의료정보정책 공개포럼

개인(의료)정보의 보호와 활용 :
4차산업혁명위원회 해커톤에서의 논의를 중심으로

고학수 교수
서울대학교 법학전문대학원

CONTENTS

논의의 법적 배경

개인정보 수집에 대한 규제

- 개인정보보호 법제도의 기본 원칙
  1. '개인정보'의 개념 규정
  2. 개인정보의 수집, 이용에 관하여 고지 및 동의 요건이 적용됨
- 개인정보
  - "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 알아볼수 있는 것을 포함한다)" (개인정보보호법 제2조)
- 비식별화된 개인정보?
  - 비식별화된 개인정보는 개인정보?
    - 개인정보가 아니라면 더 이상 '개인정보'로서 별도의 법적 규제대상이 아니라 할 것인가?

개인정보 수집, 이용 등에 대한 규제

- 비식별화된 개인정보?
  - 개인정보 보호법 제18조(개인정보의 목적 외 이용・제공 제한)
    - 제2항 : "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"
      - "4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우"
  - 개인정보 보호법 제18조 제2항 4호의 해석?
    - 비식별화? (제2조의 "알아볼"과 같은 뜻?)
    - 가명처리?

국내 실정법상 비식별화/익명화 개념

- 국내 실정법에 비식별화, 익명화 개념이 명시적으로 제시된 경우는 찾기 어려움
- 「생명윤리 및 안전에 관한 법률」
  - "익명화"(匿名化)란 개인식별정보를 영구적으로 삭제하거나, 개인식별정보의 전부 또는 일부를 해당 기관의 고유식별기호로 대체하는 것을 말한다. (제2조(정의))
  - "개인식별정보"란 연구대상자와 배아·난자·정자 또는 인체유래물의 기증자(이하 "연구대상자등"이라 한다)의 성명·주민등록번호 등 개인을 식별할 수 있는 정보를 말한다. (제2조(정의))

국내 논의동향 : 4차산업혁명위 해커톤(2018)

1. 개인정보 관련 법적 개념체계 정비
   - 개인정보와 관련된 법적 개념체계는 개인정보, 가명정보, 익명정보로 구분하여 정비하기로 하였다. 그리고 익명정보는 개인정보보호법의 적용대상이 아니라고 합의하여 개인정보와 구분하였다.
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.
3. '가명정보'에 대한 법적 근거 마련
   - '가명정보'의 정의 및 활용에 관한 법적 근거를 마련하기로 하였다.
4. 개인정보의 보호와 활용에 대한 지속적 논의 진행
   - 개인정보 보호와 활용에 관한 주요 이슈들에 대해서 추가적인 논의를 진행하기로 하였다.

식별, 비식별화, 재식별...

익명화/비식별화의 경우

- 개인정보 비식별화 방법론?
  - 다양한 가명처리 방법들
    - 가명처리(Pseudonymisation)
    - 총계처리(Aggregation); 마스킹(Data masking); 데이터 값 삭제(Data reduction); 범주화(Data suppression) 등등

