

# 개인정보 비식별화에 대한 적정성 자율평가 안내서



한국정보화진흥원은

원활한 개인정보 비식별화 조치 지원을 위하여

개인정보보호지원센터를 운영하고 있습니다.

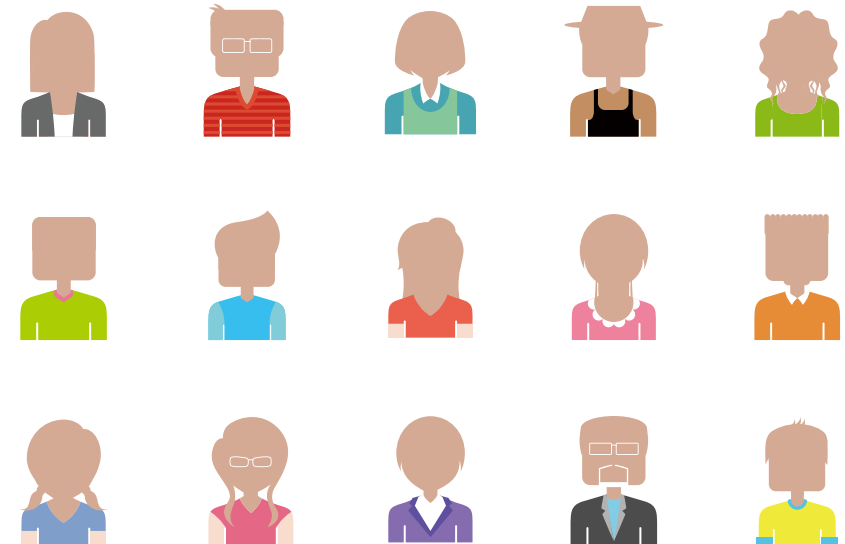
## 개인정보보호지원센터

법령 해석 지원 / 비식별화 기술 컨설팅 /

지침 운영 및 모니터링 지원 / 교육 콘텐츠 개발 보급

Tel. 02-2131-0111~2

<http://privacy.nia.or.kr>



행정자치부



한국정보화진흥원



행정자치부



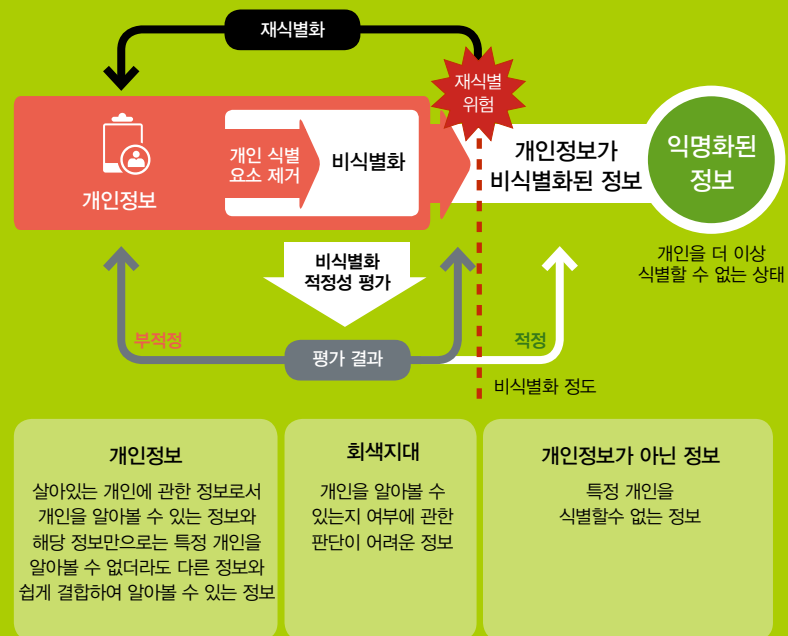
한국정보화진흥원



개인정보 비식별화에 대한  
**적정성 자율평가 안내서**

**요약문 :** 본 안내서는 공공 및 민간의 개인정보처리자가 개인정보를 포함하고 있는 데이터에서의 개인식별요소 제거 및 개인정보 비식별화에 대한 적정성 평가, 재식별 위험 관리 등 일련의 개인정보보호 조치에 대해 안내하기 위해 개발함

