//hyu.wiki/%EA%B0%80%EB%AA%85%EC%B2%98%EB%A6%AC

가명처리

2020.10 총무팀에서 제공한 '개인정보취급자를 위한 가명처리 가이드라인'을 바탕으로 한 문서이다.

- 파일:가명처리 가이드라인.pdf

□

목차

가명처리의 정의 2쪽

법률적 정의

- 가명처리
    - 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 **추가 정보없이는 특정개인을 알아볼 수 없도록 처리**하는 것
- 가명정보
    - 개인정보를 가명처리 함으로써 원래의 상태로 복원하기 위한 추가정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보
    - 가명정보도 개인정보의 범주에 포함
- 추가정보
    - 개인정보의 전부 또는 일부를 대체하는데 이용된 수단이나 방식(알고리즘, Salt 값 등), 가명정보와의 비교·대조 등을 통해 삭제 또는 대체된 개인정보 부분을 복원할 수 있는 정보(매핑 테이블 정보, 가명처리에 사용된 개인정보 등)
    - 추가정보(원본정보와 알고리즘·매핑테이블 정보 등)와 가명정보는 시행령 제30조 또는 제48조의2에 따른 안전성 확보조치 및 각각 정보의 분리보관, 접근 권한의 분리를 하여야 함
- 특이정보

    - 다른 데이터와 확연히 구분되거나 비정상적으로 데이터의 분포를 벗어나 측정이 되는 값으로서, 개인정보 식별과 관련하여 특정 개인의 식별 가능성이 매우 높은 정보
- 가명정보처리자
    - 업무를 목적으로 개인정보를 가명처리하여 활용 또는 제공하는 공공기관, 법인, 단체 및 개인 등

비식별조치의 개념

- 비식별조치의 정의
    - 개별 데이터 또는 정보집합물에서 개인을 식별할 수 있는 요소(식별자, 속성자)의 전부 또는 일부를 삭제하거나 대체하는 등의 방법을 통해 개인을 알아볼 수없도록 하는 조치[1]
- 식별정보와 비식별정보의 구분[2]
    1. 식별정보
        - 살아있는 개인을 식별할 수 있는 정보로서, 특정 개인을 직접 또는 다른 정보와 결합하여 알아볼 수 있는 정보
        - 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보로, 다른 정보의 입수 가능성 등 개인을 알아보는데 소요되는 시간, 비용, 기술 등을 합리적으로 고려해야 함
    2. 개인정보
        1. 비식별 정보
            - 원상태로 복원하기 위한 추가 정보의 사용 결합 없이는 특정 개인을 알아볼 수 없는 정보
            - 통계작성, 연구, 공익적 기록보존의 목적으로 처리 및 개인정보처리자 간 정보집합물 결합이 가능함
        2. 익명 정보
            - 더 이상 개인을 식별할 수 없는 정보
            - 개인정보보호 관련 법령의 적용을 받지 않으며, 해당 데이터에 대한 활용이 광범위하게 가능함

- 식별정보와 비식별정보(가명, 익명)의 예시
    1. 개인정보
        - 홍길동/35세/남성/경기도 분당시 불정로 6거주/식당운영/월소비액 150만원
    2. 가명정보
        - 20번 손님/35세/자영업/경기도 분당시 거주/월소비액 150만원
    3. 익명정보
        - 30대/남성/경기도 분당시 거주/월 소비액 100~200만원
    - 개인정보와 가명정보의 비교
        - **개인정보의 일부를 삭제**하거나(성별 제거), **일부를 대체**하거나(홍길동→20번 손님) 또는 **개인정보**를 범주화하는 방법(식당운영→자영업) 등으로 처리된 개인정보가 추가 정보 없이는 특정 개인을 식별할 수 없는 겨우 가명정보라고 정의함
        - ‘20번 손님 = 홍길동’이라는 추가 정보와 결합하게 된다면 개인을 식별 할 수 있는 개인정보가 되므로 개인정보 보호법(2020.02.04.개정)에 의하여 가명정보를 활용하여야 함
        - [가명정보 처리 준수사항] : 법 제28조의2부터 제28조의7까지 가명정보의 처리에 관한 특례 조항
    - 가명정보와 익명정보의 비교
        - 추가 정보와 결합 시 개인을 식별할 수 있는 것이 가명정보라면 추가 정보 와 결합하더라도 식별 불가능한 것이 익명정보이므로 개정된 개인정보 보호 법(2020.02.04.개정) 제58조의2에 따라 개인정보 보호법의 적용을 받지 않음

