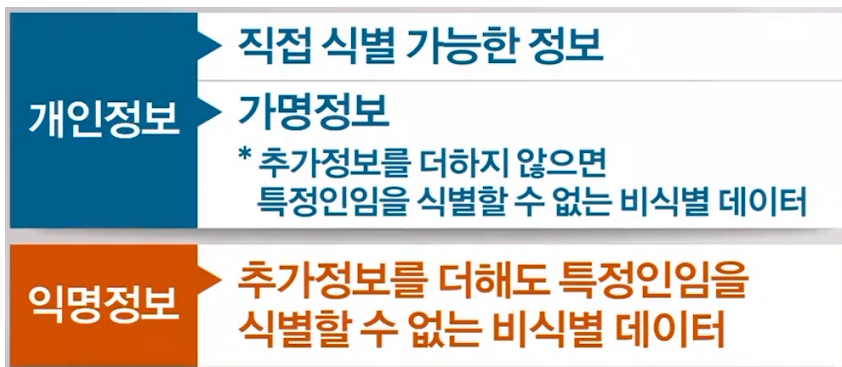


[팩트체크] 데이터 3법, 왜 개인정보 도둑법인가?

Q1. 개정안에서의 가명처리는 안전하다?

1. 개인정보, 가명정보, 익명정보란?



(출처 : SBS CNBC 용감한 토크쇼 직설 898회, 2019.11.28)

- 개인정보보호법 개정안 제2조 1의2
 - “가명처리”란 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것을 말한다.
- 익명정보는 개인정보가 아님. 따라서 개인정보보호법의 적용을 받지 않고 자유롭게 활용 가능함.
- 가명정보는 개인정보임. 따라서 개인정보보호법의 적용을 받음. 유럽연합 GDPR도 가명정보를 개인정보로 보고 있으며, 한국 정부도 인정하고 있음.

익명정보	성명	전화번호	성별	나이	보험 가입 건수
	권율	02-2345-6789	D	20대	1
	이순산	010-4567-9876	D	20대	2
	김두권	010-3698-5214	D	20대	3
	강수자	010-4562-7895	C	30대	4
	신은화	010-7412-5896	C	30대	5
	이영애	010-9871-6543	C	30대	6

▪ '성명'과 '전화번호'를 삭제
 ▪ '성별'은 직접적으로 알아볼 수 없도록 코드 형태로 변환(여→C/남→D)

ID(대체값)	성명	전화번호	성별	생년월일	보험 가입 건수
9A00F1155584BA5DDFFC4B6D DDBA	김희선	010-1234-5678	여	1974.10.1.	1
C2E6376B9035D7067C8B68F25 FA39	권율	02-2345-6789	남	1990.3.26.	2
DACE2CCC9F459387EAE890D8 5344F	강수자	010-3456-7891	여	1969.5.28.	3
...

▪ 성명, 전화번호를 조합하여 가명조치 기법 중 하나인 해시함수(SHA-256)를 통해 암호화

(출처 : 정무위, 신용정보법 개정안 심사자료)

2. 가명정보는 안전한가?

- 가명정보는 가명처리되지 않은 원래의 개인정보보다는 안전함. 따라서 가능하다면 개인식별이 가능한 개인정보의 형태로 처리, 보관하는 것보다는 가명처리해서 보관하는 것이 바람직함.
- 가명정보는 다른 정보와 결합하여 여전히 재식별될 위험성이 있음. 즉, 재식별이 불가능한 익명정보보다는 위험함. 따라서 가능하다면, 익명처리해서 활용하는 것이 가장 바람직함. 재식별의 위험성은 가명처리의 방법 및 수준에 따라 달라지며, 현재 가명처리의 기술, 재식별 기술 모두 발전하고 있는 상황임.&n!

bsp;
- 개인식별자를 삭제하더라도 여전히 개인식별의 위험성이 있음은 이미 2016년 정부가 발표한 <개인정보 비식별조치 가이드라인>에서도 인정하는 바임.

