이화여자대학교
생명의료법연구소
Ewha Institute for Biomedical Law & Ethics

이슈페이퍼 2022-1

가명처리와 가명정보의 이해

김 병 필

I. 서론

- 보건의료에서 환자의 신원이 밝혀지지 않도록 보장하면서 연구 목적으로 보건의료 데이터를 활용하는 양자 간의 타협점을 찾는 일은 중요함
- 원칙적으로는 정보주체의 동의를 얻는 것이 바람직하지만 동의를 얻는 것이 어렵다면 데이터를 적절히 가공하여 더 이상 환자의 신원이 드러나지 않도록 할 수 있는데, 이를 '비식별 조치'라고 함. 즉, 비식별 조치란 데이터에서 개인을 알아볼 수 없도록 이를 변형하는 것을 의미함
- 가명처리는 데이터를 어떻게 변형하는 것인가에 관한 것으로 이를 알아보기 전, 1997년 미국 매사추세츠 주지사의 의료정보 재식별 사건을 살펴보아야 함.[1] 이 사건은 데이터의 비식별 처리를 고민하는 이들에게 중대한 숙제를 주는 계기가 되었음

미국 매사추세츠 주지사의 의료정보 재식별 사건

미국 매사추세츠 주정부는 단체보험위원회(Group Insurance Commission, GIC)를 통해 주 공무원들에게 건강보험을 제공하고 있었음. 1990년대 중반 주정부는 보건의료 연구 목적으로 이름, 주소, 사회보장번호 등 개인을 식별할 수 있을 것으로 보이는 정보를 삭제하여 GIC 의료정보를 연구자들에게 공개함. 매사추세츠 주지사는 GIC 의료정보에서 식별자를 삭제함으로써 환자의 프라이버시를 보호하고 있다고 공중을 안심시킴. 당시 MIT 컴퓨터 공학 대학원생이던 Latanya Sweeney는 GIC 의료정보로부터 주지사에 관한 정보를 찾아낼 수 있는지 시도하였고,[2] 유권자 명부를 구입하여 공개된 의료정보를 비교하여 주지사의 의료정보를 식별해낸 사건

- 이 사건 이후 이어져 온 문제의식의 연장선에서 '가명처리'라는 개념이 등장하였음
- 우리나라 개인정보 보호법은 2020년 2월 개정되어 가명처리와 가명정보 개념을 도입하였으나 현재까지 이 개념들은 널리 이해되지 못하고 있음
- 가명처리가 무엇인지, 가명정보를 어디까지 활용할 수 있는지 살펴볼 예정임

※ 이 글은 이화여자대학교 생명의료법연구소가 박영사를 통하여 출판한 『보건의료와 개인정보』 단행본의 총론 제4장을 요약한 것입니다.

1) Sweeney, L., "k-Anonymity: A Model for Protecting Privacy", International Journal on Uncertainty, Fuzziness and Knowledge-based System, 2002, 10.; 한편 고학수 · 최경진, "개인정보의 비식별화 처리가 개인정보 보호에 미치는 영향에 관한 연구", 개인정보보호위원회 연구보고서, (2015), 36-38면은 위 사례를 상세히 소개하고 있다.

2) Ohm, Paul, "Broken promises of privacy: Responding to the surprising failure of anonymization", UCLA l. Rev. 57 (2009).

II. 가명처리와 가명정보의 개념

A. 개인정보 보호법상 '가명처리'와 '가명정보'의 정의

- 2020년 개정된 개인정보 보호법은 '가명처리'와 '가명정보'에 대한 명문의 규정을 두고 있음
  - 가명처리는 "개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것"임[3]
  - 가명정보는 개인정보를 위와 같은 방법으로 "가명처리함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용 · 결합 없이는 특정 개인을 알아볼 수 없는 정보"라고 정의함[4]
- 위 조항에 대한 해석상 쟁점
  - 가명처리 방법을 "개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법"으로 설명하고 있음. 이 때 어느 범위까지의 개인정보에 대해 "삭제 또는 대체 등"을 해야 하는지가 쟁점이 됨
  - 가명처리의 결과물은 "추가 정보가 없이는 특정 개인을 알아볼 수 없도록" 해야 하는 요건을 정하고 있음. 하지만 '추가 정보'의 의미가 무엇인지 정의되어 있지 않음. 따라서 추가정보의 개념을 이해하기 위해서 가명처리의 기술적 방법을 이해할 필요가 있음