재식별의 개념

- 재식별이
  - 비식별화된 개인정보를 다른 정보와 조합, 분석, 또는 처리하여 특정 개인을 다 시 식별할 수 있게 하는 일련의 과정 또는 방법
  - 가명정보는 추가 정보와의 결합 시 식별이 될 수 있는 정보이므로 가명처리 시 에는 재식별의 위험을 줄이기 위한 보안조치를 위해야 함 (보안조치 방법은 '가명처리의 방법'에 설명되어 있음)
- 재식별 관련 제재조치
  1. 법 제28조의 6
     - 위반사항 : 제28조의5젭항을 위반하여 특정 개인을 알아보기 위한 목적으로 가명정보를 처리하는 경우
     - 제제조치 : 전체 매출액의 100분의 3 이하의 과징금
  2. 법 71조
     - 위반사항 : 제28조의5제1항을 위반하여 특정 개인을 알아보기 위한 목적으로 가명정보를 처리하는 경우
     - 제제조치 : 5년 이하의 징역 또는 5천 만원 이하의 벌금

가명처리의 범위 5쪽

가명처리가 가능한 범위

가명처리 목적

- 개인정보처리자는 ‘’’통계작성, 과학적 연구, 공익적’’’’ 기록보존 등을 위하여 정보주 체의 동의 없이 가명정보를 처리할 수 있음 (법 제28조의2 제1항)

통계작성

- 통계란 특정 집단이나 대상 등에 관하여 작성한 수량적인 정보
- 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, 연령, 성별 등)를 생 성 ·분석하여 적합한 지역에 신규 편의시설을 선정하고자 하는 경우

과학적 연구

- 기술의 개발과 실증, 기초연구, 응용연구 및 민간투자 연구 등 과학적 방법을 적용하는 연구
- 과학적 연구는 과학적 방법을 적용하는 연구를 말하며, 자연과학, 사회과 학, 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, 민간투자 연구, 기업 등이 수행하는 연구도 가능
- 예시

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

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

청소년 대상 온라인 상담서비스를 운영하는 민간 기업이 우울증 지수 ， 자살 충동지수가 높은 청소년의 상담정보를

가명처리가 가능한 범위

특정목적에 부합하지 않는 예시

- 누구인지 알아보지 못하도록 이름을 가명처리하고 연락처를 수집하여 홍보를 위해 사용하는 경우

목적을 벗어난 가명처리 시 제재조치

- 법 제71조
  - 위반사항 : 제28조의2제1항을 위반하여 통계작성, 과학적 연구, 공익적 기록보존 등의 목적 이 아님에도 정보주체의 동의 없이 가명 정보를 처리한 경우
  - 제제조치 : 5년 이하의 징역 또는 5천 만원 이하의 벌금

가명처리의 원칙

- 가명처리 대상은 법률에서 허용한 목적 내에서 개인정보를 정보주체의 추가적인 동의 없이 수집 목적 외로 이용 가능
  - 민감정보와 고유식별정보(주민등록번호 제외)도 가명 처리할 수 있음
  - ·고유식별정보는 직접 식별자에 해당하므로 고유식별정보가 남아있거나 역추적이 가능하도록 해서는 안 됨
  - 주민등록번호는 법률 또는 시행령에 구체적인 근거가 없으면 사용이 안 됨
- 개인식별정보(식별자)는 삭제하여야 하나, 결합 등 데이터 이용 목적 상 필요한 경우 안전한 방식으로 대체값을 생성하여 개인식별정보를 대체
- 개인정보를 가명처리하여 발생한 추가정보 삭제
  - 다만, 불가피힌 경우 추가정보는 법령에서 요구하는 분리보관 등 안전성확보조치를 취하여 보관 가능
- 가명처리는 개인정보 보유부서 또는 총괄부서*에서 처리하도록 함

가명정보 이용·제공 시 필요충족요건

가명정보를 추가적으로 이용 또는 제공하려는 경우 다음의 4가지 사항을 고려해야 함

1. 당초 수집 목적과 관련성이 있는지 여부
2. 개인정보를 수집한 정황 또는 처리 관행에 비추어 볼 때 개인정보의 추가적인 이용 또는 제공에 대한 예측 가능성이 있는지 여부
3. 정보주체의 이익을 부당하게 침해하는지 여부
4. 가명처리 또는 암호화 등 안전성 확보에 필요한 조치를 하였는지 여부

가명처리의 절차 8쪽

가명처리 절차도

사전준비 > 가맹처리 > 적정성 검토 및 추가 가명처리 > 활용 및 사후관리

사전준비

가명처리 목적 명확화 및 대상(최소처리원칙 준수) 선정

- 사안에 따라 가명처리 수행에 관한 내부 승인이 필요할 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