- 온라인 영화대여 회사인 넷플릭스(Netflix)는 고객의 기호에 맞는 영화를 추천하는 알고리즘의 정확성을 높이기 위해 경연대회를 개최
 - 1999년 12월부터 2005년 12월까지 50만명의 이용자들이 영화에 대한 평점을 내린 1억건의 시청 이력 데이터를 공개
 - ※ 사용자를 식별할 수 있는 이름 등은 삭제하였으나, 데이터 처리 내용을 연결하기 위해 독특한 식별자, 영화에 대한 평가 내용, 평가 일시 등을 공개
- 텍사스 대학의 한 그룹이 넷플릭스사가 공개한 시청 이력 데이터와 영화정보 사이트 IMDb(Internet Movie Database)에 공개된 사용자 리뷰를 결합하여 일부 개인을 식별해냄
 - ※ IMDb는 웹 사이트 상에서 아이디와 평가점수를 게시
- 미국연방거래위원회(FTC)가 프라이버시에 관한 문제를 지적하여 제2회 경연은 중지됨

(출처 : 개인정보 비식별조치 가이드라인, 2016)

- [과학 저널 <네이처 커뮤니케이션스>에 실린 논문. \(관련 기사\)](#)
 - 영국 임페리얼 칼리지 런던의 입상드르 드 몽조이(Yves-Alexandre de Montjoy) 교수를 비롯한 연구진은 미국, 터키 등의 공개된 데이터로 실험한 결과 일부 속성만 알아도 비식별화된 데이터에서 특정 개인을 정확히 찾아낼 수 있다는 것을 알았다.
 - 이 모델은 15개 인구통계적 속성(나이, 성별, 결혼 여부 등)만 알아도 그 어떤 익명화된 데이터라 할지라도 99.98% 정확도로 개인을 구분해 낼 수 있었다.

3. 가명정보의 활용 범위는?

- 개정안은 가명처리만 하면 ‘과학적 연구, 통계 작성, 공익목적의 아카이브’를 목적으로 활용할 수 있도록 하고 있는데, 과학적 연구는 신상품 개발 등 기업 내부적인 상업적 연구도 포함하고 있음. 과학적 연구를 과학적 방법을 사용하는 연구로 동어반복적으로 정의하고 있기 때문임.
 - 예를 들어, A 포털은 신상품 개발 연구 목적으로 B 통신사의 가명처리된 고객정보를 제공받을 수 있음. 또한, A 포털은 자신의 고객정보를 C 신용정보업체의 고객정보와 전문기관을 통해 결합할 수 있음. 이런 식으로 A 포털, B 통신사, C 신용정보업체, D 유통사, E 병원 등은 자신의 고객정보를 서로 판매, 공유, 결합할 수 있음.

- 유럽연합 GDPR은 'scientific research' 를 정의하고 있지 않음. 다만, 'scientific research' 여부를 판단할 때, 과학적 지식과 기술의 자유로운 유통, 유럽에서의 과학적, 기술적 기반 강화라는 목적을 고려하도록 하고 있음.
- 영국의 개인정보 감독기구인 ICO는 과학적 연구는 상업적 연구(commercial research)에 적용되지 않는다고 해석하고 있음.
- 가명정보는 안전하니 상업적인 목적으로 활용할 수 있도록 열어두어야 한다는 것이 정부와 업계의 입장.
 - 개정안에서 가명정보의 재식별을 금지한다고 하지만, 서로 다른 기업 사이에 가명처리된 고객정보를 공유, 결합할 수 있게 된다면 현재보다 개인정보의 유출 및 남용이 증가할 수밖에 없음.
- 가명정보 역시 개인정보이며 재식별의 위험성이 있기 때문에, 정보주체의 동의없는 (가명처리된) 개인정보의 활용은 공익적 가치가 있는 학술 연구, 통계 등의 제한되어야 한다는 것이 시민사회의 입장.
- 해외에서는 학술 연구자에 한해, 공공기관이 보유한 개인정보를 가명처리한 후 결합하거나 접근할 수 있도록 하고 있음. 폐쇄적인 안전 공간에서 가명정보에 접근하여 연구를 수행한 후 연구결과만 갖고 나갈 수 있도록 하고 있음. 반면, 개정안은 가명처리한 기업의 개인정보를 전문기관에서 결합하도록 지원하며, 결합된 가명정보를 원래의 기업에 반출할 수 있도록 허용하고 있음.

