가명정보 처리 가이드라인

개인정보보호위원회

가명정보 처리 가이드라인

개인정보보호위원회

목차

Ⅰ 추진 배경

◈ 빅데이터, AI 등 다양한 융·복합 산업에서의 데이터 이용 수요가 급증하고 있으며, 데이터 활용의 핵심인 가명정보 활용에 대한 법적 근거가 마련되어 체계적인 데이터 활용 기반이 조성되었음

◈ 이에 데이터 활용에 필요한 가명처리 기술·절차·관리체계 등을 구체적으로 안내하여 개인정보보호는 더욱더 강화하고 안전한 데이터 활용기반을 마련하고자 함

- 4차산업혁명 시대 신성장 동력인 '데이터' 활용에 대한 시대적 요구를 반영한 데이터3법*이 개정('20.2.4.)되어 시행('20.8.5.)

  * 개인정보 보호법(이하, '보호법'), 정보통신망 이용촉진 및 정보보호 등에 관한 법률, 신용정보의 이용 및 보호에 관한 법률

  – 가명정보 처리에 관한 특례(보호법 제3장제3절)가 신설되어 개인정보처리자가 통계작성, 과학적 연구, 공익적 기록보존 등을 위한 목적으로 개인정보를 가명처리하여 활용할 수 있는 기반이 마련됨

- 본 가이드라인은 보호법 제28조의2에 따른 동의없는 가명정보의 처리 과정에서의 개인정보 오·남용을 방지하고, 데이터 산업 활성화를 위한 안전한 가명정보 활용 방안 안내

  – 가명정보 처리 목적 및 범위, 처리 절차, 가명정보의 안전성 확보를 위한 방안 및 주요 가명처리 기술 등에 대한 예시를 제공하여, 실무자가 이해하는데 도움을 주고자 함

  ※ 개인정보처리자는 적법하게 수집한 정보의 안전조치(시행령 시행령 제30조) 등을 위하여 가명처리하는 경우에도 본 가이드라인을 참고할 수 있음

- 또한, 보호법 제28조의3에 따라 서로 다른 개인정보처리자가 보유한 가명정보를 결합 및 반출하여 활용하고자 하는 경우 개인정보처리자가 참고할 수 있도록 결합·반출에 대한 일반적인 절차와 방법을 안내하고자 함

Ⅱ 가이드라인 개요

1 가명정보 처리 대상

○ **(가명처리)** 가명정보는 개인정보처리자의 정당한 처리 범위 내에서 통계작성, 과학적 연구, 공익적 기록보존 등의 목적으로 정보주체의 동의 없이 처리할 수 있음

가. 통계작성 : 통계란 특정 집단이나 대상 등에 관하여 작성한 수량적인 정보를 의미

– 시장조사와 같은 상업적 목적의 통계 처리도 포함

※ 직접(1:1) 마케팅 등을 위해 특정 개인을 식별할 수 있는 형태의 통계는 해당하지 않음

〈예시〉

● 지자체가 연령에 따른 편의시설 확대를 위해 편의시설(문화센터, 도서관, 체육시설 등)의 이용 통계(위치, 방문자수, 체류시간, 나이대, 성별 등)를 생성 · 분석하여 적합한 지역에 신규 편의시설을 선정하고자 하는 경우

나. 과학적 연구 : 과학적 연구는 기술의 개발과 실증, 기초연구, 응용연구 및 민간 투자 연구 등 과학적 방법을 적용하는 연구를 의미

– 과학적 연구는 과학적 방법을 적용하는 연구를 말하며 자연과학, 사회과학, 의료 등 다양한 분야에서 가능

– 여기서, 과학적 방법은 체계적이고 객관적인 방법으로 검증 가능한 질문에 대해 연구하는 것을 의미

– 과학적 연구는 기술의 개발과 실증, 기초 연구, 응용 연구뿐만 아니라 새로운 기술 · 제품 · 서비스 개발 등 산업적 목적을 위해서도 수행이 가능하며, 민간 투자 연구, 기업 등이 수행하는 연구도 가능

〈예시〉

- 코로나19 위험 경고를 위해 생활패턴과 코로나19 감염율의 상관성에 대한 가설을 세우고, 건강관리용 모바일앱을 통해 수집한 생활습관, 위치정보, 감염증상, 성별, 나이, 감염원 등을 가명처리하고 감염자의 데이터와 비교 · 분석하여 가설을 검증하는 경우

다. 공익적 기록보존 : 공공의 이익을 위하여 지속적으로 열람할 가치가 있는 정보를 기록하여 보존하는 것을 의미

– 공공기관이 처리하는 경우에만 공익적 목적이 인정되는 것은 아니며, 민간기업, 단체 등이 일반적인 공익을 위하여 기록을 보존하는 경우도 공익적 기록보존 목적이 인정 됨

〈예시〉

- 연구소가 현대사 연구 과정에서 수집한 정보 중에서 사료가치가 있는 생존 인물에 관한 정보를 기록 · 보관하고자 하는 경우

○ **(가명정보 결합)** 서로 다른 개인정보처리자가 보유한 가명정보는 개인정보보호위원회(이하, 보호위원회) 또는 관계 중앙행정기관의 장이 지정한 결합전문기관을 통해서만 결합하여 처리할 수 있으며, 개인정보처리자 내부에서 보유한 가명정보의 결합은 별도의 결합전문기관을 통하지 않아도 됨 (p.52, [참고3] 참조)

2 가이드라인 구성

○ 본 가이드라인은 개인정보를 가명처리하여 활용하고자 하는 개인정보처리자가 가명처리에 관한 사항 및 결합 · 반출에 관한 일반적인 절차 및 방법 등을 안내하고자 함

– 개인정보의 유형, 성격 등을 고려하여 법령을 준수하는 범위 내에서 개인정보처리자가 가명처리 절차와 방법을 자율적으로 판단하여 처리할 수 있음

○ 주요 수록사항은 다음과 같음

– 가명정보 처리 목적 및 대상에 관한 사항

– 개인정보처리자가 개인정보를 활용하여 가명처리를 수행하기 위한 절차 안내

– 가명정보 결합 및 반출에 관한 절차 안내

– 가명정보 처리에 따른 관리적 · 기술적 · 물리적 보호조치에 관한 사항

– 개인정보 가명처리의 주요 기술 안내 및 적용 예시

– 특이정보*의 처리 사례에 관한 사항 등

* 데이터의 평균치에서 크게 벗어나서 다른 데이터와 확연히 구분되거나 또는 데이터의 분포에서 비정상적으로 분포를 벗어나 측정이 되는 값

● 특정 산업 분야의 개인정보 가명처리에 관하여 보호위원회와 소관 부처가 공동으로 발간한 분야별 가이드라인이 있는 경우, 분야별 가이드라인을 우선하여 활용할 수 있으며,

– 별도의 분야별 가이드라인이 없는 경우 본 가이드라인을 활용할 수 있음

3 용어 정리

○ 개인정보 : 살아있는 개인에 관한 정보로서 다음의 정보를 포함함

– 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 정보 (이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 함)

※ 개인정보에 대한 판단기준은 개인정보처리자가 보유한 정보 또는 접근 가능한 권한 등 상황에 따라 달리 판단하여야 함

○ 가명처리 : 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것

○ 가명정보 : 개인정보를 가명처리함으로써 원래의 상태로 복원하기 위한 추가정보의 사용 · 결합 없이는 특정 개인을 알아볼 수 없는 정보

○ 익명정보 : 시간 · 비용 · 기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보

○ 추가정보 : 개인정보의 전부 또는 일부를 대체하는 데 이용된 수단이나 방식(알고리즘 등), 가명정보와의 비교 · 대조 등을 통해 삭제 또는 대체된 개인정보 부분을 복원할 수 있는 정보

(매핑 테이블 정보, 가명처리에 사용된 개인정보 등) 등

※ 추가정보(원본정보와 알고리즘 · 매핑테이블 정보)와 가명정보는 관리적 또는 기술적으로 각각 분리하고, 접근권한을 분리하여야 함

- 개인정보파일 : 개인정보를 쉽게 검색할 수 있도록 일정한 규칙에 따라 체계적으로 배열하거나 구성한 개인정보의 집합물

- 가명정보처리자 : 업무를 목적으로 개인정보를 가명처리하여 활용 또는 제공하는 공공기관, 법인, 단체 및 개인 등

- 가명정보취급자 : 가명정보를 처리하는 개인정보처리자의 지휘 · 감독을 받아 가명정보를 처리하는 임직원, 파견근로자, 시간제근로자 등

- 적정성 검토 : 본 가이드라인에서 제시하고 있는 절차를 기반으로 사전에 정의한 가명처리 기준에 따라 적절히 가명처리가 되었는지 확인하는 절차

- 재식별 : 추가정보 또는 행위자가 달리 보유하고 있는 다른 정보나 공개된 정보와의 결합 또는 대조 · 비교 등을 통해 특정 개인을 알게 되거나, 알아보려하는 상태 또는 행위

- 결합키 : 결합키관리기관이 결합키연계정보를 생성할 때 임시적으로 사용되는 정보

- 결합키연계정보 : 동일 정보주체에 대해 가명정보를 결합할 수 있도록 서로 다른 결합신청자간의 결합키를 연계한 정보

- 결합신청자 : 가명정보의 결합을 신청하는 개인정보처리자

  * 가명정보를 제공만 하는 자, 가명정보를 제공하고 결합한 정보를 이용하는 자 또는 가명정보의 제공 없이 결합한 정보를 이용하는 자를 모두 포함

- 결합전문기관 : 법 제28조의3제1항에 따라 서로 다른 개인정보처리자 간의 가명정보 결합을 수행하기 위해 개인정보 보호위원회 또는 관계 중앙행정기관의 장이 지정하는 전문기관

- 결합키관리기관 : 시행령 제29조의3제2항에 따라 결합키연계정보를 생성하여 결합전문기관에 제공하는 등 가명정보의 안전한 결합을 지원하는 업무를 하는 한국인터넷진흥원 또는 보호위원회가 지정하여 고시하는 기관을 말함

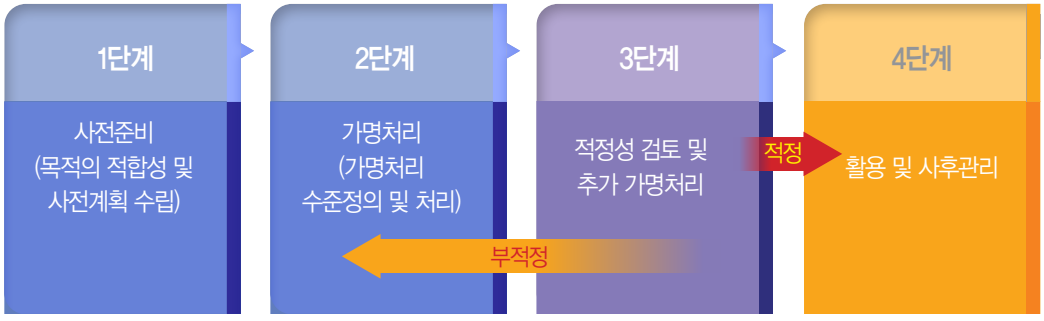
Ⅲ 가명처리

- 가명처리는 '개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가정보 없이는 특정 개인을 알아볼 수 없도록 처리하는 것'을 의미
  - 가명처리 시 가명정보 자체만으로 특정 개인을 알아볼 수 있는 지와 추가정보 또는 다른 정보의 결합가능성을 고려할 필요가 있음
    ※ 가명정보처리자가 보유한 다른 정보 등을 통해 개인이 식별 가능한 경우 가명처리가 잘못된 경우라고 할 수 있음

