개인정보 비식별 조치 가이드라인

-비식별 조치 기준 및 지원·관리체계 안내-

국무조정실 행정자치부 방송통신위원회
금융위원회 미래창조과학부 보건복지부

빅데이터, IoT(사물인터넷) 등 새로운 IT 기술과 융합산업의 출현은 세계 최고 수준의 IT강국으로 자리매김한 우리나라에게 또 다른 도약의 기회가 되고 있으나, 한편으로 그러한 기술 활용과정에서 발생할 수 있는 개인정보 침해 우려는 신산업 발전과 개인정보의 보호를 동시에 조화롭게 모색해야 하는 과제를 제기하고 있습니다.

이에 국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부 등 관계부처가 합동으로 현행 개인정보 보호 법령의 틀 내에서 빅데이터가 안전하게 활용될 수 있도록 하는데 필요한 개인정보의 비식별 조치 기준과 비식별 정보의 활용 범위 등을 명확히 제시함으로써 기업의 불확실성을 제거하여 기업투자와 산업 발전을 도모하는 한편, 국민의 개인정보인권 보호에도 소홀함이 없도록 하고자 합니다.

아울러, 이 가이드라인에 따라 정보주체를 알아볼 수 없도록 비식별 조치를 적정하게 한 비식별 정보는 개인정보가 아닌 것으로 추정되며, 따라서 빅데이터 분석 등에 활용이 가능합니다.

다만, 비식별 정보도 기술발전, 데이터 증가 등에 따른 재식별 가능성이 있음을 고려하여 재식별 방지를 위한 관리적 · 기술적 안전조치 등을 통해 안전하게 활용되고 관리되어야 함을 알려드립니다.

2016. 6. 30. 관계부처 합동

개인정보 비식별 조치 가이드라인

– 비식별 조치 기준 및 지원 · 관리체계 안내 –

Contents

개인정보 비식별 조치 가이드라인

- 비식별 조치 기준 및 지원 · 관리체계 안내 -

I 추진 배경

- 빅데이터, IoT 등 IT 융합기술 발전으로 데이터 이용 수요가 급증함에 따라 미국 · 영국 등 주요 선진국은 데이터 산업 활성화를 위한 정책 추진 중
- 이에 빅데이터 활용에 필요한 비식별 조치 기준 · 절차 · 방법 등을 구체적으로 안내하여 안전한 빅데이터 활용기반 마련과 개인정보 보호 강화를 도모

1 정부 3.0 및 빅데이터 활용 확산에 따른 데이터 활용가치 증대

- 공공정보 개방 · 공유는 투명하고 효율적인 정부 운영에, 빅데이터 활용은 과학적 정책 집행 및 맞춤형 서비스 제공에 필수적인 수단
- 특히, 빅데이터 분석, IoT 기술 등을 통한 새로운 서비스 창출과 신산업 활성화에 데이터의 활용가치 증대

2 개인정보 보호 강화에 대한 사회적 요구 지속

- 크고 작은 개인정보 유출 사고가 지속되어 개인정보 보호 정책을 강화해야 한다는 사회적 요구가 계속
- 다양한 데이터 활용을 필요로 하는 새로운 산업과 기술 발전으로 개인정보 침해 위험도 증가 추세

3 '보호와 활용'을 동시에 모색하는 세계적 정책변화에 적극 대응

- 미국 · 영국 등 주요 선진국은 개인정보 침해가능성을 최소화하면서 데이터 산업 활성화를 위한 정책 추진 중
- 사생활 침해 방지를 위한 안전장치 마련과 동시에 비식별 조치된 정보는 산업적으로 활용할 수 있도록 구체적인 가이드 제시 필요

Ⅱ 비식별 조치 기준

1 조치 개요

○ 본 가이드라인은 개인정보를 비식별 조치하여 이용 또는 제공하려는 사업자 등이 준수하여야 할 조치 기준을 제시한 것임

※ 통계법 등 관련법령에 따라 개인정보를 수집 · 이용하는 경우에는 당해 법령에 따라 처리

○ 단계별 조치사항

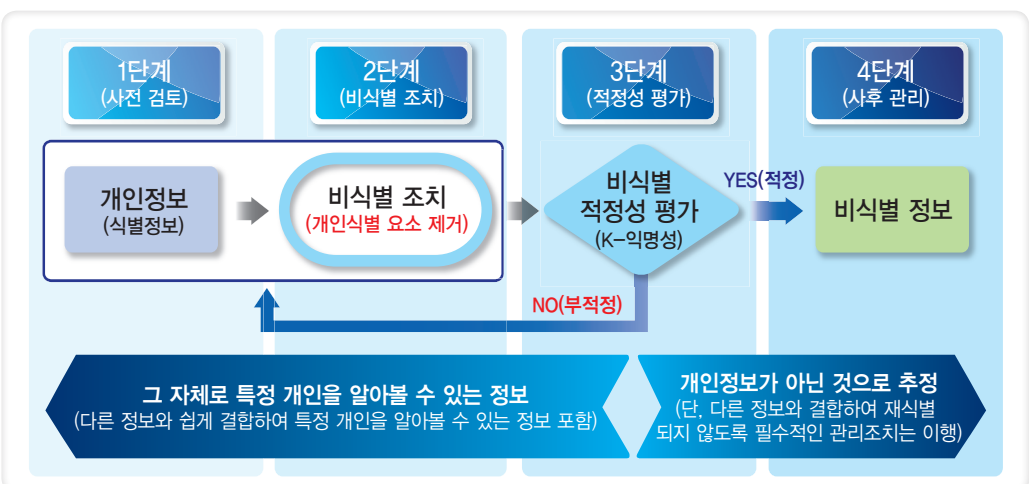
① **(사전 검토)** 개인정보에 해당하는지 여부를 검토 후, 개인정보가 아닌 것이 명백한 경우 법적 규제 없이 자유롭게 활용(4쪽 참조)

② **(비식별 조치)** 정보집합물(데이터 셋)에서 개인을 식별할 수 있는 요소를 전부 또는 일부 삭제하거나 대체하는 등의 방법을 활용, 개인을 알아볼 수 없도록 하는 조치(5~8쪽 참조)

③ **(적정성 평가)** 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는지를 「비식별 조치 적정성 평가단」을 통해 평가(9~13쪽 참조)

④ **(사후관리)** 비식별 정보 안전조치, 재식별 가능성 모니터링 등 비식별 정보 활용 과정에서 재식별 방지를 위해 필요한 조치 수행(14~16쪽 참조)

● 비식별 조치 및 사후관리 절차 ●



2 단계별 조치 기준

2-① 사전 검토 단계 : 개인정보 해당 여부 검토

- 빅데이터 분석 등을 위해 정보를 처리하려는 사업자 등은 해당 정보가 개인정보인지 여부에 대해 아래 기준을 참조하여 판단

- 해당 정보가 개인정보에 해당하지 않는 것이 명백한 경우에는 별도 조치 없이 빅데이터 분석 등에 활용 가능

  ⇨ 개인정보에 해당한다고 판단되는 경우 다음 단계의 조치 필요

〈 참고 〉 개인정보 해당 여부 판단 기준

가. 개인정보 보호법 등 관련 법률에서 규정하고 있는 개인정보의 개념은 다음과 같으며, 이에 해당하지 않는 경우에는 개인정보가 아님

나. 개인정보는 ⅰ)살아 있는 ⅱ)개인에 관한 ⅲ)정보로서 ⅳ)개인을 알아수 있는 정보이며, 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 ⅴ)다른 정보와 쉽게 결합하여 알아볼 수 있는 정보를 포함

ⅰ) (살아있는) 자에 관한 정보이어야 하므로 사망한 자, 자연인이 아닌 법인, 단체 또는 사물 등에 관한 정보는 개인정보에 해당하지 않음

ⅱ) (개인에 관한) 정보이어야 하므로 여럿이 모여서 이룬 집단의 통계값 등은 개인정보에 해당하지 않음

ⅲ) (정보)의 종류, 형태, 성격, 형식 등에 관하여는 특별한 제한이 없음

ⅳ) (개인을 알아볼 수 있는 정보)이므로 특정 개인을 알아보기 어려운 정보는 개인정보가 아님

- 여기서 '알아볼 수 있는'의 주체는 해당 정보를 처리하는 자(정보의 제공 관계에 있어서는 제공받은 자를 포함)이며, 정보를 처리하는 자의 입장에서 개인을 알아볼 수 없다면 그 정보는 개인정보에 해당하지 않음

ⅴ) (다른 정보와 쉽게 결합하여)란 결합 대상이 될 다른 정보의 입수 가능성이 있어야 하고, 또 다른 정보와의 결합 가능성이 높아야 함을 의미

- 즉, 합법적으로 정보를 수집할 수 없거나 결합을 위해 불합리한 정도의 시간, 비용 등이 필요한 경우라면 "쉽게 결합"할 수 있는 상태라고 볼 수 없음

※ 자세한 내용은 부록 「개인정보 보호 관련 법령 통합해설서」 참조

2-② 비식별 조치 단계 : 비식별 조치기법 적용

식별자(Identifier) 조치 기준

- 정보집합물에 포함된 식별자*는 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

〈 예시 〉 식별자

- 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, 회사전화, 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

※ 美 「HIPAA 프라이버시 규칙」을 참고하여 작성

속성자(Attribute value) 조치 기준

- 정보집합물에 포함된 속성자*도 데이터 이용 목적과 관련이 없는 경우에는 원칙적으로 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

〈 예시 〉 속성자

개인 특성	• 성별, 연령(나이), 국적, 고향, 시 · 군 · 구명, 우편번호 병역여부, 결혼여부, 종교, 취미, 동호회 · 클럽 등 • 흡연여부, 음주여부, 채식여부, 관심사항 등
신체 특성	• 혈액형, 신장, 몸무게, 허리둘레, 혈압, 눈동자 색깔 등 • 신체검사 결과, 장애유형, 장애등급 등 • 병명, 상병(傷病)코드, 투약코드, 진료내역 등
신용 특성	• 세금 납부액, 신용등급, 기부금 등 • 건강보험료 납부액, 소득분위, 의료 급여자 등
경력 특성	• 학교명, 학과명, 학년, 성적, 학력 등 • 경력, 직업, 직종, 직장명, 부서명, 직급, 전직장명 등
전자적 특성	• 쿠키정보, 접속일시, 방문일시, 서비스 이용 기록, 접속로그 등 • 인터넷 접속기록, 휴대전화 사용기록, GPS 데이터 등
가족 특성	• 배우자 · 자녀 · 부모 · 형제 등 가족 정보, 법정대리인 정보 등

비식별 조치 방법

- 가명처리, 총계처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹 등 여러 가지 기법을 단독 또는 복합적으로 활용

  ※ '가명처리' 기법만 단독 활용된 경우는 충분한 비식별 조치로 보기 어려움

- 각각의 기법에는 이를 구현할 수 있는 다양한 세부기술이 있으며, 데이터 이용 목적과 기법별 장 · 단점 등을 고려하여 적절한 기법 · 세부기술을 선택 · 활용

  * (참고 3) 「개인정보 비식별 조치 방법」 참조

  ⇨ 비식별 조치가 완료되면 다음 단계의 조치 필요

● 〈 예시 〉 비식별 조치 방법 ●

처리기법	예시	세부기술
가명처리 (Pseudonymization)	• 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대, 서울 거주, 국제대 재학	① 휴리스틱 가명화 ② 암호화 ③ 교환 방법
총계처리 (Aggregation)	• 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm	④ 총계처리 ⑤ 부분총계 ⑥ 라운딩 ⑦ 재배열
데이터 삭제 (Data Reduction)	• 주민등록번호 901206–1234567 → 90년대 생, 남자 • 개인과 관련된 날짜정보(합격일 등)는 연단위로 처리	⑧ 식별자 삭제 ⑨ 식별자 부분삭제 ⑩ 레코드 삭제 ⑪ 식별요소 전부삭제
데이터 범주화 (Data Suppression)	• 홍길동, 35세 → 홍씨, 30~40세	⑫ 감추기 ⑬ 랜덤 라운딩 ⑭ 범위 방법 ⑮ 제어 라운딩
데이터 마스킹 (Data Masking)	• 홍길동, 35세, 서울 거주, 한국대 재학 → 홍○○, 35세, 서울 거주, ○○대학 재학	⑯ 임의 잡음 추가 ⑰ 공백과 대체

● 〈 예시 〉 비식별 조치 기법 적용 ●

원본데이터

주민등록번호	성별	입원날짜	연령	병명
770914–1234567	남	2015/06/23	39	독감
850930–1234567	남	2015/10/01	31	독감
710119–2345678	여	2014/01/21	45	고혈압
770619–2345678	여	2014/09/23	39	고혈압
830425–1234567	남	2015/04/16	33	간염
860804–2345678	여	2014/11/11	30	간염

비식별 데이터

① 데이터 삭제(주민등록번호)

~~주민등록번호~~	성별	입원날짜	연령	병명
	남	2015/06/23	39	독감
	남	2015/10/01	31	독감
	여	2014/01/21	45	고혈압
	여	2014/09/23	39	고혈압
	남	2015/04/16	33	간염
	여	2014/11/11	30	간염

② 데이터 마스킹(주민등록번호, 입원날짜), 총계처리(평균 연령)

주민등록번호	성별	입원날짜	연령	병명
7*****–1******	남	2015/**/**	35	독감
8*****–1******	남	2015/**/**	35	독감
7*****–2******	여	2014/**/**	35	고혈압
7*****–2******	여	2014/**/**	35	고혈압
8*****–1******	남	2015/**/**	35	간염
8*****–2******	여	2015/**/**	35	간염

2–③ 적정성 평가 단계 : k–익명성 모델 활용

적정성 평가 필요성

- 비식별 조치가 충분하지 않은 경우 공개 정보 등 다른 정보와의 결합, 다양한 추론 기법 등을 통해 개인이 식별될 우려

- 개인정보 보호책임자 책임 하에 외부전문가가 참여하는 「비식별 조치 적정성 평가단(이하, '평가단')」을 구성, 개인식별 가능성에 대한 엄격한 평가 필요

- 적정성 평가 시 프라이버시 보호 모델 중 k–익명성을 활용
  - k–익명성은 최소한의 평가수단이며, 필요시 추가적인 평가모델(ℓ –다양성, t–근접성) 활용
    * (참고 3) 「개인정보 비식별 조치 방법」 참조

넷플릭스 사례(2006, 미국)

- 온라인 영화대여 회사인 넷플릭스(Netflix)는 고객의 기호에 맞는 영화를 추천하는 알고리즘의 정확성을 높이기 위해 경연대회를 개최
  - 1999년 12월부터 2005년 12월까지 50만명의 이용자들이 영화에 대한 평점을 내린 1억건의 시청 이력 데이터를 공개
    ※ 사용자를 식별할 수 있는 이름 등은 삭제하였으나, 데이터 처리 내용을 연결하기 위해 독특한 식별자, 영화에 대한 평가 내용, 평가 일시 등을 공개

- 텍사스 대학의 한 그룹이 넷플릭스사가 공개한 시청 이력 데이터와 영화정보 사이트 IMDb(Internet Movie Database)에 공개된 사용자 리뷰를 결합하여 일부 개인을 식별해냄
  ※ IMDb는 웹 사이트 상에서 아이디와 평가점수를 게시

- 미국연방거래위원회(FTC)가 프라이버시에 관한 문제를 지적하여 제2회 경연은 중지됨

적정성 평가 절차

① **(기초자료 작성)** 개인정보처리자는 적정성 평가에 필요한 데이터 명세, 비식별 조치 현황, 이용기관의 관리 수준 등 기초자료 작성

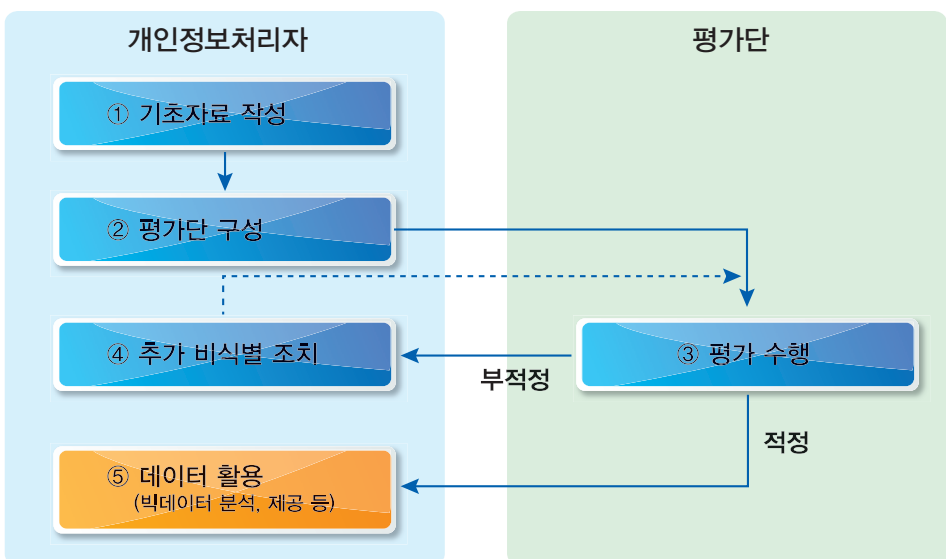
② **(평가단 구성)** 개인정보 보호책임자가 3명 이상으로 평가단을 구성(외부전문가는 과반수 이상)