Q2. 개정안이 통과되면 발생하는 문제_보건의료정보

1. 유전정보, 의료정보, 건강정보는 가명처리해도 재식별 가능

- 개인 유전정보, 건강정보는 민감정보로 현행 개인정보보호법 23조에 따라 정보주체의 별도 동의 없이는 원칙적 처리를 금지하고 있음. 이는 정보주체의 사생활을 현저히 침해할 우려가 있는 개인정보이기 때문.
- 유전정보의 경우 가명처리 자체가 불가능한 정보이며, 이미 과학계에서는 개인 의료/건강정보의 경우 99.9% 재식별이 가능한 정보라는 점에 이견이 없음. 따라서 개인과 그 가족 정보를 포함하는 유전정보, 의료/건강정보는 보다 민감한 개인정보로 보호되어 함.
- 유럽 GDPR도 건강정보에 대해 원칙적 처리 금지를 명시함. 개인의 건강정보는 한 사람의 과거, 현재, 미래의 건강 상태의 집합이라는 점에서 보다 엄격한 동의 규정, 고지 의무 등을 법제화함. 단 '명시적 동의' 혹은 '치료행위 및 공중보건을 위한 공공의 이익(Public interest)을 위한 경우, 학술 연구로 제한적 활용을 권고함. 또한 이러한 경우에도 자동화된 데이터 처리나 알고리즘에 의해 어떠한 의사 결정이 이루어질 때 이에 대해 환자가 알고 개입할 권리를 보장하도록 함. 자신의 건강정보 삭제를 요청할 권리 등도 이에 포함됨.

2. 의료민영화를 위한 보건의료 빅데이터 활용

- 개인정보 규제완화가 의료 공공성을 보호하는 안전핀 제거 역할을 할 것이라는 점임. 의료민영화의 핵심 중 공보험의 역할 축소와 민간의료보험 활성화 방향은 지난 20여년간 보험업법, 의료법 등의 개정 시도를 통해 추진돼 왔으나, 시민 다수의 반대를 통해 국회에서 법개정이 중단된 바 있음. 그 내용 중 핵심은 '개인 의료/건강정보를 민간보험사 등이 공유할 수 있도록 해달라는 요구'였음. 박근혜정부 시기 '비식별 가이드라인'을 통해 이를 허용하고자 했으나 법제화 단계◆
 권◆까지 못함. 이번 문재인 정부의 개인정보보호법 개정안은 이러한 민간의료보험사 등의 오랜 민원을 해결하는 내용을 담고 있음.
 - 병원+IT기업 합작 의료데이터회사 설립 : 아산병원-카카오인베스트먼트-현대중공업, 네이버-분당서울대병원-대웅제약사
 - 마이헬스데이터 시범사업 : 세브란스-CJ프레시웨이-에쓰푸드 (건강 검진·처방전 등의 진료 데이터를 이용한 식단 구매업), 서울대병원-웰트-삼성화재(환자 개인 의료정보 및 라이프 로그 데이터와 융합, 개인 맞춤 건강관리사업)

