라인」이 발표되면서, 정부에서 발표한 기존의 가이드라인들을 모두 대체하였다.

2016년 「개인정보 비식별 조치 가이드라인」은 '이 가이드라인에 따라 정보주체를 알아볼 수 없도록 비식별 조치를 적정하게 한 비식별 정보는 개인정보가 아닌 것으로 추정되며, 따라서 빅데이터 분석 등에 활용이 가능합니다'(전문)라면서 '가명처리 기법만 단독 활용된 경우는 충분한 비식별 조치로 보기 어렵'다고 한다. 그리고 'anonymization, 익명화'와 '비식별 조치'가 사실상 같은 개념이라고 명시하고, 식별자 삭제 이외에 속성자에 대하여도 가명처리, 총계처리 등 기법으로 대응관계를 제거하도록 하였다. 또한, ① 개인정보 해당성을 검토하는 사전검토, ② 식별자 및 속성자에 대한 비식별 조치, ③ 적정성 평가, ④ 사후관리의 네 단계 모델을 제시하였다. 적정성 평가와 관련하여서는 외부전문가를 포함한 평가단과 k-익명성 모델을 구체적으로 설명한다.

그러나 이미 본 바와 같이 가명처리는 비식별 조치의 기술 중 하나로 언급되고 있을 뿐, 정보 보호 관점에서 독자적 의의를 부여받고 있지는

못하다. 단지 단순 가명처리만으로는 비식별화되었다고 보기 어렵다는 점을 확인하였을 뿐이다.

2. 문제점

제1장에서도 언급한 바와 같이 현재의 법 상태에 대하여는 각계에서 전혀 다른 방향의 비판이 이루어지고 있다.

먼저, 2016년 「개인정보 비식별 조치 가이드라인」에 법적 효력이 없다는 점이 거듭하여 지적되고 있다. 2016년 「개인정보 비식별 조치 가이드라인」은 '가이드라인에 따라 [중략] 비식별 조치를 적정하게 한 비식별 정보는 개인정보가 아닌 것으로 추정되며, 따라서 빅데이터 분석 등에 활용이 가능합니다'라고 설명하나, 위 가이드라인이 그 자체 법규범의 지위를 갖지 아니하므로 2016년 「개인정보 비식별 조치 가이드라인」을 따랐다고 하여 적법한 처리라고 법적으로 '추정'될 수도 없다는 주장이 우선 제기된다. 그리하여 정보 활용을 원하는 산업계 등에서는 위 가이드라인을 믿고 비식별정보를 이용하기 어렵다고 비판하고 있다. 정보 보호를 강조하는 시민단체 등도 위 가이드라인을 믿고 비식별정보를 처리하여서는 아니 된다고 주장하고 있다.

다음, 2016년 「개인정보 비식별 조치 가이드라인」의 내용에 대한 비판이 있다. 그 내용에는 가령 재식별 방지의무가 법적인 근거가 있는 것인지, 재식별이 어렵게 하는 것만으로 비식별화되었다고 할 수 있는지, 재식별이 불가능할 정도에 이르러야 하는 것은 아닌지(일부 시민단체의 주장, 절대설에 상응한다), 재식별 위험성 판단과 관련하여 k-익명성을 지나치게 강조하는 것은 아닌지, 외부위원의 관여와 평가를 요구하는 것이 적절한지, 외부위원과 개인정보처리자 간 책임위험 분배는 어떻게 되는지 등의 다양한 이론적, 실무적 이슈들이 포함된다.