가명처리

처리 환경(내부활용, 제3자제공 등) 등 위험도를 고려한 가명처리 수준 설계 및 가명처리 수행

- 내부활용 : 개인정보처리자가 보유한 개인정보를 가명처리 또는 내부 결합하여 직접 활용 또는 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
- 항목별 위험도 분석 : 가명처리 대상에서 개인 식별 가능성이 높은 정보 등을 분류하여 항목별 위험도 분석
- 개인식별 가능성이 높은 정보 예시
    1. 식별정보 : 고유식별정보(여권번호, 외국인등록번호, 운전면허번호), 성명, 전화번호, 전자우편주소, 의료기록번호, 건강보험번호, 자동차 등록번호 등 외 부 연계(식별)를 목적으로 생성된 정보 등
    2. 식별가능정보
        1. 성별, 연령(나이), 국적, 혈액형, 신장, 몸무게, 직업, 위치정보, 세부주소 등 가명정보를 처리하는 자의 입장에서 개인을 알아볼 수 있는 정보
        2. 특이정보
            - 국내 최고령, 최장신, 고액체납금액, 고액급여수급자 등 전체적인 패 턴에서 벗어나 극단값이 발생할 수 있는 정보
            - 희귀 성씨, 희귀 혈액형, 희귀 눈동자 색깔, 희귀 병명, 희귀 직업 등 정보 자체로 특이한 값을 가지고 있는 정보

가명정보 적정성 검토 및 추가 가명처리

설계에 따른 가명처리 결과의 적정성 판단, 미흡한 사항에 대한 추가 가명처리 (2,3단계 반복수행) # 목적달성 가능성과 특이정보를 통한 재식별 가능성 등을 검토

- 개인정보처리자의 판단에 따라 외부전문가로 구성된 적정성 평가단을 구성하여 검토할 수 있음 # 개인을 알아볼 수 없게 처리했더라도 ‘특이정보’를 통해 개인 식별이 가능한 경우 추가처리 검토

활용 및 사후관리

적정성 검토 결과 가명처리가 적정하다고 판단되면 가명정보를 본래 활용 목적을 위해서 처리할 수 있으며, 법령에 따라 기술적·관리적·물리적 안전조치를 이행하여야함

- 가명정보처리자는 가명정보취급자에게 금지행위, 안전조치 등에 관한 사항을 안내하여 가명정보를 안전하게 처리하여야 함

가명정보 제공 시 절차도

가명정보 요청 > 제공 가능 여부 검토 > 가명처리 > 안전성 확보 조치 후 제공 > 가명정보 처리 기록

제공 가능 여부

- ”가명정보 이용·제공 시 필요충족요건” 참조
- 이용목적 및 방법, 재식별 위험관리 등 안전성 확보를 위하여 필요한 조치 를 마련하도록 하는 내용의 계약을 체결할 수 있음
    - 제공이 불가능한 조건의 경우, 가명처리하지 않고 종결

가명처리

“가명처리 절차도”에 의거하여 가명처리

안전성 확보 조치 후 제공

- 암호화 등을 하여 전달
- 원래의 상태로 복원하기 위한 추가 정보를 별도로 분리하고 가명정보만 전달

가명정보 처리 기록

가명정보의 처리 목적, 제3자 제공 시 제공받는 자 등 가 명정보의 처리 내용을 관리하기 위하여 관련 기록을 작성하여 보관하여야 함

- 가명정보 관리 대장 양식 : 해당문서 6.2 별지2. 가명정보 관리대장 참고

개인 식별이 가능한 가명정보의 이용·제공 시 제재조치

1. 법 71조
    - 위반사항 : 제28조의2제2항을 위반하여 가명정보를 제3자에게 제공 시 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함 한 경우
    - 제제조치 : 5년 이하의 징역 또는 5천 만원 이하의 벌금
2. 법 75조
    - 위반사항 : 제28조의5제2항을 위반하여 개인을 알아 볼 수 있는 정보가 생성되었음에도 이용 을 중지하지 않거나 이를 회수, 파기하지 않은 경우
    - 제제조치 : 3천만원 이하의 과태료

가명정보 결합 시 절차도

수요기관 신청 > 결합 > 분석(전문기관) > 반출

- 전문기관
    - 특정 개인을 알아볼 수 없도록 보호위가 고시하는 절차와 방법에 따라 가명정보를 결합하고, 이 과정에서 안전한 결합을 위한 지원 가능
    - 한국인터넷진흥원 등이 결합에 필요한 연계정보를 생성하고 결합기관에 제공

결합 및 반출 절차

- 수요기관이 결합신청서를 제출하면 결합 전문기관에서 가명정보 결합 후 안전성이 확보되니 분석 공간 내 처리 가능, 반출이 필요한 경 우 전문기관 승인 후 반출
- 전문기관에 반출 적정성 심사 위원회(3명 이상)를 구성하여, 반출 여부와 적정한 반출 수준 을 심사
- 가명정보 결합 신청서 양식 : 해당문서 6.1 별지1. 가명정보 결합신청서 참고