- 가명정보처리자는 개인정보를 가명처리할 때에는 일반적으로 사용할 수 있는 절차와 단계별 유의사항 제시
  - 가명정보의 특성 · 목적 및 분야별 가이드라인 등을 고려하여 추가절차를 포함하거나 일부 절차를 간략화 할 수 있음
    ※ 통계법 등 관련법령에 따라 개인정보를 수집 · 이용하는 경우에는 당해 법령에 따라 처리

1 가명처리 절차

〈가명처리 단계별 절차도〉



단계1 사전준비

- 가명처리 대상 항목 및 처리수준을 정의하기 위해서는 처리 목적이 적합한지 여부를 확인하고 필요한 서류를 작성하여야 함
  - 가명처리의 목적을 명확히 하고, 내부 승인절차를 별도로 마련한 경우 이를 위한 추가업무를 수행할 수 있음
  - 가명정보를 제3자에게 제공하는 경우 이용목적 및 방법, 재식별 위험관리 등 가명정보의 안전성 확보를 위하여 필요한 조치를 마련하도록 하는 내용을 포함한 계약을 체결할 수 있음

단계2 가명처리

- 가명정보 처리 시에도 개인정보의 최소처리원칙을 준수하여야 하며, 가명처리 방법을 정할 때에는 처리목적, 처리(이용 또는 제공)환경, 정보의 성격 등을 종합적으로 고려하여야 함
  - 목적에 필요한 최소한의 항목만을 가명처리 대상으로 선정하고, 개인정보파일에서 가명처리 대상 항목을 우선 추출하여야 함
  - 추출한 결과 정보의 '항목별 위험도 측정'은 가명정보처리자의 안전조치 수준이나 정보 자체의 재식별 가능성을 고려하여 판단하여야 하므로 내부 활용 · 제공과 제3자 제공 시 고려하여야 할 사항이 달라질 수 있음(p.13~18 참조)
  - 항목별 위험도 측정이 완료되면 이를 고려하여 항목별 가명처리 방법과 수준을 먼저 정의하고, 이에 따라 가명처리 수행

단계3 적정성 검토 및 추가 가명처리

- 목적달성을 위해 적절한 수준으로 가명처리가 이루어졌는지, 재식별 가능성은 없는지 등에 대한 최종적인 판단절차를 수행하여야 함
  - 가명처리한 결과, 목적을 달성하기 어렵거나 재식별 가능성이 있다고 판단한 경우 '2단계 (가명처리)'를 반복하거나 부분적으로 추가적인 가명처리를 할 수 있음
  - 데이터의 분포와 값을 살펴보았을 때, 특이정보가 있다고 판단한 경우 재식별 가능성을 낮추기 위한 적절한 조치를 취하여야 함

단계4 활용 및 사후관리

- 적정성 검토 결과 가명처리가 적정하다고 판단되면 가명정보를 본래 활용목적을 위해서 처리할 수 있으며, 법령에 따라 기술적 · 관리적 · 물리적 안전조치를 이행하여야 함
  - 가명정보처리자는 가명정보취급자에게 금지행위, 안전조치 등에 관한 사항을 안내하여 가명정보를 안전하게 처리하여야 함

〈기타 참고사항〉

- 가명정보처리자는 더욱 안전한 가명정보 처리를 위해 다음의 사항을 참고하여 업무에 반영할 수 있음
  - 가명처리 관련 업무의 총괄 · 관리 및 의사결정을 위한 총괄부서(또는 담당자)를 지정할 수 있으며, 주요 업무는 다음과 같음

1) 가명처리 신청(목적)에 대한 적합성 검토

2) 가명처리

3) 가명처리 적정성 검토

4) 가명정보취급자에 대한 관리 · 감독

5) 그 외 안전하고 효율적인 가명정보 처리를 위해 필요한 사항

※ 1), 3)의 경우 외부전문가를 포함한 심의위원회를 구성 · 운영할 수 있음

- 가명처리 관련 업무 담당자의 분리
  - 가명처리를 수행한 자와 가명정보취급자(활용 등), 가명정보의 적정성을 검토하는 자*는 관리적 또는 기술적으로 권한을 분리(p.28 참조)

* 추가정보의 내용을 알고 있는 자가 가명정보의 검토를 수행하거나 취급(활용)하는 경우 처리하는 과정에서 특정 개인을 알아볼 우려가 있음

2 가명처리 세부절차

〈가명처리 단계별 세부 절차도〉

단계1 사전준비 : 처리 목적의 적합성 검토 및 가명처리 준비

가명처리를 위한 사전준비 단계에서는 가명정보 활용 목적을 명확히 하고 가명처리를 수행할 것인지를 결정하여야 하며, 가명처리하기로 결정한 경우 처리를 위하여 필요한 서류를 작성할 수 있음

- 가명정보의 처리목적 명확화 : 법률에서 허용하는 목적* 내에서 가명정보를 처리하는 목적을 최대한 명확히 작성하여야 함

  * 통계작성, 과학적 연구, 공익적 기록보존 등에 한 함

  - (적절하지 않은 예시) 신제품 개발을 위한 과학적 연구 수행

    ※ 목적이 구체적으로 명시되지 않아 적절하지 않음

  - (적절한 예시) ○○제품의 성능 개선을 위해 개인별 ○○○특성에 대한 설문조사를 토대로 개인별 특성과 성능 요인의 연관성에 대한 과학적 연구

- 가명처리 적합성 검토(개인정보 보유부서 또는 전담부서) : 개인정보의 수집 목적 및 성격, 가명정보 활용 목적 등을 고려하여 가명처리 여부를 결정할 수 있음

  ※ 필요 시 심의위원회 구성 또는 외부전문가 평가 등을 통해 결정할 수 있음

- 필요서류 작성 : 가명정보의 처리 또는 가명처리를 위탁(보호법 제26조에 따라 수행)하거나 제3자에게 제공하는 경우 필요에 따라 재식별 금지에 관한 사항, 기타 처리에 있어 유의해야 할 사항* 등을 포함한 계약서를 작성할 수 있음

  * (예시) 가명정보의 재제공 금지, 가명정보 재식별 금지, 가명정보의 안전성확보조치, 가명정보의 처리기록 작성 및 보관, 가명정보의 파기, 재식별 시 손해배상 등

- 가명정보 처리에 관한 내부관리계획이 없는 경우, 계획 수립 필요

  [Ⅴ.가명정보의 안전한 관리] 참조

단계2 가명처리 : 환경에 따른 수준정의 및 처리

가명처리 단계는 세부적으로 ①대상선정, ②위험도 측정, ③가명처리 수준정의, ④ 가명처리를 하는 4가지 단계로 구성되어 있음

〈위험도 측정 세부 절차도〉

1. 대상선정

- [단계1. 사전준비]에서 수립한 목적을 달성하기 위해 개인정보파일에서 가명처리에 필요한 항목 추출

  ※ 목적달성에 필요한 최소 항목을 처리하여야 함

■ **추출한 결과 정보(예시) :**

✔ **이름, 휴대폰번호, 성별, 이메일, 주소, 구매상품, 구매액, 장바구니목록**

- 가명처리 목적 : 성별과 지역에 따른 구매액 상관관계를 분석하고자 함
- 가명처리 대상 항목 :
  - 이름, 휴대폰번호, 성별, 이메일, 주소(시군구), 구매액

  * 분석목적과 상관없는 정보는 대상선정에서 제외

2. 위험도 측정

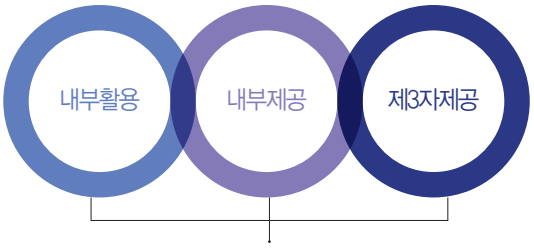
가명정보처리자의 개인정보 보호수준 및 다른 정보 보유여부 등을 검토하고, 항목별 위험도 분석을 통해 위험도를 측정

– 가명정보는 처리(제공) 환경에 따라 가명정보처리자 내부에서 활용(자체활용 또는 내부 제공 · 결합)하는 경우와 제3자에게 제공하는 경우로 구분할 수 있으며, 이에 따라 위험도 측정 결과가 달라질 수 있음

가. 처리(제공)환경 검토

○ 처리 목적에 따라 처리(제공)환경과 제공받는 자의 개인정보 보호수준 및 다른 정보 보유여부 등을 검토하여야 함

※ 불특정 제3자(공개 등)에게 제공하는 경우 익명정보로 처리하는 것을 원칙으로 함



제공받는 부서의 개인정보 보호수준 및 다른정보* 보유여부

* 제공받는 부서가 다른(개인)정보를 보유한 경우, 제3자로부터 다른 정보를 받아 함께 활용하는 경우

– (내부 활용) 가명정보처리자가 보유한 개인정보를 가명처리 또는 내부 결합하여 직접 활용 또는 다른 부서에 제공하는 경우를 의미

※ 수탁자를 통하여 가명정보를 처리하는 경우는 내부 활용에 해당하며, 제3자로부터 제공받은 가명정보와 본인이 보유한 가명정보를 결합하는 경우는 내부결합에 해당하지 않으므로 전문기관을 통하여야 함

○ 동일 처리자 내 활용이므로 [Ⅴ. 가명정보의 안전한 관리]에 따라 접근권한 분리, 취급자 교육 등을 실시하여 보호수준 통제 가능

- 이 경우에도 가명정보취급자의 소속 부서에서 이미 보유하고 있는(접근 가능한) 정보 및 처리 시점을 기준으로 제공받는 다른(개인)정보를 고려하여 검토하여야 함

  * 가명처리를 수행한 자와 가명정보취급자가 동일할 경우 원본정보에 대한 접근 가능성도 고려하여야 함

잘못된 내부활용 사례) ○○화장품 회사의 A팀은 화장품 판매 정보를 관리하는 팀으로서, 가명처리 접근권한을 분리하지 않고 해당 정보를 가명처리하여 신상품 수요조사 예측 모델 개발을 목적으로 활용

■ (처리현황) A팀은 판매정보 내 개인식별 가능성이 있는 이름, 성별, 승인번호를 가명처리하고, 희귀 지역의 판매내역을 삭제하여 A팀 가명정보 분석담당자에게 제공

✔ 가명정보 분석담당자는 A팀의 판매정보 관리 업무를 병행하여 업무를 수행하고 있음

■ (문제점) 가명정보 분석담당자는 가명정보 분석을 통해 최고가 화장품의 금액과 판매지역을 파악할 수 있으며, 판매정보 관리 업무를 병행하고 있어 해당 금액과 지역을 통해 특정개인을 식별할 가능성이 있음