나아가 이러한 문제는 개인정보 보호법 자체의 흠으로 이어진다. 개인정보 보호법이, 특히 제17조 이하에서 입법기술적으로 거칠게 입법되어

불필요한 모순, 충돌을 일으키고 있고(특히 비식별화와 제18조 제2항 제4호의 관계에 관하여 오해의 소지가 있고, 실제로 오해가 벌어지고 있다는 점은 위 서울중앙지방법원 2017. 9. 11. 선고 2014가합508066, 538302 판결이 잘 보여주는 바이다), 개정이 필요하다는 지적이 있다. 비식별 조치 자체가 개인정보의 처리 중 하나인 가공에 해당하는가 하는 점도 현행법상 논란이 있는 문제이다. 일각에서는 법이 실무상 중요한 비식별화, 익명화, 가명화 등을 모두 정면에서 규정하지 아니하여 개념상 혼란이 생기고 있다고도 한다. 일부 시민단체는 비식별화가 익명화와 가명화를 포함하는 개념이라는 전제하에 제대로 된 익명화가 아닌 가명화만 가지고 개인정보를 마음대로 이용하게 해줄 위험이 있어 입법이 필요하다고 한다. 정보 활용을 위해서도 비식별 조치에 관한 법률적 근거의 확보가 필요하다는 주장이 있다.

제 2 절 정책적 제언

1. 해외 법제도의 시사점

해외의 비식별정보에 관한 논의를 살펴보면 다음 몇 가지 시사점을 얻을 수 있다.

첫째, 법령과 가이드라인 양쪽 모두에서 비식별정보/비식별화/비식별 조치와 익명정보/익명화 사이에 명확한 경계는 확인되지 아니한다. 두 용어는 같은 뜻으로 혼용되거나 필요에 따라 다른 위계를 갖는 용어로 구분되나, 어느 경우든 각 용어에 확고하게 부여된 의미가 있는 것은 아님을 전제한다. 적어도 용어를 둘러싼 국내의 논쟁은, 그것이 해외의 논의에 (대한 잘못된 또는 편향된 이해에) 터 잡고 있는 한, 대체로 근거가 명확하지 않거나 불필요한 경우가 적지 않다고 할 수 있다.

둘째, 비식별 조치에 관하여는 법령 수준에서 직접 입법을 하는 예와 그렇지 아니한 예가 모두 발견된다. 그러나 프랑스와 일본 정도를 제외

하면 비식별 조치에 대하여 법률 등에 규정을 둔다 하더라도 비식별정보가 개인정보에 반대되는 개념이고 비식별 조치된 정보는 더는 개인정보가 아님을 확인하는 데 그치고 있다(미국 HIPAA, GDPR, 호주). 비식별 조치에 관하여 명문 규정을 두지 아니하였을 때에도(독일, 영국, 중국) 개인정보의 개념을 규정함으로써 간접적으로 비식별정보와 비식별 조치의 가능성을 열어둔다는 점에는 차이가 없다.

셋째, 어느 나라에서나 비식별 조치의 적법성을 현실적으로 주도하는 것은 기본적으로 가이드라인이다. 대부분의 가이드라인은 엄밀한 의미의 법적 효력은 없는 권고적 성격의 문서이지만 높은 권위를 인정받고 있다. 그러나 일본과 같이 다른 나라에서는 가이드라인에 규정되곤 하는 각종 관리적 조치의무를 법률상 의무로 끌어올린 예도 보이고, 프랑스와 같이 일정한 경우 비식별 조치의 적정성에 관하여 CNIL이 심사하는 예도 보인다.

넷째, 비식별 조치 가이드라인은 어느 나라에서나 다양한 기술적 대안과 평가방법을 제시하고 있고, k-익명성과 같은 한두 개의 평가기준을 강조하지 아니한다. 특히 다수의 가이드라인은 잠재적 공격을 보호대상인 정보의 내용과 상관적으로 고려하여 개별적인 평가기준을 수립할 것을 요구한다. 또한, 근래의 가이드라인들은 점차 원자료와 비식별정보 양자에 대하여 동일한 통계처리를 하여 오차의 정도를 보는 등 정보의 유용성에 대한 평가절차를 통합하는 경향을 보이고 있다. 반면 평가단 등의 관여를 필수적인 절차로 요구하는 예는 잘 보이지 아니한다. 그밖에 비식별 조치의 방법과 절차를 문서화할 것을 강조하고 있는 점도 눈에 띈다.