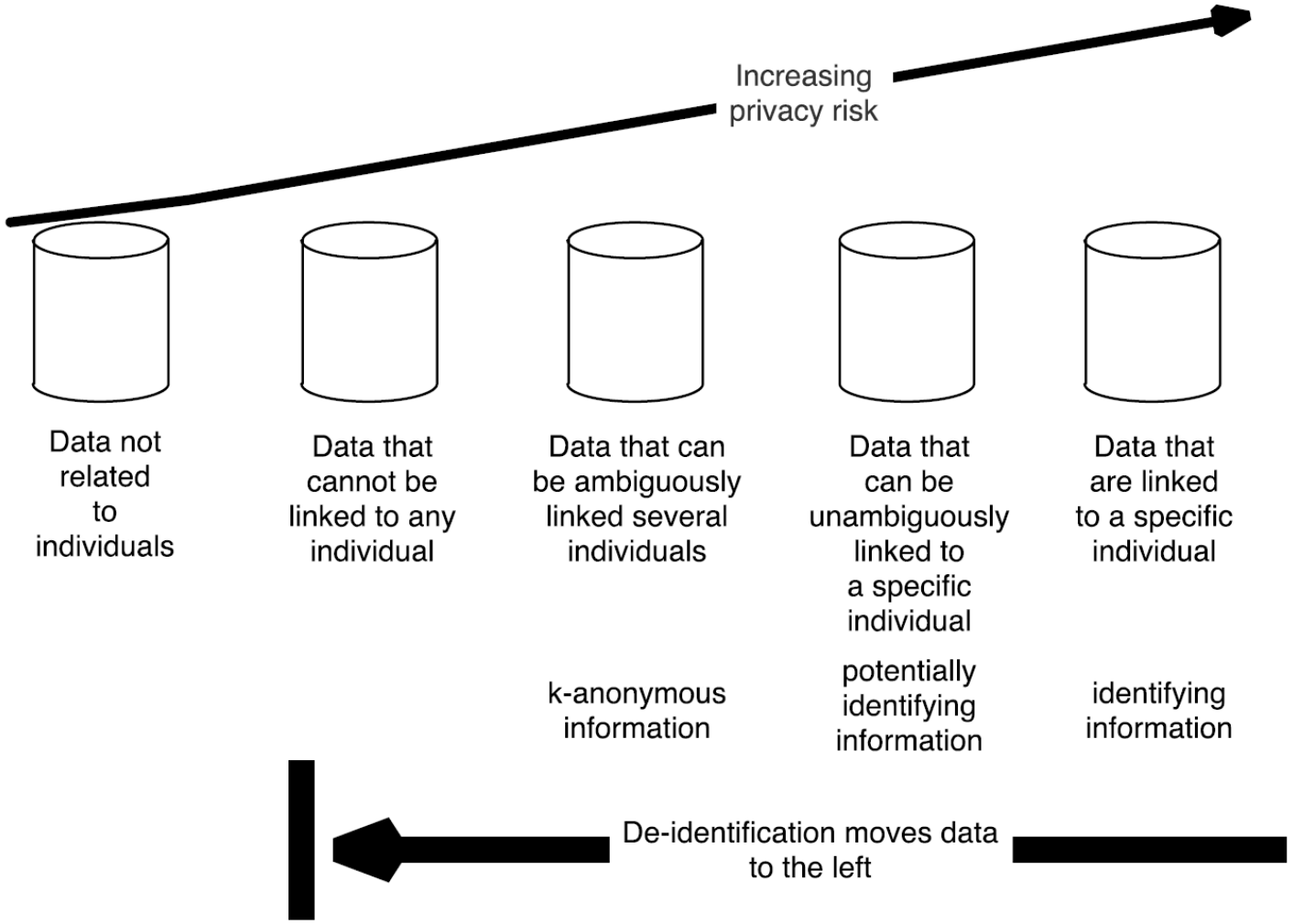
익명화/비식별화가 해답? 재식별 가능성

- 비식별화를 한 이후에 남아있을 수 있는 재식별 가능성?
  - 많이 알려진 해외 재식별 사례
    - 메사추세츠 주지사 Governor Weld 사례
    - AOL: Thelma Arnold (Id No. 4417749) 식별사례
    - Netflix 영화추천 알고리듬 사례
  - 재식별 가능성은 여러 요소에 의해 영향을 받음
    - 데이터의 내용 및 구조, 공개의 형태, 공개된 데이터의 익명화 정도, 데이터에 대한 접근(access)을 둘러싼 사후적 관리, 준식별자(quasi-identifier)의 존재 및 관리, 공격자(adversary)의 현실적 존재가능성 및 기술적 능력 등등
    - → 데이터 환경이나 맥락(context)에 대한 고려가 중요

익명화/비식별화가 해답? 재식별 가능성

- 재식별 가능성 평가를 위한 기법
  - k-익명성(k-anonymity), l-다양성(l-diversity), t-근접성(t-closeness) 등의 통계학적 개념을 통하여 익명화 수준을 평가하고 재식별의 가능성을 평가하는 기법들이 근래에 개발됨
    - k-익명성 : 주어진 데이터 집합에서 준식별자의 속성값들이 동일한 레코드가 적어도 k 개 존재하는 것
  - 또한 차등 프라이버시(differential privacy) 개념을 통하여 질의에 기반한(query-based) 익명화 기법도 논의, 개발중

익명화/비식별화가 해답? 재식별 가능성



NIST (2015), “De-Identification of Personal Information” (NISTIR 8053), p.6, Figure 1

개인정보 비식별화 관련 해외 논의
- 미국 및 영국 사례 등

해외의 논의동향 [미국 HIPAA]

- HIPAA (Health Insurance Portability and Accountability Act) Privacy Rule
  - 비식별화된 개인건강정보(De-identified personal health information):
    - 개인식별 가능한 건강정보가 아님("Not individually identifiable health information")