■ (해결방안) 가명정보처리자는 가명정보를 처리(분석)하는 담당자가 다른 정보(위의 예시에서는 판매정보 관리)에 동시에 접근할 수 없도록 접근권한을 명확히 분리하고, 접근통제를 실시하여야 함

잘못된 내부제공 사례) □□공사는 A부서에서 관리하고 있는 고속도로 이용차량 빅데이터 분석 결과를 고속도로 통행요금을 관리하는 B부서에서 처리하는 개인정보를 확인하지 않고 교통서비스 개선을 위한 연구 목적으로 내부 제공

■ (처리현황) A부서는 개인식별가능성이 있는 차량번호, 차종 등을 가명처리하고, 이동시간, 이동량, 사고정보 등의 정보를 B부서에 제공

✔ B부서는 고속도로 통행요금 관리를 위해 고객번호와 차량번호, 톨게이트 입출시간 및 결제금액 정보를 보유하고 있음

■ (문제점) B부서는 A부서에서 제공받은 정보의 이동시간 정보와 B부서가 보유한 톨게이트 입출시간을 활용하여 특정시간에 통과한 차량의 번호를 알 수 있으며, 해당 차량번호를 통해 특정 개인을 식별할 가능성이 있음

■ (해결방안) 가명정보처리자는 가명정보 생성 시 가명정보를 처리할 부서에서 보유하고 있는 정보를 고려하여 가명처리를 수행하여야 함

– (제3자 제공) 개인정보처리자가 보유한 개인정보를 가명처리하여 특정 제3자에게 제공하는 경우를 의미

- 제3자의 개인정보보호수준을 고려*하여 가명정보 제공으로 인하여 발생할 수 있는 재식별 위험을 최소화하기 위하여 노력하여야 함

  * 보호수준이 낮은 기관에는 상대적으로 높은 수준의 가명처리 수준을 적용하는 방법 등

- 또한, 제3자가 사전에 보유하고 있는 정보 및 처리 시점을 기준으로 제공받는 다른(개인)정보 등을 고려하여야 하고, 이를 파악하기 위해 관련 정보*를 요청할 수 있음

  * 제3자의 가명정보처리자가 관리하고 있는 개인정보 중 제공받는 가명정보와 결합 가능성이 있는 개인정보 목록 등

- 사전준비 단계에서 재식별에 대하여 계약서에 명시한 사항이 있다면 이를 고려할 수 있음

잘못된 제3자제공 사례) ○○호텔에서는 최고급 객실을 이용한 VIP등의 특이정보를 삭제하지 않고 호텔 투숙 및 서비스 금액 등을 ○○분석회사에 제공하고, ○○분석회사는 해당 정보를 분석하여 시간에 따른 객실이용현황 및 서비스이용에 대한 조사 연구를 수행

■ (처리현황) ○○분석회사는 온라인 SNS정보 및 다양한 기업의 정보를 수집하여 다양한 연구조사를 실시하는 회사로써 내부관리계획을 수립하고, 관리적 · 기술적 보호조치를 준수하고 있음

✔ ○○호텔은 회원번호와 이름을 가명처리하고, 나이, 성별, 등급, 예약방법, 객실정보, 체크인, 체크아웃, 서비스 이용금액을 제공

■ (문제점) ○○분석회사의 분석담당자는 특정일에 최고급 객실을 이용한 내용을 분석과정에서 인지할 수 있으며, 기존 업무(온라인 SNS정보 수집)를 수행하며 공개된 정보(예 : 개인이 SNS에 올리는 정보, 여행후기 등)를 통해 특정 개인을 식별할 가능성이 있음

■ (해결방안) ○○호텔은 제공하는 가명정보에 포함된 특이정보(최고급 객실)를 삭제 또는 가명처리 등을 수행하여야 함

나. 항목별 위험도 분석

- 추출한 결과 정보의 항목별 위험도를 분석하여야 함
  - 개인식별 가능성이 높은 항목을 분류하여 가명처리 방법 및 수준을 결정하는데 참고할 수 있도록 항목별 위험도를 분석

* 규모(레코드, 항목의 크기), 구체성(정보의 정확성 수준)

〈개인식별 가능성이 높은 정보 (예시)〉

■ 식별정보

✔ 고유식별정보(여권번호, 외국인등록번호, 운전면허번호), 성명, 전화번호, 전자우편주소, 의료기록번호, 건강보험번호, 자동차 등록번호 등 외부 연계(식별)를 목적으로 생성된 정보 등

■ 식별가능정보

✔ 성별, 연령(나이), 국적, 혈액형, 신장, 몸무게, 직업, 위치정보 등 가명정보처리자의 입장에서 개인을 알아볼 수 있는* 정보

* 개인을 '알아볼 수 있는지'는 해당 정보를 처리하는 자(정보의 제공 관계에 있어서는 제공받은 자를 포함)를 기준으로 판단하여야 함

✔ 특이정보

- 국내 최고령, 최장신, 고액체납금액, 고액급여수급자 등 전체적인 패턴에서 벗어나 극단값이 발생 할 수 있는 정보
- 희귀 성씨, 희귀 혈액형, 희귀 눈동자 색깔, 희귀 병명, 희귀 직업 등 정보 자체로 특이한 값을 가지고 있는 정보

다. 결과보고서 작성

○ 가명정보처리자는 처리(제공)환경과 항목별 위험도 분석을 참고하여 가명처리에 대한 위험도 평가 결과를 도출하여야 함

– 필요한 경우 다음의 예시와 같은 검토 결과 보고서를 작성하여 관리할 수 있음

〈가명처리 검토 결과보고서 (예시)〉

가명정보 활용목적	• A사가 보유한 부동산 시세정보를 가명처리하여 B기관에 제공하여, 부동산 임대소득 계산 및 인근 지역 시세자료 파악을 위한 연구 수행	
가명정보 항목	• 소유자명, 연락처, 주택구분, 시도, 시군구, 읍면동, 지번, 전용면적, 공급면적, 전세, 보증금, 월세 ※ 항목을 나열하지 못하는 경우 '별지' 사용 가능	
처리(제공) 환경 검토	처리 환경	• 특정 제3자(B기관) 제공 – A사는 B기관과 계약체결을 통해 가명정보를 제공
	제공 받는 자의 환경	• 가명정보를 제공받는 기관은 부동산 관련 다른(개인)정보를 보유하고 있지 않음
	제공받는 자의 보호수준	• B기관은 가명정보처리시스템에 대한 ISMS인증을 취득하고 있으며, 내부관리를 통해 관리적, 기술적 보호조치를 수행하고 있음
항목별 위험도 분석	• '소유자명', '연락처'는 식별정보, • '지번'은 식별가능정보, '시세정보(전세, 보증금, 월세)'는 특이정보 가능성 존재	
최종 검토 의견*	• 해당 연구는 특정 제3자와의 계약서 체결을 통해 가명정보를 활용하는 경우에 해당하며, 제공받는 자가 별도의 다른(개인)정보를 통해 가명정보를 재식별 할 가능성이 낮음 – '소유자명', '연락처'는 활용 목적상 반드시 필요한 경우가 아니라면 삭제 또는 목적 달성에 필요한 경우 가명처리 필요 – '지번' 및 '시세정보(전세, 보증금, 월세)'의 경우 다른(공개된 정보* 등) 정보를 통해 재식별 가능성이 있어 삭제 또는 목적 달성에 필요한 경우 가명처리 필요 • 그 외의 정보들은 재식별 가능성이 낮으며 목적달성을 위해 필요하다고 판단되므로 가명처리하지 않음	

※ 최종 검토의견은 외부전문가를 활용하여 자문 및 작성을 요청할 수 있음

3. 가명처리 수준 정의

가명정보처리자는 '가명처리 검토 결과보고서'를 기반으로 가명정보의 활용목적 달성에 필요한 수준을 고려하여 가명처리 수준 정의를 하여야 함

- 가명처리 기법 등은 '[참고1] 개인정보 가명처리 기술 및 예시' 참조

〈가명처리 수준 정의표 (작성예시)〉

• '가명처리 검토 결과보고서'에 분석한 개인정보에 대한 가명처리 수준 정의

순번	항목명	처리수준	비고
1	소유자명	– 가명처리 (암호화 : SHA2+Salt)	– 소유자명과 연락처는 추후 시계열 분석을 위해 가명처리 수행
2	연락처		
3	지번	– 가명처리(삭제)	– 세부 지번의 정보는 분석목적에 필요하지 않음
4	전세	– 기타기술 (라운딩 : 만원 단위)	– 만원 단위의 금액만 분석목적에 필요
5	보증금		
6	월세		
7	주택구분	– 처리하지 않음 ※ 항목이 다수여서 작성이 어려운 경우 '별지'를 활용하여 목록만 제시	– 처리하지 않는 항목을 작성
8	시도		
9	시군구		
10	읍면동		
11	전용면적		
12	공급면적		

4. 가명처리 수행

■ '가명처리 수준 정의표'를 기반으로 가명처리를 수행하여야 함

〈가명처리 절차 (예시)〉

원본정보

소유자명	연락처	주택구분	법정동코드	시도	시군구	읍면동	지번	건물명	전세(천원)	보증금(천원)	월세(천원)	전용면적	공급면적
김철수	090-1234-5678	아파트	2635010700	서울특별시	동작구	사당동	1388-4	대우마리나	-	25,000	750	104.00	84.00
이영희	090-2468-3579	오피스텔	3611011000	대전광역시	서구	둔산동	656	푸른지오시티	81,250	-	-	56.45	24.32
박민호	090-9876-5432	아파트	4311410100	부산광역시	해운대구	우동	111-13	평화	125,000	-	-	100.00	84.00

추출 추출 추출

(대상선정)

– 목적 : 부동산 임대소득 계산 및 인근지역 시세자료 파악을 위한 연구

소유자명	연락처	주택구분	시도	시군구	읍면동	지번	전세(천원)	보증금(천원)	월세(천원)	전용면적	공급면적
김철수	090-1234-5678	아파트	서울특별시	동작구	사당동	1388-4	-	25,000	750	104.00	84.00
이영희	090-2468-3579	오피스텔	대전광역시	서구	둔산동	656	81,250	-	-	56.45	24.32
박민호	090-9876-5432	아파트	부산광역시	해운대구	우동	111-13	125,000	-	-	100.00	84.00

(위험도 측정)

– 처리환경 검토와 개인정보 항목별 위험도 분류에 따라 가명처리 수준 정의

처리환경

A사의 부동산 시세정보를 B기관에 제공(계약)

제공되는 항목의 '지번'의 경우 등기부열람을 통해 특정개인식별 가능성 존재

+

분류

소유자명, 연락처는 개인정보로 분류하고 가명처리(암호화)

구체적인 지번은 분석목적에 관계 없어 삭제 조치 및 시세정보는 분석에 필요한 단위(만원)로 가명처리

식별정보 / 식별가능정보

소유자 명	연락처	주택 구분	시도	시군구	읍면동	지번	전세 (천원)	보증금 (천원)	월세 (천원)	전용 면적	공급 면적
김철수	090-1234-5678	아파트	서울 특별시	동작구	사당동	1388-4	-	25,000	750	104.00	84.00
이영희	090-2468-3579	오피스텔	대전 광역시	서구	둔산동	656	81,250	-	-	56.45	24.32
박민호	090-9876-5432	아파트	부산 광역시	해운대구	우동	111-13	125,000	-	-	100.00	84.00