- ◎ 정보기술의 발전, 페이스북 등 SNS 서비스의 증가, 공공정보 개방·공유 정책 시행 확대 등에 따라 개인을 식별 할 수 있는 위험 증가
- ◎ 정부는 데이터 분석·활용에 따른 개인정보 침해 위험을 최소화하고 관련 데이터의 안전한 이용을 위해 개인식별요소 제거 방법 제시
  - ※ 공공정보 개방·공유에 따른 개인정보보호 지침, 2013.9, 행정자치부

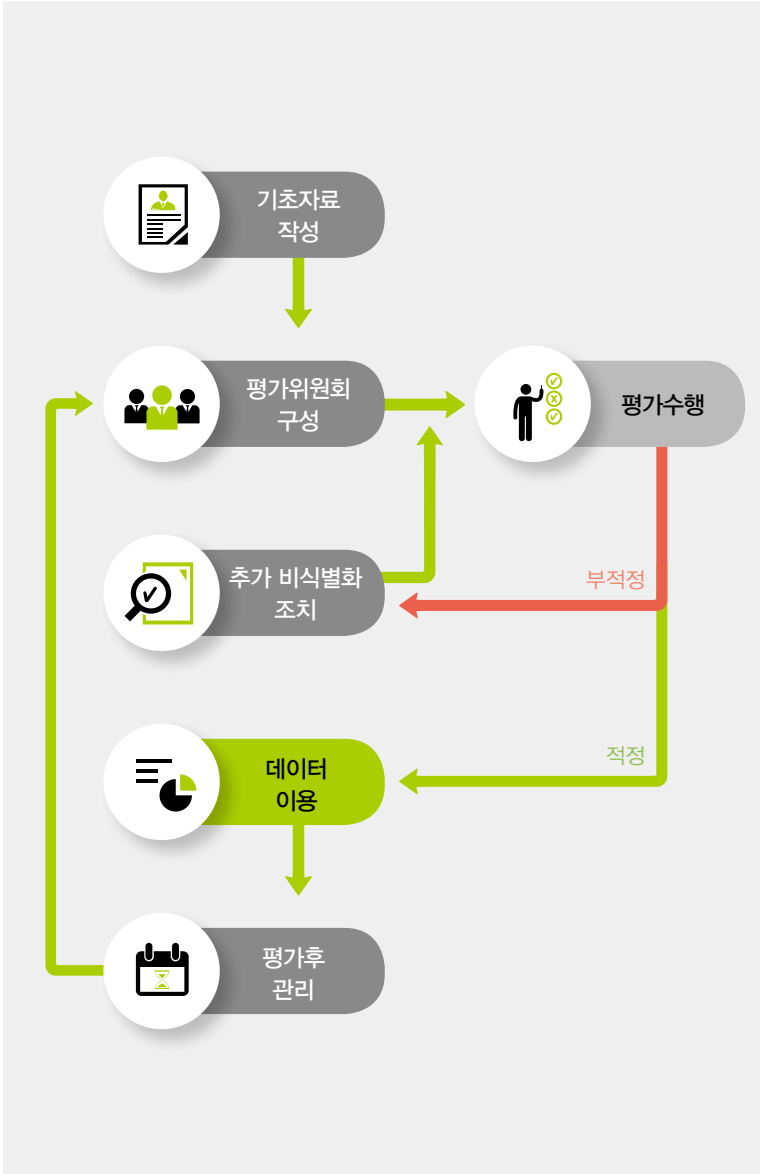


- ◎ 그러나 개인식별요소를 제거한 데이터라도 재식별 위험을 내포하고 있으며, 비식별화가 적절하게 수행 되었는지에 대해 검토할 필요성이 있음
  - ※ 개인식별요소를 제거한 데이터가 공개된 이후 특정 개인이 재식별된 사례 다수 존재 (AOL, Netflix, 미국 매사추세츠 주 사례 등, p.21 참고)