③ **(평가 수행)** 평가단은 개인정보처리자가 작성한 기초자료와 k-익명성 모델을 활용하여 비식별 조치 수준의 적정성을 평가

④ **(추가 비식별 조치)** 개인정보처리자는 평가결과가 '부적정'인 경우 평가단의 의견을 반영하여 추가적인 비식별 조치 수행

⑤ **(데이터 활용)** 비식별 조치가 적정하다고 평가받은 경우에는 빅데이터 분석 등에 이용 또는 제공이 허용

● 비식별 조치에 대한 적정성 평가 절차 ●



① 기초자료 작성

○ 개인정보처리자는 평가 대상 데이터 명세, 비식별 조치현황, 이용기관의 관리수준 등 적정성 평가에 필요한 기초자료를 작성

● 비식별 조치 적정성 평가에 필요한 기초 자료 ●

구분	기초자료	비 고
데이터 명세	• 데이터 특성(크기, 생성 및 관리 환경 등), 세부 항목별 명세, 원본 예시	필수사항
	• 비식별 조치된 평가 대상 데이터 셋 및 세부 항목별 명세	필수사항
비식별 조치 현황	• 비식별 조치에 적용한 기법 · 세부기술	필수사항
	• 평가 대상 데이터 셋에 대한 k-익명성 값 산출 결과	필수사항
이용기관의 관리수준	• 데이터 이용 기관의 이용 목적 및 방법, 이용기간, 데이터 접근 가능자 현황 등 활용에 관한 사항	필수사항
	• 데이터를 제공하는 경우 데이터를 제공받는 방법 및 데이터 보호를 위한 일련의 조치에 대한 현황	필수사항
	• 데이터 이용 및 제공과 관련이 있는 계약서 또는 협약서 사본 ※ 계약서가 없는 경우 그 사유 제시	필수사항
	• 데이터 이용 기관에서 보유하거나, 보유할 수 있는 개인정보 관련 데이터(세부 내용)에 관한 사항	선택사항
	• 데이터 이용기관의 개인정보보호 또는 정보보호 관련 인증서 사본	선택사항

② 평가단 구성

- 평가단은 해당 기관의 개인정보 보호책임자가 3명 이상의 관련 분야 전문가로 구성 (외부전문가를 과반수 이상으로 위촉*)

  * 외부전문가 위촉시 각 분야별 전문기관에서 운영하는 전문가 풀에서 법률 전문가, 비식별조치 기법 전문가를 각 1명 이상 활용(17~18쪽 참조)

- 평가단은 데이터 이용목적과 직접적인 이해관계가 없는 자로 구성

- 평가단의 단장은 외부전문가 중에서 호선으로 선출하며, 평가단 운영과 관련된 전반적인 사항을 관장

- 평가단은 착수회의를 포함해서 최소 2회 이상 운영

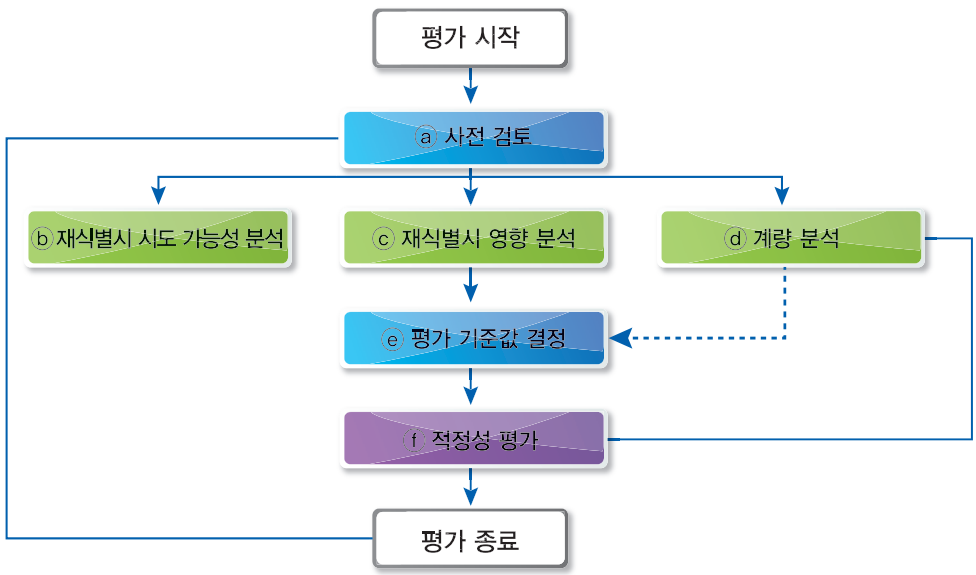
- 평가단은 필요한 경우 비식별 조치 관련 규정 마련 및 보완 등 권고

  – 컴퓨팅 환경의 발전, 이용 데이터 특성, 결합 가능한 정보의 입수 가능성 등을 고려, 필요한 경우 1년, 3년 등 재검토 일정을 제시

③ 평가 수행

- 평가단은 개인정보처리자가 제공한 기초자료와 k-익명성 모델 등을 활용하여 비식별 수준의 적정성 여부 평가

● 적정성 평가 수행 방법 ●



※ 「비식별 의료정보 공개 관리 가이드라인(캐나다, 2010.10월)」을 참조하여 절차 마련

ⓐ **(사전 검토)** 개인정보처리자가 제출한 기초자료와 인터뷰 등을 통해 평가대상 데이터의 개인 식별요소 포함 여부, 데이터 이용 목적, 비식별 조치 기법 등 검토

ⓑ **(재식별 시도 가능성)** 데이터를 이용 또는 제공받는 자의 재식별 의도와 능력, 개인정보 보호 수준 등 재식별 시도 가능성 분석

ⓒ **(재식별시 영향 분석)** 데이터가 의도적 또는 비의도적으로 재식별될 경우 정보주체 등에게 미칠 수 있는 영향 분석

ⓓ **(계량 분석)** 개인정보처리자가 제출한 k값의 정확성 여부 검증

ⓔ **(평가 기준값 결정)** 평가단에서 '재식별 시도 가능성', '재식별시 영향', '계량 분석' 결과와 데이터 이용 목적 등을 종합적으로 고려하여 평가 기준값(k-익명성 값) 결정

〈미국 교육부「프라이버시 보호 기술지원센터」의 안전도 기준〉

- 'k=3'은 안전도를 보장하는 최소한의 수준
- '5≦k≦10'은 안전도가 높은 수준

※ k-익명성 값은 데이터의 제공을 합법적으로 허용하기 위해 제시된 기준

ⓕ **(적정성 평가)** '평가 기준값'과 '계량 분석'에서 산출된 값을 비교하여 비식별 조치의 적정성 여부를 최종 결정

- 계량분석 결과의 k값이 4이고 평가 기준값이 3인 경우 '적정'으로 평가
- 계량분석 결과의 k값이 4이고 평가 기준값이 6인 경우 '부적정'으로 평가

– '적정'한 경우 ⇨ 데이터 이용 및 제공 가능

– '부적정'한 경우 ⇨ 추가적인 비식별 조치 및 재평가 수행

④ 추가 비식별 조치

- 개인정보처리자는 평가 결과가 '부적정'인 경우 평가단의 의견을 반영하여 해당 데이터에 대한 비식별 조치를 추가 실시
- 추가 비식별 조치가 완료된 경우에는 평가단에 비식별 조치가 적정히 수행되었는지에 대한 재평가를 요청

⑤ 데이터 활용

- 비식별 조치가 적정하다고 (재)평가받은 경우에는 해당 데이터를 빅데이터 분석 등에 이용하거나 제3자에게 제공이 허용
- 불특정 다수에게 공개하는 것은 식별 위험이 크므로 원칙적으로 금지

※「공공데이터의 제공 및 이용 활성화에 관한 법률」 등 법령에 따른 공개는 제외

- 데이터 이용 목적을 달성하거나 해당 데이터가 불필요하게 된 경우에는 지체 없이 파기 조치
- 데이터 활용 과정에서도 아래의 사후관리 단계의 조치사항을 준수하여야 비식별 정보로서 유효하게 활용 가능

2–④ 사후관리 단계

1 비식별 정보 안전 조치

○ 비식별 조치된 정보가 유출되는 경우 다른 정보와 결합하여 식별될 우려가 있으므로 필수적인 보호조치 이행

– **(관리적 보호조치)** 비식별 정보파일에 대한 관리 담당자 지정, 비식별 조치 관련 정보공유 금지, 이용 목적 달성시 파기 등의 조치가 필요함

– **(기술적 보호조치)** 비식별 정보파일에 대한 접근통제, 접속기록 관리, 보안 프로그램 설치 · 운영 등의 조치 필요

● 비식별 정보에 대한 관리적 · 기술적 보호조치 ●

구분	비식별 정보 보호 조치
관리적 보호조치	① 비식별 정보파일 관리담당자 지정 ② 비식별 정보파일 대장관리 ③ 원본정보 관리부서(기관)와 비식별 정보 관리부서(기관) 간 비식별 조치 관련 정보공유 금지 ④ 이용목적 달성시 지체없이 파기 ⑤ 비식별 정보파일 유출시 대응계획 수립
기술적 보호조치	⑥ 비식별 정보파일에 대한 접근권한 관리 및 접근통제 ⑦ 비식별 정보 보관시스템에 대한 접속기록 관리 ⑧ 악성 코드 방지 등을 위한 보안프로그램 설치 · 운영

○ 비식별 정보 유출 시 보호조치

– 유출 원인 분석 및 추가 유출 방지를 위한 관리적·기술적 보호조치

– 유출된 비식별 정보의 회수 · 파기

2 재식별 가능성 모니터링

- 비식별 정보를 이용하거나 제3자에게 제공하려는 사업자 등은 해당 정보의 재식별 가능성을 정기적으로 모니터링을 해야 함

- 모니터링 결과, 다음의 점검 항목 중 어느 하나에 해당되는 경우에는 추가적인 비식별 조치 강구

〈예시〉 재식별 가능성 모니터링 시 점검항목

구분	점검 항목
내부 요인의 변화	• 비식별 조치된 정보와 연계하여 재식별 우려가 있는 추가적인 정보를 수집하였거나 제공받은 경우
	• 데이터 이용과정에서 생성되는 정보가 비식별 정보와 결합해서 새로운 정보가 생성되는 경우
	• 이용부서에서 비식별 정보에 대한 비식별 수준을 당초보다 낮추어 달라고 하는 요구가 있는 경우
	• 신규 또는 추가로 구축되는 시스템이 비식별 정보에 대한 접근을 관리 · 통제하는 보안체계에 중대한 변화를 초래하는 경우
외부 환경의 변화	• 이용 중인 데이터에 적용된 비식별 조치 기법과 유사한 방법으로 비식별 조치한 사례가 재식별 되었다고 알려진 경우
	• 이용 중인 데이터에 적용된 비식별 기법과 기술을 무력화 하는 새로운 기술이 등장하거나 공개된 경우
	• 이용 중인 데이터와 새롭게 연계 가능한 정보가 출현하거나, 공개된 것으로 알려진 경우

- 비식별 정보를 제공 · 위탁한 자가 재식별 가능성을 발견한 경우에는 이를 즉시 그 정보를 처리하고 있는 자에게 통지하고 처리 중단 요구 및 해당 정보를 회수 · 파기하는 등 필요한 조치를 하여야 함

3 비식별 정보 제공 및 위탁계약 시 준수사항

- 비식별된 정보를 제3의 기관에 제공하거나, 처리 위탁하는 경우 재식별 위험관리에 관한 내용을 계약서에 포함
  - **(재식별 금지)** 비식별 정보를 제공받거나 처리를 위탁 받은 사업자 등은 다른 정보와 결합을 통해 재식별 시도가 금지됨을 명시
  - **(재제공 또는 재위탁 제한)** 비식별 정보를 제공하거나 처리를 위탁하는 자는 재제공 또는 재위탁 가능 범위를 정하여 계약서에 명시
  - **(재식별 위험 시 통지)** 재식별이 되거나 재식별 가능성이 높아지는 상황이 발생한 경우에는 데이터 처리 중지 및 비식별 정보 제공자 또는 위탁자에게 통지 의무 등 명시

계약서 특수조건 반영 내용 사례

제00조(재식별 금지)

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

② △으로부터 제공받은 정보를 ○이 제3자에게 제공하거나 처리를 위탁하고자 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.

※ 비식별 정보를 제공받은 기업은 "○", 제공한 기업은 "△"로 표시

4 재식별 시 조치요령

- 비식별 정보가 재식별된 경우에는 신속하게 그 정보의 처리를 중단하고 해당 개인정보가 유출되지 않도록 필요한 조치를 하여야 함
- 재식별된 정보는 즉시 파기 조치하되, 해당 정보를 다시 활용하려면 비식별 조치 절차를 다시 거쳐야 함

Ⅲ 지원 및 관리체계

1 개인정보 비식별 조치 지원

지원 필요성

- 비식별 조치를 통해 개인정보를 안전하게 활용할 수 있는 지원체계 필요
  - 개인정보처리자가 수행하는 비식별 조치 적정성 평가 지원 등
- 비식별 조치에 필요한 컨설팅과 전문교육 등을 통해 중소기업 및 스타트업의 빅데이터 활용 지원
- 인공지능, 새로운 결합기술 출현 등에 따른 재식별 위험에 적극 대응

지원체계 및 지원내용

분야별 전문기관

- 각 소관부처 책임 하에 분야별 전문기관을 정하여 운영

  ※ 분야별 전문기관은 한국인터넷진흥원, 한국신용정보원, 금융보안원, 사회보장정보원, 한국정보화진흥원 중에서 소관부처가 공문으로 지정 · 공표하여 운영하고 필요시 추가 지정 가능

- 분야별 전문기관의 역할
  - 비식별 조치 적정성 평가단 풀(비식별 조치 기법 전문가, 법률 전문가 등) 구성·운영
  - 산업별로 필수적인 비식별 조치 이행 권고(k-익명성 수치 등)

    * 의료, 복지, 교육, 금융 · 신용, 통신, 유통, 공공 · 기타 분과

  - 비식별 조치 적정성 실태 점검 등
- '16년 8월 중 분야별 전문기관 지정 및 운영 개시

개인정보 비식별 지원센터

- 개인정보 보호 전담기관인 한국인터넷진흥원(KISA)에 「개인정보 비식별 지원센터」 설치 · 운영

- 개인정보 비식별 지원센터의 역할
  - 분야별 전문기관 운영 가이드라인 마련 및 실태 점검
  - 분야별 전문기관 실무협의회 운영
  - 분야별 평가단 풀 관리 및 교육, 중소기업 및 스타트업 컨설팅 · 교육
  - 「개인정보 비식별 조치 가이드라인」 업데이트 및 활용 지원
  - 국내외 관련 정책 · 기술 동향 조사 및 연구 등

- '16년 8월 중 설치 및 운영 개시

2 전문기관을 통한 기업 간 정보집합물 결합 지원

지원 필요성

- 빅데이터 분석에 활용하기 위해 서로 다른 사업자가 보유하고 있는 정보집합물을 결합하는 경우 개인별로 부여된 식별자가 매칭키로 사용
  - 이 경우, 정보주체를 알아볼 수 있는 식별자 그 자체를 매칭키로 사용하는 것은 현행법 위반 소지(개인정보 보호법 제18조, 개인정보의 목적 외 이용 · 제공 제한)

- 따라서, 정보집합물 간 결합 · 분석을 위해서는 결합 과정에서만 임시로 매칭키 역할을 하는 '임시 대체키'의 활용이 필요

- 임시 대체키를 활용한 결합을 허용하는 경우에도 무분별한 결합을 통한 개인정보 침해 소지를 방지하기 위해 전문기관(제3의 공공기관)에서만 결합을 하도록 하는 등 지원 및 관리체계 필요

지원 및 관리체계

- 기업 간 정보집합물 결합 지원은 분야별 전문기관에서 수행

- 전문기관 선택 기준
  - 산업내 기업간 결합은 해당 분야 전문기관에서 결합 지원
  - 이종산업 간 결합은 대량의 정보집합물을 결합하고자 하는 기업이 속해 있는 분야별 전문기관에서 수행
  - 당해 산업을 지원해 주는 전문기관이 없는 경우에는 한국인터넷진흥원 또는 한국정보화진흥원에서 지원

- 전문기관의 주요 역할 및 책임
  - 임시 대체키를 활용, 기업 간 정보집합물 결합 지원
  - 업무처리 전반에 있어 개인을 식별하려는 일체의 시도 금지
  - 정보집합물 결합 및 정보 제공 완료 후 모든 정보 지체 없이 파기

- 전문기관에 대한 세부 이용기준은 각 부처에서 마련 · 시행

정보집합물 결합 절차 및 유의사항

결합 절차

① A社와 B社가 같은 알고리즘을 적용하여 식별자를 임시 대체키로 전환하고, 결합대상 정보집합물도 비식별 조치 및 적정성 평가 수행

※ '임시 대체키' 생성시 동 대체키에 잡음을 추가하거나, 2개 이상의 식별자를 활용할 경우 식별자 중 일부를 조합하여 불법적 복호화 또는 원본 정보와 결합시에도 개인을 식별할 수 없도록 조치