(소유자 명, 연락처)
+Salt
암호화

삭제 / 라운딩

가명처리

ID	주택 구분	시도	시군구	읍면동	전세 (천원)	보증금 (천원)	월세 (천원)	전용 면적	공급 면적
wd4e85D2C1qe89rwqe	아파트	서울 특별시	동작구	사당동	-	25,000	800	104.00	84.00
r5w1e2SXzi4wd64qwz	오피스텔	대전 광역시	서구	둔산동	81,300	-	-	56.45	24.32
ghe6W15Z5ax4Qe24jx	아파트	부산 광역시	해운대구	우동	125,000	-	-	100.00	84.00

- 가명처리 단계에서 생성되는 추가정보는 가명정보와 분리하여 별도로 저장하여야함
  - 추가정보의 분리보관은 [Ⅴ. 가명정보의 기술적 보호조치](p.28)를 참조할 수 있음

※ 추가정보는 원칙적으로 파기하되, 필요한 경우 분리보관을 할 수 있음

단계3 적정성 검토 및 추가 가명처리

[단계2. 가명처리]에서 작성한 '가명처리 수준 정의표'의 기준에 따라 적절히 가명처리가 되었는지 확인하고, 가명정보의 활용목적을 달성할 수 있는지와 재식별 가능성이 없는지를 검토

■ 적정성 검토 사항

가명정보의 적정성 검토는 개인정보처리자의 판단에 따라 외부전문가로 구성된 적정성 평가단을 구성하여 검토할 수 있음

- (가명처리의 적정성) 가명정보처리자가 정의한 가명처리 수준에 따라 적절히 가명처리가 되었는지 확인
  ※ 가명정보 항목 전체를 검토하여 가명처리가 제대로 되었는지 확인(특히 대용량의 정보의 경우 중간에 처리되지 않은 부분이 있을 수 있으므로 확인 필요)

- (목적달성 가능성 검토) 생성한 가명정보가 초기 가명정보 활용 목적을 달성할 수 있는지 여부 검토
  ※ 생성한 가명정보가 활용 목적을 달성하지 못하는 경우 [단계2. 가명처리] 절차를 재수행하여 목적을 달성할 수 있는 수준으로 가명처리 수준을 다시 설계하여 처리

■ 추가 가명처리

○ (특이정보 처리) 항목별 위험도를 바탕으로 가명처리한 경우에도 '특이정보'를 통해 개인식별이 가능할 수 있으므로 추가로 가명처리를 할 필요가 있음

〈특이정보 (예시)〉

사례1) 국회의원 같이 특정 지역에서 소수만 존재하는 직업의 경우 지역구 국회의원 명단 등을 통해 개인이 식별될 수 있음

※ (가명처리 예시) 특정 지역을 인접 지역과 병합* 하거나, 직업을 일반화(정치인)

* 국가통계기관의 경우 세부 지역단위 통계 시 2천명이 되지 않는 경우 인접 지역에 병합

사례2) 호텔, 렌터카 등 여행업종에서 보유중인 최고급 객실이용정보, 특정 차량이용정보는 개인(공인 등)이 SNS등 온라인에 공개하는 정보와 결합되어 개인이 식별될 수 있음

※ (가명처리 예시) 특정 차량(슈퍼카)의 이름을 일반화(스포츠카)하여 게시하거나, 호텔 최고급 객실정보를 일반객실 정보 대체

사례3) 공인이 희귀질병을 앓고 있는 경우 해당 병명만으로 개인이 식별될 수 있음

※ (가명처리 예시) 희귀질병을 일반화(일반 질병명)하거나, 직업을 일반화(회사원)

※ 특이정보 처리사례는 [참고2] 특이정보 정의 및 처리사례 참고

단계4 활용 및 사후관리

■ 가명정보처리자는 누구든지 특정 개인을 알아보기 위한 목적으로 가명정보를 처리*해서는 안 됨 (법 제28조의5제1항)

– 가명정보 처리 중 우연히 특정 개인이 식별되는 경우 등 목적성이 없는 경우는 해당되지 않음

■ 가명정보처리자는 가명정보 처리 과정에서 개인식별 가능성이 증가하는지 여부 등을 지속적으로 모니터링 하여 안전하게 처리하여야 함(법 제28조의5제2항)

– 특정 개인이 식별되는 경우 즉시 처리중지, 회수, 파기 등 위와 같은 위험을 제거하기 위해 적절한 조치를 수행하여야 함

■ 가명정보는 추가정보의 분리 보관, 접근권한의 분리, 기록 작성/보관 및 공개의 의무를 준수하여야 하며, 구체적인 사항은 [Ⅴ.가명정보의 안전한 관리] 참조할 수 있음

Ⅳ 가명정보 결합

V 가명정보의 안전한 관리

개인정보처리자가 가명정보를 처리하는 경우에는 원래의 상태로 복원하기 위한 추가정보를 별도로 분리하여 보관 · 관리하여야 하고, 가명정보가 분실 · 도난 · 유출 · 위조 · 변조 또는 훼손되지 않도록 안전성 확보에 필요한 관리적 · 기술적 · 물리적 보호조치를 적용하여야 함

1 가명정보의 관리적 보호조치

① 개인정보처리자는 가명정보 및 추가정보를 안전하게 관리하기 위한 내부 관리계획을 수립 · 시행하여야 함 (시행령 제29조의5)①항1호

- 내부 관리계획에는 추가정보의 별도 분리 보관 및 이에 대한 접근권한 분리에 대한 사항 등을 포함하여야 함

〈가명정보처리 내부 관리계획에 포함될 사항 (예시)〉

가. 가명정보 또는 추가정보의 관리책임자 지정에 관한 사항
나. 추가정보 별도 분리 보관
다. 가명정보 또는 추가정보의 안전성확보조치에 관한 사항
라. 가명정보처리자의 교육에 관한 사항
마. 가명정보 처리 기록 작성 및 보관에 관한 사항
바. 개인정보 처리방침 공개에 관한 사항
사. 가명정보의 재식별 금지에 관한 사항
※ 상기 내용에 포함되지 않은 항목은 '개인정보 안전성 확보조치' 해설서 참조

- 가명정보처리자는 내부 관리계획에서 정한 사항에 중요한 변경이 있는 경우 이를 즉시 반영하여 내부관리계획을 수정 · 시행하고, 관리책임자는 연 1회 이상 내부 관리계획 이행 실태를 점검 · 관리 하여야 함

② 수탁자 관리 · 감독의 의무 (시행령 제28조)

- 가명정보 처리업무를 외부에 위탁하는 경우, 가명정보도 개인정보에 해당하므로 법 제26조에 따라 위탁업무 수행 목적 외 가명정보의 처리 금지에 관한 사항 등을 포함한 문서를 작성하여야 함

- 또한, 위탁하는 업무의 내용과 가명정보 처리업무를 위탁받아 처리하는 자를 공개하여야하며, 업무 위탁으로 인하여 정보주체의 가명정보가 분실 · 도난 · 유출 · 위조 · 변조 · 훼손 또는 재식별되지 아니하도록 수탁자를 교육하고, 처리현황 점검 등 수탁자가 가명정보를 안전하게 처리하는지를 감독하여야 함

〈가명정보 처리업무 위탁계약서에 포함되어야 할 사항 (예시)〉

구분	위탁계약서에 포함되어야 할 사항
재식별 금지	가명정보를 제공받거나 처리를 위탁 받은 사업자 등은 다른 정보와 결합을 통해 재식별 시도가 금지됨을 명시
재제공 또는 재위탁 제한	가명정보를 제공하거나 처리를 위탁하는 자는 재제공 또는 재위탁 가능 범위를 정하여 계약서에 명시
재식별 위험 발생시 통지	가명정보가 재식별 되었거나, 재식별 가능성이 높아지는 상황이 발생한 경우에는 가명정보 처리 중지 및 위탁자에게 통지 의무 명시

〈가명정보 처리업무 위탁계약서 특수조건 반영 사례 (예시)〉

제○○조(재식별 금지)

① ○은 △으로부터 제공받은 가명정보를 ××의 목적으로 안전하게 이용하고, 이를 이용해서 개인을 재식별하기 위한 어떠한 행위도 하여서는 아니 된다.

② △으로부터 제공받은 정보의 재위탁은 원칙적으로 금지한다. 다만 불가피한 사유로 이를 재위탁하고자 하는 경우에는 사전에 △의 동의를 얻어야 하며, 이 경우 ○는 재식별 방지를 위해 필요한 조치를 하여야 한다.

③ ○은 △으로부터 제공받은 정보가 재식별 되거나 재식별 가능성이 현저하게 높아지는 상황이 발생하면 즉시 해당 정보의 처리를 중단하고 관련 사항을 △에게 알리며, 필요한 협조를 하여야 한다.

④ ○은 제1항에서 제3항까지의 사항을 이행하지 않아 발생하는 모든 결과에 대해 형사 및 민사상 책임을 진다.

※ 가명정보를 제공받은 기업은 "○", 제공한 기업은 "△"로 표시

2 가명정보의 기술적 보호조치

■ 가명정보처리자는 추가정보의 분리 보관, 접근권한 관리, 접근통제 및 접속기록의 보관 및 점검 등의 기술적 보호조치를 하여야 함

① 추가정보의 분리 보관 (시행령 제29조의5)①항2호

- 추가정보는 가명정보와 분리하여 별도로 저장 · 관리하고 가명정보와 불법적으로 결합되어 재식별에 악용되지 않도록 접근 권한을 최소화하고 접근통제를 강화하는 등 필요한 조치를 적용하여야 함
  - 추가정보와 가명정보는 분리하여 보관하는 것을 원칙으로 하고, 불가피한 사유로 물리적인 분리가 어려운 경우 DB 테이블 분리 등 논리적으로 분리*하는 것도 가능 함
    * 논리적으로 분리할 경우 엄격한 접근통제를 적용하여야 함
    ※ 추가정보의 활용 목적달성 및 불필요한 경우에는 추가정보를 파기할 수 있으며, 이 경우 파기에 대한 기록을 작성하고 보관할 필요가 있음

② 접근권한의 분리 (시행령 제29조의5)①항3호

- 가명정보처리자는 가명정보 또는 추가정보에 접근할 수 있는 담당자를 가명정보 처리 업무 목적달성에 필요한 최소한의 인원으로 엄격하게 통제하여야 하며, 접근권한도 업무에 따라 차등부여 하여야 함
  - 가명정보를 취급할 자를 추가로 둘 여력이 없는 경우 등 접근권한의 분리가 어려운 정당한 사유가 있는 경우*에는 업무 수행에 필요한 최소한 접근 권한 부여 및 접근권한의 보유 현황을 기록으로 보관하는 등 접근권한을 관리 · 통제하여야 함
    * 「소상공인 보호 및 지원에 관한 법률」 제2조에 따른 소상공인 등
  - 가명정보취급자가 가명처리하는 시스템(이하 '가명정보처리시스템') 외의 특정 개인을 알아 볼 수 있는 다른 개인정보처리시스템에 접근할 수 없도록 권한을 제한할 필요가 있음