B. 가명처리의 대상 – 식별자와 속성값의 구분

- 가명처리를 위해 삭제 · 대체 등의 조치를 취해야 할 정보를 정하는 데 있어서 유용한 방법은 '식별자(identifier)'와 '속성값(attribue value)'을 구분하는 것임

1. 식별자(직접 식별자)에 대한 가명처리

- 식별자란 개인이나 그의 가족을 유일하게 식별할 수 있는 정보를 의미함
- 식별자는 '직접 식별자(direct identifier)'와 '간접 식별자(indirect identifier)'로 구분됨
  - '직접 식별자'는 그 자체로 개인을 식별할 수 있는 경우를 의미함
  - '간접 식별자'는 여러 속성값을 조합하여 개인을 식별할 수 있는 경우를 의미함
- 국내 비식별 조치 가이드라인(2016)은 '직접 식별자'를 '식별자'라고 칭하고,[5] 보건의료 데이터 활용 가이드라인(2020)도 '식별자'라고 하고 있음[6]. 개인정보보호위원회의 가명정보 처리 가이드라인(2020)은 이를 '식별정보'라고 하지만 같은 뜻으로 이해할 수 있음
- 식별자에 대한 가명처리 방법은 원칙상 간단함. 이를 삭제하거나 원래 정보를 확인할 수 없는 일련번호로 대체하거나 필요한 경우 암호화 기법을 이용하여 원래 정보를 복원할 수 없도록

3) 개인정보 보호법(2020. 2. 4. 개정) 제2조 제1의2호.
4) 개인정보 보호법(2020. 2. 4. 개정) 제2조 제1호 다목.
5) 관계부처 합동, 개인정보 비식별 조치 가이드라인(2016) 5면.
6) 개인정보보호위원회 · 보건복지부, 보건의료 데이터 활용 가이드라인(2020) 12면.

암호화할 수 있음. 이러한 일련번호나 암호화된 정보는 '가명'에 해당함. 이는 가명처리의 가장 기본적인 방법이고, 가명처리라는 용어의 유래이기도 함

2. 속성값(간접 식별자)에 대한 가명처리

- 서론에서 언급한 매사추세츠 주지사 사건에서의 해당 정보는 그 자체로는 개인을 식별할 수 없더라도 개인을 식별하는데 사용될 수 있으므로 간접 식별자에 해당함
- 간접 식별자를 지칭하는 용어는 다양함. 국내 비식별조치 가이드라인(2016)은 속성자(attribute value)라 칭하고 있고,[7] 가명정보 처리 가이드라인(2020)은 '식별가능정보'라 칭하고 있음[8]
- 간접 식별자와 연관된 개념으로 '특이정보'가 있음. 특이정보란 "관측된 데이터의 범위에서 많이 벗어난 아주 작은 값이나 아주 큰 값"을 의미함[9]
- 특이정보를 파악하기 위해서는 데이터셋을 관찰해야 함. 즉, 데이터의 주된 분포 범위 이외에 존재하는 항목들이나 빈도가 매우 적은 항목을 찾아내야 함
- 실무상으로는 데이터에 포함된 속성값 중 어느 항목까지 간접 식별자로 볼 것인가에 대한 논란이 있음. 이러한 판단이 어려운 이유는 데이터가 사용되는 맥락에 따라 그 판단이 달라지기 때문임
- 간접 식별자를 가명처리 할 때, 다양한 방법이 활용될 수 있음[10]
- 간접 식별자를 가명처리 하는데 사용하는 가장 흔한 방법은 데이터의 정밀도를 낮추어서 재식별 위험을 낮추는 것으로 이를 '일반화' 또는 '범주화'라고 함[11]
- 데이터를 일반화하더라도 특이정보에 대하여 여전히 재식별 위험이 남아 있을 수 있는데, 이 경우 특이정보를 삭제하거나 일반화시켜야 함

