

가명정보의 개념과 가명처리의 쟁점

고려대학교 정보보호대학원

김 법 연

목차

I. 가명정보의 개념과 도입배경

II. 가명정보 및 가명처리의 쟁점과 이슈

III. 가명정보의 활용과 가명처리에 있어 고려사항

I. 가명정보의 개념과 도입배경

가명정보의 개념과 의미

개인정보, 가명정보, 가명처리, 익명정보

- **가명정보:** 가명처리함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보(법 제2조 제1호 다목)
- **가명처리:** 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보 없이는 특정 개인을 알아볼 수 없도록 처리한 것(법 제2조 제1의2호)
- **익명정보:** 시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보(법 제58조의2)

구분	개념	활용가능 범위
개인정보	특정 개인에 관한 정보 개인을 알아볼 수 있게 하는 정보	수집목적과 합리적으로 연관된 범위 내에서 정보주체 동의없이 개인정보 추가 이용·제공 가능
가명정보	추가정보의 사용없이 특정 개인을 알아볼 수 없게 조치한 정보	다음 목적에 동의 없이 활용 가능 ① 통계작성 ② 연구 ③ 공익적 기록보존 목적 등
익명정보	다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없게 조치한 정보	개인정보가 아니기 때문에 제한없이 자유롭게 활용

가명정보의 개념과 의미

가명정보 처리에 관한 특례

- 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있음(법 제28조의2 제1항)
- 가명처리된 정보의 처리자는 '정보주체 이외로부터 수집한 개인정보의 수집 출처 등 고지 의무(법 제29조)', '보유기간 경과, 처리목적 달성에 따른 개인정보 파기의무(법 제21조)', '영업양도 등에 따른 개인정보의 이전 제한(법 제27조)', '개인정보 유출 통지 의무(법 제34조 제1항)'가 배제되며
- 개인정보가 가명화된 경우 해당 가명정보에 대해서 개인정보주체는 '열람권(법 제35조)', '정정·삭제권(법 제36조)', '처리정지권(법 제37조)'을 행사할 수 없음
- 정보통신서비스제공자 특례조항 중 유출시 감독기관에 신고, 휴면이용자 개인정보파기, 개인정보 이용내역 주기적 통지, 정보주체의 동의철회권 규정이 적용 제외됨

가명정보의 개념과 의미

가명정보의 보호

근거조항	대상정보	내용
법 제29조	가명정보 추가정보	<ol style="list-style-type: none"> 1. 내부 관리계획의 수립·시행 2. 접근통제 및 접근권한의 제한 조치 3. 안전한 저장·전송을 위한 암호화 기술 4. 침해사고대응을 위한 접속기록보관 및 위·변조 방지 5. 보안프로그램의 설치 및 갱신 6. 보관시설 or 잠금장치설치 등 물리적 조치
법 제29조의4 제1항	추가정보	별도 분리 보관 및 접근 권한 분리
	가명정보 추가정보	접근권한 관리 및 물리적·기술적 안전조치에 관한 내부 관리계획의 수립·시행
령 제29조의5 제3항	가명정보	처리목적이 달성되거나 보유기간 경과시 지체없이 파기
법 제28조의4 제2항	가명정보	처리시 목적, 처리 및 보유기간, 추가정보의 이용 및 파기내역 작성·보관

가명정보의 개념과 의미

개인정보, 가명정보, 익명정보 개념의 구조



개인정보

개인정보보호법 규율대상



가명정보

개인정보에 해당
단, 처리에 관한 특례 적용



익명정보

개인정보보호법 적용배제

가명정보의 도입배경

데이터 활용 수요 증가와 사전동의제의 한계

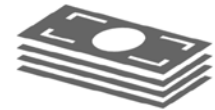
- 가명처리된 정보는 여전히 식별가능한 자연인에 관한 정보, 곧 개인정보로서 개인정보 규범의 적용범위에 편입되어 보호대상이 됨
- 그러나 다른 한편으로는 가명처리된 정보는 별도로 보관되고 안전조치가 적용된 것으로
 - ① 프라이버시 침해 위험을 감소시키고
 - ② 개인정보처리자의 의무이행에 도움이 되어 일정한 범위에서 활용하는 것이 허용
- 즉, 프라이버시 보호를 향상(강화)시키는 기술을 사용하여 침해 위험이 감소된 가명정보는 익명정보보다 가치가 있는 것으로 활용할 수 있도록 접근방법을 취하는 것