- 전보 또는 퇴직 등 인사이동이 발생하여 가명정보취급자가 변경되었을 경우 지체 없이 가명정보를 처리하는 시스템의 접근 권한을 변경 또는 말소하여야 함

- 가명정보처리시스템의 접근 권한 부여, 변경 또는 말소에 대한 내역을 기록하고, 그 기록을 최소 3년간 보관하여야 함

- 가명정보처리시스템에 접속할 수 있는 사용자계정을 발급하는 경우 가명정보취급자 별로 사용자계정을 발급하여야 하며, 다른 가명정보취급자 및 개인정보취급자와 공유되지 않도록 하여야 함

- 가명정보취급자가 안전한 비밀번호를 설정하여 이행할 수 있도록 비밀번호 작성규칙을 수립하여 적용하여야 함

- 권한 있는 가명정보취급자만이 가명정보처리시스템에 접근할 수 있도록 계정정보 또는 비밀번호를 일정 횟수 이상 잘못 입력한 경우 가명정보처리시스템에 대한 접근을 제한하는 등 필요한 기술적 조치를 하여야 함

③ 가명정보 기록 작성 · 보관 및 공개

- 가명정보처리자는 가명정보의 처리목적, 가명처리한 개인정보 항목, 가명정보의 이용내역, 제3자 제공 시 제공받는 자를 작성하여 보관하여야 함(시행령 제29조의5)②

- 가명정보처리자는 가명정보 처리와 관련하여 아래와 같은 내용을 개인정보 처리방침에 포함하여 공개하여야 함(법 제30조)

〈가명정보 활용 관련 개인정보처리방침에 포함될 사항 (예시)〉

1. 가명정보 처리 목적
2. 가명정보 처리 및 보유 기간(선택)
3. 가명정보 제3자 제공에 관한 사항(해당되는 경우에만 정한다)
4. 가명정보 처리의 위탁에 관한 사항(해당되는 경우에만 정한다)
5. 처리하는 가명정보의 항목
6. 법 제28조의4(가명정보에 대한 안전조치의무 등)에 따른 가명정보의 안전성 확보 조치에 관한 사항

3 가명정보의 물리적 보호조치

■ 가명정보처리자는 가명정보 또는 추가정보의 안전한 관리를 위하여 물리적 안전조치를 취하여야 함

- 가명정보 또는 추가정보가 전산실이나 자료보관실에 보관되어 있는 경우에는 비인가자의 접근으로부터 보호하기 위하여 출입 통제 등의 절차를 수립하여야 함

- 또한 가명정보 또는 추가정보가 보조저장매체 등에 저장되어 있는 경우 잠금장치가 있는 안전한 장소에 보관하여야 하며, 이러한 보조저장매체 등에 대한 반 · 출입 통제를 위한 보안대책을 마련하여야 함

참고자료

1. 개인정보 가명처리 기술 및 예시
2. 특이정보 정의 및 처리사례
3. 가명정보 내부결합 절차

참고 1 개인정보 가명처리 기술 및 예시

개인정보의 가명 · 익명처리 기술 종류

※ 아래 분류는 이해를 돕기 위해 2016년 개인정보 비식별조치 가이드라인, ISO/IEC 20889, 그리고 EU ENISA에서 발간한 보고서[1] 등 국내 · 외 자료들을 참고하여 작성했으며 표준이 아닙니다.

분류	기술	세부기술	설명
개인정보 삭제	삭제기술	삭제 (Suppression)	• 원본정보에서 개인정보를 단순 삭제
		부분삭제 (Partial suppression)	• 개인정보 전체를 삭제하는 방식이 아니라 일부를 삭제
		행 항목 삭제 (Record suppression)	• 다른 정보와 뚜렷하게 구별되는 행 항목을 삭제
		로컬 삭제 (Local suppression)	• 특이정보를 해당 행 항목에서 삭제
개인정보 일부 또는 전부 대체		마스킹 (Masking)	• 특정 항목의 일부 또는 전부를 공백 또는 문자(' * ', ' _ ' 등이나 전각 기호)로 대체
	통계도구	총계처리 (Aggregation)	• 평균값, 최댓값, 최솟값, 최빈값, 중간값 등으로 처리
		부분총계 (Micro aggregation)	• 정보집합물 내 하나 또는 그 이상의 행 항목에 해당하는 특정 열 항목을 총계처리. 즉, 다른 정보에 비하여 오차 범위가 큰 항목을 평균값 등으로 대체
	일반화 (범주화) 기술	일반 라운딩 (Rounding)	• 올림, 내림, 반올림 등의 기준을 적용하여 집계 처리하는 방법으로, 일반적으로 세세한 정보보다는 전체 통계정보가 필요한 경우 많이 사용
		랜덤 라운딩 (Random rounding)	• 수치 데이터를 임의의 수인 자리 수, 실제 수 기준으로 올림(round up) 또는 내림(round down)하는 기법
		제어 라운딩 (Controlled rounding)	• 라운딩 적용 시 값의 변경에 따라 행이나 열의 합이 원본의 행이나 열의 합과 일치하지 않는 단점을 해결하기 위해 원본과 결과가 동일하도록 라운딩을 적용하는 기법
		상하단코딩 (Top and bottom coding)	• 정규분포의 특성을 가진 데이터에서 양쪽 끝에 치우친 정보는 적은 수의 분포를 가지게 되어 식별성을 가질 수 있음 • 이를 해결하기 위해 적은 수의 분포를 가진 양 끝단의 정보를 범주화 등의 기법을 적용하여 식별성을 낮추는 기법

1) EU ENISA(European Union Agency for Network and Information Security), Recommendations on shaping technology according to GDPR provisions, An overview on data pseudonymisatin, November 2018

EU ENISA(European Union Agency for Network and Information Security), Pseudonymisation and best practices, November 2019

개인정보 일부 또는 전부 대체	일반화 (범주화) 기술	로컬 일반화 (Local generalization)	• 전체 정보집합물 중 특정 열 항목(들)에서 특이한 값을 가지거나 분포상 의 특이성으로 인해 식별성이 높아지는 경우 해당 부분만 일반화를 적용하여 식별성을 낮추는 기법
		범위 방법 (Data range)	• 수치 데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로, 해당 값의 범위 또는 구간(interval)으로 표현
		문자데이터 범주화 (Categorization of character data)	• 문자로 저장된 정보에 대해 보다 상위의 개념으로 범주화 하는 기법
	암호화	양방향 암호화 (Two-way encryption)	• 특정 정보에 대해 암호화와 암호화된 정보에 대한 복호화가 가능한 암호화 기법 • 암호화 및 복호화에 동일 비밀키로 암호화하는 대칭키(Symmetric key) 방식과 공개키와 개인키를 이용하는 비대칭키(Asymmetric key) 방식으로 구분
		일방향 암호화 – 암호학적 해시함수 (One-way encryption – Cryptographic hash function)	• 원문에 대한 암호화의 적용만 가능하고 암호문에 대한 복호화 적용이 불가능한 암호화 기법 • 키가 없는 해시함수(MDC, Message Digest Code), 솔트(Salt)가 있는 해시함수, 키가 있는 해시함수(MAC, Message Authentication Code)로 구분 • 암호화(해시처리)된 값에 대한 복호화가 불가능하고, 동일한 해시 값과 매핑(mapping)되는 2개의 고유한 서로 다른 입력값을 찾는 것이 계산상 불가능하여 충돌 가능성이 매우 적음
		순서보존 암호화 (Order-preserving encryption)	• 원본정보의 순서와 암호값의 순서가 동일하게 유지되는 암호화 방식 • 암호화된 상태에서도 원본정보의 순서가 유지되어 값들 간의 크기에 대한 비교 분석이 필요한 경우 안전한 분석이 가능
		형태보존 암호화 (Format-preserving encryption)	• 원본 정보의 형태와 암호화된 값의 형태가 동일하게 유지되는 암호화 방식 • 원본 정보와 동일한 크기와 구성 형태를 가지기 때문에 일반적인 암호화가 가지고 있는 저장 공간의 스키마 변경 이슈가 없어 저장 공간의 비용 증가를 해결할 수 있음 • 암호화로 인해 발생하는 시스템의 수정이 거의 발생하지 않아 토큰화, 신용카드 번호의 암호화 등에서 기존 시스템의 변경 없이 암호화를 적용할 때 사용
		동형 암호화 (Homomorphic encryption)	• 암호화된 상태에서의 연산이 가능한 암호화 방식으로 원래의 값을 암호화한 상태로 연산 처리를 하여 다양한 분석에 이용가능 • 암호화된 상태의 연산값을 복호화 하면 원래의 값을 연산한 것과 동일한 결과를 얻을 수 있는 4세대 암호화 기법
		다형성 암호화 (Polymorphic encryption)	• 가명정보의 부정한 결합을 차단하기 위해 각 도메인별로 서로 다른 가명처리 방법을 사용하여 정보를 제공하는 방법 • 정보 제공 시 서로 다른 방식의 암호화된 가명처리를 적용함에 따라 도메인별로 다른 가명정보를 가지게 됨
	무작위화 기술	잡음 추가 (Noise addition)	• 개인정보에 임의의 숫자 등 잡음을 추가(더하기 또는 곱하기)하는 방법

개인정보 일부 또는 전부 대체	무작위화 기술	순열(치환) (Permutation)	• 분석 시 가치가 적고 식별성이 높은 열 항목에 대해 대상 열 항목의 모든 값을 열 항목 내에서 무작위로 순서를 변경하여 식별성을 낮추는 기법 • 개인정보를 다른 행 항목의 정보와 무작위로 순서를 변경하여 전체정보에 대한 변경 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 방법
		토큰화 (Tokenisation)	• 개인을 식별할 수 있는 정보를 토큰으로 변환 후 대체함으로써 개인정보를 직접 사용하여 발생하는 식별 위험을 제거하여 개인정보를 보호하는 기술 • 토큰 생성 시 적용하는 기술은 의사난수생성 기법이나 양방향 암호화, 형태보존 암호화 기법을 주로 사용
		(의사)난수생성기 ((P)RNG, (Pseudo) Random Number Generator)	• 주어진 입력값에 대해 예측이 불가능하고 패턴이 없는 값을 생성하는 메커니즘으로 임의의 숫자를 개인정보와 대체
가명 · 익명처리를 위한 다양한 기술 (기타 기술)		표본추출 (Sampling)	• 데이터 주체별로 전체 모집단이 아닌 표본에 대해 무작위 레코드 추출 등의 기법을 통해 모집단의 일부를 분석하여 전체에 대한 분석을 대신하는 기법
		해부화 (Anatomization)	• 기존 하나의 데이터셋(테이블)을 식별성이 있는 정보집합물과 식별성이 없는 정보집합물로 구성된 2개의 데이터셋으로 분리하는 기술
		재현데이터 (Synthetic data)	• 원본과 최대한 유사한 통계적 성질을 보이는 가상의 데이터를 생성하기 위해 개인정보의 특성을 분석하여 새로운 데이터를 생성하는 기법
		동형비밀분산 (Homomorphic secret sharing)	• 식별정보 또는 기타 식별가능정보를 메시지 공유 알고리즘에 의해 생성된 두 개 이상의 쉐어(share)*로 대체 *기밀사항을 재구성하는데 사용할 수 있는 하위 집합
		차분 프라이버시 (Differential privacy)	• 특정 개인에 대한 사전지식이 있는 상태에서 데이터베이스 질의(Query)에 대한 응답 값으로 개인을 알 수 없도록 응답 값에 임의의 숫자 잡음(Noise)을 추가하여 특정 개인의 존재 여부를 알 수 없도록 하는 기법 • 1개 항목이 차이나는 두 데이터베이스간의 차이(확률분포)를 기준으로 하는 프라이버시 보호 모델

개인정보의 가명 · 익명처리 예시

※ 아래 모든 예시는 각 기법의 적용에 대한 예시이며 전체 데이터에 대한 가명 · 익명처리에 대한 예시가 아닙니다.