C. 보건의료 데이터의 가명처리

- 보건의료 데이터의 가명처리 방법에 관하여 개인정보보호위원회와 보건복지부가 2020년 발행한 '보건의료 데이터 활용 가이드라인'을 참고할 수 있음
- 이는 개인정보보호위원회의 가명처리에 관한 가이드라인을 기본으로 하여 보건의료 분야의 특수성을 반영하였으며, 데이터 유형별 가명처리 방법에 대하여 비교적 상세한 지침을 제공하고 있음[12]
- 이 가이드라인은 12가지 유형의 속성값에 대한 지침을 제공하고 있으며, 크게 세 가지로 분류할 수 있음

7) 관계부처 합동, 개인정보 비식별 조치 가이드라인(2016) 5면. '속성자'는 국어 상으로도 어색할 뿐만 아니라, 그 의미를 명확하게 전달하고 있지 못하므로 적절하지 못한 용어이다.
8) 개인정보보호위원회, 가명정보 처리 가이드라인(2020) 17면.
9) 개인정보보호위원회, 가명정보 처리 가이드라인(2020) 56면.
10) 간접 식별자에 대한 가명처리의 구체적 방법들은 개인정보보호위원회, 가명정보 처리 가이드라인(2020) 42-55면 참조.
11) 이재훈, "데이터 3법 개정에 따른 바이오 · 의료정보 활용방향과 시사점", BioINpro Vol. 71, 생명공학정책연구센터, (2020) 6면.
12) 개인정보보호위원회 · 보건복지부, 보건의료 데이터 활용 가이드라인(2020) 13-16면.

- 첫째, 해당 속성값을 이용한 개인식별 가능성이 낮다고 보아 원칙적으로 별도의 조치가 필요하지 않다고 정한 것들임
- 둘째, 체외 촬영 영상정보, 체내 촬영 영상정보, 단층촬영 · 3D 이미지 정보, 영상정보와 같은 것으로, 영상 자체나 DICOM 헤더 등 메타 데이터상에 식별자가 포함된 것은 삭제하거나 마스킹해야 함. 체내 영상을 통해 신체의 외양이나 실루엣을 확인하거나 복원할 수 있는 경우에는 모자이크 처리, 마스킹, 가장자리 삭제 등의 조치를 이행해야 함[13]
- 셋째, 가이드라인에서 가명처리 가능 여부를 유보한 것임. 정형화되지 않은 자유입력 정보, 음성정보, 일부 예외를 제외한 유전체 정보, 지문 등 생체인식정보 등이 이에 해당함. 가이드라인은 이와 같은 속성값들에 대해서 적절한 가명처리 방법이 아직 개발되어 있지 않은 것으로 판단하여 안전한 가명처리 방법이 있을 때까지는 본인의 동의를 얻어서 사용하도록 정하고 있음.[14] 한편 가이드라인에서 제시한 방식 이외의 신기술 등 다른 방법이나 이를 채용한 소프트웨어 등을 활용하려 가명처리를 할 때에는 적절성 · 효과성 · 안전성 등을 외부 전문가에게 평가받은 뒤 데이터 심의위원회[15] 승인을 얻도록 하고 있음