다섯째, 위 사항과 관련하여, 많은 나라에서 일괄적이고 기계적인 기준을 마련하는 것이 적절하지 않음을 명시적으로 밝히고 있다. 오히려 데이터 맥락(context)이나 환경(situation)을 고려하여 개별 상황에 따라 개별적으로 판단할 필요가 있음을 강조하는 경우가 많다. 그러한 접근 하에서는 리스크(risk)에 대한 고려가 중요해진다. 리스크를 완벽하게 제거하는 것은 현실적이지 않다고 하는 한편, 어떻게 리스크를 최소화

할 것인지에 대한 고려가 핵심이 되어야 한다는 것이다.

여섯째, 리스크에 대한 고려는, 위 네 번째 사항으로 언급한 것과 같이 어느 한두 가지 기준이나 방법에 의존하지 말아야 함을 의미하기도 한다. 그와 동시에, 리스크를 통제하기 위해서는, 통계적, 공학적 방법론과 절차적, 관리적 방법론이 서로 보완적인 역할을 해야 함을 의미한다. 데이터에 대한 일회적인 비식별 기법의 적용 및 평가를 통해 문제가 해결될 수 있는 것이 아니라, 데이터가 어떤 절차를 거쳐 어떻게 통제되고 관리되고 있는지에 대한 고려가 동시에 이루어져야 한다는 것이다. 이는, 예를 들어, 데이터에 대한 접근통제의 수준 및 절차가 어떠한지를 고려하여, 그에 따라 데이터에 대해 요구되는 비식별의 수준과 기법이 차등적으로 적용될 수 있음을 시사한다.

2. 국내 규제 환경하에서 가명처리 개념의 도입가능성

GDPR의 큰 특징은, 익명화에 대하여는 별다른 명확한 규정을 포함하지 않는 한편, 오히려 가명처리를 정면에서 규정하고 여러 규정에서 이를 활용하고 있다는 점이다. 비식별화 내지 익명화와 가명처리 사이의 경계가 유동적이고, 정보처리자로서는 완전한 비식별화는 곤란하다 하더라도 가명처리는 가능한 경우가 적지 아니한데, 그 결과 정보 보호가 제고되는 측면이 있다. 개인정보와 비식별정보 개념의 구분이 동태적이고 비식별정보가 되는 경우 개인정보 보호법의 적용이 배제되어 그 효과가 강력하다는 점에서, 향후의 발전에 적응하는 데 어려움이 있는 입법적 고정이 저어되는 바가 있는 것과 달리, 가능한 한 가명처리를 하도록 유도하는 것은 정보 보호를 위하여 그 범위에서도 언제나 의미 있고 입법적 고정에 큰 위험이 따르지도 아니한다. 나아가 법률상 가명정보는 원칙적으로 개인정보임을 명확히 하여, 비식별화에 가명처리가 포함되는지 여부를 둘러싼 논란과 비식별화의 허용이 가명처리로 개인정보 보호법을 피할 수 있게 해주는 것은 아닌가 하는 의심을 해소하는

데 기여할 수 있다. 즉, 직접적으로도 최소수집을 포함한 기존의 대원칙을 구체화하여 규범준수를 강화하고 정보 보호를 제고할 수 있을 뿐 아니라, 간접적으로는 비식별 조치라는 타협점을 현실화하는 데도 도움이 되는 것이다. 이 점에서 가명처리 개념을 명문으로 도입하고, 비식별 조치 가이드라인을 비식별 조치 이외에 가명처리까지 포함하여 체계화하는 것은 고려해볼 만한 대안이라고 보인다.

제 6 장 결론

비식별 조치는 개인정보 보호법 체제하에서 정보 보호와 정보 이용의 조화를 꾀하는 대표적인 방법이다. 그러나 비교적 새로운 접근법으로 해외에서도 빠른 속도로 발전하고 있어 늘 업데이트가 필요한 영역이기도 하다.