      → 별도 동의 없이 수집, 이용 가능 (C.F.R. § 164.502(d)(2))
  - 비식별화는 크게 두가지 방법으로 가능
    1. 통계적 방법 [Statistical Standard] 이 충족된 경우 (C.F.R. § 164.502(b)(1)), 또는
    2. 세이프 하버 [Safe Harbor]에 따라 18개 항목이 제거된 정보 (C.F.R. § 164.502(b)(2))

해외의 논의동향 [미국 HIPAA]

- HIPAA (Health Insurance Portability and Accountability Act) Privacy Rule
  - 통계적 방법 [Statistical Standard]
    1. 전문가에 의한 결정 : someone with “appropriate knowledge of and experience with generally accepted statistical and scientific principles” determines
    2. 판단기준 : 매우 낮은 재식별 확률 “risk is very small.”
    3. 판단 방법과 결과의 기록 : the methods and results of the analysis that justifies this determination
  - 세이프 하버 [Safe Harbor] : 2가지 요건
    1. 18개 항목의 식별자(identifying data elements) 제거
    2. 재식별 상황 부재 확인 : “no actual knowledge”

해외의 논의동향 [미국 HIPAA]

- 전문가에 의한 방법 : 재식별의 위험성 판단 기준:
  - HIPAA Privacy Rule § 164.514(b)

> (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information;

  - "매우 작은 위험" : 특정한 수치나 기계적인 기준이 있는 것은 아님. 개별 상황에 따라 데이터를 식별화하려는 '예상주체(anticipated recipient)'의 능력을 고려하여 판단
  - 위험성 평가를 위해 기본적으로 적용해야 할 기본적인 고려사항
    - 반복성(replicability) : 동일한 결과값의 재현 가능성
    - 데이터 소소에의 접근가능성(data source availability)
    - 구별성(distinguishability) : 특정인의 결과값이 타인의 결과값과 구분되는 정도
    - 연결가능성(linkability)

해외의 논의동향 [미국 HIPAA]

- HIPAA (Health Insurance Portability and Accountability Act) Privacy Rule

  [Safe Harbor] 18개 항목이 제거된 정보 :

1. 이름
2. '주' 보다 작은 행정구역 (우편번호는 앞 3자리; 20,000 거주자 미만일때는 우편번호 모두 제거)
3. 대부분의 날짜 (연도는 가능)
4. 전화번호
5. 팩스번호
6. 이메일 주소
7. 사회보장번호 (SSN)
8. 의료기록번호(medical record numbers)
9. 의료보험가입자번호(health plan beneficiary numbers)
10. 계좌번호(account numbers)
11. 증명서/면허증 번호(certificate/license numbers)
12. 자동차 번호판 등(vehicle identifiers, such as serial numbers and license plate numbers)
13. 기기고유번호(device identifiers and serial numbers)
14. URLs
15. IP 주소
16. 생체정보(biometric identifiers)
17. 얼굴사진(full face photos)
18. 기타 번호 등(any other unique identifying number, characteristic or code

해외의 논의동향 : 영국

- 보건의료정보에 대한 활발한 논의
  - National Health Service(NHS) 시스템을 통한 건강서비스 제공
  - NHS Digital(구 HSCIC)을 중심으로 건강정보의 수집과 활용이 이루어지는 체계 구축
- 주요 보고서

[1] ICO(영국 개인정보보호기구)의 익명화 가이드라인 (2012)

- 잠재적 공격자의 관점을 적용한 "motivated intruder test"의 개념

[2] UKAN(영국 익명화네트워크)의 익명화 보고서 (2016)

- 맥락의존성을 강조하는 데이터환경적 접근법(data situation approach) 강조

[3] Caldicott 보고서

- 1997, 2013, 2016의 세 차례 보고서 발표
- 2013년 & 2016년 보고서에서 데이터의 활용가능성 강조

해외의 논의동향 [영국 : ICO Code of Conduct]

- 재식별의 위험성에 대한 판단 기준의 전제
  - 허용되는 익명화의 수준은 재식별 위험성이 영(zero)인 수준은 아니라는 전제하에 논의
  - 재식별의 위험성 판단 기준으로 '합리적 가능성(reasonably likely test)' 기준 채택
- '의도적 공격자(motivated intruder)' 기준
  - 이 기준은 재식별의 가능성이 무한대로 확장되는 것을 제한하는 기능을 함
    - '공격자'는 아무런 사전적 지식이 없는 상태
    - 합리적 수준의 능력을 가지고 인터넷, 도서관 등에서 정보에 접근해서 탐색할 수 있는 수준을 상정
      - 따라서 해당 영역에 대한 전문가적 지식이나 해킹 능력을 포함한 고도의 기술적 능력을 갖춘 공격자에 의한 재식별의 위험성은 현실적으로 용인