전문기관을 통하지 않은 결합 시 제재조치

1. 법 제71조
    - 위반사항 : 제28조의3을 위반하여 전문기관을 통하 지 않고 기관간 가명정보를 결합하는 경 우
    - 제제조치 : 5년 이하의 징역 또는 5천 만원 이하의 벌금

가명처리의 R&A 13쪽

대원칙

원칙적으로 원본 데이터에 관리·책임이 있는 부서만 가명처리 수행 가능

- (예시) 서울캠퍼스 공과대학 학생(학부생)과 관련한 개인정보의 가명처리는 서울 학사팀에서 수행

데이터별 가명처리 가능 부서

학생 데이터

1. 학부생 : 학사팀(서울, ERICA)
2. 대학원생
    - 일반대학원-대학원팀
    - 전문/특수대학원 -각 대학원 행정팀

- 학부생, 일반대학원생에 대한 가명처리는 단 과대학에서 처리할 수 없음
- 창구의 일원화 및 중앙화를 통하여 일관성 있는 판단으로 가명처리 오남용 위험을 방지하고 오류를 최소화하기 위함임.

교원 데이터

- 교무팀(서울, ERICA)

직원 데이터

- 인사팀

(산단) 연구원 데이터

- 산학기회팀(서울)
- 연구진흥팀(ERICA)

기타 데이터

- 상기 분류에 속하지 않는 원본 데이터를 보유하는 기관(또는 팀) 에서 해당 데이터에 대한 가명처리 가능)
- (예) 평생교육시설(미래인재교육원, 사회교육원) 학생 데이터 등

가명처리 가능 부서의 부서장의 역할과 권한

- 가명처리 신청(목적)에 대한 적합성 검토
    1. 절차에 의거한 가명처리
    2. 가명처리 적정성 검토
    3. 가명정보취급자에 대한 관리·감독
    4. 안전성 확보조치의 이행
    5. 기타 본교『개인정보 보호 규정』 제7조 제2항 “분임책임자”의 업무에 준 함
    - 1번, 3번 사항은 외부전문가를 포함한 심의위원회를 구성, 운영할 수 있음
- 권한
    - 관리 데이터에 대하여 별도의 내부 승인 절차 없이 가명정보를 이용·제공 결합 등을 할 수 있다.

가명처리의 방법 15쪽

비식별조치 기준

식별자 조치기준

- 정보집합물에 포함된 식별자는 원칙적으로 삭제 조치
    - ‘식별자’란 개인 또는 개인과 관련한 사물에 고유하게 부여된 값 또는 이름
- 다만, 데이터 이용 목적상 반드시 필요한 식별자는 비식별 조치 후 활용

식별자 예시

- 고유식별정보(주민등록번호, 여권번호, 외국인등록번호, 운전면허번호)
- 성명(한자, 영문 성명, 필명 등 포함)
- 상세 주소(구 단위 미만까지 포함된 주소)
- 날짜정보 : 생일(양/음력), 기념일(결혼, 돌, 환갑 등), 자격증 취득일 등
- 전화번호(휴대전화번호, 집전화, 회사번호, 팩스번호)
- 의료기록번호, 건강보험번호, 복지 수급자 번호
- 통장계좌번호, 신용카드번호
- 각종 자격증 및 면허 번호
- 자동차 번호, 각종 기기의 등록번호 & 일련번호
- 사진(정지사진, 동영상, CCTV 영상 등)
- 신체 식별정보(지문, 음성, 홍채 등)
- 이메일 주소, IP 주소, Mac 주소, 홈페이지 URL 등
- 식별코드(아이디, 사원번호, 고객번호 등)
- 기타 유일 식별번호 : 군번, 개인사업자의 사업자 등록번호 등

속성자 조치기준

- 정보집합물에 포함된 속성자도 데이터 이용 목적과 관련이 없는 경우에는 원칙 적으로 삭제

    - ‘속성자’란 개인과 관련된 정보로서 다른 정보와 쉽게 결합하는 경우 특정 개인을 알아볼 수도 있는 정보
- 데이터 이용 목적과 관련이 있는 속성자 중 식별요소가 있는 경우에는 가명처리, 총계처리 등의 기법을 활용하여 비식별 조치
- 희귀병명, 희귀경력 등의 속성자는 구체적인 상황에 따라 개인 식별 가능성이 매우 높으므로 엄격한 비식별 조치 필요