② 비식별 조치된 정보를 전문기관에 제공 및 결합 요청

※ 이 경우 전문기관은 제공받은 비식별 정보를 통해 특정 개인 식별 불가

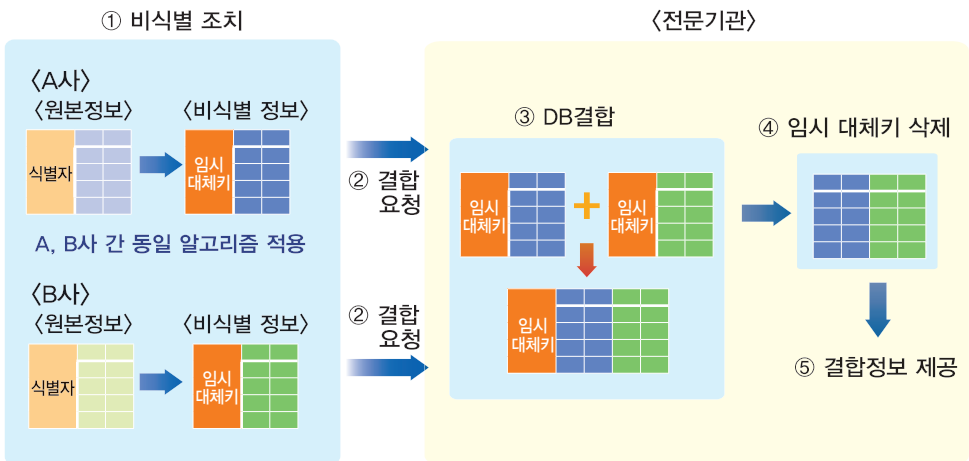
③ 임시 대체키를 활용, 전문기관에서 결합 수행

④ 임시 대체키 삭제

⑤ 결합 DB를 필요한 기업에게 제공(전문기관은 제공 후 파기 조치)

※ 임시 대체키가 삭제된 결합 DB가 제공되어 A와 B도 결합 DB를 통해 특정 개인의 식별이 어려움

● 임시 대체키를 통한 기업 간 정보집합물 결합 절차 ●



결합 시 유의사항

- A와 B는 분야별 전문기관과 임시 대체키 생성 알고리즘에 대한 정보공유 금지
- 임시 대체키 생성을 위해 주민등록번호를 활용하는 것은 금지
  (개인정보 보호법 제24조의2, 주민등록번호 처리의 제한)
- 다른 정보와의 결합을 위해 임시 대체키를 활용하는 경우, k-익명성 값은 임시 대체키를 제외하고 산출*

  * 임시 대체키를 제외하지 않으면, 'k=1'로 산출되어 객관적 평가 불가
- 전문기관은 결합 과정에서 재식별 발생시 해당 정보를 즉시 파기
- 결합 DB를 제공받은 기관은 이용 전에 반드시 적정성 평가 수행

3 재식별 시 법적 제재

형사처벌

○ 비식별 정보를 재식별하여 이용하거나 제3자에게 제공한 경우

> 예시 1 연구자에게 비식별 정보를 제공하면서 비식별 조치 요령을 공유하여 결과적으로 개인정보를 목적 외로 제공한 경우
>
> 예시 2 이름, 생년월일, 전화번호 등 주요 식별정보를 공개된 알고리즘으로 암호화하는 등 쉽게 재식별 될 수 있도록 하여 제3자에게 제공한 경우
>
> 예시 3 비식별 정보를 의도적으로 재식별하여 보관하고 있거나 1:1 마케팅 등에 활용한 경우

– 개인정보의 목적 외 이용 · 제공에 해당(개인정보 보호법 제18조제1항 위반, 정보통신망법* 제24조 및 제24조의2 위반, 신용정보법** 제32조 및 제33조 위반)

* 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, ** 「신용정보의 이용 및 보호에 관한 법률」

– 5년 이하의 징역 또는 5천만원 이하의 벌금

※ 정보통신망법 적용 사업자는 위반행위 관련 매출액의 3% 이하 과징금이 추가 부과될 수 있음

행정처분

○ 비식별 정보를 활용하여 재식별하고도 즉시 파기 조치하지 않고 보관하고 있는 경우

> 예시 비식별 정보를 제공받은 자가 업무처리 과정에서 특정 개인을 재식별하였으나 재식별된 정보를 즉시 파기하지 않고 보관하고 있는 경우

– 정보주체의 동의없이 개인정보를 수집한 경우에 해당(개인정보 보호법 제15조제1항 위반, 정보통신망법 제22조제1항 위반, 신용정보법 제15조제2항 위반)

– 5천만원 이하의 과태료가 부과

※ 정보통신망법 적용 사업자는 5년 이하 징역 또는 5천만원 이하 벌금형에 처해질 수 있으며 위반행위 관련 매출액의 3% 이하 과징금이 추가 부과될 수 있음

개인정보 비식별 조치 가이드라인

– 비식별 조치 기준 및 지원 · 관리체계 안내 –

참고

1. 비식별 정보의 산업적 활용(예시)
2. 국내외 동향
3. 개인정보 비식별 조치 방법
4. 비식별 조치 적정성 평가단 세부 평가수행 방법

참고 1 비식별 정보의 산업적 활용(예시)

기업 내부에서 비식별 정보 활용

업무처리 과정에서 수집한 고객정보, 거래내역, 민원처리 내역 등 개인정보가 포함된 각종 자료를 비식별 조치한 후 시장조사, 신상품 및 서비스 개발, 마케팅 전략 수립, 업무 프로세스 개선, 위험관리 등 다양한 용도로 분석 및 활용 가능
다만, 개인 식별을 전제로 하는 1:1 마케팅 용도로 사용할 수 없으며 재식별 시도는 금지

❶ ㅁㅁ공사는 고속도로 이용차량 빅데이터 분석을 통해 고객서비스 개선 전략 수립

▸ ㅁㅁ공사는 최근 5년간 톨게이트 진출입 데이터를 비식별 조치한 후 월별 · 시간대별 차량 평균속도, 상습 정체구간, 사고구간 및 원인 등 빅데이터 분석을 실시하여 도로 구조 개선 및 휴게공간 추가 설치 등 고객서비스 개선

❷ ○○통신사는 무선사업부 고객정보를 비식별 조치하여 단말기 판매부서에서 활용

▸ 단말기 판매부서는 요금제별 단말기 교환주기 및 단말기 선호 가격을 분석하여 단말기 판촉 전략을 수립

❸ △△보험사는 보험사기 사례 정보를 비식별 조치한 후 보험사기 방지시스템 개발

▸ 동 시스템을 통해 보험계약 체결, 유지, 보험금 지급 등 거래 전 단계에서 보험사기 징후를 자동으로 추출 · 예방함으로서, 보험사기 발생률 및 보험관리 비용 절감

다른 기업으로부터 비식별 정보를 제공받아 활용

본 가이드라인에 따라 비식별 조치한 정보는 제3자에게 제공할 수 있으며, 당해 비식별 정보를 제공받은 기업은 이를 시장조사, 신상품 및 서비스 개발, 마케팅 전략 수립 등 다양한 용도로 분석 및 활용 가능
다만, 이 경우 재식별 금지, 재식별 위험시 통지 등의 내용을 해당 비식별 정보 제공과 관련한 계약서에 반드시 포함하여야 하며, 제공받은 정보 또한 개인 식별을 전제로 하는 1:1 마케팅 용도로 사용할 수 없음

❶ ○○증권은 △△은행, ◇◇보험 등에서 비식별 조치한 자료를 제공받아 신상품 개발에 활용

▸ △△은행, ◇◇보험 등은 보유하고 있는 다양한 신용 정보를 비식별 조치한 후 ○○증권에게 제공

▸ ○○증권은 제공받은 자료를 빅데이터 분석하여 '로보어드바이저', 'ISA' 등 다양한 신상품 개발에 활용하고 국내 및 해외시장 개척을 추진

▸ '로보어드바이저'를 통해 개인이 문의할 경우 온라인 투자자문, 자산운용 상담 지원

❷ 신생 스타트업인 ◇◇사가 □□은행으로부터 비식별 정보를 제공받아 새로운 비즈니스 모델 개발에 활용

▸ □□은행은 보유하고 있는 학력 · 연령 · 성별 첫 직장, 이직경로, 연봉 등의 정보를 비식별 조치하여 신생 기업인 ◇◇사에 제공

▸ ◇◇사는 기존의 헤드헌팅 회사와 차별화된 '첫 직장부터 퇴직 후까지 커리어 관리 프로그램'을 제공하는 비즈니스 모델을 개발하여 활용

❸ ○○제약회사는 △△심사평가원으로부터 제공받은 비식별 정보를 ××신약개발 연구에 활용

▸ △△심사평가원이 특정 질병 환자의 연령과 성별에 따른 진료기록을 충분히 비식별 조치한 후, ○○제약회사에게 제공

▸ ○○제약회사는 해당 정보를 활용하여 ××병의 발병 원인 및 치유 원인을 분석하여 신약을 개발, 수입 약품 대비 20% 저렴한 가격으로 판매

❹ □□홈쇼핑은 ◇◇카드사로부터 구매금액 상위 10% 고객의 결제 내역에 대한 비식별 정보를 제공받아 우수고객 마케팅 전략 수립에 활용

▸ □□홈쇼핑과 ◇◇카드사는 고객 전화번호와 카드 결제정보를 각각 복원되지 않는 알고리즘으로 비식별 조치하여 A전문기관에 제공하고 A전문기관은 두 정보를 결합한 후, □□홈쇼핑에게 제공

▸ 비식별 조치된 고객의 결제정보를 통해 □□홈쇼핑은 우수고객이 선호하는 물품을 특정 시간대에 할인 행사를 실시하는 마케팅 전략 수립

참고 2 국내외 동향

1 미국

일반 동향

- 개인정보 보호에 관한 일반법이 없으며, 개별 법령에서 제한하지 않는 한 자유로운 데이터의 이용이 보장
  - 의료 · 교육 등 분야별로 개인정보 보호에 관한 개별 법령 운영 중
- 의료정보는 「건강보험 이전과 책임에 관한 법(HIPAA)」에 따른 「HIPAA 프라이버시 규칙」에서 비식별 조치 기준 제시
  ※ 비식별 조치된 의료정보는 제한 없이 이용 가능
- 「경제적 · 임상적 보건에 대한 건강 정보기술법(Health Information Technology for Economic and Clinical Health Act; HITECH Act)」에서는 비식별 조치된 건강정보에 대해 프라이버시 관련 규제 미적용
- 「가족의 교육적 권리 및 프라이버시 법(Family Educational Rights and Privacy Act; FERPA)」은 비식별 조치된 학생기록에 대해 별도의 동의없이 배포 가능(k-익명성 모델 활용)

HIPAA 프라이버시 규칙(HIPAA Privacy Rule)

- 「보호대상 의료정보의 비식별 가이드(Guidance on De-identification of Protected Health Information)」에 따라 비식별 조치된 의료정보를 규제대상에서 제외
  - 비식별 조치 방법에 따라 '전면적 규율면제' 또는 '부분적 규율면제' 방식 적용
- '전면적 규율면제'가 적용되는 비식별 조치방법에는 '전문가 결정방식'과 '세이프 하버 방식'이 있음
  - '전문가 결정방식'은 통계적, 과학적 원칙과 방법에 대한 지식과 경험을 보유한 전문가가 개인식별 위험 최소화 방법 적용

● 개인 위험 평가시 고려사항 ●

구 분	설 명
반복 가능성 (Replicability)	특정 개인과 관련하여 반복적으로 계속되는 정보(eg, 생일 등)는 개인식별 위험이 높다고 판별
데이터 소스 이용가능성 (Data Source Availability)	외부에 많이 공개되어 있는 정보는 덜 공개된 정보보다 상대적으로 개인식별에 활용될 가능성이 높음
구별 가능성 (Distinguishability)	특정인을 구별할 수 있는 정보(eg, 주소정보)는 그렇지 않은 정보(생년)보다 상대적으로 더 위험함

– '세이프 하버 방식'은 이름, 주소정보 등 18가지 주요 식별자*를 제거하여 개인식별이 가능하지 못하도록 하는 방식

* 18가지 식별자 : ①이름 ②주소정보 ③개인과 직접 관련된 날짜정보(생일, 합격일 등) ④전화번호 ⑤팩스번호 ⑥이메일 주소 ⑦사회보장번호 ⑧의료기록번호 ⑨건강보험번호 ⑩계좌번호 ⑪자격취득번호 ⑫자동차번호 ⑬각종 장비 식별번호 ⑭URL정보 ⑮IP주소 ⑯생체정보 ⑰전체 얼굴사진과 이와 유사한 이미지 ⑱기타 특이한 식별 번호 또는 코드

○ '부분적 규율면제'는 위의 18가지 식별자 중 '③ 날짜정보', '⑱ 기타 식별번호 또는 코드' 외의 16개 식별자를 제거한 경우에 적용

– 데이터 제공자와 제공받는 자 간에 데이터 이용 및 제공목적 등을 담은 계약을 체결하여 진행

2 EU

EU 동향

○ EU 역내 개인정보 보호를 규율하는 일반규정인 「EU 개인정보 보호 지침*(EU Data Protection Directive)」 서문에서는 익명화된(Anonymized) 정보는 동 지침의 규제 대상이 아니라고 명시

* 지침은 각 회원국이 국내법을 제정 · 시행하는데 적용되는 가이드로서 강제성은 없음

○ 지침에는 가명처리된(Pseudonymized) 정보는 과학적 연구, 역사연구, 통계 목적으로 사용 가능함을 명시

– 구체적인 처리 기준 등 자세한 사항은 회원국이 정하도록 함

○ 최근 EU에서 단일한 개인정보 보호법(General Data Protection Regulation; GDPR)이 EU 의회를 통과('16.4.14., 2년 후 시행 예정)

– 기존 지침의 가명정보 규정이 법적 구속력이 있는 GDPR에 명문화

– 다만, 가명정보를 동의없이 처리할 수 있는 목적의 범위가 공익, 과학적 연구, 역사연구, 통계 목적으로 일부 변경

영국 사례

- 영국 정보보호 커미셔너(ICO)는 EU 지침 서문에 근거하여 '12년에 EU 최초로 익명화 규약*을 출간

  * ICO, "Anonymisation: managing data protection risk code of practice, 2012"

  – 요구되는 익명화의 정도는 식별 위험성이 '0(zero) 수준'은 아니나, 식별 위험성이 매우 낮은(remote) 수준이어야 함을 명시

- 식별 위험성의 판단 기준으로 '합리적 가능성(reasonably likely test)' 기준을 채택

  – 식별의 위험성이 합리적으로 가능할 경우에는 규제의 대상이 되는 개인정보에 해당하며 이 기준은 EU 지침과 동일

3 일본

- 최근 일본 정부는 IT종합전략본부를 중심으로 개인정보의 합리적 활용을 촉진하기 위해 개인정보 보호법 개정('15.9월 개정, '17.1월 시행)

  – 개인정보의 빅데이터 활용 확대를 위해 익명가공정보 개념 신설

  – 익명가공정보는 복원 불가능도록 안전 조치를 함을 전제로 정보주체의 동의없이 활용할 수 있도록 허용

  – 익명가공정보 취급 사업자*에게 일정한 기술적 · 관리적 조치 의무** 부여

  * 특정의 익명가공정보를 전자장치를 이용하여 검색할 수 있도록 체계적으로 구성하여 다른 사업자에게 판매 또는 제공하는 자

  ** 복원불능 정보 작성, 정보누설 방지, 정보항목 공개, 제3자 제공 시 공표, 식별행위 금지, 안전조치 의무 등 부담

- 일본 정부는 관계 전문가와 함께 비식별 가이드라인 마련 추진 중

  – 이와 관련, 전문가들은 '완전한 익명화'는 있을 수 없음을 인정하고, 익명화의 수준을 단계별로 구분 · 제시하는 방안 논의 중

  – 그중 익명화 수준을 가장 높도록 조치하기 위한 방법으로 k–익명성 모델을 활용하는 방안을 검토하고 있음

> 우리나라의 경우에는 k–익명성 모델을 기본적으로 적용하고, 필요시 추가적인 평가모델인 ℓ –다양성 모델과 t–근접성 모델까지 적용

4 우리나라

구 분	공공정보 개방 · 공유에 따른 개인정보 보호지침	개인정보 비식별화에 대한 적정성 자율평가 안내서	빅데이터 개인정보보호 가이드라인	빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서
발간	2013.9	2014.12	2014.12	2015.6
주관 부처	행정자치부	행정자치부	방송통신위원회	미래창조과학부
대 상	공공	공공, 민간	민간	공공, 민간
목적	공공정보 개방 · 공유 등에 따른 개인정보 보호조치 요령 안내	개인정보 비식별화 적정성 평가, 재식별 위험 관리 및 조치요령 안내	빅데이터 관련 개인정보 보호법령 적용 방안 설명	빅데이터 활용을 위한 비식별화 기술 활용방법 및 관련 법규 안내
주요 내용	(1단계) 공공정보 개방 · 공유 근거 확인 (2단계) 개인식별 요소 제거 (3단계) 개인식별 가능성 검토 (4단계) 비식별 기법 및 재식별 가능성에 관한 주기적 모니터링 실시	개인정보가 포함된 정보는 개인정보 식별요소 제거를 통해 개인을 식별할 수 없는 형태로 정보를 변경 후 이용 재식별 위험에 대한 관리적 조치 및 재식별시 대응 조치 공공정보 개방 · 공유 등에 따른 개인정보 보호지침과 유사한 조치 및 프라이버시 보호모델에 따른 검토를 하도록 함	공개된 정보, 이용내역 정보, 생성된 정보는 비식별화 조치 후 처리 비식별화 조치한 경우 이용자의 동의 없이 수집 · 이용 가능 비식별 조치 후 저장 시 보호조치 필요 비식별화 조치된 개인정보는 이용자 동의 없이 제3자 제공 가능	개인식별이 가능한 정보는 삭제 다른 정보와 결합으로 재식별될 위험 최소화 및 사후관리 철저 가명처리, 총계처리, 데이터 값 삭제, 범주화, 데이터 마스킹 등 18가지 비식별화 기술 및 적용사례 제시