해외의 논의동향 : EU

- 법개정
  - 개인정보보호지침(DPD: Data Protection Directive)
  - → 개인정보보호 일반규정(GDPR: General Data Protection Regulation, 2016 입법, 2018 시행)
- DPD나 GDPR에 비식별화/익명화에 대한 명시적인 조항은 없음
  - GDPR에 '가명화(pseudonymization)' 관련 규정은 다수
  - 개인정보의 안전한 처리를 위하여 가명화 권고

가명처리 (Pseudonimization) [EU GDPR]

- 가명처리 vs. 익명화(pseudonimization vs. anonymisation)
  - 가명화된 정보는 법적으로 개인정보
    - 익명화된 정보는 개인정보 아님
  - 익명정보 (GDPR 전문(recital) (26))
    - 이 원칙은 식별되었거나 또는 식별될 수 있는 개인과 관련되지 않는 정보 또는 그런 방식으로 익명처리되어 더 이상 식별될 수 없는 정보주체에는 적용되지 않는다. 따라서 이 법은 통계목적 및 연구 목적 등을 위한 익명정보의 처리에는 적용되지 않는다.
      - (The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.)
    - 다른 한편, 비식별화/익명화 방법 등에 대해서는 명시적 규정은 없음

가명처리 (Pseudonimization) [EU GDPR]

- 가명처리의 개념 (GDPR제4조 (정의) (5))
  - ‘추가정보’ 없이는 개인정보를 특정 정보주체에 귀속되는 것인지 파악할 수 없도록 하는 것
  - 추가정보에 대해서는 (1) 별도 보관이 필요, (2) 기술적, 관리적 조치가 필요
- 가명처리란 추가적으로 정보를 사용하지 않고서는 더 이상 특정 정보주체를 알아볼 수 없도록 개인정보를 처리하는 것이다. 이러한 추가 정보는 별도로 보관되어야 하고 해당 개인정보가 자연인을 식별하거나 식별할 수 없도록 하기 위해서 기술적·조직적인 조치가 적용되어야 한다
  - (“‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person”)

가명처리 (Pseudonimization) [EU GDPR]

- 가명처리의 의의/유인
  1. 제89조의 활용 가능
     - (1) 공익을 위한 유지보존의 목적, (2) 과학이나 역사적 연구의 목적 또는 (3) 통계 목적에서의 개인정보 처리
       - (Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes)
     - "안전장치" (safeguards) 및 가명처리를 전제로, 정보주체의 권리에 대한 일정한 제한
       - 안전장치 : 데이터 최소화를 포함한, 기술적 관리적 조치의 적용
       - 제15조(열람권), 제16조(수정권), 제18조(처리에 대한 제한권), 제21조(반대할 권리) 등의 제한
       - "subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation"

가명처리 (Pseudonimization) [EU GDPR]

- 가명처리의 의의/유인
  2. 제6조 제4항에 따른 양립가능성(compatibility) 고려시 기준중 하나
     - 목적외 이용의 맥락에서, 애초에 정보가 수집된 목적과의 양립가능성 판단
     - 5가지 예시된 판단기준
       - 목적 사이의 링크 (“link between the purposes”)
       - 맥락 (“context in which the personal data have been collected”)
       - 개인정보의 본질 (“nature of the personal data”)
       - 가능성 있는 영향 (“possible consequences of the intended further processing for data subjects”)
       - “암호처리(encryption) 및 가명처리(pseudonymisation) 등 적절한 보호수단의 유무” (“the existence of appropriate safeguards, which may include encryption or pseudonymisation”)

Data Governance Practice (El Emam & Malin, 2015)

- 기록 마련
  - Developing and maintaining global anonymization documentation
- 정보 제공/공개 절차 마련
  - Process and tools for tracking all data releases
- 정보이용 기간만료 관리
  - Process and tools for triggering alerts for data use expirations
- 정보 제공/공개 건별 기록 관리
  - Ensuring that documentation for the de-identification for each data release is complete and indexed
- 재식별 가상공격 주기적 실행
  - On occasion, commissioning controlled re-identification attacks
- 감사
  - Implementing audit process
- 윤리위원회
  - Ensuring that there is ethics review that covers protections against attribute disclosure