- ◎ 개인식별 요소를 제거한 데이터가 비식별화된 데이터로서 현 상태로 이용·제공 가능한지, 추가 조치가 필요한지를 판단할 수 있는 방안 제시



03  
평가 절차



◎ 기초자료 작성  
개인정보 비식별화 적정성 평가 수행에 필요한 기초자료 작성

구분	기초자료	구분
데이터	데이터 원본 예시 및 세부 항목별 명세	필수
	개인식별요소가 제거된 평가 대상 데이터 및 세부 항목별 명세	필수
개인식별요소 제거 현황	개인식별요소 제거 기법 적용 기준	필수
	평가 대상 데이터에 대한 $k$ -익명성, $t$ -다양성, $t$ -근접성 값 산출 결과	선택
이용기관의 관리수준	데이터 이용 기관의 데이터 이용 상세 목적, 이용 방법, 이용 기간, 데이터 접근 가능자 현황 등 데이터 활용에 관한 사항	필수
	데이터를 제공하는 경우 데이터를 제공받는 방법 및 데이터를 보호하기 위해 취하는 일련의 조치에 대한 현황	필수
	데이터 이용 및 제공과 관련이 있는 계약서 또는 협약서 사본	필수
	데이터 이용 기관에서 보유하거나 보유할 수 있는 개인정보 관련 데이터(세부 내용 및 항목 등)에 관한 사항	선택
	데이터 이용기관의 PIPL, PIMS 등 개인정보보호 인증 사본	선택

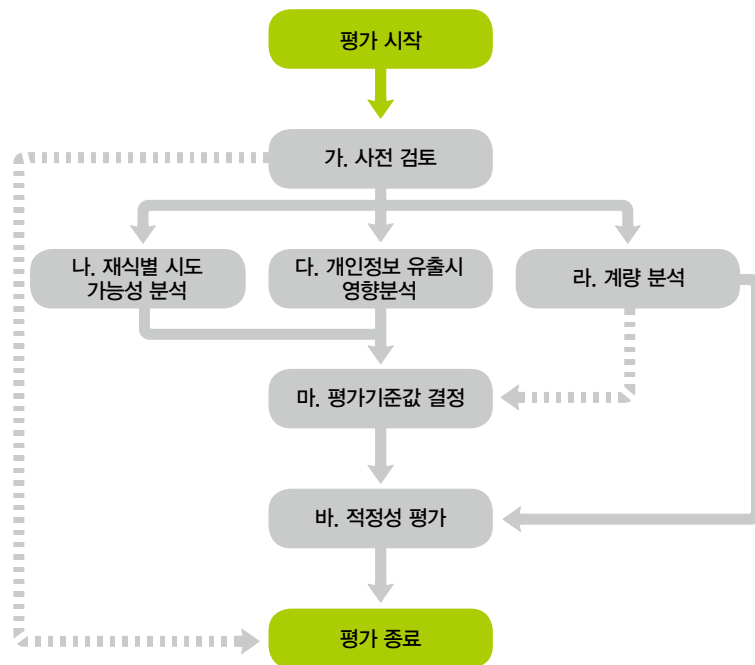
- ◎ 평가위원회 구성  
개인정보처리자의 개인정보보호책임자(CPO)가 지정하는 3인 이상의 평가위원으로 구성  
※ 평가위원의 과반수 이상은 외부의 전문가로 지정하고, 업무영역 전문가 1인, 개인정보 비식별화 전문가 1인, 법률전문가 1인은 필수적으로 포함하여 구성
- ◎ 평가 수행  
개인정보처리자에서 작성한 '기초자료'와 '세부 평가 방법'을 기반으로 평가를 수행하고 '적정' 또는 '부적정' 의견 제시  
※ 적정 : 다른 정보와 결합해도 특정 개인을 식별할 수 없는 상태를 의미  
부적정 : 추가적인 개인식별요소 제거 조치가 필요한 상태를 의미
- ◎ 추가 비식별화 조치  
평가위원회의 의견에 기초하여 평가 대상 데이터에 대해 개인식별요소 제거 조치 등을 추가적으로 수행  
※ 추가 비식별화 조치 이후 비식별 조치가 적정하게 수행되었는지 평가 수행