D. 가명처리의 기술적 수단과 '추가 정보'의 의미

- 데이터셋에서 직접 식별자를 삭제할 수 없는 상황도 존재함. 가령 패널 데이터를 가명처리하여 연구자에게 제공한 다음, 이듬해 데이터를 업데이트하려면 기존 데이터와 새 데이터를 연결 지을 방법이 필요함. 또는 가명처리된 데이터를 다른 데이터와 결합하는 경우에도 필요함
- 위와 같은 상황에서는 데이터셋에서 개인을 유일하게 식별할 수 있는 필드(식별자)를 남겨둘 필요가 있으나 식별자를 원본 그대로 유지할 수 없기 때문에 이를 암호화하는 것이 일반적임
- 가명처리에 활용할 수 있는 암호화 기법에는 일방향 암호화와 쌍방향 암호화 2가지가 있음
- 일방향 암호화는 수학적 기법을 활용하여 원본값을 복원될 수 없는 다른 일련의 숫자 값으로 변형하는 방법으로, 널리 활용되는 기법은 해시(hash)임. 이 방법의 단점은 추후 원본값으로 복원해야 할 경우를 대비해서 원본값과 암호화된 값을 대응시키는 표(매핑 테이블, mapping table)를 유지해야 하는 것임
- 쌍방향 암호화는 암호 키를 이용하여 암호화함으로써 암호 키를 이용하면 원본값으로 복원할 수 있는 방법임. 이를 이용하면 매핑 테이블을 유지할 필요 없이 암호 키만 있으면 원본값을 복원할 수 있으나 암호 키를 안전하게 보관하여야 한다는 부담이 발생하게 됨
- 암호화를 하면 긴 숫자열이 생성되어 원본 필드와는 형태가 달라지게 되어 IT 부서에서 소프

13)개인정보보호위원회 · 보건복지부, 보건의료 데이터 활용 가이드라인(2020) 14-15면.
14)개인정보보호위원회 · 보건복지부, 보건의료 데이터 활용 가이드라인(2020) 13-16면.
15)데이터 심의위원회는 위 가이드라인이 정하고 있는 특징적 제도로서, 가명정보의 기관 내 활용, 기관외 제공, 결합신청, 가명처리 적정성 검토 등을 실시할 수 있는 독립 위원회를 의미한다. 개인정보보호위원회 · 보건복지부, 보건의료 데이터 활용 가이드라인(2020) 10면 참조.
16)https://en.wikipedia.org/wiki/Format-preserving_encryption 참조. 형태 보존 암호화는 동형 암호화(homomorphic encryption)와는 전혀 다른 것이다. 동형 암호화는 데이터를 암호화된 상태에서 연산할 수 있도록 하는 방법이다.

트웨어 개발 목적으로 가명정보를 활용하는 경우 데이터의 포맷이 달라지는 불편함이 발생할 수 있음. 이러한 문제를 해결하기 위해 형태 보존(Format-Preserving Encryption) 암호화 기법을 사용할 수 있음[16]

- 일방향 암호화와 쌍방향 암호화 모두 가명처리 후 원래 정보로 복원하기 위한 추가적인 정보가 남아 있게 되는데, 일방향 암호화(해시함수)에서의 매핑 테이블, 쌍방향 암호화에서의 암호 키가 이에 해당함. 이처럼 원래 정보로 복원하기 위한 추가적인 정보를 흔히 '가명처리 비밀(pseudonymisation secret)'이라 함
- 기술적 의미에서의 가명처리는 직접 식별자를 다른 값(즉, 가명)으로 대체하는 것만을 의미하며, 이 글에서 논의하는 법적인 의미의 가명처리와는 구분되는 점을 유의해야 함[17]
- 법적인 의미의 가명처리는 단지 식별자를 대체하는 것과 직접 식별자는 삭제 또는 일련번호로 대체하고, 속성값 중 개인식별 위험성이 높은 정보(특이정보 포함)는 일반화(범주화)하거나 삭제하는 등의 조치를 포함하는 것을 의미함