개인정보 비식별화 관련 국내 논의
- 4차 산업혁명위원회 해커톤에서의 논의

국내 논의동향 : 4차산업혁명위 해커톤(2018)

1. 가명정보의 활용과 보호

- 가명정보의 활용 목적과 범위
  - 가명정보는 ① 공익을 위한 기록 보존의 목적, ② [학술 연구 / 학술 및 연구]* 목적, ③ 통계 목적을 위하여 당초 수집 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다.
    - [학술 연구 / 학술 및 연구] : 연구의 범위에 관하여 이견이 있어 참석자 일부는 '학술 연구'라는 표현을, 다른 일부는 '학술 및 연구'라는 표현을 지지하였다.
- 그리고 이를 위해서는 가명처리를 포함한 기술적, 관리적 조치 등 안전조치가 취해져야 한다.
- 위 [학술 연구 / 학술 및 연구] 목적에는 산업적 연구 목적이 포함될 수 있고, 통계 목적에는 상업적 목적이 포함될 수 있다.

국내 논의동향 : 4차산업혁명위 해커톤(2018)

1. 가명정보의 활용과 보호

- 가명정보의 활용 목적과 범위 : 학술 연구 / 학술 및 연구?
  - Where personal data are processed for scientific research purposes, this Regulation should also apply to that processing. For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research. (GDPR Recital 159)
- 가명정보의 활용 목적과 범위 : 통계목적?
  - [1] Statistical purposes mean any operation of collection and the processing of personal data necessary for statistical surveys or for the production of statistical results. Those statistical results may further be used for different purposes, including a scientific research purpose. [2] The statistical purpose implies that the result of processing for statistical purposes is not personal data, but aggregate data, and that [3] this result or the personal data are not used in support of measures or decisions regarding any particular natural person. (GDPR Recital 162)

국내 논의동향 : 4차산업혁명위 해커톤(2018)

1. 가명정보의 활용과 보호

- 최초 수집목적과 양립되는 추가적인 개인정보 처리
  - 정부는 유럽연합 일반개인정보보호법(EU GDPR) 등 해외 입법례를 참조하여, 가명처리 여부 등 여러 사정을 고려하여 개인정보를 당초 수집한 목적과 상충되지 아니하는 목적으로 활용할 수 있도록 하는 제도를 마련한다.

국내 논의동향 : 4차산업혁명위 해커톤(2018)

2. 익명처리의 절차, 기준, 평가, 등

- 정부는 익명처리의 적정성을 평가하기 위한 절차와 기준을 마련할 수 있고, 이러한 절차와 기준은 기술적 중립성에 입각한 것이어야 하며, 강제적인 것이거나 최종적인 것으로 해석되어서는 아니된다
- 그리고, 정부는 적정성 평가를 위해 정보의 속성과 산업별 특성을 반영하여 신뢰할 수 있는 제3의 기관(Trusted Third Party)이나 전문가를 활용하는 등 다양한 제도적 장치를 마련할 수 있다