참고 3 개인정보 비식별 조치 방법

1 개요

일반적 기법 : 개인 식별요소 삭제 방법

처리기법	예시	세부기술
가명처리 (Pseudonymization)	• 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대, 서울 거주, 국제대 재학	① 휴리스틱 가명화 ② 암호화 ③ 교환 방법
총계처리 (Aggregation)	• 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm	④ 총계처리 ⑤ 부분총계 ⑥ 라운딩 ⑦ 재배열
데이터 삭제 (Data Reduction)	• 주민등록번호 901206-1234567 → 90년대 생, 남자 • 개인과 관련된 날짜정보(합격일 등)는 연단위로 처리	⑧ 식별자 삭제 ⑨ 식별자 부분삭제 ⑩ 레코드 삭제 ⑪ 식별요소 전부삭제
데이터 범주화 (Data Suppression)	• 홍길동, 35세 → 홍씨, 30~40세	⑫ 감추기 ⑬ 랜덤 라운딩 ⑭ 범위 방법 ⑮ 제어 라운딩
데이터 마스킹 (Data Masking)	• 홍길동, 35세, 서울 거주, 한국대 재학 → 홍○○, 35세, 서울 거주, ○○대학 재학	⑯ 임의 잡음 추가 ⑰ 공백과 대체

프라이버시 보호 모델 : 재식별 가능성 검토 기법 * k, ℓ, t값은 전문가 등이 검토하여 마련

기법	의미	적용례
k-익명성	• 특정인임을 추론할 수 있는지 여부를 검토, 일정 확률수준 이상 비식별 되도록 함	• 동일한 값을 가진 레코드를 k개 이상으로 함. 이 경우 특정 개인을 식별할 확률은 1/k임
ℓ-다양성	• 특정인 추론이 안된다고 해도 민감한 정보의 다양성을 높여 추론 가능성을 낮추는 기법	• 각 레코드는 최소 l개 이상의 다양성을 가지도록 하여 동질성 또는 배경지식 등에 의한 추론 방지
t-근접성	• ℓ-다양성 뿐만 아니라, 민감한 정보의 분포를 낮추어 추론 가능성을 더욱 낮추는 기법	• 전체 데이터 집합의 정보 분포와 특정 정보의 분포 차이를 t이하로 하여 추론 방지

2 일반적 기법

가명처리(Pseudonymization)

- (개념) 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

실무적용 방법

① **휴리스틱 가명화(Heuristic Pseudonymization)**

- 식별자에 해당하는 값들을 몇 가지 정해진 규칙으로 대체하거나 사람의 판단에 따라 가공하여 자세한 개인정보를 숨기는 방법
  (ex) 성명을 홍길동, 임꺽정 등 몇몇 일반화된 이름으로 대체하여 표기하거나 소속기관명을 화성, 금성 등으로 대체하는 등 사전에 규칙을 정하여 수행
- 식별자의 분포를 고려하거나 수집된 자료의 사전 분석을 하지 않고 모든 데이터를 동일한 방법으로 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 세심한 고려 필요
- 적용정보 : 성명, 사용자 ID, 소속(직장)명, 기관번호, 주소, 신용등급, 휴대전화번호, 우편번호, 이메일 주소 등

② **암호화(Encryption)**

- 정보 가공시 일정한 규칙의 알고리즘을 적용하여 암호화함으로써 개인정보를 대체하는 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 암호화에 비해 더욱 안전하고 효과적인 비식별 기술에 해당
- 적용정보 : 주민등록번호, 여권번호, 의료보험번호, 외국인등록번호, 사용자 ID, 신용카드번호, 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)

- (개념) 통계값(전체 혹은 부분)을 적용하여 특정 개인을 식별할 수 없도록 함
- (대상) 개인과 직접 관련된 날짜 정보(생일, 자격 취득일), 기타 고유 특징(신체정보, 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

실무적용 방법

④ 총계처리(Aggregation)

– 데이터 전체 또는 부분을 집계(총합, 평균 등)

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

(예시) 집단에 소속된 전체 인원의 평균 나이값을 구한 후 각 개인의 나이값을 평균 나이값(대푯값)으로 대체하거나 해당 집단 소득의 전체 평균값을 각 개인의 소득값으로 대체

– 적용정보 : 나이, 신장, 소득, 카드사용액, 유동인구, 사용자수, 제품 재고량, 판매량 등

⑤ 부분총계(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 소득만 선별하여 평균값을 구한 후 40대에 해당하는 각 개인의 소득값을 해당 평균값으로 대체

– 적용정보 : 나이, 신장, 소득, 카드사용액 등

⑥ 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

⑦ 재배열(Rearrangement)

– 기존 정보값은 유지하면서 개인이 식별되지 않도록 데이터를 재배열하는 방법으로, 개인의 정보를 타인의 정보와 뒤섞어서 전체 정보에 대한 손상 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 방법

(예시) 데이터 셋에 포함된 나이, 소득 등의 정보를 개인별로 서로 교환하여 재배치하게 되면 개인별 실제 나이와 소득과 다른 비식별 자료를 얻게 되지만, 전체적인 통계 분석에 있어서는 자료의 손실 없이 분석을 할 수 있는 장점이 있음

– 적용정보 : 나이, 신장, 소득, 질병, 신용등급, 학력 등

데이터 삭제(Data Reduction)

- (개념) 개인 식별이 가능한 데이터 삭제 처리
- (대상) 개인을 식별 할 수 있는 정보(이름, 전화번호, 주소, 생년월일, 사진, 고유식별정보(주민등록번호, 운전면허번호 등), 생체정보(지문, 홍채, DNA 정보 등), 기타(등록번호, 계좌번호, 이메일주소 등))
- (장점) 개인 식별요소의 전부 및 일부 삭제 처리가 가능
- (단점) 분석의 다양성과 분석 결과의 유효성 · 신뢰성 저하

실무적용 방법

⑧ **식별자 삭제**

– 원본 데이터에서 식별자를 단순 삭제하는 방법

(예시) 성명, 생년월일(yy–mm–dd)이 나열되어 있는 경우 분석 목적에 따라 생년월일을 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 결합하였을 경우에도 개인을 식별할 수 없어야 함

– 적용정보 : 성명, 전화번호, 계좌번호, 카드번호, 요양기관번호, 이메일 주소 등

⑨ **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

⑩ **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 제거할 때에도 사용 가능

– 적용정보 : 키, 소득, 질병, 카드지출액 등

⑪ **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 정보의 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)

- (개념) 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 추론 가능성 있음

실무적용 방법

⑫ **감추기**

– 명확한 값을 숨기기 위하여 데이터의 평균 또는 범주값으로 변환하는 방식

– 단, 특수한 성질을 지닌 개인으로 구성된 단체 데이터의 평균이나 범주값은 그 집단에 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

⑬ **랜덤 라운딩(Random Rounding)**

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

– '⑥ 라운딩(rounding)과 달리 수치 데이터 이외의 경우에도 확장 적용 가능

(예시) 나이, 우편번호 등과 같은 수치 정보로 주어진 식별자는 일의 자리, 십의 자리 등 뒷자리 수를 숨기고 앞자리 수만 나타내는 방법(나이 : 42세, 45세 → 40대로 표현)

– 적용정보 : 나이, 소득, 카드지출액, 우편번호, 유동인구, 사용자 등

⑭ **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, 사용자 수, 분석 시간/기간 등

⑮ **제어 라운딩(Controlled Rounding)**

– '⑬랜덤 라운딩' 방법에서 어떠한 특정값을 변경할 경우 행과 열의 합이 일치하지 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)

- (개념) 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 변형이 적음
- (단점) 마스킹을 과도하게 적용할 경우 데이터 필요 목적에 활용하기 어려우며 마스킹 수준이 낮을 경우 특정한 값에 대한 추론 가능

실무적용 방법

⑯ **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, 잡음값은 데이터 값과는 무관하기 때문에, 유효한 데이터로 활용하기 곤란

– 적용정보 : 사용자 ID, 성명, 생년월일, 키, 나이, 병명 코드, 전화번호, 주소 등

⑰ **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**-**-**'

– 적용정보 : 성명, 생년월일, 전화번호, 주소, 사용자 ID 등

3 프라이버시 보호 모델

k-익명성(k-anonymity) : 프라이버시 보호를 위한 기본 모델

○ 공개된 데이터에 대한 연결공격(linkage attack) 등 취약점*을 방어하기 위해 제안된 프라이버시 보호 모델

＊ 공개 데이터의 취약점

- **개인정보를 포함한 공개 데이터**
  - 일반적으로 활용하는 데이터에는 이름, 주민등록번호 등과 같이 개인을 직접 식별할 수 있는 데이터는 삭제(예: 〈표 1〉)
  - 그러나 활용 정보의 일부가 다른 공개되어 있는 정보 등과 결합하여 개인을 식별하는 문제(연결공격)가 발생 가능(예: 〈표 2〉)
- **연결공격(linkage attack)**
  - 예를 들어, 〈표 1〉의 의료데이터가 〈표 2〉의 선거인명부와 지역 코드, 연령, 성별에 의해 결합되면, 개인의 민감한 정보인 병명이 드러날 수 있음
    (ex) 김민준 (13053, 28, 남자)→ 환자 레코드 1번→ 전립선염
  - 미국 매사추세츠 주, '선거인명부'와 '공개 의료데이터'가 결합하여 개인의 병명 노출 사례

○ (정의) 주어진 데이터 집합에서 같은 값이 적어도 k개 이상 존재하도록 하여 쉽게 다른 정보로 결합할 수 없도록 함

- 데이터 집합의 일부를 수정하여 모든 레코드가 자기 자신과 동일한(구별되지 않는) k-1개 이상의 레코드를 가짐
- 예를 들어, 〈표 1〉의 의료 데이터가 비식별 조치된 〈표 3〉에서 1~4, 5~8, 9~12 레코드는 서로 구별되지 않음

● 〈표 1〉 공개 의료데이터 사례 ●

구분	지역 코드	연령	성별	질병
1	13053	28	남	전립선염
2	13068	21	남	전립선염
3	13068	29	여	고혈압
4	13053	23	남	고혈압
5	14853	50	여	위암
6	14853	47	남	전립선염
7	14850	55	여	고혈압
8	14850	49	남	고혈압
9	13053	31	남	위암
10	13053	37	여	위암
11	13068	36	남	위암
12	13068	35	여	위암

● 〈표 2〉 선거인명부 사례 ●

구분	이름	지역코드	연령	성별
1	김민준	13053	28	남
2	박지훈	13068	21	남
3	이지민	13068	29	여
4	최현우	13053	23	남
5	정서연	14853	50	여
6	송현준	14850	47	남
7	남예은	14853	55	여
8	성민재	14850	49	남
9	윤건우	13053	31	남
10	손윤서	13053	37	여
11	민우진	13068	36	남
12	허수빈	13068	35	여

● 〈표 3〉 k-익명성 모델에 의해 비식별된 의료데이터 사례 ●

구분	지역 코드	연령	성별	질병	비고
1	130**	< 30	*	전립선염	다양한 질병이 혼재되어 안전
2	130**	< 30	*	전립선염	
3	130**	< 30	*	고혈압	
4	130**	< 30	*	고혈압	
5	1485*	> 40	*	위암	다양한 질병이 혼재되어 안전
6	1485*	> 40	*	전립선염	
7	1485*	> 40	*	고혈압	
8	1485*	> 40	*	고혈압	
9	130**	3*	*	위암	모두가 동일 질병(위암)으로 취약
10	130**	3*	*	위암	
11	130**	3*	*	위암	
12	130**	3*	*	위암	

※ '*' 표시는 임의의 글자를 나타낸다. 가령, 지역코드 '130**'은 '13000~13099' 범위 안에 존재하는 하나의 지역코드 값을 의미한다.

- 따라서, 비식별된 데이터 집합에서는 공격자가 정확히 어떤 레코드가 공격 대상인지 알아낼 수 없음

  ※ (예시) 〈표 2〉 김민준→ 〈표 3〉 레코드 1~4→ 전립선염 또는 고혈압

- 여기서, 같은 속성자 값들로 비식별된 레코드들의 모임을 '동일 속성자 값 집합(equivalent class, 이하 동질 집합)'이라고 함

  ※ (예시) 〈표 3〉 레코드1~4, 5~8, 9~12

ℓ-다양성(ℓ-diversity) : k-익명성의 취약점*을 보완한 프라이버시 보호 모델

- k-익명성에 대한 두 가지 공격, 즉 동질성 공격 및 배경지식에 의한 공격을 방어하기 위한 모델
- (정의) 주어진 데이터 집합에서 함께 비식별되는 레코드들은 (동질 집합에서) 적어도 ℓ개의 서로 다른 민감한 정보를 가져야 함
  - 비식별 조치 과정에서 충분히 다양한(ℓ개 이상) 서로 다른 민감한 정보를 갖도록 동질 집합을 구성
- 정보가 충분한 다양성을 가지므로 다양성의 부족으로 인한 공격에 방어가 가능하고, 배경지식으로 인한 공격에도 일정 수준의 방어능력

* k-익명성의 취약점

- **취약점 1. 동질성 공격 (Homogeneity attack)**
  - k-익명성에 의해 레코드들이 범주화 되었더라도 일부 정보들이 모두 같은 값을 가질 수 있기 때문에 데이터 집합에서 동일한 정보를 이용하여 공격 대상의 정보를 알아내는 공격
  - 〈표 3〉에서 범주화의 기초가 되는 정보(지역코드, 연령, 성별)에 대해서는 여러 다양한 값들이 혼재되어 있어서 연결 공격에 의한 식별이 어렵지만, 이 정보와 연결된 정보(질병)는 'k-익명성'의 기초가 아니기 때문에 발생할 수 있는 현상
  - 예를 들어, 〈표 3〉에서 레코드 9~12의 질병정보는 모두 '위암'이므로 k-익명성 모델이 적용되었음에도 불구하고 그 질병정보가 직접적으로 노출됨

- **취약점 2. 배경지식에 의한 공격 (Background knowledge attack)**
  - 주어진 데이터 이외의 공격자의 배경 지식을 통해 공격 대상의 민감한 정보를 알아내는 공격
  - 〈표 2〉와 〈표 3〉에서 공격자가 '이지민'의 질병을 알아내려고 하면 정보의 결합(13068, 29, 여)에 따라 '이지민'은 〈표 3〉의 1~4 레코드 중 하나이며 질병은 전립선염 또는 고혈압임을 알 수 있음
  - 이 때, '여자는 전립선염에 걸릴 수 없다'라는 배경 지식에 의해 공격 대상 '이지민'의 질병은 고혈압으로 쉽게 추론 가능함

- **k-익명성의 취약점의 원인**
  - 다양성의 부족 (lack of diversity)
    - 비식별 조치 할 때 정보의 다양성을 고려하지 않음
    - 동일한 정보를 가진 (다양하지 않은) 레코드가 비식별되어 하나의 '동질 집합'으로 구성될 경우 동질성 공격에 무방비
  - 강한 배경지식 (strong background knowledge)
    - k-익명성은 '여자는 전립선염에 걸리지 않는다' 또는 '남자는 자궁암에 걸리지 않는다'와 같은 공격자의 배경지식을 고려하지 않아 이를 이용한 공격에 취약

- 예를 들어, 〈표 4〉에서 모든 동질 집합은 3-다양성(ℓ =3)을 통해 비식별되어 3개 이상의 서로 다른 정보를 가짐
  - 〈표 3〉과 같이 동일한 질병으로만 구성된 동질 집합이 존재하지 않음
  - 공격자가 질병에 대한 배경지식(예: 여자는 전립선염에 걸리지 않음)이 있더라도 어느 정도의 방어력을 가지게 됨(예: 여성 이지민이 속한 동질 집합 2, 3, 11, 12에서 전립선염을 제외하더라도 고혈압, 위암 중 어느 질병이 이지민의 것인지 여전히 알 수 없음)

● 〈표 4〉 ℓ -다양성 모델에 의해 비식별된 의료데이터의 예 ●

구분	지역 코드	연령	성별	질병	비고
1 4 9 10	1305* 1305* 1305* 1305*	≤ 40 ≤ 40 ≤ 40 ≤ 40	* * * *	전립선염 고혈압 위암 위암	다양한 질병이 혼재되어 안전
5 6 7 8	1485* 1485* 1485* 1485*	> 40 > 40 > 40 > 40	* * * *	위암 전립선염 고혈압 고혈압	다양한 질병이 혼재되어 안전
2 3 11 12	1306* 1306* 1306* 1306*	≤ 40 ≤ 40 ≤ 40 ≤ 40	* * * *	전립선염 고혈압 위암 위암	다양한 질병이 혼재되어 안전

t-근접성(t-closeness) : 값의 의미를 고려하는 프라이버시 모델

- ℓ -다양성의 취약점*(쏠림 공격, 유사성 공격)을 보완하기 위해 모델

* ℓ -다양성의 취약점

- **쏠림 공격 (skewness attack)**
  - 정보가 특정한 값에 쏠려 있을 경우 ℓ -다양성 모델이 프라이버시를 보호하지 못함

〈쏠림 공격의 예〉
- 임의의 '동질 집합'이 99개의 '위암 양성' 레코드와 1개의 '위암 음성' 레코드로 구성되어 있다 가정
- 공격자는 공격 대상이 99%의 확률로 '위암 양성'이라는 것을 알 수 있음

• **유사성 공격 (similarity attack)**

– 비식별 조치된 레코드의 정보가 서로 비슷하다면 ℓ –다양성 모델을 통해 비식별 된다 할지라도 프라이버시가 노출될 수 있음

〈유사성 공격의 예〉

• 〈표 5〉는 3–다양성(ℓ =3) 모델을 통해 비식별 된 데이터

• 레코드 1,2,3이 속한 동질 집합의 병명이 서로 다르지만 의미가 서로 유사함(위궤양, 급성 위염, 만성 위염)

• 공격자는 공격 대상의 질병이 '위'에 관련된 것이라는 사실을 알아낼 수 있음

• 또 다른 민감한 정보인 급여에 대해서도 공격 대상이 다른 사람에 비해 상대적으로 낮은 급여 값을 가짐을 쉽게 알아낼 수 있음(30 ~ 50백만원)