속성자 예시

- 개인특성
    - 성별, 연력(나이), 국적, 고향, 시/군/구명, 우편번호, 변역여부, 결혼여부, 종교, 취미, 동호회/클럽 등
    - 흡연여부, 음주여부, 채식여부, 관심사항 등
- 신체특성
    - 혈액형, 신장, 몸무게, 허리둘레, 혈압, 눈동자 색깔 등
    - 신체검사 결과, 장애유형, 장애등급 등
    - 병명, 상병( )코드, 투약코드, 진료내역 등
- 신용특성
    - 세금 납부액, 신용등급, 기부금 등
    - 건강보험료 납부액, 소득분위, 의료 급여자 등
- 경력특성
    - 학교명, 학과명, 학년, 성적, 학력 등
    - 경력, 직업, 직종, 직장명, 부서명, 직급, 전직장명 등
- 전자적특성
    - 쿠키정보, 접속일시, 방문일시, 서비스 이용 기록, 접속로그 등
    - 인터넷 접속기록, 휴대전화 사용기록, GPS 데이터 등
- 가족특성
    - 배우자/자녀/부모/형제 등 가족정보, 법정대리인 정보 등

비식별 조치방법

비식별 조치의 대표적 기법은 아래와 같으며 가명처리 시 데이터 속성, 처리 상황에 따 라 5가지 기법을 적절히 구사해야 함.

가명처리 (Pseudonymization)

개인 식별이 가능한 데이터를 직접적으로 식별할 수 없는 다른 값으로 대체하는 기법

- 대상 : 성명, 기타 고유특징(출신학교, 근무처 등)
- 장점 : 데이터의 변형 또는 변질 수준이 적음
- 단점 : 대체 값 부여 시에도 식별 가능한 고유 속성이 계속 유지
- 예시 : 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대, 서울 거주, 국제대 재학
- 세부기술 및 실무적용 방법
    1. 휴리스틱 가명화 (Heuristic Pseudonymization)
        - 식별자에 해당하는 값들을 몇 가지 정해진 규칙으로 대체하거나 사람의 판단에 따 라 가공하여 자세한 개인정보를 숨기는 방법
        - 예시) 성명을 홍길동, 임꺽정 등 몇몇 일반화된 이름으로 대체하여 표기하거나 소 속기관명을 화성, 금성 등으로 대체하는 등 사전에 규칙을 정하여 수행
        - 식별자의 분포를 고려하거나 수집된 자료의 사전 분석을 하지 않고 모든 데이터를 동일한 방법

으로 가공하기 때문에 사용자가 쉽게 이해하고 활용 가능
        - 활용할 수 있는 대체 변수에 한계가 있으며, 다른 값으로 대체하는 일정한 규칙이 노출되는 취약점이 있음. 따라서 규칙 수립 시 개인을 쉽게 식별할 수 없도록 세 심한 고려 필요
        - 적용정보 : 성명, 사용자 ID, 소속(직장)명, 기관번호, 주소, 신용등급, 휴대전화번 호, 우편번호, 이메일 주소 등
    2. 암호화(Encryption)
        - 정보 가공시 일정한 규칙의 알고리즘을 적용하여 암호화함으로써 개인정보를 대체 하는 방법, 통상적으로 다시 복호가 가능하도록 복호화 키(key)를 가지고 있어서 이에 대한 보안방안도 필요
        - 일방향 암호화(one-way encryption 또는 hash)를 사용하는 경우는 이론상 복호화가 원천적으로 불가능
            - 일방향 암호화는 개인정보의 식별성을 완전히 제거하는 것으로, 양방향 암호화 에 비해 더욱 안전하고 효과적인 비식별 기술에 해당
        - 적용정보 : 주민등록번호, 여권번호, 의료보험번호, 외국인등록번호, 사용자 ID, 신 용카드번호, 생체정보 등
    3. 교환 방법(Swapping)
        - 기존의 데이터베이스의 레코드를 사전에 정해진 외부의 변수(항목)값과 연계하여 교환
        - 적용정보 : 사용자 ID, 요양기관번호, 기관번호, 나이, 성별, 신체정보(신장, 혈액형 등), 소득, 휴대전화번호, 주소 등

총계처리 (Aggregation)