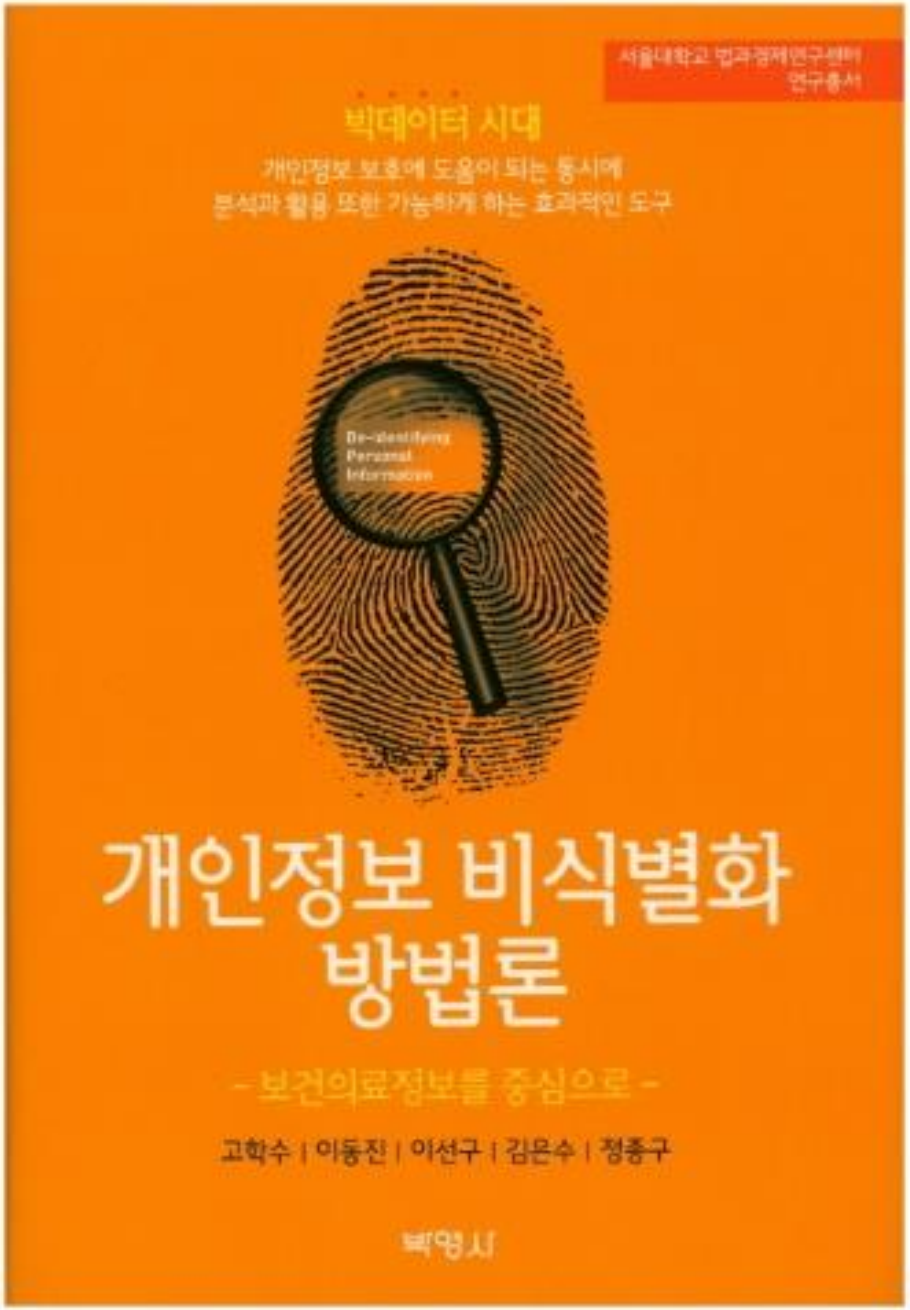
국내 논의동향 : 4차산업혁명위 해커톤(2018)

3. 데이터 결합

- 데이터 결합은 사회적 후생을 증진하는 중요한 역할을 할 수 있으나 그 과정에서 발생할 수 있는 개인정보 침해의 위험성도 간과되어서는 안된다.
- 그리고 정부는 데이터 결합의 법적 구성방식들을 구체화하고 개인정보 침해 위험에 비례하여 사전적 또는 사후적 통제방안을 마련하도록 노력해야 한다.
- 데이터 결합과 관련한 구체적인 방안에 대해서는 시민단체와 산업계가 서로 다른 의견을 제시하여 합의에 이르지 못하였다.

의료영역의 과제

- 데이터 유형별 비식별 방법론 모색
  - 영상정보, 유전자 정보 등
  - Genetic privacy에 관한 개념 정리 및 필요한 안전장치 정립
- 데이터 표준화, 상호운용성 확보
  - EMR 등
- 데이터 거버넌스 체계를 고려한 비식별화 제도의 정비
  - 건강보험 시스템, 대형 및 중소형 병원, 제약회사, IT 기업 등등 국내의 데이터 '생태계' 고려
    - 미국의 HIPAA Privacy Rule은 미국식 의료체계(HMO)를 전제로 한 것임
    - 영국식 정보공유 체계는 NHS를 통한 의료제공 시스템이 배경임


서울대학교 법과경제연구센터
연구총서
빅데이터 시대
개인정보 보호에 도움이 되는 동시에
분석과 활용 또한 가능하게 하는 효과적인 도구
De-identify
Personal
Information
개인정보 비식별화
방법론
- 보건의료정보를 중심으로 -
고학수 | 이동진 | 이선구 | 김은수 | 정종구
박영사

감사합니다!