○ (정의) 동질 집합에서 특정 정보의 분포와 전체 데이터 집합에서 정보의 분포가 t이하의 차이를 보여야 함

– 각 동질 집합에서 '특정 정보의 분포'가 전체 데이터집합의 분포와 비교하여 너무 특이하지 않도록 함

– 〈표 5〉에서 전체적인 급여 값의 분포는 30 ~ 110이나 레코드 1, 2, 3이 속한 동질 집합에서는 30 ~ 50으로 이는 전체 급여 값의 분포(30 ~ 110)와 비교할 때 상대적으로 유사한 수준이라 볼 수 있음

→ 공격자는 근사적인 급여 값을 추론할 수 있음

– t–근접성 모델은 이러한 동질 집합과 전체 데이터 집합 사이의 분포의 과도한 차이를 ℓ –다양성 모델의 취약점으로 규정함

● 〈표 5〉 ℓ –다양성 모델에 의해 비식별되었지만 유사성 공격에 취약한 사례 ●

구 분	속성자		민감한 정보		비고
	지역 코드	연령	급여(백만원)	질병	
1	476**	2*	30	위궤양	모두가 '위'와 관련한 유사 질병으로 취약
2	476**	2*	40	급성 위염	
3	476**	2*	50	만성 위염	
4	4790*	≥ 40	60	급성 위염	다양한 질병이 혼재되어 안전
5	4790*	≥ 40	110	감기	
6	4790*	≥ 40	80	기관지염	
7	476**	3*	70	기관지염	다양한 질병이 혼재되어 안전
8	476**	3*	90	폐렴	
9	476**	3*	100	만성 위염	

○ ‘정보의 분포’를 조정하여 정보가 특정 값으로 쏠리거나 유사한 값들이 뭉치는 경우를 방지

– 〈표 6〉에서 t-근접성 모델에 따라 레코드 1, 3, 8은 하나의 동질 집합

– 이 경우, 레코드 1, 3, 8의 급여의 분포는 (30 ~ 90)으로 전체적인 급여의 분포(30 ~ 110)와 큰 차이가 나지 않음

– 또한, 레코드 1, 3, 8의 질병 분포는 위궤양, 만성위염, 폐렴으로 병명이 서로 다르고 질병이 ‘위’와 관련된 것 이외에 ‘폐’와 관계된 것도 있어 특정 부위의 질병임을 유추하기 어려움

– 따라서 〈표 5〉의 경우와 비교하여 공격자가 공격 대상의 정보를 추론하기가 더욱 어려워짐

● 〈표 6〉 t-근접성 모델에 의해 비식별 조치된 데이터 사례 ●

구 분	속성자		민감한 정보		비고
	지역 코드	연령	급여(백만원)	질병	
1	4767*	≤ 40	30	위궤양	급여의 분포와 다양한 질병으로 안전
3	4767*	≤ 40	50	만성 위염	
8	4767*	≤ 40	90	폐렴	
4	4790*	≥ 40	60	급성 위염	급여의 분포와 다양한 질병으로 안전
5	4790*	≥ 40	110	감기	
6	4790*	≥ 40	80	기관지염	
2	4760*	3*	40	급성 위염	급여의 분포와 다양한 질병으로 안전
7	4760*	3*	70	기관지염	
9	4760*	3*	100	만성 위염	

○ t수치가 0에 가까울수록 전체 데이터의 분포와 특정 데이터 구간의 분포 유사성이 강해지기 때문에 그 익명성의 방어가 더 강해지는 경향

– 익명성 강화를 위해 특정 데이터들을 재배치해도 전체 속성자들의 값 자체에는 변화가 없기 때문에 일반적인 경우에 정보 손실의 문제는 크지 않음

참고 4 비식별 조치 적정성 평가단 세부 평가수행 방법

사전 검토

- 평가 수행기관에서 제출한 기초자료와 인터뷰 등을 통해 평가대상 데이터에 개인 식별요소(식별자, 속성자) 포함 여부, 데이터 이용 목적, 적용된 비식별 조치 기법 등 검토
- 첫째, 평가 수행기관에서 작성 · 제출한 기초자료가 필수사항을 모두 포함하고 있고, 적절히 작성되었는지 검토
  - 기초자료가 충분하지 않은 경우 평가수행기관에 추가적인 자료 제출 및 보완을 요구
- 둘째, 평가 대상 데이터의 특성에 대해 확인하고 개인을 식별할 수 있는 식별 요소를 포함하고 있는지 확인
  - 평가 대상 데이터의 생성 및 관리되는 환경, 데이터의 크기, 시간 흐름에 따른 축적 여부 등 데이터의 특성에 대해 확인
  - 평가 대상 데이터의 식별자 또는 속성자에 식별요소를 포함하고 있는지 검토
  - 평가 대상 데이터가 개인 식별요소를 포함하고 있는 경우 개인 식별요소 제거 조치가 '부적정' 한 것으로 판단하고 비식별 조치 보강 요청
- 셋째, 기초자료로 제출된 '비식별 조치에 적용한 기법 · 세부기술'에 따라 비식별 조치가 적절히 수행되었는지 검토
  - '데이터 원본 예시', '비식별 조치된 평가 대상 데이터 셋 및 세부 항목별 명세', '비식별 조치에 적용한 기법 · 세부기술' 등 검토
  - '비식별 조치에 적용한 기법 · 세부기술'에 따라 개인 식별요소 제거 조치가 충분히 되지 않은 경우 '부적정' 한 것으로 판단하고 비식별 조치 보강 요청

재식별 시도 가능성 분석

- 데이터를 이용 또는 제공받는 자의 개인정보 재식별 의도와 능력, 개인정보 보호 수준 등을 통해 재식별 시도 가능성을 분석

1) 재식별 의도 및 능력 분석

- 데이터 이용자 또는 요청자의 재식별 의도 및 능력에 대한 검토 실시
- 평가단은 〈표 1〉 평가지표의 세부 질문에 대해 평가하고 개별 평가 지표별로 '예' 또는 '아니오'로 평가를 실시

● 〈표 1〉 재식별 의도 및 능력 분석 평가 지표 ●

구 분	세부 지표	평가
재식별 의도	• 데이터 이용자 또는 요청자가 데이터 제공자와 기존에 함께 업무를 수행하면서 상호 신뢰관계를 구축한 경험이 없음	예/아니오
	• 데이터 이용자 또는 요청자가 데이터를 재식별 하는 경우 경제적인 이익이 있음	예/아니오
	• 데이터 이용자 또는 요청자가 데이터를 재식별 하는 경우 비경제적인 이익이 있음	예/아니오
	• 데이터 이용자 또는 요청자가 데이터를 제3의 이용자에게 사전 허가 없이 제공할 가능성이 있음	예/아니오
	• 데이터 이용자 또는 요청자가 데이터 이용(제공) 관련 계약서에 재식별 금지 및 제3자에게 데이터 제공 제한 등의 문구를 반영하고 있지 않음	예/아니오
재식별 능력	• 데이터 이용자 또는 요청자가 개인정보 재식별을 시도 할 수 있는 전문 지식을 보유하고 있음	예/아니오
	• 데이터 이용자 또는 요청자가 개인정보 재식별을 시도 할 수 있는 재원(자금)을 보유 또는 조달할 수 있음	예/아니오
	• 데이터 이용자 또는 요청자가 개인정보 재식별을 위해 연계할 수 있는 다른 데이터베이스를 직접 보유하고 있거나 접근 할 수 있음	예/아니오
외부 정보 연계 가능성	• 인터넷, SNS 등에 평가대상 데이터와 결합 가능한 데이터가 존재할 수 있음	예/아니오

- 평가점수는 개인별로 각 평가지표에 대해 '예'로 평가한 지표의 개수를 합산해서 산출 (개인별 점수는 최대 9점, 점수가 높을수록 재식별 의도 및 능력이 큼)
- 개인별 점수를 합산한 후 전체 평가인원의 수로 나누어 '재식별 의도 및 능력 분석'의 평가 점수를 구하고, 〈표 2〉 평가 기준표에 따라 '높음', '중간', '낮음'으로 1차 평가결과 도출
- 1차 평가결과는 평가단 토의를 거쳐 확정하되, 1차 평가결과를 기준표와 달리 적용하는 경우에는 이에 대한 사유를 명확히 문서로 남겨야 함

● 〈표 2〉 재식별 의도 및 능력 분석 평가 기준표 ●

구 분	평가 기준
높음	• 평균 점수가 5점 이상인 경우
중간	• 평균 점수가 3점 이상, 5점 미만인 경우
낮음	• 평균 점수가 3점 미만인 경우

2) 개인정보 보호 수준 분석

- 데이터 이용자 또는 요청자의 개인정보 보호 수준을 검토하고 평가 실시
- 평가단 개인별로 〈표 3〉 평가 지표의 세부 질문에 대해 검토하고 개별 평가 지표별로 '예' 또는 '아니오'로 평가를 실시

〈표 3〉 개인정보 보호 수준 평가 지표

구 분	세부 지표	평가
개인정보 보호 능력	• 데이터에 접근할 수 있는 인력에 대해 보안각서를 받고 있음	예/아니오
	• 데이터에 접근할 수 있는 인력에 대해 정기적으로 보안 교육을 실시하고 있음	예/아니오
	• 데이터 이용자 또는 요청자가 데이터의 보관 및 처리를 위한 관리계획을 수립하고 있음	예/아니오
	• 데이터 이용자 또는 요청자가 데이터의 보관 및 처리를 위한 관리계획에 따라 운영하고 있음	예/아니오
	• 데이터는 물리적, 기술적 보호 조치가 마련된 안전한 방법을 이용해서 제공하고 제공 받음	예/아니오
	• 침입차단 및 침입탐지 시스템이 설치된 서버, PC 등에서 이용됨	예/아니오
	• 데이터에 접근할 수 있는 인력의 접근권한 부여 및 접근 이력이 관리되고 있음	예/아니오
	• 데이터 이용자 또는 요청자가 보안 관리부서로부터 정기적으로 보안 점검을 받고 있음	예/아니오
	• 데이터 이용자 또는 요청자가 ISO27001, ISMS, PIMS 등의 인증을 받음	예/아니오

- 평가점수는 개인별로 각 평가지표에 대해 '예'로 평가한 지표의 개수를 합산해서 산출 (개인별 점수는 최대 9점, 점수가 높을수록 보호수준이 높음)
- 개인별 점수를 합산한 후 전체 평가인원의 수로 나누어 '개인정보 보호 수준'의 평균 점수를 구하고, 〈표 4〉 평가 기준에 따라 '높음', '중간', '낮음', '없음'으로 1차 평가결과를 도출
- 1차 평가결과는 평가단 토의를 거쳐 확정하되, 1차 평가결과를 기준표와 달리 적용하는 경우에는 이에 대한 사유를 명확히 문서로 남겨야 함

〈표 4〉 개인정보 보호 수준 분석 평가 기준표

구 분	평가 기준
높음	• 평균 점수가 6점 이상인 경우
중간	• 평균 점수가 4점 이상, 5점 미만인 경우
낮음	• 평균 점수가 4점 미만인 경우
없음	• 인터넷 등 일반에 공개하는 경우

3) 재식별 시도 가능성 분석

- ‘1) 재식별 의도 및 능력 분석’, ‘2) 개인정보 보호 수준 분석’의 결과를 고려해서 비식별 조치된 데이터에 대한 재식별 시도 가능성을 평가
- 재식별 시도 가능성에 대한 평가는 ‘빈번한’, ‘가능한’, ‘가끔’, ‘거의 없는’ 등 4단계로 평가
- 아래 그림과 같이 ‘1) 재식별 의도 및 능력 분석’의 결과 값과, ‘2) 개인정보 보호 수준 분석’의 결과 값이 교차하는 지점의 평가값으로 재식별 시도 가능성 분석

● 재식별 시도 가능성 분석표 ●

2)개인정보 보호 수준				
없음	빈번한	빈번한	빈번한	
낮음	가능한	가능한	빈번한	
중간	가끔	가끔	가능한	
높음	거의 없는	거의 없는	가끔	
	낮음	중간	높음	1) 재식별 의도 및 능력

재식별 시 영향 분석

- 데이터가 의도적 또는 비의도적으로 재식별 되었을 때 정보주체에게 미치는 영향에 대해 분석
  - 특히, 경제적 피해 또는 비경제적인 피해(개인정보 또는 프라이버시 침해)를 줄 수 있는 가능성에 대해 평가를 실시
- 평가단 개인별로 〈표 5〉 평가지표 세부 질문에 대해 검토하고 지표별로 ‘예’ 또는 ‘아니오’로 평가를 실시함
- 평가점수는 개인별로 각 평가지표에 대해 ‘예’로 평가한 지표의 개수를 합산해서 산출함(개인별 점수는 최대 4점, 점수가 높을수록 재식별시 영향이 큼)
- 개인별 점수를 합산한 후 전체 평가인원의 수로 나누어 ‘재식별시 영향 분석’의 평균 점수를 구하고, 〈표 6〉 평가 기준에 따라 ‘높음’, ‘중간’, ‘낮음’으로 1차 평가결과를 도출함

○ 1차 평가결과는 평가단 토의를 거쳐 확정하되, 1차 평가결과를 기준표와 달리 적용하는 경우에는 이에 대한 사유를 명확히 문서로 남겨야 함

● 〈표 5〉 재식별시 영향 분석 평가 지표 ●

구 분	세부 지표	평가
재식별시 영향	• 데이터가 재식별되었을 때 법적, 도덕적, 기술적 이슈로 사회적인 혼란을 가져올 가능성이 있음	예/아니오
	• 데이터가 재식별되었을 때 관련 정보주체의 개인정보 또는 프라이버시를 침해할 수 있음	예/아니오
	• 데이터가 재식별되었을 때 관련 정보주체에게 경제적 또는 비경제적 손실을 발생시킬 수 있음	예/아니오
	• 데이터가 재식별되었을 때 데이터 이용자 또는 요청자에게 경제적 또는 비경제적 손실을 발생시킬 수 있음	예/아니오

● 〈표 6〉 재식별시 영향 분석 평가 기준표 ●

구 분	평가 기준
높음	• 평균 점수가 2점 이상인 경우
중간	• 평균 점수가 1점 이상, 2점 미만인 경우
낮음	• 평균 점수가 1점 미만인 경우

계량 분석

○ 평가 대상 데이터의 특성을 고려하여 평가 대상 데이터에 대한 비식별 수준을 분석할 수 있는 분석 기법을 선정하고 분석 값(예시:k=5) 도출

※ 평가단에서 데이터의 특성, 비식별 정도 등을 고려해서 분석 기법 선정

– 비식별 정도를 분석하기 위한 기법에는 k–익명성(k–anonymity), ℓ –다양성(ℓ –diversity), t–근접성(t–closeness) 등의 프라이버시 보호 모델이 있음

○ 분석결과는 '평가 기준값' 결정시 참고할 수 있으며, 필요시 재분석 할 수 있음

○ 평가 대상 데이터에 대한 비식별 정도에 대한 계량 분석은 평가단에서 직접 수행 하거나, 외부의 공신력 있는 전문기관에 의뢰하여 수행할 수 있음

평가 기준값 결정

- 평가단은 비식별 조치의 적정성을 평가하기 위하여 'k–익명성', ' ℓ – 다양성', 't–근접성' 값 등을 단독 또는 복수개 이상으로 설정할 수 있음

- 평가 기준 값 설정시 고려 사항
  - 평가 대상 데이터의 속성자 항목 수, 규모, 시간 흐름에 따른 누적 데이터 존재 여부 등의 데이터 특징
  - 기초자료
  - 사전검토 결과
  - 재식별 시도 가능성 분석 결과
  - 재식별시 영향 분석 결과
  - 계량 분석 결과

- 필요시 계량 분석을 재실시 할 수 있으며, 이때 분석 기준 등에 대해서도 재검토 및 설정할 수 있음

● 평가 기준 값 사례 *●

재식별시 영향					
침해위험 높음	k = 5 ℓ = 2	k = 10 ℓ = 3	k = 10 ℓ = 4	k = 20 ℓ = 5 t 〈 0.3	
침해위험 중간	k = 3 ℓ = 2	k = 5 ℓ = 2	k = 10 ℓ = 3	k = 10 ℓ = 4	
침해위험 낮음	k = 3 ℓ = 2	k = 5 ℓ = 2	k = 5 ℓ = 2	k = 10 ℓ = 3	
	거의 없는	가끔	가능한	빈번한	재식별 시도 가능성

* 세부 검토 기준 값은 단순 사례이며, 실제 적용시 일반적인 기준 값으로 이용하는 것은 적정하지 않을 수 있음. 기준값에 대한 결정은 평가단의 검토 및 논의에 따라 적용 프라이버시 모델 및 기준을 정하여 사용해야 함

적정성 평가

- 평가단은 평가 기준 값 결정에서 도출된 평가 기준 값과 계량 분석에서 계산된 분석 값을 비교하여 비식별 조치에 대한 1차 평가 결과를 도출

- 최종적인 평가는 1차 평가결과를 기초로 평가단 토의를 거쳐 최종 확정하며, 1차 평가 결과와 다른 결과를 도출한 경우에는 이에 대한 근거와 사유를 명확히 문서로 남겨야 함

1) k-익명성 값을 이용한 비식별 적정성 평가

- 계량 분석에서 분석된 평가 대상 데이터의 k-익명성 분석값이 평가단에서 결정한 '평가 기준값' 보다 작은 경우에는 개인 식별요소 제거 조치가 '부적정'한 것으로 평가