통계값(전체 혹은 부분)을 적용하여 특정 개인을 식별할 수 없도록 함

- 대상 : 개인과 직접 관련된 날짜 정보(생일, 자격 취득일), 기타 고유 특징(신체 정보, 진료기록, 병력정보, 특정소비기록 등 민감한 정보)
- 장점 : 민감한 수치 정보에 대하여 비식별 조치가 가능하며, 통계분석용 데이터 셋 작성에 유리함
- 단점 : 정밀 분석이 어려우며, 집계 수량이 적을 경우 추론에 의한 식별 가능성 있음
- 예시 : 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165츠
- 세부기술 및 실무적용 방법
    1. 총계처리(Aggregation)
        - 데이터 전체 또는 부분을 집계(총합, 평균 등)
            - 단, 데이터 전체가 유사한 특징을 가진 개인으로 구성되어 있을 경우 그 데이 터의 대푯값이 특정 개인의 정보를 그대로 노출시킬 수도 있으므로 주의
            - 예시) 집단에 소속된 전체 인원의 평균 나이값을 구한 후 각 개인의 나이값을 평 균 나이값(대푯값)으로 대체하거나 해당 집단 소득의 전체 평균값을 각 개 인의 소득값으로 대체
        - 적용정보 : 나이, 신장, 소득, 카드사용액, 유동인구, 사용자수, 제품 재고량, 판매량 등
    2. 적용정부분총계(Micro Aggregation)
        - 데이터 셋 내 일정부분 레코드만 총계 처리함. 즉, 다른 데이터 값에 비하여 오차 범위가 큰 항목을 통계값(평균 등)으로 변환
        - 예시) 다양한 연령대의 소득 분포에 있어서 40대의 소득 분포 편차가 다른 연령대에 비하여 매우 크거나 특정 소득 구성원을 포함하고 있을 경우, 40대의 소 득만 선별하여 평균값을 구한 후 40대에 해당하는 각 개인의 소득값을 해당 평균값으로 대체
        - 적용정보 : 나이, 신장, 소득, 카드사용액 등
    3. 라운딩(Rounding)
        - 집계 처리된 값에 대하여 라운딩(올림, 내림, 사사오입) 기준을 적용하여 최종 집계 처리하는 방법으로, 일반적으로 세세한 정보보다는 전체 통계정보가 필요한 경우 많이 사용

        - 예시) 23세, 41세, 57세, 26세, 33세 등 각 나이값을 20대, 30대, 40대, 50대 등 각 대표 연령대로 표기하거나 3,576,000원, 4,210,000원 등의 소득값을 일부 절삭하여 3백만원, 4백만원 등으로 집계 처리하는 방식
        - 적용정보 : 나이, 신장, 소득, 카드지출액, 유동인구, 사용자 수 등
    4. 재배열(Rearrangement)
        - 기존 정보값은 유지하면서 개인이 식별되지 않도록 데이터를 재배열하는 방법으로, 개인의 정보를 타인의 정보와 뒤섞어서 전체 정보에 대한 손상 없이 특정 정 보가 해당 개인과 연결되지 않도록 하는 방법
        - 예시) 데이터 셋에 포함된 나이, 소득 등의 정보를 개인별로 서로 교환하여 재배 치하게 되면 개인별 실제 나이와 소득과 다른 비식별 자료를 얻게 되지만, 전체적인 통계 분석에 있어서는 자료의 손실 없이 분석을 할 수 있는 장점 이 있음
        - 적용정보 : 나이, 신장, 소득, 질병, 신용등급, 학력 등

데이터 삭제 (Data Reduction)

개인 식별이 가능한 데이터 삭제 처리

- 대상 : 개인을 식별 할 수 있는 정보(이름, 전화번호, 주소, 생년월일, 사진, 고유식별 정보(주민등록번호, 운전면허번호 등), 생체정보(지문, 홍채, DNA 정보 등), 기타(등록번호, 계좌번호, 이메일 주소 등))
- 장점 : 개인 식별요소의 전부 및 일부 삭제 처리가 가능
- 단점 : 분석의 다양성과 분석 결과의 유효성·신뢰성 저하
- 예시 : 주민등록번호 901206-1234567 → 90년대 생, 남자 / 개인과 관련된 날짜정보(합격일 등)는 연단위로 처리
- 세부기술 및 실무적용 방법
    1. 식별자 삭제
        - 원본 데이터에서 식별자를 단순 삭제하는 방법
        - (예시) 성명, 생년월일(yy-mm-dd)이 나열되어 있는 경우 분석 목적에 따라 생년월 일을 생년(yy)으로 대체 가능하다면 월일(mm-dd) 값은 삭제 ※ 이때 남아 있는 정보 그 자체로도 분석의 유효성을 가져야 함과 동시에 개인을 식별할 수 없어야 하며, 인터넷 등에 공개되어 있는 정보 등과 결합 하였을 경우에도 개인을 식별할 수 없어야 함
        - 적용정보 : 성명, 전화번호, 계좌번호, 카드번호, 요양기관번호, 이메일 주소 등
    2. 식별자 부분삭제
        - 식별자 전체를 삭제하는 방식이 아니라, 해당 식별자의 일부를 삭제하는 방법
        - (예시) 상세 주소의 경우 부분 삭제를 통하여 대표지역으로 표현 (서울특별시 송파구 가락본동 78번지 → 서울시 송파구)
        - 수치 또는 텍스트 데이터 등에도 폭넓게 활용 가능(‘⑫감추기’는 주로 수치데이터에 적용)
        - 적용정보 : 주소, 위치정보(GPS), 전화번호, 계좌번호 등
    3. 레코드 삭제(Reducing Records)
        - 다른 정보와 뚜렷하게 구별되는 레코드 전체를 삭제하는 방법
        - (예시) 소득이 다른 사람에 비하여 뚜렷이 구별되는 값을 가진 정보는 해당 정보 전체를 삭제 *#* 이 방법은 통계분석에 있어서 전체 평균에 비하여 오차범위를 벗어나는 자료를 제 거할 때에도 사용 가능
        - 적용정보 : 키, 소득, 질병, 카드지출액 등
    4. 식별요소 전부삭제
        - 식별자뿐만 아니라 잠재적으로 개인을 식별할 수 있는 속성자까지 전부 삭제하여 프라이버시 침해 위험을 줄이는 방법
        - (예시) 연예인·정치인 등의 가족정보(관계정보), 판례 및 보도 등에 따라 공개되어 있는 사건과 관련되어 있음을 알 수 있는 정보 등 잠재적 식별자까지 사전에 삭제함으로써 연관성 있는 정보