이 보고서는 그러한 관점에서 해외의 비식별 조치 관련 동향을 조사, 분석하였다. 먼저 개인정보의 개념에 터 잡아 비식별정보/비식별 조치/비식별화, 익명정보/익명화, 가명정보/가명화를 구분하였다. 비식별정보와 익명정보의 개념의 이동(異同)과 정부(正否)에 대하여는 논란이 상당하나, 어느 하나의 정답은 없고, 적어도 정부와 일부 법령, 하급심 판례는 '비식별화'를 개인정보로서의 식별가능성을 제거하는 행위 또는 그 결과 식별가능성이 제거된 정보를 뜻하는 용어로 쓰는데, 그렇게 이해하는 한 '익명화'와 같은 의미라고 할 수 있다. 반면 가명정보는 식별자를 직접 식별할 수 없는 다른 정보로 대체하는 것, 나아가 GDPR 등의 정의를 참고한다면 그 복원을 위한 '키' 등을 별도로 보관하는 것을 말하는데, 그것만으로 당연히 개인정보가 아니라고 할 수는 없고, 본래는 정보 보호를 위한 제도로서 비식별화와 기능이 일치하지는 아니하나, 비식별정보와의 경계가 유동적이라는 점에서 양자 사이에 일정한 관련이 있다고 결론 내렸다. 이어 비식별 조치의 기술적 및 관리적 방법과 그 활용사례를 국내외에 걸쳐 개관하였다. 특히 기술적 차원에서 국제적 표준이 수립되어가는 단계임을 지적하고, 그 주요내용을 살펴보았고, 오픈소스 소프트웨어를 통한 비식별화를 지적하였으며, 기술적 조치의 개요를 정리하는 이외에, 국내 및 해외에서 비식별정보를 활용한 사례를 소개함으로써 비식별 조치에 어떠한 잠재력이 있는지를 가늠해볼 수 있게 하였다.

이 보고서의 핵심은 비식별 조치에 관한 해외 동향의 소개, 분석이다. 먼저, 유럽과 관련하여서는 EU를 아우르는 법인 GDPR과 개별국가로 독일, 프랑스, 영국, 핀란드의 예를 살펴보았다. 올해 발효한 GDPR의 경우 본문에서 직접 익명화를 규정하고 있지는 아니하나 서문(recital)이 익명화의 개념과 효과에 대한 일정한 이해를 드러내고 있고, 무엇보다도 서문과 본문 각조에서 가명처리에 관한 개념규정과 적절한 보호장치로서 가명처리의 활용 및 그러한 활용을 촉진할 수 있게 하는 각종 제도적 유인책을 규정하고 있다는 특징이 있다. 반면 익명화 방법과 절차에 대하여는 WP 29의 의견서(Opinion 05/2014) 이후 유럽연합 수준에서 이루어진 진전은 없는 것으로 보인다. 독일, 프랑스, 영국의 (국내) 정보보호법은 GDPR의 시행에 맞추어 개정되면서 주요 규율 대부분을 GDPR을 원용하는 방식으로 바꾸었다(영국은 이른바 Brexit에도 불구하고 그러한 태도를 택하였다). 2017년 개정 전 독일 연방데이터보호법(BDSG)은 본래 익명화(Anonymisierung)와 가명화(Pseudonymisierung)에 관한 개념정의 규정을 두고 특히 그중 가명화는 각칙 본조에서 종종 의무화하는 등 활용하고 있었는데, 이들 규정도 2017년 개정에서 GDPR을 원용하고 기본규정 대부분을 삭제한 결과 익명화와 가명화에 대한 정의 규정을 포함한 관련 규정 대부분이 삭제되었다. 프랑스 정보처리자유법(Loi informatique et libertés)에서는 처음부터 정의규정은 없었고 익명화 개념을 전제로 CNIL이 일정한 경우 익명화의 적정성에 관하여 승인할 권한을 갖는 것으로 규정하는 것이 있을 뿐이다. 영국의 정보보호법(Data Protection Act)에는 관련 규정이 없다. 그러나 이들 규정은 모두 GDPR을 원용하여 국내법으로 삼는 규정을 두므로 GDPR의 익명화 및 가명처리에 관한 규정이 동시에 국내법적인 근거규정이 된다. 독일에서는 종래 통계청의 익명화 관련 보고서 이외에 최근 발간된 가명처리 백서가 눈에 띄고, 프랑스는 익명화에 관한 WP 29 의견서에 대해 CNIL이 발표한 의견이 주목된다. 영국의 경우 ICO 가이드라인과 UKAN 보고서가 여전히 중요한 참고자료라고 할 수 있다. 핀란드에서도 가이드라인이 상세한 절차를 제시하고 있고,