E. 가명정보의 개념과 익명정보와의 차이

- 가명정보는 "가명처리함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용 · 결합 없이는 특정 개인을 알아볼 수 없는 정보"로, 핵심적 판단 지표는 '원래 상태로 복원하기 위한 추가 정보의 사용 · 결합 없이는 특정 개인을 알아볼 수 없다'는 것임
- 추가 정보는 원본 데이터 그 자체이거나 또는 앞서 설명한 가명처리 비밀[18] 두 가지로 구분됨
- 가명정보를 활용하는 연구자가 추가 정보(즉, 원본 데이터나 가명처리 비밀)에 접근할 수 없도록 해야 함. 이처럼 연구자가 추가 정보에 접근할 수 없는 상황에서 개인을 식별할 수 없도록 데이터를 변형하는 것이 가명처리이고, 그와 같은 결과물이 가명정보임
- 가명정보는 추가 정보가 있으면 개인을 알아볼 수 있는 정보이기에 가명정보도 여전히 개인정보에 해당함. 하지만 가명정보는 가명처리를 통해 개인을 식별할 위험성을 크게 낮춘 것임. 이에 관해 개인정보 보호법은 가명정보 처리에 관한 특례를 두고 있음
- '익명정보'는 가명정보와 구별되는 개념으로, 2020년 개정 개인정보 보호법 제58조의 2는 "이 법은 시간 · 비용 · 기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보에는 적용하지 아니한다."는 규정을 신설함
  - "시간 · 비용 · 기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보"를 흔히 '익명정보'라 부름

17) 이러한 개념상 차이는 개인정보 보호법상 가명처리의 개념이 도입되기 이전에 제정된 개인정보 비식별 조치 가이드라인(2016)의 설명에서도 확인된다. 위 가이드라인은 가명처리를 "개인 식별이 가능한데이터를 직접적으로 식별할 수 없는 다른 값으로 대체하는 기법"이라고 설명하고 있는데, 이는 기술적 의미의 가명처리를 지칭하는 것으로 이해될 수 있다.

18) 가명처리 비밀에는 가명처리에 사용된 알고리즘이 무엇인지에 관한 정보도 포함될 수 있다. 약한 암호화 기법이 활용된 경우에는, 공격자가 어떤 필드에 대해 어떤 해시함수를 적용했는지 알고 있다면원래 값을 찾아낼 수 있는 상황이 발생할 수 있다. 이러한 위험성을 평가하여 강력한 암호화 기법을 사용하거나 어떠한 알고리즘을 사용했는지에 관한 정보도 안전하게 보관할 필요가 있다.

- 익명정보에 대해서는 개인정보 보호법이 적용되지 않는다고 명시하고 있으므로 익명정보는 개인정보가 아님

• 가명정보와 익명정보 모두 “개인을 알아볼 수 없게” 한 것이 공통점
  - 가명정보는 그 목적 활용이 통계작성, 과학적 연구, 공익적 기록보존 등으로 제한되어 있음. 따라서 가명정보는 해당 목적을 수행하기 위한 이들에게 한정적으로 제공되고, 그 목적으로만 사용됨
  - 익명정보는 개인식별 가능성이 없는 정보를 의미함. 이는 활용 목적의 제한이 없고, 특히 익명정보는 데이터를 공중에 공개하는 상황에서 활용 가능함. 따라서 개인정보를 익명정보로 변형하기 위해서는 더욱 강력하고, 엄격한 비식별 조치가 취해져야 함

III. 가명정보의 활용과 결합 :
가명정보의 활용과 과학적 연구

A. 가명정보 활용의 장점

- 가명정보의 활용은 중요한 장점이 있음
  - 첫째, 가명처리는 정보주체의 프라이버시를 보호하는 수단
  - 둘째, 정보주체의 동의를 얻지 않고 활용하거나 제3자에게 제공할 수 있다는 것. 다만, 만약 정보주체의 동의를 얻는 것이 어렵지 않다면 정보주체의 동의를 얻어 사용하거나 제3자에게 제공하는 것을 우선적으로 고려할 필요가 있음