의 식별 및 결합을 예방
    - 개인정보 유출 가능성을 최대한 줄일 수 있지만 데이터 활용에 필요한 정보까지 사전에 모두 없어지기 때문에 데이터의 유용성이 낮아지는 문제 발생
    - 적용정보 : 나이, 소득, 키, 몸무게 등 개별적으로는 단순한 정보이지만 분석 목적에 따라 추후 개인 식별이 가능성이 있다고 판단되는 정보

데이터 범주화 (Data Suppression))

특정 정보를 해당 그룹의 대푯값으로 변환(범주화)하거나 구간값으로 변환(범주화)하여 개인 식별을 방지

- 대상 : 개인을 식별할 수 있는 정보(주소, 생년월일, 고유식별정보(주민등록번호, 운전면허번호 등), 기관·단체 등의 이용자 계정(등록번호, 계좌번호))
- 장점 : 통계형 데이터 형식이므로 다양한 분석 및 가공 가능
- 단점 : 정확한 분석결과 도출이 어려우며, 데이터 범위 구간이 좁혀질 경우 추 론 가능성 있음
- 예시 : 홍길동, 35세 → 홍씨, 30~40세
- 세부기술 및 실무적용 방법
    1. 감추기
        - 명확한 값을 숨기기 위하여 데이터의 평균 또는 범주값으로 변환하는 방식
        - 단, 특수한 성질을 지닌 개인으로 구성된 단체 데이터의 평균이나 범주값은 그 집 단에 속한 개인의 정보를 쉽게 추론할 수 있음
        - (예시) 간염 환자 집단임을 공개하면서 특정인물 ‘갑’이 그 집단에 속함을 알 수 있도록 표시하는 것은 ‘갑’이 간염 환자임을 공개하는 것과 마찬가지임
    2. 랜덤 라운딩(Random Rounding)
        - 수치 데이터를 임의의 수 기준으로 올림(round up) 또는 내림(round down)하는 기법 - ‘라운딩(rounding)과 달리 수치 데이터 이외의 경우에도 확장 적용 가능
        - (예시) 나이, 우편번호 등과 같은 수치 정보로 주어진 식별자는 일의 자리, 십의 자 리 등 뒷자리 수를 숨기고 앞자리 수만 나타내는 방법(나이 : 42세, 45세 → 40대로 표현)
        - 적용정보 : 나이, 소득, 카드지출액, 우편번호, 유동인구, 사용자 등
    3. 범위 방법(Data Range)
        - 수치데이터를 임의의 수 기준의 범위(range)로 설정하는 기법으로, 해당 값의 범위 (range) 또는 구간(interval)으로 표현
        - (예시) 소득 3,300만원을 소득 3,000만원 4,000만원으로 대체 표기
        - 적용정보 : 서비스 이용 등급, 처방정보(횟수, 기간 등), 위치정보, 유동인구, 사용 자 수, 분석 시간/기간 등
        - 제어 라운딩(Controlled Rounding
        - ‘랜덤 라운딩’ 방법에서 어떠한 특정값을 변경할 경우 행과 열의 합이 일치하 지 않는 단점 해결을 위해 행과 열이 맞지 않는 것을 제어하여 일치시키는 기법
        - 그러나 컴퓨터 프로그램으로 구현하기 어렵고 복잡한 통계표에는 적용하기 어려우며, 해결할 수 있는 방법이 존재하지 않을 수 있어 아직 현장에서는 잘 사용하지 않음
        - 적용정보 : 나이, 키, 소득, 카드지출액, 위치정보 등