미국의 경우 NIST 보고서가 계속 발간되고 있다. 캐나다는 연방법 이외에 특히 온타리오 주에서 발간한 보고서들이 상세한 가이드라인을 제시한다. 호주는 명문의 규정을 두고 있고, UKAN 보고서를 대폭 참고하여 (그러나 유럽연합적 맥락에 있는 'anonymisation' 용어는 의식적으로 'de-identification' 용어로 대체하였다) 거의 같은 내용의 가이드라인을 마련하였다. 아시아에서는 일본이 특히 주목된다. 일본은 최근 개인정보보호법을 전면개정하여 이른바 익명가공정보라는 개념을 명문으로 도입하였고, 익명가공정보를 작성함에 있어 준수하여야 할 법적 의무와 절차도 규정하였다. 또한, 개인정보 보호위원회가 익명가공정보에 관한 가이드라인을 발표하여 좀 더 구체적인 작성방법과 준수사상을 설명하고 있다. 이러한 입법적 조치 등이 익명가공정보의 활용을 활성화하기 위한 정책적 결단의 산물이라는 점 또한 주목하여야 할 부분이다. 중국은 사이버보안법이 일부 규율하고 있는 개인정보 개념에 터 잡아 가이드라인 형식으로 비식별 조치를 허용하고 있고, 싱가포르도 개인정보보호법상의 개인정보 개념에 터 잡아 비식별정보에 관한 사항을 포함한 가이드라인을 마련하였다.

2018년 상반기에 열린 두 차례의 해커톤을 비롯하여 비식별 조치에 관한 국내적 논의도 활발하였다. 이미 국회에는 비식별 조치에 관한 명문의 규정을 포함하는 다수의 입법안들이 제출되어 있고, 지난 9월 20일 국회를 통과한 이른바 규제프리존법 중에도 규제프리존 내 IoT에서 수집한 정보의 '비식별화'에 관한 명문 규정이 포함되어 있다. 2016년 「개인정보 비식별 조치 가이드라인」 발표에도 불구하고 이 문제를 둘러싼 논란이 해소되거나 더 진전하지 못하고 서로 다른 입장이 평행선을 달리면서 복잡한 법적 분쟁으로 비화한 데에는 우리의 규제체계가 비식별정보를 둘러싼 복잡한 이해관계와 충돌하는 가치를 충분히 고려하여 설계되지 못하였다는 점도 일정 부분 기여하였을 것이다. 비식별 조치는 해외에서도 한해가 다르게 발전하는 영역이자, 점점 더 규제내용과 그 수준의 국제적 조화가 중요해지는 영역이기도 하다. 해외 비식별 조치에 관한 동향을 추적하는 이 보고서의 작업은 이와 같은 비식별 조치

관련 규제의 계속발전에 기초자료로서 기여할 수 있을 것이다.

<참 고 문 헌>

제2장 개념 정의 및 비교

[국내 자료]

이동진, “개인정보 보호법 제18조 제2항 제4호, 비식별화, 비재산적 손해 - 이른바 약학정보원 사건을 중심으로 -”, 정보법학 제21권 제3호(2017).