**Difficult to identify
the subject of big data**



**Difficult to receive
consent to access
an enormous quantity
of new data**



**The cost is
extremely high**

국가별 가명정보 및 익명정보의 개념

EU

Term	Definitions
<p>Pseudonymisation</p>	<p>『General Data Protection Regulation』, Article 4(5)</p> <ul style="list-style-type: none"> • 가명화(pseudonymisation)란 추가정보(additional information)를 이용하지 않고는 더 이상 특정 정보주체를 식별할 수 없는 방식으로 개인정보를 처리하는 것으로서 • 개인정보가 식별되거나 식별될 수 있는 자연인을 특정하지 않도록 그 추가적 정보를 별도로 보관하고, 기술적·관리적 안전조치를 취한 것으로 정의 • 가명처리는 별도로 보관되고 있고 기술적·관리적 안전조치가 적용되어 있는 추가적인 정보 없이는 특정 개인과 연계되지 않도록 데이터로부터 직접 식별자(direct identifier)를 분리시키는 과정으로 데이터를 익명화시키는 것에 해당하지 않으면서 개인정보주체를 직접 식별하는 것에 해당하지 않도록 처리하는 것
<p>Anonymisation</p>	<p>WP 29, 『Opinion on Anonymisation Techniques』</p> <ul style="list-style-type: none"> • 익명화(Anonymisation)란 현행 기술상 개인식별 수단으로는 원래의 개인정보를 알아볼 수 없는 상태로 만드는 것 • 익명화된 정보는 특정 개인과 연관지을 수 없거나 그러한 식별 가능성이 완전히 제거된 정보로서 GDPR의 규제대상이 아님 • 익명화를 위해서는 식별자가 삭제 또는 복원될 수 없는 형태로 치환되어야 하고, 식별자 이외의 나머지 정보의 경우에도 이를 통해 개인을 식별할 수 없을 정도로 개인의 고유 속성이 남아있지 않아야 함

국가별 가명정보 및 익명정보의 개념

미국

Term	Definitions
De-identified	<p>『캘리포니아 소비자 프라이버시보호법 California Consumer Privacy Act』, Article 1798.140(h)</p> <ul style="list-style-type: none"> • 비식별(De-identified)이란 특정정보가 특정 소비자를 합리적으로 식별하거나 관련시키거나 묘사하거나 결부될 수 없거나 또는 직간접적으로 연결되지 않으면 그 정보를 비식별 정보라고 함 • 비식별화는 재식별을 방지하는 기술적 보호조치(재식별 금지를 위한 기술적 안전장치의 구현 등), 관리적 보호조치(재식별을 금지하는 구체적 업무 프로세스의 구현 등) 등을 사용하는 것을 의미
	<p>『소비자 온라인 프라이버시권 법(Consumer Online Privacy Rights Act)』, S.2968 — 116th Congress</p> <ul style="list-style-type: none"> • 특정 개인, 가정, 또는 개인기기에 관한 정보를 알아내기 위해 합리적으로 사용될 수 없는 정보를 비식별 데이터(De-identified Data)라고 정의했고, 적용대상 데이터에서 제외
	<p>『온라인 프라이버시권 법(Online Privacy Act of 2019)』, H.R.4978 — 116th Congress</p> <ul style="list-style-type: none"> • 특정 개인 또는 장치를 합리적으로 식별하거나 관련시키거나 묘사하거나, 참조, 연관, 직간접적으로 연결될 수 없는 정보를 비식별 데이터(De-identified Data)라고 정의함
Pseudonymisation (pseudonymize)	<p>『캘리포니아 소비자 프라이버시보호법 California Consumer Privacy Act』, Article 1798.140(r)</p> <ul style="list-style-type: none"> • 가명화(Pseudonymisation)란 추가 정보가 별도로 보관되어 있고, 개인정보가 확실하게 제공될 수 있도록 기술적·조직적 조치의 대상이 되는 경우 추가 정보를 사용하지 않고 더 이상 특정 소비자에게 귀속되지 않는 방식으로 개인정보를 처리하는 것을 의미