B. 과학적 연구의 범위

- 가명정보를 정보주체의 동의 없이 활용할 수 있는 것은 통계작성, 과학적 연구, 공익적 기록보존 등의 목적을 달성하기 위한 경우로 제한됨[19)]
- 이러한 목적 범위에 어디까지 포함되는가가 쟁점이며, 특히 '과학적 연구'의 범위에 대한 것이 문제가 됨
- 가명정보의 활용 범위를 좁게 보아야 한다는 입장에서는 '과학적 연구'의 범위에서 상업적 · 산업상 연구를 제외시켜야 한다는 주장을 제기하기도 함[20)]
- 개인정보 보호법 제2조 제8호는 '과학적 연구'를 "기술의 개발과 실증, 기초연구, 응용연구 및 민간 투자 연구 등 과학적 방법을 적용하는 연구"라고 정의하고 있음. 따라서 현행 법조문의 해석상 상업적 연구라고 해서 과학적 연구에 포함되지 않는다고 해석하기 어려움. 또한 현실적으로 대학, 공공연구기관과 기업들 간에 긴밀한 협업 연구가 이루어지고 있는 상황에서 어떠한 연구가 상업적인지 아닌지 나누는 것은 어려움
- 보건의료 데이터 활용 가이드라인(2020)도 과학적 연구는 "새로운 기술 · 제품 · 서비스의 연구개발 및 개선 등 산업적 목적의 연구를 포함"하는 것이라고 설명하고 있음[21)]

C. 가명정보의 결합

- 가명정보의 또 다른 주요한 활용 가능성은 다른 정보와 결합하여 활용되는 것임. 서로 다른 시점에 여러 기관에서 다양한 보건의료 데이터가 수집되는데, 여러 출처로부터 얻어진 보건의료 데이터를 결합함으로써 유용한 분석을 수행할 수 있음[22)]

---

19)개인정보 보호법 제28조의2 제1항.
20)진보네트워크센터 외, "가명처리 가이드라인에 대한 시민사회 의견", (2020. 8. 27), https://act.jinbo.net/wp/43320/.
21)개인정보보호위원회 · 보건복지부, 보건의료 데이터 활용 가이드라인(2020) 5면.
22)El Emam, K. & Arbuckle L., "Anonymizing Health Data", O'reilly, (2014).

- 서로 다른 데이터셋들을 결합하기 위해서는 양자를 연결시킬 수 있는 식별자가 필요한데, 식별자는 단일한 필드일 수도 있고, 여러 필드를 합친 것일 수도 있음
- 가명정보에 대하여 데이터를 결합하는 방법은 두 데이터셋에 대해 식별자를 암호화한 다음, 암호화한 식별자를 서로 대조하여 일치하는 항목을 찾는 것으로 간단함
- 두 데이터셋을 결합하려면 데이터셋 둘 다를 확보하고 양자를 비교해야 하는데, 안전하게 데이터를 결합하기 위해서는 '신뢰할 수 있는 제3자(Trusted Third Party)'를 활용할 수 있음
- 2020년 개정 개인정보 보호법은 개인정보보호위원회나 관계 중앙행정기관의 장이 지정하는 '전문기관'이 결합을 수행하도록 정하였음.[23] 보건복지부는 2020년 10월 건강보험심사평가원, 국민건강보험공단, 한국보건산업진흥원 3곳을 보건의료 분야 결합전문기관으로 지정하였음
- 데이터 결합과 관련하여 유의할 사항은 데이터를 결합한 결과 개인식별 위험성이 증가하게 됨. 따라서 결합된 데이터는 결합전문기관 내에 설치된 별도의 공간에서 반출을 위한 추가적인 가명처리 또는 익명처리가 이루어져야 함.[24] 그 후 결합전문기관에 반출 신청을 하고 승인을 얻어야 함[25]
- 가명정보의 결합 및 반출 절차가 지나치게 복잡하다는 비판이 제기되지만 앞으로 실무상 어떻게 운용될 것인지 지켜봐야 함

23)개인정보 보호법 제28조의3 제1항.
24)개인정보 보호법 시행령 제29조의3 제3항.
25)개인정보 보호법 제28조의3 제2항.