이은우, “비식별화, 개인정보보호법이 맞은 최대의 위기”, 빅데이터 활용과 다가올 위험 - 개인정보 비식별화 문제를 중심으로 - (2015. 8. 19. 국회토론회 자료집).

[외국 자료]

Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Technique (2014).

NIST, Guide to Protecting the Confidentiality of Personally Identifiable Information (Special Publication 800-122) (2010).

[참조 사이트]

http://www.mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&nttId=55287

https://www.4th-ir.go.kr/topic/6/detail/11

제3장 비식별 조치의 이론적 현황 및 사례

[국내 자료]

공공데이터포털, “국민건강 알람서비스 고도화” 완료보고, https://www.data.go.kr/comm/file/download.do?atchFileId=FILE_000000001315675&fileDetailSn=0

공공데이터포털 기업탐방인터뷰, “공공데이터 활용한 창업으로 현대판 나이팅게일을 꿈꾸는 남자”, https://www.data.go.kr/useCase/interview/1000666/show.do

뉴스투데이, CJ헬스케어, 빅데이터로 ‘니즈’를 찾아내 신약개발, http://www.news2day.co.kr/92887

디지털타임스, 삼성카드, 빅데이터 기반 ‘링크’ 서비스, http://www.dt.co.kr/contents.html?article_no=2014101502100558785001

매일경제, http://news.mk.co.kr/newsRead.php?year=2017&no=72979

매일경제, 매경 핀테크 어워드 2018 / 최우상, http://news.mk.co.kr/newsRead.php?year=2018&no=544729

삼성SDS, 차세대금융! 금융분야 빅데이터(Bigdata) 분석 활용 사례, https://www.samsungsds.com/global/ko/support/insights/1196790_2284.html

신한카드블로그, 신한카드 코드9 시리즈, 500만매 돌파 - 빅데이터 정확성의 결과, http://www.shinhancardblog.com/291

이현승, 송지환, 「개인정보 비식별화기술의 쟁점 연구」, 소프트웨어정책연구소(2016)

중앙일보, 보험금 청구 즉시 빅데이터로 “사기” 적발, https://news.joins.com/article/11536647

청년의사, 라인웍스, 의료 빅데이터 기계학습을 통해 재입원 예측, http://www.docdocdoc.co.kr/news/articleView.html?idxno=1058491

한국일보, CJ헬스케어, 위식도 역류질환 신약 ‘케이캡정’ 허가받아, http://hankookilbo.com/v/b184f001cc1246a18c97ed641ba8a8ee

ZDNet Korea, 서울시, 심야버스노선 어떻게 만드나, http://www.zdnet.co.kr/news/news_view.asp?artice_id=20130702115100

[외국 자료]

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

BoardingArea, A Complete Guide To Amex Sync (American Express Sync), https://milestomemories.boardingarea.com/a-complete-guide-to-amex-sync-american-express-sync/

DataFloq, T-Mobile USA Cuts Downs Churn Rate By 50% With Big Data, https://datafloq.com/read/t-mobile-usa-cuts-downs-churn-rate-with-big-data/512

El Emam, K., et al., “De-identification Methods for Open Health Data: The Case of the Heritage Health Prize Claims Dataset”, Journal of Medical Internet Research, Vol. 14, No. 1 (2012).

GIGAOM, Credit scores, with a little help from your friends, https://gigaom.com/2012/05/16/credit-scores-with-a-little-help-from-your-friends/

Ginsberg, J. et al., “Detecting influenza epidemics using search engine query data”, Nature, Vol. 457, pp. 1012-1014 (2009).

International Organization for Standardization, ISO 25237:2017 Health informatics — Pseudonymization (2017).