데이터 마스킹 (Data Masking))

데이터의 전부 또는 일부분을 대체값(공백, 노이즈 등)으로 변환

- 대상 : 쉽게 개인을 식별할 수 있는 정보(이름, 전화번호, 주소, 생년월일, 사진, 고유식별정보(주민등록번호, 운전면허번호 등), 기관·단체 등의 이용자 계정(등록번호, 계좌번호, 이메일 주소 등))
- 장점 : 개인 식별 요소를 제거하는 것이 가능하며, 원 데이터 구조에 대한 변형 이 적음

- 단점 : 마스킹을 과도하게 적용할 경우 데이터 필요 목적에 활용하기 어려우며 마스킹 수준이 낮을 경우 특정한 값에 대한 추론 가능
- 예시 : 홍길동, 35세, 서울 거주, 한국대 재학 → 홍 , 35세, 서울 거주, 대학 재학
- 세부기술 및 실무적용 방법
    1. 임의 잡음 추가(Adding Random Noise)
        - 개인 식별이 가능한 정보에 임의의 숫자 등 잡음을 추가(더하기 또는 곱하기)하는 방법
        - (예시) 실제 생년월일에 6개월의 잡음을 추가할 경우, 원래의 생년월일 데이터에 1일부터 최대 6개월의 날짜가 추가되어 기존의 자료와 오차가 날 수 있도록 적용
        - 지정된 평균과 분산의 범위 내에서 잡음이 추가되므로 원 자료의 유용성을 해치지 않으나, 잡음 값은 데이터 값과는 무관하기 때문에, 유효한 데이터로 활용하기 곤란
        - 적용정보 : 사용자 ID, 성명, 생년월일, 키, 나이, 병명 코드, 전화번호, 주소 등
    2. 공백(blank)과 대체(impute)
        - 특정 항목의 일부 또는 전부를 공백 또는 대체문자(‘ * ’, ‘ _ ’ 등이나 전각 기호)로 바꾸는 기법
        - (예시) 생년월일 ‘1999-09-09’ ⇒ ‘19 - - ’ 또는 ‘19**-**-**’
        - 2적용정보 : 성명, 생년월일, 전화번호, 주소, 사용자 ID

안전한 가명처리를 위한 방법

안전조치

- 내부관리계획 수립, 추가 정보 별도 분리보관, 접근권한 분리, 물리적·기술적 안전조치 실시
    - 개인정보 보호책임자 지정, 취급자에 대한 교육, 접근 권한 관리 및 접근 통 제, 접속기록 보관 및 점검에 관한 사항 등 포함(안전성 확보조치 기준 제4 조) (→ 총무팀 개인정보 담당자 관할)

기록 작성·보관

- 가명정보의 처리 목적, 처리 및 보유기간, 제3자 제공 시 제공받는 자, 추가정보의 이용 및 파기 등에 관한 사항을 작성하여 보관하여야 함(별지2 참조)

정보주체의 권리

- 가명정보는 적법하게 수집된 개인정보를 법에서 정한 목적 범위 내에서 가명처리하여 활용하는 것으로, 가명정보에 대한 정보주체의 열람, 정정 삭제, 처리정지권에 대한 규정(제35조~제37조)은 적용되지 않음
- 가명정보 처리 목적을 제한하고 있으며, 가명정보 특성상 어떠한 정보주체에 대한 정보인 지 확인할 수 없음

안전한 가명처리를 하지 않은 경우의 제재조치

1. 법 제75조
    - 위반사항 : 제28조의4제1항을 위반하여 안전조치 의 무 미준수
    - 제제조치 : 3천만원 이하 과태료
2. 법 제75조
    - 위반사항 : 제28조의4제2항을 위반하여 가명처리 관련 기록 작성 의무 미준수
    - 제제조치 : 1천만원 이하 과태료
3. 법 제75조
    - 위반사항 : 제28조의4제2항을 위반하여 안전조치 의 무 미준수로 개인정보(가명정보 포함)를 분실·도난·유출·위조·변조 또는 훼손당한 경우
    - 제제조치 : 2년 이하 징역 또는 2천만 원 이하 벌금

첨부자료

별지1. 가명정보 결합신청서 25쪽

별지2. 가명정보 관리대장29쪽

붙임1. 가명처리 관련 제재조치 일람표 30쪽

1. ↑ <출처> 정부통합전산센터 개인정보 비식별 조치 운영지침 제2조(정의)
2. ↑ <자료> 차연철, 『데이터 경제와 개인정보 비식별 기술 동향』