IV. 결론

- 개인정보 보호법상의 가명처리와 가명정보의 개념을 보건의료 정보를 중심으로 개괄적으로 살펴보았음
- 매사추세츠 주지사의 의료정보 재식별 사건은 1997년의 일로 비식별 조치의 중요성에 대하여 국제적으로 관심을 기울인 지는 20년 남짓밖에 되지 않음. 현재 전 세계적으로 여러 가지 아이디어가 제안되고 입법적 시도가 이루고 지고 있는 단계임
- 보건의료 데이터의 가명처리나 익명처리는 두 마리 토끼를 모두 쫓는 일임. 보건의료 데이터로부터 환자의 신원이 드러나지 않도록 보장하면서 데이터를 활용하여 보건의료 분야의 연구를 활성화할 필요도 있음. 따라서 양자를 적절히 저울질하여 좋은 균형점을 찾아야 함
- 앞으로 구체적 적용 사례가 차츰 쌓이면서 좋은 실무 관행이 정착될 것으로 기대함[26)]

26)이 글에서 상세히 다루지 못한 세부 사항이 적지 않으므로, 실무상 가명처리를 직접 적용하고자 하는 분들은 여러 가이드라인과 해설서의 구체적 사항을 다시금 챙겨 볼 것을 당부드린다.

참고 문헌

참고문헌

- 개인정보보호위원회, “가명정보 처리 가이드라인”, (2020)
- 개인정보보호위원회 · 보건복지부, “보건의료 데이터 활용 가이드라인”, (2020)
- 고학수 · 최경진, “개인정보의 비식별화 처리가 개인정보 보호에 미치는 영향에 관한 연구”, 개인정보보호위원회 연구보고서, (2015)
- 고학수 외, “개인정보 비식별화 방법론: 보건의료정보를 중심으로”, 박영사, (2017)
- 고학수 · 구본효 · 정종구, “가명정보의 ‘과학적 연구’ 목적을 위한 활용”, 서울대학교 인공지능정책 이니셔브 이슈 페이퍼, (2020)
- 고학수 · 백대열 · 구본효 · 정종구 · 김은수, “개정 개인정보 보호법상 가명정보의 개념 및 가명처리에 관하여”, 서울대학교 인공지능정책 이니셔티브 이슈페이퍼, (2020)
- 관계부처 합동, “개인정보 비식별 조치 가이드라인”, (2016)
- 김은수, “비식별화 방식을 적용한 개인정보보호에 대한 연구”, 서울대학교 박사학위 논문, (2018)
- 이동진, “개인정보 보호법 제18조 제2항 제4호, 비식별화, 비재산적 손해 - 이른바 약학정보원 사건을 계기로 -”, 정보법학 제21권 제3호, (2017)
- 이원복, “유전체 연구와 개정 개인정보 보호법의 가명처리 제도”, 법학논집 제25권 제1호, (2020)
- 이재훈, “데이터 3법 개정에 따른 바이오 · 의료정보 활용방향과 시사점”, BioINpro Vol. 71, 생명공학정책연구센터, (2020)

외국문헌

- Arbuckle, L. & El Emam, K. “Building an Anonymization Pipeline”, O’Reilly, (2020)
- El Emam, K. & Arbuckle L., “Anonymizing Health Data”, O’reilly, (2014)
- European Union Agency for Cybersecurity (ENISA), “Recommendations on shaping technology according to GDPR provisions – An overview on data pseudonymisation”, (2018)
- El Emam, K. & Arbuckle L., “Pseudonymisation techniques and best practices”, (2019)
- Finck, M., & Pallas, F., “They who must not be identified-distinguishing personal from non-personal data under the GDPR”, International Data Privacy Law, (2020)
- Garfinkel, Simon L., National Institute of Standards and Technology(US), “De-Identification of Personally Identifiable Information”, (2015)
- Hintze, M., “Viewing the GDPR through a de-identification lens: A tool for ompliance, clarification, and consistency”, International Data Privacy Law, 8(1), (2018)
- Hintze, M., & El Emam, K., “Comparing the benefits of pseudonymisation and anonymisation under the GDPR”, Journal of Data Protection and Privacy, 2(2), (2018)
- Information Commissioner’s Office, “Anonymisation: Managing Data Protection Risk Code of Practice”, (2012)
- Ohm, Paul, “Broken promises of privacy: Responding to the surprising failure of anonymization”, UCLA l. Rev. 57 (2009)
- Sweeney, L., “k-Anonymity: A Model for Protecting Privacy”, International Journal on Uncertainty, Fuzziness and Knowledge-based System, (2002)