International Organization for Standardization, ISO/IEC 20889 IT

Security techniques — Privacy enhancing data de-identification terminology and classification of techniques (2018).
mobihealthnews, Google Flu Trends website shuts down; will send data to Boston Children's, Columbia, CDC. https://www.mobihealthnews.com/46248/google-flu-trends-website-shuts-down-will-send-data-to-boston-childrens-columbia-cdc
Prasser, F. and Kohlmayer, F., "Putting Statistical Disclosure Control Into Practice: The ARX Data Anonymization Tool", in Gkoulalas-Divanis, Aris, Loukides, Grigorios (Eds.), Medical Data Privacy Handbook, Springer (2015).
PR Newswire, Carpe Data: The Next Generation Insurance Data Company Announces Launch, https://www.prnewswire.com/news-releases/carpe-data-the-next-generation-insurance-data-company-announces-launch-300337756.html
Shin, S.-Y., et al., "A De-identification for Bilingual Clinical Texts of Various Note Types", Journal Korean Medical Science, Vol. 30, No. 1, pp. 7-15 (2015).
The Balance, Progressive Snapshot Review, https://www.thebalance.com/progressive-snapshot-review-4141266

[참조 사이트]

https://arx.deidentifier.org/overview/related-software/
http://forecast.nhis.or.kr/menu.do
https://www.data.go.kr/useCase/interview/1000666/show.do
https://www.kaggle.com/c/hhp
http://www.law.go.kr/행정규칙/개인정보의기술적·관리적보호조치기준
http://sync.americanexpress.com/

ISO/TC 215 Health informatics, https://www.iso.org/committee/54960.html
ISO/IEC jTC 1/SC 27 IT Security techniques, https://www.iso.org/committee/45306.html

제4장 비식별 조치에 대한 해외의 제도적 현황

[외국 자료]

Act on the Protection of Privacy in Electronic Communications (516/2004; amendments up to 125/2009 included) 제13조 참조
Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques (2014)
Bergt, BGH: Speicherung von IP-Adressen durch die Bundesrepublik Beschluss vom (2014)
Bevott, Collins, PDP (2016).
Canada Information and Privacy Commissioner of Ontario, De-identification Guidelines for Structured Data (2016).
_____, Big Data and Innovation, Setting the Record Straight: De-identification Does Work (2014).
_____, Big Data Guideline (2017).
Colin J. Bennett, Regulating Privacy: Data Protection and Public Policy in Europe and the United States (Cornell University Press, 1992)
Dammann, in Simitis [Hrsg.], Bundesdatenschutzgesetz, Nomos (2011).
David H. Flaherty, Protecting Privacy in Surveillance Societies: The Federal Republic of Germany, Sweden, France, Canada, and the

United States (1989).
Elliot, Mark, Mackey, Elaine, O'Hara, Kieron, Tudor, Caroline, The Anonymisation Decision-Making Framework, UKAN Publications, 2016
Maldoff, The Privacy Advisor (2016).
Office of the Australian Information Commissioner, De-identification Decision-Making Framework (2016).
_____, De-identification and the Privacy Act (2018).
Singapore Personal Data Protection Commission, Proposed Advisory Guidelines on the Personal Data Protection Act for Selected Topics(2013).
_____, Guide to Basic Data Anonymisation Techniques (2018).
The Finnish Social Science Data Archive
UK Information Commissioner's Office, Anonymisation: Managing Data Protection Risk Code of Practice (2012).
UK Anonymisation Decision-Making Framework, The Anonymisation Decision-Making Framework (2016).
US Office of Human Rights, Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule (2012).
US National Institute of Standards and Technology, De-Identification of Personal Data (2015).
_____, De-Identifying Governmental Datasets (2nd draft) (2016).
WP 29 Opinion on anonymization techniques, Clinical Trial Data Sharing: Methods and Experiences with De-Identification (2015)

日本 個人情報保護委員会,
個人情報の保護に関する法律についてのガイドライン（匿名加工情報

編）(2016).
_____, (匿名加工情報- パーソナルデータの利活用促進と消費者の信頼性確保の両立に向けて) (2017).
中国 国家标准化管理委员会, 信息安全技术-揭秘个人信息指南(草案) (2017).
_____, 信息安全技术 个人信息安全规范(GB/T 35273-2017) (2018).