## 04

세부  
평가 방법

## ◎ 평가후 관리

평가위원회 또는 개인정보처리자에서 인정한 경우 일정 기간이 경과한 이후 개인정보 비식별화에 대한 적정성을 재평가



## ◎ 사전검토

개인정보처리자가 제출한 기초자료와 인터뷰 등을 통해 평가 대상 데이터가 개인식별요소를 포함하고 있는지, 개인식별요소 제거 기법·기준에 따라 관련 조치가 수행되었는지 등 검토

## ◎ 재식별 시도 가능성

데이터를 이용하거나 제공받는 자의 개인정보 재식별 의도와 능력, 개인정보 보호 수준 등을 통해 재식별 시도 가능성을 '빈번한, 가능한, 가끔, 거의 없는' 등 4단계로 분석

## ◎ 개인정보 유출시 영향

데이터가 의도적 또는 비의도적으로 재식별된 경우 정보주체 및 관련자에게 유형·무형의 개인적 피해, 사회적 손해 등의 피해를 줄 수 있는 가능성에 대해 '높음, 중간, 낮음' 등 3단계로 분석

## ◎ 계량 분석

평가위원회에서 선정한 프라이버시 보호 모델( $k$ -익명성,  $l$ -다양성,  $t$ -근접성 등)을 이용하여 계량 분석 실시

## ◎ 평가기준값 결정

평가위원회에서 '재식별 시도 가능성'과 '개인정보 유출시 영향' 분석 결과, 데이터 특성 등을 종합적으로 고려해서 적정성 평가에 필요한 평가 기준 값의 종류 및 기준 결정

개인정보 유출시 영향	〈평가 기준 값 (예시)〉				
침해위험 높음	$k=5, l=2$	$k=10, l=3$	$k=10, l=4$	$k=20, l=5$	
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