국가별 가명정보 및 익명정보의 개념

일본

Term	Definitions
<p>익명가공정보 (匿名加工情報)</p>	<p>『個人情報の保護に関する法律』, 제2조 제9호</p> <ul style="list-style-type: none"> 개인정보와 개인식별부호의 구분에 따라 다음의 조치를 취하여 특정 개인을 식별할 수 없도록 개인정보를 가공하여 얻어지는 개인에 관한 정보로 당해 개인정보를 복원할 수 없도록 한 것 ① 개인정보: 당해 개인정보에 포함된 기술 등의 일부를 제거하는 것(당해 일부의 기술 등을 복원할 수 있는 규칙성을 갖지 않는 방법에 의해 다른 기술 등으로 대체하는 것을 포함) ② 개인식별부호: 당해 개인정보에 포함된 개인식별부호의 전부를 제거하는 것(당해 개인식별부호를 복원할 수 있는 규칙성을 갖지 않는 방법에 의해 다른 기술 등으로 대체하는 것을 포함)
<p>가명가공정보 (仮名加工情報)</p>	<p>『個人情報の保護に関する法律等の一部を改正する法律案』</p> <ul style="list-style-type: none"> 가명화란 데이터 내의 성명 등 특정인을 직접 식별할 수 있는 기술(記述)을 다른 형태로 작성하거나 삭제함으로써 가공하여 특정인을 식별할 수 없도록 하는 것 가명가공정보란 개인정보의 구분에 따라 다음의 조치를 강구하여 다른 정보와 조합하지 아니하는 한 특정인을 식별할 수 없도록 개인정보를 가공하여 얻을 수 있는 개인에 관한 정보 ① 개인정보: 당해 개인정보에 포함되는 기술 등의 일부를 제거하는 것(당해 일부의 기술 등을 복원할 수 있는 규칙성을 갖지 않는 방법에 의해 다른 기술 등으로 대체하는 것을 포함) ② 개인식별부호: 당해 개인정보에 포함된 개인식별부호의 전부를 제거하는 것(당해 개인식별부호를 복원할 수 있는 규칙성을 갖지 않는 방법에 의해 다른 기술 등으로 대체하는 것을 포함)

Ⅱ. 가명정보 및 가명처리의 쟁점과 이슈

가명정보의 관념

가명정보와 가명처리의 관념상의 혼동




- 개인정보와 관련한 비식별정보, 가명정보, 익명정보, 비식별조치, 가명조치, 익명조치 등은 모두 **일정 한도에서 정보를 활용(data reuse, data service 등)하게 위한 개념**
- 프라이버시 침해의 위험성에 따라 정보에 대한 규범의 적용 내지 정보의 활용을 달리하려는 입법으로 활용가치가 떨어지는 익명화된 데이터의 보완방안으로서 가명정보는 의미가 있음
- 어떠한 정보가 특정 개인을 전혀 식별할 수 없다는 것은 불가능한 상황에서 식별가능성을 염두에 두고 가명화된 데이터를 관념화 하여 법적 개념으로 포섭시키는 것은 진일보한 것으로 평가
- 개정안의 '가명정보'는 여전히 '개인정보'에 해당하며 '개인정보'에 해당하는 정보를 개인정보가 아닌 것으로 하여 활용하도록 하는 것은 개인정보에 대한 정의와 배치되기도 하며, 현실적으로 활용이 가능할 수 있느냐에 대한 물음에도 답변이 어렵게 됨
- 가명처리는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 처리를 허용하는 것이 아닌 **처리를 위한 수단과 안전조치 중의 하나에 해당**(GDPR과 CCPA, HIPPA 등은 일정한 경우에 있어 처리요건을 이행하는 수단으로서 가명조치를 규정, 개인정보에 대하여 비식별조치와 수단을 통해 활용을 한정적으로 허용하려는 것으로 판단)