- 계량 분석에서 분석된 평가 대상 데이터의 k-익명성 분석값이 평가단에서 결정한 '평가 기준값' 보다 크거나 같은 경우에는 개인 식별요소 제거 조치가 '적정'한 것으로 평가

● k-익명성 기반 적정성 평가 사례표 ●

k-익명성 값을 이용한 비식별 조치에 대한 적정성 평가	
계량분석의 k-익명성 값 < 평가 기준값(k-익명성 값)	계량분석의 k-익명성 값 >= 평가 기준값(k-익명성 값)
⇩	⇩
'부적정' (개인 식별요소 제거 조치 필요)	'적정' (개인 식별요소 제거 조치 불필요)

2) ℓ-다양성 값을 이용한 비식별 적정성 평가

- 계량분석에서 분석된 평가대상 데이터의 ℓ-다양성 분석값이 평가단에서 결정한 '평가 기준값(ℓ-다양성)' 보다 작은 경우에는 개인 식별요소의 제거 조치가 '부적정'한 것으로 평가

- '계량 분석'에서 분석된 평가 대상 데이터의 ℓ-다양성 분석 값이 평가단에서 결정한 '평가 기준값(ℓ-다양성)' 보다 크거나 같은 경우에는 개인 식별요소 제거 조치가 '적정'한 것으로 평가

● ℓ –다양성 기반 적정성 평가 사례표 ●

ℓ –다양성 값을 이용한 비식별 조치에 대한 적정성 평가	
계량분석의 ℓ –다양성 값 < 평가 기준값(ℓ –다양성)	계량분석의 ℓ –다양성 값 >= 평가 기준값(ℓ –다양성)
⇩	⇩
'부적정' (개인 식별요소 제거 조치 필요)	'적정' (개인 식별요소 제거 조치 불필요)

3) t–근접성 값을 이용한 비식별 적정성 평가

- 계량 분석에서 분석된 평가 대상 데이터의 t–근접성 분석 값이 평가단에서 결정한 '평가 기준값(t–근접성)' 보다 작은 경우에는 개인 식별요소 제거 조치가 '적정'한 것으로 평가
  - 통상 t–근접성 값의 범위는 0에서 1 사이의 소수이며, 0에 가까울수록 개인을 식별할 가능성이 적다는 것을 의미함

- 계량 분석에서 분석된 평가 대상 데이터의 t–근접성 분석 값이 평가단에서 결정한 '평가 기준값(t–근접성)' 보다 크거나 같은 경우에는 개인 식별요소 제거 조치가 '부적정'한 것으로 평가

● t–근접성 기반 적정성 평가 사례표 ●

t–근접성 값을 이용한 비식별 조치에 대한 적정성 평가	
계량분석의 t–근접성 값 >= 평가 기준 값(t–근접성)	계량분석의 t–근접성 값 < 평가 기준 값(t–근접성)
⇩	⇩
'부적정' (개인 식별요소 제거 조치 필요)	'적정' (개인 식별요소 제거 조치 불필요)

개인정보 비식별 조치 가이드라인

– 비식별 조치 기준 및 지원 · 관리체계 안내 –

부록 1

– 개인정보의 범위 명확화 및 비식별 정보의 안전한 활용을 위한 –

개인정보 보호 관련 법령 통합 해설서

1 개인정보의 범위

개인정보보호법 제2조제1호

"개인정보"란 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 알아볼 수 있는 것을 포함한다)를 말한다.

정보통신망 이용 촉진 및 정보보호 등에 관한 법률 제2조제1항제6호

"개인정보"란 생존하는 개인에 관한 정보로서 성명 · 주민등록번호 등에 의하여 특정한 개인을 알아볼 수 있는 부호 · 문자 · 음성 · 음향 및 영상 등의 정보(해당 정보만으로는 특정 개인을 알아볼 수 없어도 다른 정보와 쉽게 결합하여 알아볼 수 있는 경우에는 그 정보를 포함한다)를 말한다.

신용정보의 이용 및 보호에 관한 법률 제2조제1호 및 제2호, 제34조제1항

"신용정보"란 금융거래 등 상거래에 있어서 거래 상대방의 신용을 판단할 때 필요한 정보로서 다음 각 목의 정보를 말한다.

가. 특정 신용정보주체를 식별할 수 있는 정보
나. 신용정보주체의 거래내용을 판단할 수 있는 정보
다. 신용정보주체의 신용도를 판단할 수 있는 정보
라. 신용정보주체의 신용거래능력을 판단할 수 있는 정보
마. 그 밖에 가목부터 라목까지와 유사한 정보

"개인신용정보"란 신용정보 중 개인의 신용도와 신용거래능력 등을 판단할 때 필요한 정보를 말한다.

"개인식별정보"란 생존하는 개인의 성명, 주소 및 주민등록번호, 여권번호, 운전면허번호, 외국인등록번호, 국내거소신고번호 및 성별, 국적 등 개인을 식별할 수 있는 정보를 말한다.

1 개인정보의 개념

- 「개인정보 보호법」과 「정보통신망 이용 촉진 및 정보보호 등에 관한 법률(이하 정보통신망법)」에서는 개인정보의 개념을 규정하고 있으며, 「신용정보의 이용 및 보호에 관한 법률(이하 신용정보법)」에서는 개인신용정보와 개인식별정보의 개념에 대해 규정하고 있습니다.

○ 우선, 개인정보 보호법과 정보통신망법에서의 개인정보 개념 정의는 법률상 표현이 조금 다르게 되어 있으나, 법률 해석상 그 내용은 사실상 동일합니다.

- 두 법에서 정의하는 개인정보는 살아 있는 개인에 관한 정보로서 개인을 알아볼 수 있는 정보이며, 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보가 포함됩니다.
- 한편, 신용정보법 상의 개인신용정보 및 개인식별정보는 개인정보 보호법과 정보통신망법에서 말하는 개인정보 개념과 다르지 않습니다.

○ 현행 신용정보법은 개인신용정보를 "개인을 알아볼 수 있는 정보"일 것을 명시적으로 요구하지는 않지만, 개인을 알아볼 수 없는 신용정보가 개인신용정보에 포함되지 않는다고 보아야 합리적입니다.

○ 또한, 개인식별정보는 생존하는 개인의 성명, 주소 및 주민등록번호, 여권번호, 운전면허번호, 외국인등록번호, 국내거소신고번호 등 개인을 식별할 수 있는 정보를 의미하기 때문입니다.

2 개인정보의 구체적 판단 기준

1 '생존하는' 개인에 관한 정보이여야 합니다.

가. **(법인의 정보)** 개인정보의 주체는 자연인이어야 하며, 법인 또는 단체의 정보는 개인정보에 해당하지 않습니다.

○ 따라서 법인 또는 단체의 이름, 소재지 주소, 대표 연락처(이메일 주소 또는 전화번호), 업무별 연락처, 영업실적 등은 개인정보에 해당하지 않습니다.

○ 그러나 법인 또는 단체의 정보에 해당하면서 동시에 개인에 관한 정보인 대표자를 포함한 임원진과 업무 담당자의 이름 · 주민등록번호 · 자택주소 및 개인 연락처, 사진 등은 개인정보에 해당합니다.

나. **(개인사업자에 관한 정보)** 개인사업자의 상호명, 사업장 주소, 전화번호, 사업자등록번호, 매출액, 납세액 등은 사업체의 운영과 관련한 정보로서 원칙적으로 개인정보에 해당하지 않습니다.

○ 다만, 예외적으로 해당 정보가 사업자 개인의 직업 · 소득수준 · 활동영역 · 사회적 지위 등을 나타내는 정보로 이용되는 경우 개인정보로 보아야 하며, 개인사업자의 사업과 관련된 정보이면서 동시에 사업자 개인의 이름 · 주민등록번호 · 집주소 · 휴대전화번호 등은 개인정보에 해당합니다.

- 또한, 개인사업자의 고유식별정보 및 민감정보는 법령에 근거가 있거나 별도의 동의가 있어야 수집 · 이용이 가능하며, 고유식별정보 중 주민등록번호는 동의에 의하여서는 수집할 수 없으며 법령에 구체적인 근거가 있어야 처리할 수 있습니다.

다. **(사망자의 정보)** 개인정보 보호 법령상 개인정보는 '살아있는' 자연인에 관한 정보이므로 사망했거나 실종선고 등 관계 법령에 의해 사망한 것으로 간주되는 자에 관한 정보는 개인정보로 볼 수 없습니다.

- 다만, 사망자의 정보라고 하더라도 유족과의 관계를 알 수 있는 정보는 유족의 개인정보에 해당합니다.

라. **(사물에 관한 정보)** 사람이 아닌 사물에 관한 정보는 원칙적으로 개인정보에 해당하지 않습니다. 그러나 해당 사물 등의 제조자 또는 소유자 등을 나타내는 정보는 개인정보에 해당합니다.

- 예를 들어, 특정 건물이나 아파트의 소유자가 자연인인 경우, 그 건물이나 아파트의 주소가 특정 소유자를 알아보는데 이용된다면 개인정보에 해당합니다.

2 '개인에 관한' 정보이여야 합니다.

가. **(개인에 관한 정보의 범위)** '개인에 관한 정보'란 당해 개인에 대한 사실 · 판단 · 평가 등 개인과 관련된 정보를 의미하므로, 특정 개인의 신원, 성격, 행위 등에 관한 것 또는 정보주체에 관한 평가 등에 영향을 미치는 것은 개인정보에 해당합니다.

나. **(2인 이상의 관련성)** '개인에 관한 정보'는 반드시 특정 1인만에 관한 정보이어야 한다는 의미가 아니며, 직 · 간접적으로 2인 이상에 관한 정보는 각자의 정보에 해당합니다.

- SNS에 단체 사진을 올린다면 사진의 영상정보는 사진에 있는 인물 모두의 개인정보에 해당하며, 의사가 특정 아동의 심리치료를 위해 진료 기록을 작성하면서 아동의 부모 행태 등을 포함하였다면 그 진료기록은 아동과 부모 모두의 개인정보에 해당합니다.

3 '정보'의 내용 · 형태 등은 제한이 없습니다.

가. **(정보의 내용 · 형태)** 정보의 내용 · 형태 등은 특별한 제한이 없어서 개인을 알아볼 수 있는 모든 정보가 개인정보가 될 수 있습니다.

- 즉, 디지털 형태나 수기 형태, 자동처리 정보와 수동처리정보 등 그 형태 또는 처리방식과 관계없이 모두 개인정보에 해당할 수 있습니다.

나. **(정보의 주관성 또는 객관성)** 정보주체와 관련되어 있으면 키, 나이, 몸무게 등 '객관적 사실'에 관한 정보나 그 사람에 대한 제3자의 의견 등 '주관적 평가' 정보 모두 개인정보가 될 수 있습니다.

○ 또한, 그 정보가 반드시 '사실'이거나 '증명된 것'이 아닌 부정확한 정보 또는 허위의 정보라도 특정한 개인에 관한 정보이면 개인정보가 될 수 있습니다.

4 개인을 '알아볼 수 있는' 정보이여야 합니다.

○ **('알아볼 수 있는'의 의미)**는 해당 정보를 '처리하는 자'의 입장에서 합리적으로 활용될 가능성이 있는 수단을 고려하여 개인을 알아볼 수 있다면 개인정보에 해당합니다.

○ 여기서 '처리'란 개인정보 보호법 제2조제2호에 따른 개인정보의 수집, 생성, 연계, 연동, 기록, 저장, 보유, 가공, 편집, 검색, 출력, 정정(訂正), 복구, 이용, 제공, 공개, 파기(破棄), 그 밖에 이와 유사한 행위를 말합니다.

○ 현재 처리하는 자 외에도 제공 등에 따라 향후 처리가 예정된 자도 포함됩니다.

○ 한편, 주민등록번호와 같은 고유식별정보는 해당 정보만으로도 정보주체인 개인을 알아볼 수 있지만, 생년월일의 경우에는 같은 날 태어난 사람이 여러 사람일 수 있으므로 다른 정보없이 생년월일 그 자체만으로는 개인을 알아볼 수 있다고 볼 수 없습니다.

5 다른 정보와 '쉽게 결합하여' 개인을 알아볼 수 있는 정보도 포함합니다.

○ **('쉽게 결합하여'의 의미)**는 결합 대상이 될 정보의 '입수 가능성'이 있어야 하고 '결합 가능성'이 높아야 함을 의미합니다.

○ '입수 가능성'의 의미는 두 종 이상의 정보를 결합하기 위해서는 결합에 필요한 정보에 합법적으로 접근 · 입수할 수 있어야 하며, 이는 해킹 등 불법적인 방법으로 취득한 정보까지 포함한다고 볼 수는 없습니다.

○ '결합 가능성'의 의미는 합법적인 방법으로 정보를 입수하여도 현재의 기술 수준에 비추어 결합이 사실상 불가능하거나, 결합하는데 비합리적인 수준의 비용이나 노력이 수반된다면 이는 결합이 용이하다고 볼 수 없습니다.

○ 따라서, 공유 · 공개될 가능성이 희박한 정보는 합법적 입수 가능성이 없다고 보아야 하며, 일반적으로 사업자가 구매하기 어려울 정도로 고가의 컴퓨터가 필요한 경우라면 '쉽게 결합'하기 어렵다고 보아야 합니다.

〈 참고 〉 폴란드 개인정보보호법

폴란드 개인정보보호법은 '식별을 위해서 불합리한 정도의 시간, 비용 및 인력을 필요로 하는 경우에는 식별을 가능하게 하는 정보로 간주해서는 안 된다'라고 하고 있음

3 개인정보의 개념 관련 판례 및 유권해석 사례

가. 판례(判例)

〈 판례 〉 휴대전화번호 뒤 4자리

대전지법 논산지원(2013고단17 판결)은 「휴대전화번호 뒷자리 4자」에 대하여, "휴대전화번호 뒷자리 4자만으로도 그 전화번호 사용자가 누구인지를 식별할 수 있는 경우가 있고, 특히 그 전화번호 사용자와 일정한 인적 관계를 맺어온 사람이라면 더더욱 그러할 가능성이 높으며, 설령 휴대전화번호 뒷자리 4자만으로는 그 전화번호 사용자를 식별하지 못한다 하더라도 그 뒷자리 번호 4자와 관련성이 있는 다른 정보(생일, 기념일, 집 전화번호, 가족 전화번호, 기존 통화내역 등)와 쉽게 결합하여 그 전화번호 사용자가 누구인지를 알아볼 수도 있다"고 하여 「개인정보보호법」 제2조제1호에 규정된 개인정보에 해당된다고 판시하고 있다.

나. 유권해석 사례

〈 개인정보보호위원회 결정 〉 배달음식점 고객의 전화번호 및 주소

개인정보보호위원회는 2012년 1월 30일 「개인정보 보호법 관련 법령해석 요청 건(의안 제2호)」에 대한 의결 이유에서 "배달음식점 고객의 전화번호 및 주소는 그 자체로는 특정 개인을 식별할 수 없지만, 용이하게 다른 정보와 결합하여 특정 개인을 식별할 수 있으므로, 「개인정보 보호법」 제2조제1호의 '개인정보'에 해당함"이라고 해석하고 있다. 용이하게 다른 정보와 결합하여 특정 개인을 식별할 수 있기만 하면 그 자체로서 특정 개인을 식별할 수 없는 경우에도 개인정보로 보고 있으므로, 의결 이유에서 지적한 '고객의 전화번호 및 주소'이외에도 개인정보로 인정할 수 있는 정보의 범위가 확장될 수 있다고 해석할 수 있다.

2 비식별 정보의 개념 및 재식별 시 제재

1 비식별 정보의 개념

가. **(비식별 정보의 개념)** 개인정보를 비식별 조치한 정보, 즉 '비식별 정보'란 정보의 집합물에 대해 「개인정보 비식별 조치 가이드라인」에 따라 적정하게 '비식별 조치'된 정보를 말합니다.

- '비식별 조치'란 정보의 집합물에서 개인을 식별할 수 있는 요소를 전부 또는 일부 삭제하거나 대체 등의 방법을 통해 개인을 알아볼 수 없도록 하는 조치를 말합니다. (자세한 내용은 「개인정보 비식별 조치 가이드라인」 참고)
- 참고로 EU 개인정보지침은 'anonymization, 익명화'한 경우에는 지침이 적용되지 않도록 하고 있는데, 이 해설서에서 안내하는 '비식별 조치'는 EU의 익명화와 사실상 같은 개념입니다.
- 한편 비식별 정보가 개인정보에 해당하는지 여부가 의문이 있을 수 있으나, 가이드라인에 따라 적정하게 비식별 조치가 된 정보는 더 이상 특정 개인을 알아볼 수가 없으므로 개인정보가 아닌 것으로 추정됩니다.
- 개인정보가 아닌 것으로 추정된다는 의미는 개인정보에 해당한다는 반증이 없는 한 개인정보가 아니되, 개인정보라는 반증이 나오는 경우 개인정보로 본다는 뜻입니다.

나. **(비식별 정보의 활용)** 비식별 정보는 개인정보가 아닌 정보로 추정되므로 정보주체로부터의 별도 동의없이 해당 정보를 이용하거나 제3자에게 제공할 수 있습니다.

- 다만, 개인정보가 아닌 것으로 추정되더라도 불특정 다수에게 공개되는 경우에는 다른 정보를 보유하고 있는 누군가에 의해 해당 정보주체가 식별될 가능성이 있으므로 비식별 정보의 공개는 원칙적으로 금지됩니다.