① 개인정보 삭제

– 삭제기술 : 선택된 항목을 제거하는 기술

❶ 삭제(Suppression) 수치형데이터 문자형데이터

– 원본정보에서 개인정보를 단순 삭제

※ 이때 남아 있는 정보 그 자체로도 분석의 유효성을 가져야 함과 동시에 개인을 식별할 수 없어야 하며, 인터넷 등에 공개되어 있는 정보 등과 결합하였을 경우에도 개인을 식별할 수 없어야 함

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적 포인트
김철수	남	41세	010–6666–8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010–9999–2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010–2222–7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010–3333–4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010–5555–6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성별	나이	통신료	단말기금액	누적포인트
남	41세	98,700	1,198,700	356,800
여	61세	69,400	505,400	203,000
남	30세	104,400	1,604,400	198,000
여	57세	954,800	3,954,800	20,532,000
남	28세	83,600	883,600	400,900

❷ 부분삭제(Partial suppression) 수치형데이터 문자형데이터

– 개인정보 전체를 삭제하는 방식이 아니라 일부를 삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김	남	41세	8888	서울특별시 중구	98,700	1,198,700	356,800
이	여	61세	2222	부산광역시 북구	69,400	505,400	203,000
박	남	30세	7777	광주광역시 서구	104,400	1,604,400	198,000
이	여	57세	4444	전라남도 나주시	954,800	3,954,800	20,532,000
최	남	28세	6666	세종특별자치시	83,600	883,600	400,900

❸ 행 항목 삭제(Record suppression) 수치형데이터 문자형데이터

– 다른 정보와 뚜렷하게 구별되는 행 항목을 삭제

– 통계분석에 있어서 전체 평균에 비하여 오차범위를 벗어나는 자료를 제거할 때 사용

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010-3333-4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적포인트
김철수	남	41세	010-6666-8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010-9999-2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010-2222-7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
최동욱	남	28세	010-5555-6666	세종특별자치시 어진동	83,600	883,600	400,900

❹ 로컬 삭제(Local suppression) 수치형데이터 문자형데이터

– 특이정보를 해당 행 항목에서 삭제

(설명) 다른 누적포인트에 비하여 뚜렷이 구별되는 누적포인트를 항목에서 삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적 포인트
김철수	남	41세	010–6666–8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010–9999–2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010–2222–7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010–3333–4444	전라남도 나주시 빛가람동	954,800	3,954,800	20,532,000
최동욱	남	28세	010–5555–6666	세종특별자치시 어진동	83,600	883,600	400,900

삭제

성명	성별	나이	핸드폰번호	주소	통신료	단말기금액	누적 포인트
김철수	남	41세	010–6666–8888	서울특별시 중구 무교동	98,700	1,198,700	356,800
이영희	여	61세	010–9999–2222	부산광역시 북구 화명동	69,400	505,400	203,000
박민호	남	30세	010–2222–7777	광주광역시 서구 금호동	104,400	1,604,400	198,000
이윤정	여	57세	010–3333–4444	전라남도 나주시 빛가람동	954,800	3,954,800	
최동욱	남	28세	010–5555–6666	세종특별자치시 어진동	83,600	883,600	400,900

❺ 마스킹(Masking) 수치형데이터 문자형데이터

– 특정 항목의 일부 또는 전부를 공백 또는 문자(‘ * ’, ‘ _ ’ 등이나 전각 기호)로 대체

※ 분류는 개인정보 일부 또는 전부 대체로 분류되지만, 기술적으로 마스킹된 부분은 데이터로써의 가치가 없어져 일부 문건에서는 삭제로 분류되기도 함

성명	성별	나이	핸드폰번호
김철수	남	41세	010–6666–8888
이영희	여	61세	010–9999–2222
박민호	남	30세	010–2222–7777
이윤정	여	57세	010–3333–4444
최동욱	남	28세	010–5555–6666

마스킹

성명	성별	나이	핸드폰번호
김 * *	남	4 * 세	* * * – * * * * – * * * *
이 * *	여	6 * 세	* * * – * * * * – * * * *
박 * *	남	3 * 세	* * * – * * * * – * * * *
이 * *	여	5 * 세	* * * – * * * * – * * * *
최 * *	남	2 * 세	* * * – * * * * – * * * *

② 개인정보 일부 또는 전부 대체

– 통계도구 : 데이터의 전체 구조를 변경하는 통계적 성질을 가진 기법

❶ 총계처리(Aggregation) 수치형데이터

– 평균값, 최댓값, 최솟값, 최빈값, 중간값 등으로 처리

※ 단, 데이터 전체가 유사한 특징을 가진 개인으로 구성되어 있을 경우 그 데이터의 대푯값이 특정 개인의 정보를 그대로 노출시킬 수도 있으므로 주의 필요

통신료
98,700
69,400
104,400
954,800
83,600

평균값

통신료
262,180
262,180
262,180
262,180
262,180

통신료
98,700
69,400
104,400
954,800
83,600

최댓값

통신료
954,800
954,800
954,800
954,800
954,800

통신료
98,700
69,400
104,400
954,800
83,600

최솟값

통신료
69,400
69,400
69,400
69,400
69,400

통신료
98,700
69,400
104,400
954,800
104,400

최빈값

통신료
104,400
104,400
104,400
104,400
104,400

통신료
98,700
69,400
104,400
54,800
83,600

정렬

통신료
54,800
69,400
83,600
98,700
104,400

중간값

통신료
83,600
83,600
83,600
83,600
83,600

1–1. 부분총계(Micro Aggregation) 수치형데이터

– 정보집합물 내 하나 또는 그 이상의 행 항목에 해당하는 특정 열 항목을 총계처리즉, 다른 정보에 비하여 오차 범위가 큰 항목을 평균값 등으로 대체

– 동질 집합 내의 특정 항목을 총계처리 하거나 특정 조건에 너무 특이한 값이 있어 개인의 식별 가능성이 높지만 분석에 꼭 필요한 값인 경우 처리

(설명) 지역, 나이 기준으로 동질집합을 형성하고, 오차 범위가 큰 소득금액을 동질집합 내 평균값으로 대체

지역	나이	소득금액
서울	30대	5,987,900
서울	30대	28,169,700
서울	30대	3,009,600
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

지역	나이	소득금액
서울	30대	12,389,067
서울	30대	12,389,067
서울	30대	12,389,067
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

– 일반화기술 : 범주화로도 불리며, 특정한 값을 상위의 속성으로 대체

❶ 라운딩(Rounding) 수치형데이터

1–1. 일반 라운딩

– 올림, 내림, 반올림 등의 기준을 적용하여 집계 처리하는 방법

나이
33세
61세
47세
66세
40세

올림	내림	반올림
40세	30세	30세
70세	60세	60세
50세	40세	50세
70세	60세	70세
40세	40세	40세

※ 적절하지 않은 라운딩의 경우 라운딩 후에도 남은 값의 유일성이 남게 될 수 있으며, 적용하는 단위에 대한 판단이 중요

금액	백 단위 라운딩
983,116,785	983,117,000
984,715,591	984,716,000
984,932,383	984,932,000
985,660,262	985,660,000
986,047,778	986,048,000

적절하지 않은 라운딩

금액	백만 단위 라운딩
983,116,785	980,000,000
984,715,591	980,000,000
984,932,383	980,000,000
985,660,262	990,000,000
986,047,778	990,000,000

적절한 라운딩

1–2. 랜덤 라운딩(Random Rounding) 수치형데이터

– 수치 데이터를 임의의 수인 자리 수, 실제 수 기준으로 올림(round up) 또는 내림(round down)하는 기법

금액		금액
869,250	만 단위 라운딩	900,000
4,559,120	십만 단위 라운딩	4,000,000
13,601,564	십만 단위 라운딩	14,000,000
979,118	만 단위 라운딩	900,000
122,848,878	백만 단위 라운딩	120,000,000

1-3. 제어 라운딩(Controlled rounding) 수치형데이터

– 라운딩 적용 시 값의 변경에 따라 행이나 열의 합이 원본의 행이나 열의 합과 일치하지 않는 단점을 해결하기 위해 원본과 결과가 동일하도록 라운딩을 적용하는 기법

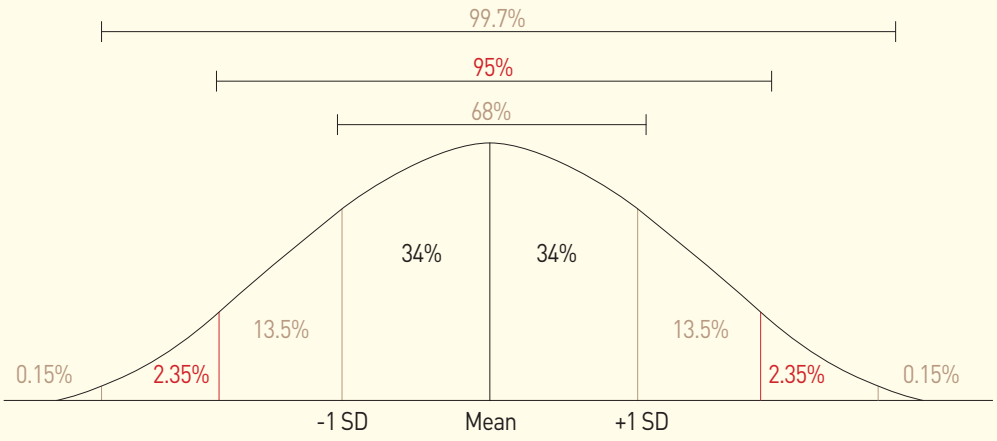
※ 컴퓨터 프로그램으로 구현하기 어렵고 복잡한 통계표에는 적용하기 어려우며, 해결할 수 있는 방법이 존재하지 않을 수 있어 아직 실무에서는 잘 사용하지 않음

(설명) 나이에 대한 평균 분석 시 원본의 경우 평균이 51세가 되나 일반 라운딩을 적용한 경우 평균이 50세가 되어 결과가 다르게 되고, 이에 일부 값을 다르게 라운딩(제어)하여 평균 나이가 원본과 일치되도록 함

원본(나이)	일반 라운딩	제어 라운딩
33세	30세	30세
61세	60세	60세
50세	50세	50세
72세	70세	70세
43세	40세	40세
44세	40세	50세
23세	20세	20세
67세	70세	70세
68세	70세	70세
49세	50세	50세
평균 : 51세	평균 : 50세	평균 : 51세
합계 : 510	합계 : 500	합계 : 510