가명정보와 가명처리의 관념



ISO/IEC 20889의 개인정보(PII)와 비(非)개인정보(Non-PII)

데이터는 식별수준에 따라 개인정보와 비(非)개인정보로 구분

Personal Identifiable Information

-  식별 데이터
(identified data)
-  가명 데이터
(pseudonymized data)
-  비연결 가명데이터
(unlinked pseudonymized data)

Non Personal Identifiable Information

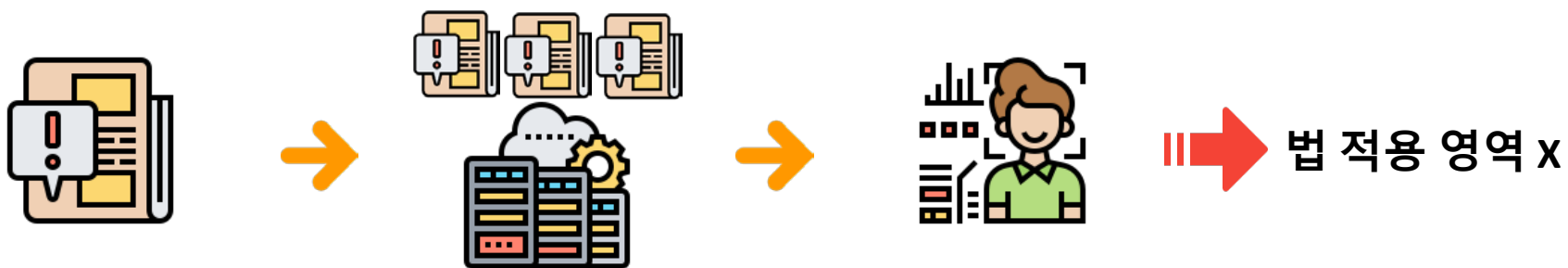
-  익명 데이터
(anonymized data)
-  통합 데이터
(aggregated data)

식별행위에 대한 통제가능성

데이터 처리 행위 전반에 대한 통제 필요성과 고려사항



당초 개인정보였던 것에서 식별성을 제거하여 활용 가능해진 것



당초 개인정보가 아니었던 것에서 식별이 가능해지는 행위를 하는 상황

가명처리의 법적 요건과 효과

가명정보 처리자에 대한 법적 책임 문제

- 누가 어떤 권한에 근거하여 가명처리를 하여야 하는가?
- 가명처리의 정도와 기준 등 요건은 어떻게 설정하여야 하는가?
 - 가명처리를 실시하는 사업자의 면책 요건
 - 가명처리를 실시하는 사업자가 고의 또는 중대한 과실로 가명처리에 소홀하여 식별을 용이하게 하여 활용하거나 제3자에게 제공한 경우의 법적 책임 문제
- 가명처리를 한 행위자체가 법률상 면책이 되는 요건이 명확히 제시되어야 하고, 이를 입증하면 면책이 될 수 있도록 되어야 함
 - 예컨대 비식별조치 수준이 K-익명성의 3이상의 수준을 갖추어야 한다면 요건이 성립되어 있을 경우 3이상 수준을 확보했음을 입증이 가능해야 함
- 가명처리를 한 행위자가 현존하는 기술의 최대치를 활용하여 선량한 관리자의 주의의무를 다하였어야만 면책이 가능

Ⅲ. 가명정보의 활용과 가명처리에 있어 고려사항

가명처리기술 적용의 수준 파악

ISO/IEC 20889의 비식별처리 기술(De-identification techniques)