다. **(비식별 정보의 보호)** 비식별 정보는 개인정보가 아닌 것으로 추정되지만, 새로운 결합 기술이 나타나거나 결합 가능한 정보가 증가하는 경우에는 정보주체가 '재식별'될 가능성이 있습니다. 따라서 비식별 정보라고 하더라도 필수적인 관리적 · 기술적 보호 조치는 이행해야 합니다. (자세한 내용은 「개인정보 비식별 조치 가이드라인」 참고)

2 재식별 시 제재

가. 비식별 정보를 재식별하여 이용하거나 제3자에게 제공한 경우에는 개인정보의 목적 외 이용 · 제공에 해당하여 5년 이하의 징역 또는 5천만원 이하의 벌금형에 처해집니다 (개인정보 보호법 제18조제1항 위반, 정보통신망법 제24조 및 제24조의2 위반, 신용정보법 제32조 및 제33조 위반)

※ 정보통신망법 적용 사업자는 위반행위 관련 매출액의 3% 이하 과징금이 추가 부과될 수 있음

- 예를 들어, 비식별 정보를 제3자에게 제공하면서 비식별 조치 요령을 공유하거나 공개되어 있는 알고리즘으로 암호화하여 쉽게 복호화될 수 있도록 정보를 제공하는 경우 등이 이에 해당 할 수 있습니다.

나. 비식별 정보를 처리하는 자(비식별 정보를 제공받은 자 포함)가 해당 정보를 이용하는 과정에서 재식별하게 된 경우에는 해당 정보를 즉시 처리중지하고 파기하여야 합니다.

- 추가적 비식별 조치없이 재식별된 정보를 보관하는 경우 5천만원 이하의 과태료가 부과됩니다 (개인정보 보호법 제15조 제1항 위반, 정보통신망법 제22조제1항 위반, 신용정보법 제15조 제2항 위반)

※ 정보통신망법 적용 사업자는 5년 이하 징역 또는 5천만원 이하 벌금형에 처해지며, 위반행위 관련 매출액의 3% 이하 과징금이 추가 부과될 수 있음

〈 참고 〉 미국의 De-identified data

- 소비자 프라이버시 권리장전 Sec4.(2)에서는 'De-identified data'의 개념을 정의하고 해당 정보는 비개인정보로 취급

〈 참고 〉 EU의 개인정보 보호지침 서문 제26조

- EU 개인정보 보호지침의 서문 제26조에서는 정보주체의 신원을 확인할 수 없는 익명정보는 보호원칙이 적용되지 않음을 명시

〈 참고 〉 일본의 익명가공정보

● 개인정보 보호법에 빅데이터 활용 목적의 '익명가공정보'라는 개념을 신설

〈일본 개인정보 보호법 제2조제9항〉
이 법률에서 정하는 "익명가공정보"란 다음의 각 호에 해당하는 개인정보 구분에 대응하고 해당 각 호에 정하는 조치를 취하여 특정 개인을 식별할 수 없도록 개인정보를 가공하여 얻는 개인에 관한 정보로서, 해당 개인정보를 복원할 수 없도록 한 것을 말한다.

● 익명가공정보의 법적취급 : 복원불가능을 전제로 하여 정보주체의 동의를 받을 필요가 없고 제3자 제공도 자유로움, 다만 일정한 기술적 · 관리적 조치를 해야 함

● 개인정보취급사업자의 익명가공정보 작성시 의무사항

– 개인정보취급사업자가 익명가공정보를 작성할 때는 특정 개인을 식별하는 것 및 그 작성에 이용되는 개인정보를 복원할 수 없도록 가공해야 한다. (제36조 제1항)

– 개인정보취급사업자가 익명가공정보를 작성할 때는 정보의 누설을 방지하기 위하여 정보의 안전관리를 위한 조치를 하여야 한다. (제36조 제2항)

3 개인정보 보호법과 다른 법률과의 관계

1 일반 원칙

가. 일반법과 특별법이 저촉되면 특별법이 먼저 적용되고, 특별법에 규정이 없는 사항에 대해서는 일반법이 적용된다. (헌법재판소 2004. 9. 23. 2004헌가12 결정 참조)

나. 법률이 상호 모순, 저촉되는 경우에는 신법이 구법에, 그리고 특별법이 일반법에 우선하나, 법률이 상호 모순되는지 여부는 각 법률의 입법목적, 규정사항 및 그 적용범위 등을 종합적으로 검토하여 판단하여야 한다. (대법원 1989. 9. 12. 선고 88누6856 판결, 대법원 1995. 2. 3. 선고 94누2985 판결 등)

2 개인정보 보호법과 정보통신망법의 관계

가. 정보통신서비스 제공자에 대하여는 정보통신망법이 우선 적용되지만, 정보통신망법에 특별한 규정이 없고 개인정보 보호법과 상호 모순·충돌하지 않는 경우에는 개인정보 보호법이 적용됩니다.

나. 개인정보 보호법 제2조제1호의 개인정보 개념과 정보통신망법 제2조제6호의 개인정보 개념 정의는 개인정보의 예시와 관련하여 일부 차이가 있을 뿐 동일한 내용을 규정하고 있으므로 사실상 동일한 개념이라고 볼 수 있습니다.

○ 따라서, 개인정보의 개념과 개인정보가 아닌 것으로 추정되는 비식별 정보의 개념 또한 차이가 없습니다.

● **개인정보 보호법과 정보통신망법 상의 "개인정보" 정의 규정 비교** ●

「개인정보 보호법」 제2조제1호	정보통신망법 제2조제1항제6호
개인정보란 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 알아볼 수 있는 것을 포함한다)를 말한다.	개인정보란 생존하는 개인에 관한 정보로서 성명·주민등록번호 등에 의하여 특정한 개인을 알아볼 수 있는 부호·문자·음성·음향 및 영상 등의 정보(해당 정보만으로는 특정 개인을 알아볼 수 없어도 다른 정보와 쉽게 결합하여 알아볼 수 있는 경우에는 그 정보를 포함한다)를 말한다.

3 개인정보 보호법과 신용정보법의 관계

가. 신용정보법 제3조의2는 "개인정보의 보호에 관하여 이 법에 특별한 규정이 있는 경우를 제외하고는 「개인정보 보호법」에서 정하는 바에 따른다"라고 하고 있으므로 신용정보법은 개인정보 보호법에 대해 특별법의 지위를 가집니다.

○ 따라서, 개인신용정보에 대해서는 신용정보법을 우선 적용하되, 신용정보법에 규정되어 있지 않는 사항은 개인정보 보호법을 적용하여야 합니다.

나. 신용정보법 상의 개인신용정보 및 개인식별정보는 금융거래 등에 사용되는 개인정보의 특수한 형태로 개인정보 보호법 상의 개인정보의 개념은 신용정보법 상의 개인신용정보 및 개인식별정보보다 포괄적인 개념입니다.

○ 따라서, 신용정보법 상의 개인신용정보와 개인식별정보는 당연히 개인정보 보호법 상의 개인정보에 해당한다고 봐야 하며, 개인정보 보호법 상 개인정보가 아닌 것으로 추정되는 비식별 정보는 신용정보법에서도 개인정보가 아닌 것으로 추정됩니다.

부록 2 질의 및 응답(Q&A)

부록 2 질의 및 응답(Q&A)

구분	질의 사항
주요개념 및 적용범위	❶ 개인정보 보호법과 정보통신망법, 신용정보법 등에서 규정하는 개인정보의 개념에 차이는 없는지?
	❷ 어떤 정보가 개인정보에 해당하는 지를 판단할 때 해당 정보를 처리하는 자의 관점에서 보아야 한다고 했는데, 이의 의미는?
	❸ 개인정보를 비식별 조치하는 경우 개인정보가 아닌 것으로 추정한다는데 이것의 법적 의미는?
	❹ 본 가이드라인이 통신 사업자나 금융기관 등에도 적용되는지?
	❺ 통계청 등 관련 법령에 따라 데이터를 수집, 연계 · 활용하는 기관들에 대해 본 가이드라인의 적용 여부는?
	❻ 개인정보 보호법 제18조제2항제4호에 따른 정보와 「개인정보 비식별 조치 가이드라인」에 따라 비식별 조치한 정보와의 차이는?
비식별 조치	❼ 본 가이드라인에서 말하는 비식별 조치는 무엇인가?
	❽ 고객정보를 제공받는 기관이 비식별 조치를 한다면, 제공하는 기관은 비식별 조치를 하지 않을 수 있는지?
	❾ 비식별 대상인 개인 식별정보의 구체적인 항목은 어떻게 되는지?
	❿ 비식별 조치가 적정한지 어떻게 알 수 있나?
	⓫ 평가단 구성 시 '데이터 이용 목적과 직접적인 이해관계가 없는 자'로 위원을 구성토록 하고 있음. 이 때, 평가단에 참여하는 내부전문가의 경우에는 이해관계자에 포섭될 수 있는데, 평가단 구성에 대한 구체적 기준은 무엇인지?
	⓬ 평가단이 단순히 k-익명성 값만을 가지고 판단할 가능성이 있는데?
	⓭ 적정성 평가 시 k-익명성을 기본으로 활용하되, 필요시 추가적인 평가모델(ℓ-다양성, t-근접성)을 활용하도록 규정하고 있는데, '필요시'에 대한 객관적인 기준은 무엇인지?
	⓮ 평가단이 '적정'으로 판단한 비식별 정보가 추후에 재식별된 경우 그 책임소재는?
비식별 정보 활용	⓯ 비식별 조치된 고객정보를 시장조사, 신상품 개발, 마케팅 전략수립 등에 활용하거나 제휴회사에 제공하고자 하는 경우 해당 고객의 동의가 필요한가?
	⓰ '적절한 수준으로 비식별 조치'된 데이터에 대해서는 제3자 제공 동의를 받지 않았더라도 다양한 비즈니스 목적으로 제3자에게 유상/무상으로 제공이 가능한지?
	⓱ 1대1 마케팅 등 맞춤형 서비스 목적으로 이용 가능한지?

비식별 정보 활용	⑱ 적절한 수준으로 비식별 조치'된 데이터에 대해서는 정보활용에 대한 동의를 받지 않았더라도 다양한 고객 분석, 신상품 기획, 세그먼트 마케팅 등의 목적에 활용할 수 있는가?
	⑲ 법령상 '민감정보'에 해당하는 건강정보 및 유전정보의 경우에도 비식별 조치를 한다면 개인의 동의없이 활용 가능한지?
	⑳ 유전정보도 다른 건강정보와 동일하게 취급해야 하는지 아니면 별도 강화된 조치가 필요한지?
	㉑ 고객 행태 분석을 위해 서비스 이용 기록이나 SNS 등에 공개된 정보를 수집하여 비식별 조치 후 이용하는 것이 가능한가?
	㉒ 해당 데이터를 가이드라인에 따라 적정성 평가를 받고 활용하다 새로운 분석을 위해 비식별 조치 방법을 변경하고자 할 때 이 경우 별도의 적정성 평가를 진행해야 하는지?
사후관리	㉓ 비식별 정보가 재식별되면 어떻게 해야 하나?
	㉔ 비식별 정보도 재식별 가능성이 있다고 하는데 비식별 조치가 제대로 안된 것이 아닌지?
	㉕ 제공한 비식별 정보의 모니터링 책임이 제공자에게 있는 것인지?
	㉖ '적절한 수준으로 비식별 조치'된 정보는 개인정보 보관기한 등과 무관하게 저장하여 활용할 수 있는지?
	㉗ 비식별 조치를 한 정보에 대한 열람, 정정 · 삭제 및 처리정지 등의 요구에 어떻게 대응해야 하는지?
	㉘ 개인이 재식별 된 경우, 개인정보 보호법 제34조의 '유출'로 보아 해당 정보주체에 대한 유출통지를 해야 하는지?
지원 및 관리체계 (재식별 법적 제재)	㉙ 다른 사업자가 보유한 DB를 결합하여 빅데이터 분석 등에 활용할 수 있는가?
	㉚ DB 결합을 위해 분야별 전문기관에 데이터를 제공하는 경우 식별자만을 제거하고 제3의 기관에서 제공한 알고리즘으로 임시대체키를 생성하여 붙인 뒤 다른 비식별 조치를 하지 않을 수 있는지?
	㉛ 업종별의 경우 전문기관이 서로 다른데 이업종의 DB결합시 각자 자신이 속한 업종의 전문기관의 지원을 받으면 되는 것인지?
	㉜ 이종 DB결합시 주민등록번호를 사용하여 임시 대체키를 만드는 것이 현행법 위반인지?
	㉝ 기업 내에서 서로 다른 부서간의 DB를 결합하여 이용하고자 하는 경우에도 반드시 외부의 전문기관을 통해야 하는지?
	㉞ 개인정보를 비식별 조치하여 활용할 경우 법적인 책임은 없는지?
	㉟ 가이드라인에서 정하는 대로 비식별 조치를 실시하였다고 가정할 때 의도하지 않은 재식별 발생시 면책이 가능한지?

주요개념 및 적용범위

문1 **개인정보 보호법과 정보통신망법, 신용정보법 등에서 규정하는 개인정보의 개념에 차이는 없는지?**

답 개인정보 보호법 제2조제1호의 개인정보 개념과 정보통신망법 제2조제1항제6호의 개인정보 개념은 일부 예시와 관련하여 차이가 있을 뿐 사실상 동일한 개념임
또한, 신용정보법 상의 개인신용정보와 개인식별정보는 금융거래 등에 사용되는 개인정보의 특수한 형태로 이는 개인정보 보호법 상의 개인정보에 해당함

문2 **어떤 정보가 개인정보에 해당하는 지를 판단할 때 해당 정보를 처리하는 자의 관점에서 보아야 한다고 했는데, 이의 의미는?**

답 개인정보 보호법 상 개인정보란 그 자체의 정보로 또는 다른 정보와 쉽게 결합하여 개인을 알아볼 수 있는 정보를 의미하는 바, 여기서 '알아볼 수 있는'의 주체는 개인정보를 처리하는 자로 한정하여야 함
이는 만약 '알아볼 수 있는'의 주체를 불특정 제3자로 확대 해석하게 되면, 모든 정보가 다른 정보와 결합하여 개인정보가 될 수 있는 불합리한 결과가 초래되기 때문임
다만, '해당 정보를 처리하는 자'는 정보를 제공하는 관계에서는 해당 정보를 제공받은 자를 포함하는 개념임

문3 **개인정보를 비식별 조치하는 경우 개인정보가 아닌 것으로 추정한다는데 이것의 법적 의미는?**

답 본 가이드라인에 따라 특정 개인을 알아볼 수 없도록 비식별 조치가 적정하게 된 경우에는 개인정보에 해당한다는 반증이 없는 한 개인정보가 아닌 것으로 보되, 개인정보라는 반증이 나오는 경우 개인정보로 본다는 뜻임

문4 **본 가이드라인이 통신 사업자나 금융기관 등에도 적용되는지?**

답 본 가이드라인은 비식별 조치 기준과 지원 · 관리체계 등 비식별 정보를 안전하게 활용하기 위한 목적으로 행정자치부(개인정보 보호법), 방송통신위원회(정보통신망법), 금융위원회(신용정보법), 보건복지부(의료법) 등 관계부처와 공동으로 마련한 것으로 본 가이드라인은 통신 사업자나 금융기관 등 모든 사업자에 적용됨

문5 **통계청 등 관련 법령에 따라 데이터를 수집, 연계 · 활용하는 기관들에 대해 본 가이드라인의 적용 여부는?**

답 통계법 등 관련 개별 법령에서 정한 바에 따라 데이터를 수집, 연계 · 활용하는 기관들의 경우에는 본 가이드라인의 내용보다 관련 법령의 규정이 우선 적용되어야 함
따라서, 통계청 등이 관계 법령에 따라 통계작성 등 고유의 공공목적을 위해 데이터를 수집, 연계 · 활용하는 경우는 해당 법령에 근거한 비식별 조치 방식을 적용해야 함

문6 **개인정보 보호법 제18조제2항제4호에 따른 정보와 「개인정보 비식별 조치 가이드라인」에 따라 비식별 조치한 정보와의 차이는?**

답 개인정보 보호법 제18조제2항제4호에 따른 정보는 개인정보가 아닌 것으로 추정한다는 점에서는 비식별 정보와 동일하지만, 법에서 허용된 통계작성 및 학술연구 등을 위한 국한된 목적으로만 제공할 수 있다는 점에서 차이가 있어서 「개인정보 비식별 조치 가이드라인」에 따른 비식별 조치 중 적정성 평가는 제외할 수 있음

비식별 조치

문7 본 가이드라인에서 말하는 비식별 조치는 무엇인가?

답 "비식별 조치"란 정보집합물(데이터 셋)에서 개인을 식별할 수 있는 요소(식별자, 속성자)를 전부 또는 일부 삭제하거나 대체하는 등의 방법으로 개인을 알아볼 수 없도록 하는 조치를 말함

비식별 조치는 우선 '가명처리', '총계처리', '데이터 삭제', '데이터 범주화', '데이터 마스킹'과 같은 기법 등을 활용하여 개인을 알아볼 수 없도록 조치하고, 또 'k-익명성' 모델 등을 활용하여 비식별 조치가 적정한지 여부에 대한 평가절차를 거쳐야 함

문8 고객정보를 제공받는 기관이 비식별 조치를 한다면, 제공하는 기관은 비식별 조치를 하지 않을 수 있는지?

답 비식별 조치되지 않은 개인정보 제공은 개인정보 제3자 제공에 해당하므로 정보주체의 별도 동의가 없었다면 현행법 위반임

따라서, 정보주체로부터 제3자 제공에 대한 별도동의를 받지 않았다면 제공하는 기관이 비식별 조치를 한 후 제공하여야 함

문9 비식별 대상인 개인 식별정보의 구체적인 항목은 어떻게 되는지?