❷ 상하단코딩(Top and bottom coding) 수치형데이터

– 정규분포의 특성을 가진 데이터에서 양쪽 끝에 치우친 정보는 적은 수의 분포를 가지게 되어 식별성을 가질 수 있으며, 이를 해결하기 위해 적은 수의 분포를 가진 양 끝단의 정보를 범주화 등의 기법을 적용하여 식별성을 낮추는 기법



❸ 로컬 일반화(Local generalization) 수치형데이터

– 전체 정보집합물 중 특정 열 항목(들)에서 특이한 값을 가지거나 분포상의 특이성으로 인해 식별성이 높아지는 경우 해당 부분만 일반화를 적용하여 식별성을 낮추는 기법

(설명) 서울 지역의 30대 중 분포 상 다른 금액에 비해 특이한 값을 동질집합 내 범주화
※ 특이한 로컬(28,169,700)에만 3,009,600 ~ 28,169,700으로 범주화 할 수 있음

지역	나이	소득금액
서울	30대	5,987,900
서울	30대	28,169,700
서울	30대	3,009,600
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

지역	나이	소득금액
서울	30대	3,009,600~28,169,700
서울	30대	3,009,600~28,169,700
서울	30대	3,009,600~28,169,700
나주	30대	4,607,300
나주	30대	3,560,800
나주	30대	2,940,100
세종	30대	6,088,400
세종	30대	2,789,200
세종	30대	5,048,300

❹ 범위 방법(Data range) 수치형데이터

– 수치 데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로, 해당 값의 범위 또는 구간(interval)으로 표현

(예시) 소득 3,300만원을 소득 3,000만원~4,000만원으로 대체 표기

❺ 문자데이터 범주화(Categorization of character data) 문자형데이터

– 문자로 저장된 정보에 대해 상위의 개념으로 범주화하는 기법

품목	품목
분유	육아용품
기저귀	육아용품
젖병	육아용품
샤워타올	육아용품
욕실화	육아용품

– 암호화 : 정보 가공 시 일정한 규칙의 알고리즘을 적용하여 대체

❶ 암호화(Encryption) 수치형데이터 문자형데이터

※ 암호화에 따른 세부적인 내용은 한국인터넷진흥원 암호이용활성화 관련 안내서 참조

1–1. 양방향 암호화(Two–way encryption)

– 특정 정보에 대해 암호화와 암호화된 정보에 대한 복호화가 가능한 암호화 기법

– 암호화 및 복호화에 동일한 비밀키로 암호화하는 AES, ARIA 등 대칭키(Symmetric key) 방식과 공개키와 개인키를 이용하는 RSA 등 비대칭키(Asymmetric key) 방식으로 구분되며, 키(key) 관리에 주의 필요

1–2. 일방향 암호화 – 암호학적 해시함수(One–way encryption – Cryptographic hash function)

– 원문에 대한 암호화의 적용만 가능하고 암호문에 대한 복호화 적용이 불가능한 암호화 기법

– 키가 없는 해시함수(MDC, Message Digest Code), 키가 있는 해시함수(MAC, Message Authentication Code), 솔트(Salt)가 있는 해시함수로 구분

– 암호화(해시처리)된 값에 대한 복호화가 불가능하고, 동일한 해시 값과 매핑(mapping)되는 2개의 고유한 서로 다른 입력값을 찾는 것이 계산상 불가능하여 충돌 가능성이 매우 적음

1–3. 순서보존 암호화(Order–preserving encryption)

– 원본정보의 순서와 암호값의 순서가 동일하게 유지되는 암호화 방식

– 암호화된 상태에서도 원본정보의 순서가 유지되어 값들 간의 크기에 대한 비교 분석이 필요한 경우 안전한 분석이 가능

1–4. 형태보존 암호화(Format–preserving encryption)

– 원본 정보의 형태와 암호화된 암호값의 형태가 동일하게 유지되는 암호화 방식

– 원본 정보와 동일한 크기와 구성 형태를 가지기 때문에 일반적인 암호화가 가지고 있는 저장 공간의 스키마 변경 이슈가 없어 저장 공간의 비용 증가를 해결할 수 있음

– 암호화로 인해 발생하는 시스템의 수정이 거의 발생하지 않아 토큰화, 신용카드 번호의 암호화 등에서 기존 시스템의 변경 없이 암호화를 적용할 때 사용

1–5. 동형 암호화(Homomorphic encryption)

– 암호화된 상태에서의 연산이 가능한 암호화 방식

– 원래의 값을 암호화한 상태로 연산 처리를 하여 다양한 분석에 이용가능

– 암호화된 상태의 연산한 값을 복호화 하면 원래의 값을 연산한 것과 동일한 결과를 얻을 수 있는 4세대 암호화 기법

1–6. 다형성 암호화(Polymorphic encryption)

– 가명정보의 부정한 결합을 차단하기 위해 각 도메인별로 서로 다른 가명처리 방법을 사용하여 정보를 제공하는 방법

– 정보 제공 시 서로 다른 방식의 암호화된 가명처리를 적용함에 따라 도메인별로 다른 가명정보를 가지게 됨

– 무작위화기술 : 속성의 값을 원래의 값과 다르게 변경

❶ 잡음 추가(Noise addition) 수치형데이터 문자형데이터

– 개인정보에 임의의 숫자 등 잡음을 추가(더하기 또는 곱하기)하는 방법

– 지정된 평균과 분산의 범위 내에서 잡음이 추가되므로 원 자료의 유용성을 해치지 않으나, 잡음값은 데이터 값과는 무관하기 때문에 유효한 데이터로 활용하기 곤란하여, 중요한 종적정보는 동일한 잡음을 사용해야함 (예시로 입원일자에 +30이라는 노이즈를 추가하는 경우 퇴원일자에도 +30이라는 노이즈를 부여해야 전체 입원일수에 변화가 없음)

생년월일	잡음추가	잡음추가생년월일
2011–12–05	+3	2011–12–08
2016–08–09	–2	2016–08–07
2009–02–11	–5	2009–02–06
1998–05–27	–6	1998–05–21
1991–06–18	+9	1991–06–27

❷ 순열(치환)(Permutation) 수치형데이터 문자형데이터

– 기존 값은 유지하면서 개인이 식별되지 않도록 데이터를 재배열하는 방법

– 개인정보를 다른 행 항목의 정보와 무작위로 순서를 변경하여 전체정보에 대한 변경 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 방법

※ 데이터의 훼손 정도가 매우 큰 기법으로 무작위로 순서를 변경하는 조건 선정에 주의 필요

(설명) 원본과 비교하여 평균 분석 시 전체 재배열은 결과가 다르며 동질집합 내 재배열 결과는 동일

지역	나이	소득금액(원본)	소득금액(전체 재배열)	소득금액(등질집합 내 재배열)
서울	30대	5,987,900	2,789,200	3,009,600
서울	30대	8,169,700	4,607,300	5,987,900
서울	30대	3,009,600	5,987,900	8,169,700
나주	30대	4,607,300	2,940,100	2,940,100
나주	30대	3,560,800	8,169,700	4,607,300
나주	30대	2,940,100	5,048,300	3,560,800
세종	30대	6,088,400	3,009,600	2,789,200
세종	30대	2,789,200	3,560,800	5,048,300
세종	30대	5,048,300	6,088,400	6,088,400

원본 분석결과	지역	서울	나주	세종
	평균소득	5,722,400	3,702,733	4,641,967
전체 재배열 분석결과	지역	서울	나주	세종
	평균소득	4,461,467	5,048,300	4,219,600
등질집합 내 재배열 분석결과	지역	서울	나주	세종
	평균소득	5,722,400	3,702,733	4,641,967

❸ 토큰화(Tokenisation) 수치형데이터 문자형데이터

– 개인을 식별할 수 있는 정보를 토큰으로 변환 후 대체함으로써 개인정보를 직접 사용하여 발생하는 개인에 대한 식별 위험을 제거하여 개인정보를 보호하는 기술

– 토큰 생성 시 적용하는 기술은 의사난수생성 기법이나 일방향 암호화, 순서보존 암호화 기법을 주로 사용

고객번호	이름	성별	핸드폰번호	나이	회원등급	연간 이용액
D1304365	이공재	남	010–1234–5678	30세	2등급	3,782,459

유사난수 생성기 | 암호화 기법 | 형태보존 암호화

고객번호	이름	성별	핸드폰번호	나이	회원등급	연간 이용액
AD921648	Wzcd88qdp ekfhandkcosekrn	남	159–6857–6384	30세	2등급	3,782,459

❹ (의사)난수생성기((P)RNG, (Pseudo) Random Number Generator) 수치형데이터 문자형데이터

– 주어진 입력 값에 대해 예측이 불가능하고 패턴이 없는 값을 생성하는 메커니즘으로 임의의 숫자를 개인정보에 할당

※ 난수는 원칙적으로 규칙적인 배열순서가 없는 임의의 수를 의미하며 컴퓨터는 원천적으로 입력에 의한 처리 결과를 반환하는 것으로 처리의 방법과 입력이 동일하면 항상 동일한 출력이 발생하기 때문에 완전한 난수의 생성은 불가능

③ 가명 · 익명처리를 위한 다양한 기술 (기타 기술)

❶ 표본추출(Sampling) 수치형데이터 문자형데이터

– 데이터 주체별로 전체 모집단이 아닌 표본에 무작위 레코드 추출 등의 기법을 통해 모집단의 일부를 분석하여 전체에 대한 분석을 대신하는 기법

– 확률적 표본추출 방법과 비확률적 표본추출 방법으로 나누어지며, 확률적 표본추출이 통계적 분석에 많이 사용

– 확률적 표본추출 : 무작위 표본추출(복원 표본추출, 비 복원 표본추출), 계통적 표본추출, 층화 표본추출, 집락 표본추출 등

– 비확률적 표본부출 : 임의 표본추출, 판단 표본추출, 할당 표본추출, 누적 표본추출 등

❷ 해부화(Anatomization) 수치형데이터 문자형데이터

– 기존 하나의 데이터셋(테이블)을 식별성이 있는 정보집합물과 식별성이 없는 정보집합물로 구성된 2개의 데이터셋으로 분리하는 기술

Record ID	이름	성별	나이	월 납입금액	총 납부금액
1	조미선	F	33	817,250	66,300,000
2	홍길병	M	61	4,559,120	327,700,000
3	김영심	F	50	13,601,564	41,300,000
4	이미정	F	70	979,118	64,600,000
5	김경태	M	40	5,501,809	23,549,000
6	유영근	M	43	609,622	13,900,000

Record ID	이름	성별	나이
1	조미선	F	33
2	홍길병	M	61
3	김영심	F	50
4	이미정	F	70
5	김경태	M	40
6	유영근	M	43

Record ID	월 납입금액	총 납부금액
1	817,250	66,300,000
2	4,559,120	327,700,000
3	13,601,564	41,300,000
4	979,118	64,600,000
5	5,501,809	23,549,000
6	609,622	13,900,000