[참조 사이트]

Greens/European Free Alliance, EU General Data Protection Regulation State of play and 10 main issues https://www.greens-efa.eu/en/article/eu-general-data-protection-regulation/
https://ec.europa.eu/programmes/horizon2020/en/
https://opengovdata.io/2014/us-federal-open-data-policy/
https://www.ncbi.nlm.nih.gov/books/NBK9573
https://stm.fi/en/article/-/asset_publisher/sosiaali-ja-terveystietojen-tietoturvallinen-hyodyntaminen
https://thl.fi/en/ajankohtaista/lausunnot-ja-kuulemiset/kuulemiset
http://tukija.fi/en/publications1
http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf
http://web.abo.fi/personal/kurskatalog/material/Varantola07042017.pdf
http://www.bfdi.bund.de/DE/BfDI/Artikel_BFDI/AufgabenBFDI.html
https://www.bfdi.bund.de/SharedDocs/Publikationen/Taetigkeitsberichte/TB_BfDI/22TB_07_08.html
https://www.chinalawblog.com/2018/02/chinas-personal-information-security-specification-get-ready-for-may-1.html

https://www.cnil.fr/sites/default/files/typo/document/Act78-17VA.pdf
https://www.cnil.fr/sites/default/files/typo/document/Guide_Security_of_Personal_Data-2010.pdf
https://www.cnil.fr/fr/loi-78-17-du-6-janvier-1978-modifiee
https://www.cnil.fr/sites/default/files/atoms/files/cnil_guide_securite_personnelle.pdf
https://www.cnil.fr/fr/PIA-privacy-impact-assessment
https://www.dataguidance.com/france-cnil-takes-pragmatic-approach-gdpr-enforcement/
https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/AllgemeinesMethoden/Anonymisierung_42003.pdf?__blob=publicationFile
http://www.finlex.fi/fi/laki/kaannokset/2004/en20040516.pdf
https://www.forbes.com/sites/forbestechcouncil/2018/08/20/how-will-californias-consumer-privacy-law-impact-the-data-privacy-landscape/#5ca73677e922
http://www.fsd.uta.fi/aineistonhallinta/en/
https://www.gdd.de/downloads/whitepaper-zur-pseudonymisierung
http://www.jdsupra.com/legalnews/gdpr-compliance-update-which-government-39098
https://www.legifrance.gouv.fr/eli/loi/2016/11/18/JUSX1515639L/jo
https://www.legifrance.gouv.fr/eli/loi/2018/6/20/JUSC1732261L/jo/texte
https://www.myhealthrecord.gov.au/for-you-your-family/secondary-uses-data
https://www.ncbi.nlm.nih.gov/books/NBK9573/
http://www.oecd.org/sti/sci-tech/38500813.pdf
http://www.pharmatimes.com/news/govt_suspends_controversial_nhs_data_sharing_deal_1235020
https://www.pdpc.gov.sg/-/media/Files/PDPC/New_DPO_Connect/nov_15/pdf/Anonymisation.pdf

https://www.reuters.com/article/us-usa-internet-privacy/trump-administration-working-on-consumer-data-privacy-policy-idUSKBN1KH2MK
https://www.service-public.fr/
https://www.smh.com.au/technology/australians-health-records-unwittingly-exposed-20171218-p4yxt2.html
https://www.straitstimes.com/singapore/new-guide-to-clear-doubts-on-sharing-customer-data
http://www.wipo.int/wipolex/en/text.jsp?file_id=343249
Security of personal data Guideline (La sécurité des données personnelles)
https://www.cnil.fr/fr/un-nouveau-guide-de-la-securite-des-donnees-personnelles
The North Rhine-Westphalia Commissioner for Data Protection and Freedom of Information, “Regulation of Data Protection”, Data Protection.
https://www.ldi.nrw.de/LDI_EnglishCorner/mainmenu_DataProtection/Inhalt2/authorities/regulation.php