구분	세부기술	내용
통계 도구 (statistical tools)	표본추출(sampling)	데이터 주체별로 전체 모집단이 아닌 표본에 일반화 또는 무작위 레코드(마이크로 데이터) 추출
	통계처리(aggregation)	속성값들의 평균 또는 합계 등으로 처리
암호화 도구 (cryptographic tools)	결정성 암호화 (deterministic encryption)	동일 속성값에 대한 암호화 값은 항상 동일하나 암호화 값의 순서는 보장되지 않음
	순서보존 암호화 (order-preserving encryption)	동일 속성값에 대한 암호화 값이 동일하고, 암호화 값의 순서는 이전의 순서가 유지됨
	형식보존 암호화 (format-preserving encryption)	원본 데이터와 같은 형식 또는 길이를 갖는 일련의 기호 형식으로 변환
	동형암호 (homomorphic encryption)	암호화 값을 복호화하지 않더라도 연산(비교·분석)이 가능하며, 원본 속성값의 노출이 없음
	동형 비밀분산 (homomorphic secret sharing)	특정 레코드 내의 민감정보를 K명의 소유자에게 나눠 갖도록 분산하여 대체하는 방식
삭제 기술 (suppression techniques)	마스킹(masking)	특정 속성값을 OO 또는 ** 등으로 대체
	로컬 삭제(local suppression)	특정 속성값을 삭제
	레코드 삭제(record suppression)	해당 레코드를 삭제

가명처리기술 적용의 수준 파악

ISO/IEC 20889의 비식별처리 기술(De-identification techniques)(계속)

구분	세부기술	내용
	가명화 기술 (pseudonymization)	해당 속성값을 암호화 또는 해시함수 등을 이용하여 가명으로 대체
	해부화 (anatomization)	하나의 데이터셋을 식별정보와 속성정보로 분리하여 별도로 데이터셋으로 분리하고, 원본 데이터셋과 매핑이 가능한 별도의 속성(매핑키)이 추가되며 용도에 따라 접근 권한 및 공개 범위 설정
일반화 기술 (generalization techniques)	라운드잉 (rounding)	정해진 기준에 따라 값을 올림 또는 반올림
	상하단코딩 (top and bottom coding)	정해진 최대값 또는 최소값으로 값을 대체
	단일속성결합 (combining a set of attributes into a single attribute)	여러 개의 속성값을 하나의 속성으로 결합
	로컬 일반화(local generation)	속성 중 희소값이 있을 때 해당 속성값을 제거
무작위화 기술 (randomization techniques)	잡음 추가(noise addition)	원본 속성의 통계적 특징을 최대한 유지하면서 해당 속성값에 임의의 랜덤값을 추가
	순열(permutation)	속성값을 수정없이 레코드들 간에 속성값을 교환
	마이크로어그리게이션(microaggregation)	총계처리의 일종으로 특정속성의 모든 값을 알고리즘에 따라 계산된 평균치로 대체
	합성데이터 (synthetic data)	무작위화 및 표본추출 기술을 사용하여 데이터를 인위적으로 가공 후 데이터셋을 생성하는 기법

가명처리기술 적용의 수준 파악

ISO/IEC 20889의 일반 프라이버시 측정 모델(Formal Privacy measurement models)

- (목적) 비식별처리 기술들은 재식별 위험성에 대한 정량적 평가에 한계가 존재, 비식별 수준 및 재식별 위험을 계산하기 위함
- (종류)
 - (K-익명성) : 데이터셋에서 같은 속성값을 가지는 레코드가 K이상 존재하도록 하는 기법
 - (L-다양성) : K-익명성의 취약점(동일성 공격, 배경지식 공격)을 차단하기 위해 동질집합에서 L개의 서로 다른 정보를 가지도록 변환
 - (T-접근성) : 비식별 값이 골고루 분포되도록 하여 추론이 용이
 - 차분 프라이버시(Differentially Private) 모델 : 통계적인 데이터 유용성을 유지하되 식별성을 제거하도록 일정량의 잡음(노이즈)를 추가
 - 선형 민감도(Linear sensitivity) 모델 : 비식별 정보가 특정인과 연관되는지, 추정할 수 있는지를 한계값, 모호성 등으로 수치화

가명처리기술 적용의 수준 결정

가명처리 기술의 적용 정도와 수준을 결정

- 가명처리자의 가명처리를 위해 필요한 기술수준과 처리기법 등에 대한 파악 필요
- 익명화, 가명화 등 비식별 처리 과정의 단계별로 현재의 비식별 처리기술의 적용 정도와 수준을 정할 수 있어야 함



감사합니다.