❸ 재현데이터(Synthetic data) 수치형데이터 문자형데이터

– 원본과 최대한 유사한 통계적 성질을 보이는 가상의 데이터를 생성하기 위해 개인정보의 특성을 분석하여 새로운 데이터를 생성하는 기법

※ 원본 데이터 포함 여부에 따라 완전 재현 데이터(Fully Synthetic Data), 부분 재현 데이터(Partially Synthetic Data), 하이브리드 재현 데이터(Hybrid Synthetic Data)로 구분

❹ 동형비밀분산(Homomorphic secret sharing) 수치형데이터 문자형데이터

– 식별정보 또는 기타 식별가능정보를 메시지 공유 알고리즘에 의해 생성된 두 개 이상의 쉐어(share)*로 대체

*기밀사항을 재구성 하는 데 사용할 수 있는 하위 집합

※ 재식별은 가명 · 익명처리된 데이터의 쉐어를 소유한 모두가 동의하는 경우만 가능

❺ 차분 프라이버시(Differential privacy) 수치형데이터 문자형데이터

– 특정 개인에 대한 사전지식이 있는 상태에서 해당정보가 포함된 데이터베이스와 포함되지 않은 데이터베이스 질의(Query)에 대한 응답 값으로 개인을 알 수 없도록 응답 값에 임의의 숫자 잡음(Noise)을 추가하여 특정 개인의 존재 여부를 알 수 없도록 하는 기법

– 1개 항목이 차이나는 두 데이터베이스간의 차이(확률분포)를 기준으로 하는 프라이버시 보호 모델

※ 질의응답 값을 확률적으로 일정 크기 이하의 차이를 갖도록 함으로써 차이에 따른 차분 공격 방지

참고 2 특이정보 정의 및 처리사례

필요성

- 개인정보를 가명처리를 통해 특정 개인을 알아볼 수 없게 처리했더라도 '특이정보'를 통해 다른 정보와 쉽게 결합하여 개인을 알아 볼 수 있음
  - 따라서, 특이정보의 유형 등을 살펴보고 가명정보 내 해당 유형의 정보가 존재하고 있는지 검토할 필요가 있음

※ 특이정보는 관측된 데이터의 범위에서 많이 벗어난 아주 작은 값이나 아주 큰 값을 의미

특이정보 사례

- 특정 기관의 급여가 2천만원에서 6천만원까지 고루 분포되어 있는데, 일부 고액 급여 수령자가 발생하는 경우
- 특정 직업의 소속인원이 전국에서 약 300명 정도로 추정되는데, 지역에 극소수(1~2인)만 존재하고 있는 경우
- 정보공개 규정에 따라 공개되는 정보에서 특정 나이대가 현저하게 적게 나타나는 경우우

특이정보 관찰 방법

- 정보의 특이정보는 3시그마규칙 또는 도수분포표 등을 이용하여 검토할 수 있음
  - 3시그마 규칙 : 68–95–99.7규칙이라고도 하며, 정보의 분포의 3시그마(표준편차) 범위에 거의 모든 값들(99.7%)가 들어가는 것을 의미
  - 도수분포표 : 항목에 대한 값을 적당한 범위로 분류하고, 각 범위에 해당하는 수량을 조사하여 표로 나타내는 것을 의미

• 급여

직원	급여(만원)
직원1	2,200
직원2	3,400
직원3	4,600
직원4	5,300
직원5	10,000
직원6	6,700

※ 3시그마 규칙을 이용 하여 표준편차에 벗어난 특이정보 검토

• 지역, 직업

주소	직업	빈도
경기	국회의원	5
경기	국회의원	5
강원	국회의원	1
경기	국회의원	5
경기	국회의원	5
경기	국회의원	5

※ 지역에 대한 도수분포 (빈도)를 이용하여 특이정보 검토

• 나이

나이(세)	빈도
10~20	4
20~30	11
30~40	21
40~50	18
50~60	5
60~70	1

※ 특정 나이에 도수분포(빈도)를 측정하여 특이정보 검토

특이정보 처리 사례

삭제 기법을 활용한 목적별 사용 예시

– 분석 목적에 해당 정보가 없어도 분석에 크게 영향이 없는 경우에만 가능한 기법, 해당 특이정보를 삭제하여 개인 식별성을 제거

가. 로컬 삭제(Local suppression) : 일반적으로 특이정보 처리에 많이 사용되는 기법으로 도수분포표를 활용하여 빈도가 적은 항목을 삭제하여 처리하는 방법

〈로컬삭제 기법 예시〉

나이	주소	직업	월소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	변호사	1,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	1,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	울릉도	변호사	200만원

나이	주소	직업	소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	변호사	1,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	1,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	Null	변호사	200만원

나. 행 삭제(Record suppression) : 특이정보로 인해 개인의 식별가능성이 있는 경우 사용되는 기법으로 특이정보를 가지고 있는 행 전체를 삭제하여 처리하는 방법

※ 통계 분석에서 특이정보는 분석 목적을 달성하기보다 분석의 목적을 저해하는 요소로 작용하는 경우가 있으며, 이 경우 행 삭제 기법이 가장 적절한 기법이 될 수 있음

〈레코드 삭제 기법 예시〉

나이	주소	직업	월소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	변호사	1,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	7,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

나이	주소	직업	소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
35	서울	변호사	1,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

○ 통계적 기법의 종류와 목적별 사용 예시

– 분석 목적에 특이정보를 가지고 있는 해당 정보가 필요한 경우 활용하는 기법으로, 해당 특이정보를 통계적인 방법을 통해 통계값으로 변경하여 사용

가. 단일 속성으로 대체(Combining a set of attributes into a single attribute) : 숫자형 정보가 아닌 경우(문자형 등) 주로 사용되는 방법으로 분류군의 상위로 묶어 처리하는 방법

※ 특정한 직업이 희귀하여 개인의 식별이 가능한 경우 상위의 분류로 변경하여 사용함으로 희귀성을 제거

〈단일속성 대체 예시〉

나이	주소	직업	월소득	나이	주소	직업	소득
35	서울	변호사	600만원	35	서울	변호사	600만원
35	서울	변호사	700만원	35	서울	변호사	700만원
35	서울	변호사	500만원	35	서울	변호사	500만원
35	서울	변호사	700만원	35	서울	변호사	700만원
35	서울	판사	1,200만원	35	서울	법조인	1,200만원
35	경기	검사	800만원	35	경기	법조인	800만원
35	경기	변호사	600만원	35	경기	변호사	600만원
35	경기	변호사	1,300만원	35	경기	변호사	1,300만원
35	경기	변호사	300만원	35	경기	변호사	300만원
35	경기	변호사	900만원	35	경기	변호사	900만원
35	경기	변호사	800만원	35	경기	변호사	800만원
35	경기	변호사	200만원	35	경기	변호사	200만원

나. 로컬 일반화(Local generalization) : 선택한 행에서 일부 특정 값을 일반화하여 활용하는 기법으로, 다른 행의 속성값은 수정하지 않고 희귀 값을 가진 속성값만 처리하여 사용

〈로컬 일반화(상단 코딩) 기법 예시〉

나이	주소	직업	월소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
36	서울	변호사	1,200만원
36	경기	변호사	800만원
36	경기	변호사	600만원
36	경기	변호사	1,300만원
37	경기	변호사	300만원
…	…	…	…
84	경기	변호사	800만원
88	경기	변호사	200만원

나이	주소	직업	소득
35	서울	변호사	600만원
35	서울	변호사	700만원
35	서울	변호사	500만원
35	서울	변호사	700만원
36	서울	변호사	1,200만원
36	경기	변호사	800만원
36	경기	변호사	600만원
36	경기	변호사	1,300만원
37	경기	변호사	300만원
…	…	…	…
80초과	경기	변호사	800만원
80초과	경기	변호사	200만원

다. 부분 총계(Micro Aggregation) : 부분 총계는 일부(특정그룹 값의 합)속성에서 정확한 통계적 값을 확인하는 기법으로, 로컬일반화 보다 일부 속성에서 정확한 값을 알 수 있음

〈부분 총계 기법 예시〉

나이	주소	직업	월소득
35	경기	변호사	600만원
35	경기	변호사	700만원
35	경기	변호사	500만원
35	경기	변호사	700만원
35	경기	변호사	6,200만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	7,300만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

나이	주소	직업	소득
35	경기	변호사	600만원
35	경기	변호사	700만원
35	경기	변호사	500만원
35	경기	변호사	700만원
35	경기	변호사	6,750만원
35	경기	변호사	800만원
35	경기	변호사	600만원
35	경기	변호사	6,750만원
35	경기	변호사	300만원
35	경기	변호사	900만원
35	경기	변호사	800만원
35	경기	변호사	200만원

참고 3 가명정보 내부결합 절차

- 개인정보처리자가 보유한 개인정보를 내부에서 가명처리하여 결합을 하는 경우 별도의 외부 결합전문기관을 거치지 않음

- 안전한 결합을 위하여 결합전문기관의 결합 절차*와 유사하게 처리하는 것을 권고

  * 개인정보를 보유한 자가 결합키 생성 및 가명처리를 하고, 제3자가 결합하는 방식

  – 내부결합에 대하여는 법령에서 별도로 정하고 있지 않지만, 결합 과정에서 가명정보가 재식별되지 않도록 유의하여야 함

내부결합 처리 절차

〈가명정보 내부결합 단계별 절차도〉

- 'Ⅲ. 가명처리'의 [단계1 (p.12)]에서 각 정보를 결합할 때 결합키로 활용될 공통 항목을 선정하고, 이 항목을 결합키로 바꾸기 위한 결합알고리즘(암호종류+salt포함)을 선정

  – 선정된 결합알고리즘을 이용하여 결합키 생성

〈내부결합 결합키 생성 절차 (예시)〉

- 결합키 생성 항목 정의
  - A부서 : 이름, 휴대전화번호, 이메일, 성별, 주소, 구매상품코드, 금액 등
  - B부서 : 이름, 휴대전화번호, 나이, 성별, 가입상품명, 연체금액 등

- 양 부서는 SHA256해시함수에 salt값을 추가하여 결합키를 생성하기로 함
  ⇨ A, B부서가 동일하게 가지고 있는 이름, 휴대전화전호, 성별을 결합키 항목으로 선정하고 SHA256+salt값을 적용하여 일방향 암호화 처리

- 결합키를 제외한 나머지 정보에 대하여 본 가이드라인 'Ⅲ. 가명처리'의 [단계1~단계3]을 참고하여참고하여 가명처리 수행
- 결합을 수행할 부서에서 결합할 가명정보를 모두 제공받아 결합키를 이용하여 결합 수행

- 가명정보를 결합한 정보에 대해 본 가이드라인 'Ⅲ. 가명처리'의 [단계3 (p.22)]에 따라 결합된 가명정보의 처리 수준이 적정한지 판단하고 필요하다고 판단한 경우 추가 가명처리를 수행할 수 있음

가명정보 처리 가이드라인

2020년 9월 인쇄

2020년 9월 발행

발행처 : 개인정보보호위원회

• 본 가이드 내용의 무단전재를 금하며, 가공 · 인용할 때는 출처를 밝혀 주시기 바랍니다.

• 본 가이드는 개인정보보호포털(http://privacy.go.kr)에서 무료로 다운받으실 수 있습니다.

KISA 한국인터넷진흥원