- 보험사 등의 건강관리서비스업 허용 : 개인 건강관련 수치, 질환 관련 정보, 건강검진 결과 등을 개인정보보호법 정하는 수준에 따라 수집 활용하도록 함.
- 5대 바이오헬스 빅데이터 사업 : 의료 빅데이터 플랫폼을 구축, 신기술개발 활용되도록 지원, 공공기관(건강보험공단, 심사평가원 등)이 보유한 국민 의료/건강 정보를 가명처리 후 개방하고 활용하도록 허용, 이러한 개인정보의 본인동의를 없어도 과학적 연구 활용을 할 수 있는 근거의 내용은 행안부가 추진하는 개인정보보호법 개정에 따르도록 명시해 둬.
- 아산병원과 카카오, 분당서울대병원과 네이버등이 설립 추진하고 있는 의료정보데이터회사의 경우는 현행 불법임. 진료 목적으로 수집된 정보의 다른 목적 이용은 재동의를 필요하며, 제3자 제공도 불법임. 그러나 현대중공업, 아산병원, 카타오 등은 전자의무기록과 임상시험 정보 및 예약 기록 등을 토대로 의료 데이터 전문회사인 '아산카카오메디컬데이터'를 만들겠다고 발표함.
- 박능후 복지부 장관은 지난 해 복지부 국정감사에서 관련 질의에 대해 “의료법에 저촉된다며 단호한 조치를 취하도록 하겠다”고 말한 바 있으나, 개보법 개정안이 통과된다면 아산병원은 가명처리된 의료정보를 '과학적 방법의 연구'라며 활용하려 할 것임. 이는 의료법으로 보호하고 있는 진료의 무기록 상 정보의 남용 문제를 불러, 사회적으로 큰 문제를 야기할 것이며, 재벌이 병원 및 민간보험 회사를 소유하고 있는 국내 의료현실에서 의료영리화를 더욱 추동하는 재앙이 될 것임.
- 진료 목적으로 제공한 환자의 내밀한 정보들이, 모든 병의원 약국은 건강보험 환자들을 진료 해야 하며, 누구나 건강보험료를 납부해야 하는 당연지정제라는 공적 제도를 통해 국민건강보험공단과 심사평가원 등에 수집된 국민들의 개인 의료/건강정보가 환자와 개인의 동의 없이 개방, 활용된다 는 것 또한 있을 수 없는 일임. 공공기관에 수집돼 있는 나의 개인정보는 공공기관의 것이 아니라 내 개인 정보임. 목적 외 이용, 제 3자 제공에 대해 개인들의 동의가 필요하며, 그 목적에 있어 공익성이 담보되어야 함. 현재 추진되고 있는 개인정보보호법은 이러한 사회적 논의와 동의 절차, 충분한 고지의 의무가 생략된 채 일방적으로 추진되고 있는 것임. 따라서 이러한 일방적이며 비민주적인 추진은 결국 건강보험공단으로 상징되는 공보험에 대한 신뢰 악화를 가져올 것이며, 환자들은 치료를 위한 목적에도 불구하고 의사에게 자신의 내밀한 건강 상태 등의 문제를 이야기하지 않음으로 인해 진료의 왜곡, 신뢰 파괴의 결과를 낼 수 있음. 이는 건강증진은커녕 오진의 확률을 높이고, 사회적으로 불필요한 재정 낭비를 가져올 수 있음.
- 국회 입법조사처는 정부 개인정보보호법 개정안은 “생체인식정보의 경우 유출 시 정보주체에게 심각한 침해가 발생할 수 있어 민감정보로 보다 엄격하게 관리하는 방안을 검토할 필요가 있다”고 지적함. 현재 라이프로그 방식과 의료정보를 연계하는 마이헬스데이터 시범 사업은 홍채, 지문, 안면 인식 등의 생체인식정보의 수집에 대한 보호는커녕, 개인정보 연계를 통해 '식별된 개인에게 맞춤형 상품을 판매하려는' 것임. GDPR 역시 생체인식정보를 '민감정보'에 상응하는 '특정범주의 개인정보'에 포함하여 엄격히 관리하도록 해야 한다고 지적함.

- 빅데이터는 결국 사람을 가리킨다는 것을 간과해선 안됨. 개인 건강정보의 경우, 이것이 유출되거나 공개되었을 시 한 개인의 사생활 침해 문제로 국한되는 것이 아니라 취업, 결혼 등 각종 선택적 사회 제도에서의 차별과 배제의 결과로 이어질 수 있음. 특히 취약자의 정보 유출은 사회적 불평등으로 인해 발생하고 있는 건강불평등을 더욱 심화시킬 것임. 알고리즘 설계 자체에서 배제된 데이터, 즉 개인은 존재하지 않거나 차별과 낙인을 정당화하는 도구로 고착화될 위험도 있음.
- ‘데이터 중심 건강관리’ 라는 데이터경제론은 근거가 없음. 넛치(nudge)이론에 근거한 개인의 행동 변화를 통한 건강증진 사업은 효과가 없음이 이미 증명됨. 거꾸로 건강증진 앱은 감시, 두려움, 죄책감을 동반해 경쟁적 자아 경영을 도모하며, 앱 사용에 있어 경제적 문화적 차별을 전제하고 있다는 문제가 제기되고 있음. 결국 건강정보 규제완화는 건강을 결정하는 사회경제적 요인 문제들을 사회적 문제로 인식하지 못하도록 만들고 건강의 개인책임화를 부추기는 경제논리임. 따라서 공적인 예산들이 다수 건강증진 효과가 없다는 것이 드러난 의료상업화로 투자되고 있는 공적자금의 왜곡도 데이터경제론의 큰 문제 중 하나임.
- 보건의료 빅데이터는 그 활용이 정보주체자, 집단, 지역사회에 주는 해보다 큰 사회적 가치가 있는가의 여부(공공의 이익), 연구 과정과 결과가 모든 이들에게 호혜적이며 사회적 연대를 갖는가의 여부(형평성), 데이터의 질과 안전, 사용에 있어 투명하게 사용되고 있음을 증명할 수 있는가(책임성) 등에 대한 사회적 논의와 공론화가 우선되어야 함.