답 본 가이드라인에 따른 비식별 조치 대상은 정보집합물(데이터 셋)에 포함되어 있는 개인 식별요소(식별자 및 속성자)이며, 각 식별요소에 대한 비식별 조치 여부와 방법은 데이터 이용 목적 등을 고려하여 결정되어야 할 것임

식별자는 원칙적으로 삭제하여야 하고, 데이터 이용 목적상 반드시 필요한 식별자는 비식별 조치 후 활용하여야 함

속성자는 데이터 이용 목적과 관련이 없는 경우 삭제하고, 이용 목적과 관련이 있는 속성자 중 식별요소가 있는 경우에는 가명처리, 총계처리 등의 조치 기법을 활용하여 비식별 조치하여야 함

문10 비식별 조치가 적정한지 어떻게 알 수 있나?

답 비식별 조치가 적정한지에 대한 여부는 프라이버시 보호 모델인 'k-익명성' 등을 활용한 평가를 거쳐 결정됨

예를 들어, k값을 5로 정하여 비식별 조치하였다면 정보집합물 내에는 특정인을 식별할 수 있는 요소가 없음은 물론이고, 최소 5개 이상의 동일한 레코드(row, 列)가 존재하여 특정 개인을 식별하기 어려우므로 비식별 조치가 적정한 것으로 봄

문11 평가단 구성 시 '데이터 이용 목적과 직접적인 이해관계가 없는 자'로 위원을 구성토록 하고 있음. 이 때, 평가단에 참여하는 내부전문가의 경우에는 이해관계자에 포섭될 수 있는데, 평가단 구성에 대한 구체적 기준은 무엇인지?

답 평가단 구성에 내부 전문가를 참여시키고자 하는 경우에는 해당 데이터 이용 목적과 직접적인 이해관계가 없는 내부 전문가를 지정하여 평가 결과의 공정성과 신뢰성을 보장하여야 할 것이며, 이 경우 간접적인 이해관계자까지 모두 배재해야 할 필요는 없음

문12 평가단이 단순히 k-익명성 값만을 가지고 판단할 가능성이 있는데?

답 '비식별 적정성 평가단'이 단순히 k-익명성 값을 만족하는지 여부만을 평가하는 것이 아니라 평가대상 데이터의 특성, 재식별 시도 가능성 등을 고려하여 현재의 비식별 조치 수준이 적정한지, 재식별 위험이 없는지 여부 등을 종합적으로 평가하는 것임

문13 **적정성 평가 시 k-익명성을 기본으로 활용하되, 필요시 추가적인 평가모델(ℓ-다양성, t-근접성)을 활용하도록 하고 있는데, '필요시'에 대한 객관적인 기준은 무엇인지?**

답 본 가이드라인에 제시된 k-익명성을 활용한 평가는 최소한의 평가 수단이며, 평가 대상 데이터의 특성, 재식별 시도 가능성 등을 평가단에서 종합적으로 판단하여 추가적인 평가모델(ℓ-다양성, t-근접성)을 결정해야 함

예를 들어, k-익명성에 의해 범주화 되었더라도 각 레코드들이 충분한 다양성을 가지지 못하거나, 특정한 값에 쏠려 있다고 판단되는 경우에는 ℓ-다양성 또는 t-근접성을 추가적으로 적용해야 함

문14 **평가단이 '적정'으로 판단한 비식별 정보가 추후에 재식별된 경우 그 책임소재는?**

답 재식별에 대한 책임 소재는 당초 평가단의 평가 내용과 재식별된 경위 등을 종합적으로 고려하여 판단해야 할 사항임

당초 평가단 평가시 '적정'으로 판단할 만한 상당한 근거가 있었다면, 추후에 재식별 되었다는 이유만으로 책임을 부과하는 것은 곤란함

비식별 정보 활용

문15 **비식별 조치된 고객정보를 시장조사, 신상품개발, 마케팅 전략수립 등에 활용하거나 제휴회사에 제공하고자 하는 경우, 해당 고객의 동의가 필요한가?**

답 개인 식별요소 삭제 등 충분한 비식별 조치가 이루어졌다면 고객의 추가 동의 없이 시장조사, 신상품 개발, 마케팅전략 수립 등의 용도로 이용할 수 있음

다만, 제휴회사에 제공하는 경우에는 다른 정보와의 결합을 통한 재식별 가능성이 있으므로 재식별 위험관리 사항을 계약서에 반영하는 등 본 가이드라인에 따른 사항을 준수해야 할 것임

문16 **'적절한 수준으로 비식별 조치'된 데이터에 대해서는 제3자 제공 동의를 받지 않더라도 다양한 비즈니스 목적으로 제3자에게 유상/무상으로 제공이 가능한지?**

답 본 가이드라인에 따라 적정한 비식별 조치가 이루어졌다면 고객의 동의를 받지 않더라도 다양한 비즈니스 용도에 활용될 수 있도록 제3자에게 제공이 가능함. 이 경우 실비 수준의 수수료를 받고 비식별 정보를 제공할 수 있음

또한, 재식별 금지 및 재제공 제한, 재식별 위험시 통지 등의 내용을 해당 비식별 정보 제공과 관련한 계약서에 반드시 포함하여야 함

문17 **1대1 마케팅 등 맞춤형 서비스 목적으로 이용 가능한지?**

답 비식별 조치된 정보는 특정 개인을 알아볼 수 없으므로 1대1 마케팅 등 맞춤형 서비스 목적으로 활용하는 것이 현실적으로 불가능함

※ 개인 식별이 가능한 정보를 이용해 상품 판매 또는 홍보 등 1대1 마케팅을 하려면 현행 법령에 따라 정보주체의 사전 동의 필요

문18 **'적절한 수준으로 비식별 조치'된 데이터에 대해서는 정보활용에 대한 동의를 받지 않았더라도 다양한 고객 분석, 신상품 기획, 세그먼트 마케팅 등의 목적에 활용할 수 있는가?**

답 비식별 정보는 개인정보가 아닌 것으로 추정되는 바, 고객 분석, 신상품 기획, 세그먼트 마케팅 등의 목적으로 활용이 가능

다만, 세그먼트 마케팅을 위한 비식별 조치의 경우 특정 개인을 알아볼 수 없도록 세그먼트를 충분한 규모로 산정해야 함

문19 **법령상 '민감정보'에 해당하는 건강정보 및 유전정보의 경우에도 비식별 조치를 한다면 개인의 동의없이 활용 가능한지?**

답 개인정보 보호법상 민감정보에 해당하더라도 가이드라인에 따라 특정개인을 알아볼 수 없도록 비식별 조치한 경우 개인의 사전동의 없이 빅데이터 분석 등에 활용이 가능함
다만, 「생명윤리 및 안전에 관한 법률」에 근거한 인간 대상의 연구 목적으로 수집된 개인정보는 동 법 제18조에 의거 별도의 제공절차에 따라야 함

문20 **유전정보도 다른 건강정보와 동일하게 취급해야 하는지 아니면 별도의 강화된 조치가 필요한지?**

답 '유전정보'는 「디엔에이신원확인 정보의 이용 및 보호에 관한 법률」, 「생명윤리 및 안전에 관한 법률」에 따라 엄격히 보호되고 있는 정보이므로 해당 법률에서 정하는 별도의 강화된 조치가 필요함

문21 **고객 행태 분석을 위해 서비스 이용 기록*이나 SNS 등에 공개된 정보를 수집하여 비식별 조치 후 이용하는 것이 가능한가?**

*인터넷 접속정보, 웹사이트 방문정보, 사용하는 단말기 정보 등

답 합법적으로 수집한 정보라면 비식별 조치 후 이용하는 것은 가능하며, 이 경우 정보주체의 동의를 받지 않아도 됨

문22 해당 데이터를 가이드라인에 따라 적정성 평가를 받고 활용하다 새로운 분석을 위해 비식별 조치 방법을 변경하고자 할 때 이 경우 별도의 적정성 평가를 진행해야 하는지?

답 본 가이드라인에 따른 적정성 평가는 데이터 마스킹, 총계처리 등 비식별 기법이 적용된 정보를 대상으로 하여 비식별 조치가 적정하게 이루어졌는지 여부를 평가하는 것임
따라서, 당초 적용된 비식별 기법을 변경하여 다른 기법을 적용하는 경우에는 기존 평가 대상의 변경이 수반되므로 추가적인 적정성 평가가 진행되어야 함

사후관리

문23 비식별 정보가 재식별되면 어떻게 해야 하나?

답 사업자 등은 비식별 정보를 활용하는 과정에서 정기적인 모니터링과 필수적인 안전조치 등을 통해 재식별 위험을 최소화해야 함
다만, 비식별 정보의 처리 과정에서 비의도적으로 특정 개인을 재식별하게 된 경우에는 즉시 그 정보의 처리를 중단하고 파기조치를 하여야 함

문24 비식별 정보도 재식별 가능성이 있다고 하는데 비식별 조치가 제대로 안된 것이 아닌지?

답 재식별 가능성이 현저하다면 이는 비식별 조치가 제대로 이루어지지 않은 것임

※ 비식별 조치가 충분히 이루어졌다면 그 시점에서 재식별은 불가능

다만, 비식별 조치가 적정하게 된 경우에도 새로운 결합기술이 출현하고 입수가능한 정보가 증가하는 경우에는 사후에 재식별이 될 수 있음
따라서, 비식별 조치가 적정하게 된 경우에도 재식별 방지를 위해 필수적인 안전조치는 이행하여야 함

문25 제공한 비식별 정보의 모니터링 책임이 제공자에게 있는 것인지?

답 비식별 정보를 이용하거나 제3자에게 제공하려는 사업자 등은 해당 정보의 재식별 가능성을 정기적으로 모니터링 해야 함

이미 제공된 비식별 정보의 모니터링 책임은 과거에 그 정보를 제공한 자가 아니라 현재 그 정보를 이용하는 사업자 등에게 있음

문26 '적절한 수준으로 비식별 조치'된 정보는 개인정보 보관기한 등과 무관하게 저장하여 활용할 수 있는지?

답 비식별 정보는 개인정보가 아닌 것으로 추정되므로 보관 및 이용 목적 · 기간을 뚜렷하게 한 후 해당 목적 및 기간종료 시까지 저장하여 활용할 수 있음

다만, 비록 비식별 정보가 특정 개인을 알아볼 수는 없더라도, 재식별 의도가 있는 제3자가 부정한 목적으로 활용하지 않도록 필수적인 안전조치는 이행하여야 함

문27 비식별 조치를 한 정보에 대한 열람, 정정 · 삭제 및 처리정지 등의 요구에 어떻게 대응해야 하는지?

답 비식별 조치된 정보는 요구자에 대한 정보를 확인할 수 없으므로 개인정보 열람, 정정 · 삭제 및 처리정지 등이 현실적으로 불가능

문28 **개인이 재식별 된 경우, 개인정보 보호법 제34조의 '유출'로 보아 해당 정보주체에 대한 유출통지를 해야 하는지?**

답 '개인정보 유출'은 '법령이나 개인정보처리자의 자유로운 의사에 의하지 않고 개인정보처리자가 통제를 상실하거나 또는 권한 없는 자의 접근을 허용한 경우'를 의미하며, 정보처리 과정에서 우연히 개인이 재식별 되었다는 사실만으로는 개인정보 유출로 볼 수 없으므로 유출통지 대상이 아님

※ '개인정보처리자의 통제 상실' 및 '제3자의 접근 허용'에 해당하지 않음

다만, 재식별된 정보를 파기하지 않고 보관하다가 해커 등 권한 없는 제3자에게 그 정보가 노출되었다면 '개인정보 유출'에 해당하므로 지체없이(5일 이내, 정보통신망법 적용 사업자는 24시간 이내) 유출통지를 해야 할 것임

지원 및 관리체계

문29 **다른 사업자가 보유한 DB를 결합하여 빅데이터 분석 등에 활용할 수 있는가?**

답 정보주체의 동의없이 당사자간 개인정보를 직접 주고받는 것은 현행법 상 허용되지 않음
다만, DB 결합 과정에서만 임시로 매칭키 역할을 하는 '임시 대체키'를 부여하고 비식별 조치한 후 신뢰할 수 있는 전문기관이 결합하는 것은 가능

문30 **DB 결합을 위해 분야별 전문기관에 데이터를 제공하는 경우 식별자만을 제거하고 제3의 기관에서 제공한 알고리즘으로 임시대체키를 생성하여 붙인 뒤 다른 비식별 조치를 하지 않을 수 있는지?**

답 직접적인 식별자만 제거하고 속성자에 대한 비식별 조치 없이 분야별 전문기관에 데이터를 제공하는 행위는 본 가이드라인에 따른 충분한 비식별 조치가 아님
분야별 전문기관에 데이터 결합을 위해 제공하는 정보는 본 가이드라인에서 정하는 비식별 조치 및 '비식별 조치 적정성 평가단'의 평가를 거쳐 비식별 조치를 적정하게 한 후에 제공해야 함

문31 업종별의 경우 전문기관이 서로 다른데 이업종의 DB결합시 각자 자신이 속한 업종의 전문기관의 지원을 받으면 되는 것인지?

답 DB 결합을 지원할 전문기관을 선택하고자 하는 경우 ① 산업내 기업간 결합은 해당 분야 전문기관에서 결합을 지원하고, ② 이종산업간 결합은 대량의 정보집합물을 결합하고자 하는 기업이 속해 있는 분야별 전문기관에서 수행

※ 분야별 전문기관은 한국인터넷진흥원, 한국신용정보원, 금융보안원, 사회보장정보원, 한국정보화진흥원 중에서 소관부처가 공문으로 지정 · 공표하여 운영하고 필요시 추가 지정 가능

당해 산업을 지원해 주는 전문기관이 없는 경우에는 한국인터넷진흥원 또는 한국정보화진흥원에서 지원을 받으면 됨

문32 이종 DB결합 시 주민등록번호를 사용하여 임시 대체키를 만드는 것이 현행법 위반인지?

답 개인정보 보호법 제24조의2에 따라 주민등록번호는 법령에서 구체적으로 요구하거나 허용하는 경우 등을 제외하고는 처리가 엄격히 제한됨

따라서, 임시 대체키 생성 시 주민등록번호를 사용하는 것은 현행 법령에서 구체적으로 요구하거나 허용하는 경우 등으로 볼 수 없으므로 현행법 위반의 소지가 있음

문33 기업 내에서 서로 다른 부서간의 DB를 결합하여 이용하고자 하는 경우에도 반드시 외부의 전문기관을 통해야 하는지?

답 기업 내에서 서로 다른 부서간 DB를 결합하여 이용하고자 하는 경우에는 반드시 외부의 전문기관을 통할 필요는 없음

다만, 기업내 DB결합 시에도 결합 전 · 후 본 가이드라인에 따른 비식별 조치 및 적정성 평가를 수행하여야 하며, 결합과정에서 임시 대체키를 활용할 경우에는 결합대상 정보를 관리하지 않는 제3의 부서가 임시 대체키를 안전하게 생성 · 관리하고 재식별 시도 금지 및 재식별시 즉시 파기 등 필수 보호조치를 엄격히 해야 함

이 경우 평가단은 동일한 평가단에서 평가를 수행할 수 있음

문34 개인정보를 비식별 조치하여 활용할 경우 법적인 책임은 없는지?

답 본 가이드라인에 따라 특정 개인을 알아볼 수 없도록 비식별 조치하여 활용하는 경우에는 개인정보 보호법 상 목적 외 이용 · 제공 등에 관련한 책임은 없음. 다만, 적정한 비식별 조치 없이 활용하는 경우에는 책임이 있음

또한, 다른 정보와 결합하여 재식별 되지 않도록 필수적인 관리조치는 이행해야함. 즉, 비식별 정보에 대한 관리적 · 기술적 보호조치*와 재식별이 되는 경우 정보처리를 즉시 중단하고 파기 조치를 하는 등 가이드라인에서 정한 조치사항을 이행하지 않아 비식별 정보가 재식별이 되면 그에 따른 법적인 책임이 있음

* 비식별 정보파일에 대한 접근권한 관리 및 접근 통제, 비식별 정보파일 유출 시 대응계획 수립 등

문35 가이드라인에서 정하는 대로 비식별 조치*를 실시하였다고 가정할 때 의도하지 않은 재식별 발생 시 면책이 가능한지?

* 비식별 기법 적용 후 전문가 평가결과(k-익명성 등) '적정'인 경우

답 본 가이드라인에 따라 특정 개인을 알아볼 수 없도록 충분한 비식별 조치를 실시하였다면 고의성 없는 재식별 사실만으로 책임을 부과할 수는 없음

다만, 의도적으로 비식별 정보가 쉽게 재식별 될 수 있도록 이용 · 제공하거나, 재식별 정보를 보호조치 없이 보관 · 이용 · 제공한 경우에는 관련 법령에 따른 벌칙*이나 과태료**가 부과될 수 있음

* 재식별 정보 이용 · 제공시 : 5년 이하 징역 또는 5천만원 이하 벌금(정보통신망법 적용 사업자는 위반행위 관련 매출액의 3% 이하 과징금이 추가 부과될 수 있음)

** 파기 조치 없이 보관 : 5천만원 이하 과태료(정보통신망법 적용 사업자는 5년 이하 징역 또는 5천만원 이하 벌금형 및 위반행위 관련 매출액의 3% 이하 과징금이 추가 부과될 수 있음)

본 가이드라인은 국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부 등 관계부처가 합동으로 작성하였습니다.

각 부처가 개인정보 비식별 조치와 관련하여 기존에 발간한 지침, 안내서, 가이드라인 등은 2016. 6. 30일 부로 일괄 폐지되고 2016. 7. 1일 부터는 본 가이드라인이 적용됨을 알려드립니다.

개인정보 비식별 조치 가이드라인
-비식별 조치 기준 및 지원 · 관리체계 안내-

2016년 6월 28일 인쇄
2016년 6월 30일 발행

발행처 | 국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부
제 작 | 호정씨앤피