※ 평가 기준 값은 프라이버시 보호 모델( $k$ -익명성,  $l$ -다양성,  $t$ -근접성 값 등)의 값으로 결정

## ◎ 적정성 평가

'평가 기준 값'과 '계량분석'에서 산출된 분석 값을 비교하여 개인식별요소 제거 수준에 대한 최종적인 적정성을 '적정' 또는 '부적정'으로 판단

# contents

02

요약문

08

목차

## I. 개요

12

1. 배경 및 필요성

13

2. 용어정의

## II. 개인정보 재식별 및 위협 요소

17

1. 개인정보 재식별 요소

17

가. 재식별 주체

18

나. 내부 데이터

18

다. 외부 데이터

19

2. 개인정보 재식별 위협 요소

19

가. 데이터 공개의 증가

19

나. 맞춤형 광고 데이터의 증가와 데이터 집종의 심화

20

다. 사물통신 환경과 비식별 정보의 증가

21

3. 개인정보 재식별 사례

21

가. 매사추세츠 주 사례

22

나. 아메리카 온라인 사례

23

다. 넷플릭스 사례

23

라. SNS에 노출된 개인정보를 이용한 개인 재식별

## III. 개인정보 비식별화에 대한 적정성 평가

24

1. 적정성 평가 개요

24

가. 평가 배경

24

나. 평가 목적

25

다. 평가 대상

25

라. 평가 추진 근거

26

마. 평가 시기

26

바. 안내서 활용을 위한 참고사항

28

2. 평가 절차

28

가. 기초자료 작성

29

나. 평가위원회 구성

30

다. 평가 수행

31

라. 추가 비식별 조치

32

마. 평가 후 관리

33

3. 세부 평가 방법

33

가. 사전 검토

34

나. 재식별 시도 가능성 분석

38

다. 개인정보 유출시 영향 분석

40

라. 계량 분석

40

마. 평가 기준 값 결정

42

바. 적정성 평가

IV. 재식별 위험 관리 방안

45

1. 재식별 위험에 대한 관리적 조치

45

가. 데이터 제공 및 위탁 계약시 재식별 금지 관련 조항 반영

46

나. 데이터 공개시 재식별 금지 관련 조항 게시

47

다. 재식별 가능성 모니터링

48

라. 개인정보 재식별시의 대응 매뉴얼 마련 및 시행

48

2. 재식별시 대응 조치

48

가. 데이터 공개 중단 및 회수

48

나. 데이터 제공 및 처리 위탁 중단

49

다. 데이터 재식별 위험 통지

49

라. 개인정보 유출통지 및 유출신고

49

마. 추가적 비식별화 조치

부 록

53

1. 비식별화 조치 방법

57

2. 프라이버시 보호 모델

73

3. 참고자료

75

4. 참여 전문가

표 목차

15

〈표 1〉 식별자 예

16

〈표 2〉 준식별자 예

29

〈표 3〉 개인정보 비식별화에 대한 적정성 평가를 위한 기초자료

35

〈표 4〉 재식별 의도 및 능력 분석 평가 지표

36

〈표 5〉 재식별 의도 및 능력 분석 평가 기준

37

〈표 6〉 정보보호 능력 분석 평가 지표

37

〈표 7〉 정보보호 능력 분석 평가 기준

39

〈표 8〉 개인정보 유출시 영향 분석 평가 지표

40

〈표 9〉 개인정보 유출시 영향 분석 평가 기준

42

〈표 10〉  $k$ -익명성 기반 적정성 평가 사례

43

〈표 11〉  $l$ -다양성 기반 적정성 평가 사례

44

〈표 12〉  $t$ -근접성 기반 적정성 평가 사례

47

〈표 13〉 개인정보 비식별화에 대한 적정성 재평가 검토 질문서

50

〈표 14〉 개인정보 유출시 통지 방법

51

〈표 15〉 개인정보 유출시 신고 방법

그림 목차

14

〈그림 1〉 개인정보 비식별 및 재식별 개념

21

〈그림 2〉 의료데이터와 투표자 명부를 이용한 개인정보 재식별

22

〈그림 3〉 뉴욕타임즈의 재식별 관련 기사

25

〈그림 4〉 개인정보 비식별화에 대한 적정성 평가 개념

28

〈그림 5〉 개인정보 비식별화에 대한 적정성 평가 절차

33

〈그림 6〉 개인정보 비식별 적정성 세부 평가 방법

38

〈그림 7〉 재식별 시도 가능성 분석표

41

〈그림 8〉 평가 기준 값 사례

## I. 개요

본 안내서는 공공 및 민간의 개인정보처리자가 개인정보를 포함하고 있는 데이터에서의 개인식별요소 제거 및 개인정보 비식별화에 대한 적정성 평가, 재식별 위험 관리 등 일련의 개인정보보호조치에 대해 안내하기 위해 개발함

- ◎ 정부3.0 추진에 따른 공공정보 개방 · 공유 정책 시행과 빅데이터 기술의 발전으로 사회 각 분야에서 데이터의 분석 및 활용이 증가
- ◎ 행정자치부는 데이터 분석 · 활용에 따른 개인정보 침해 위험을 최소화하고 관련 데이터의 안전한 이용을 활성화하기 위해
  - ‘공공정보 개방 · 공유에 따른 개인정보보호 지침(2013.9)’을 통해 개인식별요소를 제거하는 비식별화 조치 방법을 제시
- ◎ 그러나 개인식별요소를 제거한 데이터는 데이터의 품질 및 활용 가치가 낮아질 수 있기 때문에, 관련 데이터 처리자는 최소한의 비식별화 조치를 적용할 가능성이 높으며,
- ◎ 데이터 처리속도, 저장기술 등의 기술발전과 페이스북, 트위터 등 사회관계망 서비스 이용자의 증가 및 공개되는 정보의 증가에 따라, 향후 개인을 재식별할 수 있는 위험도 높아지고 있음
- ◎ 개인정보 침해위험을 최소화 하면서 데이터의 안전한 이용을 활성화하기 위해서는, 다음과 같은 사항에 대해 개인정보처리자 스스로 사전 검토 및 평가를 수행할 필요가 있음
  - ① 개인식별요소 제거 기법과 기준은 향후의 재식별 위험을 고려하여 적절한 수준으로 선택 및 계획 되었는가?

- ② 개인식별요소 제거 조치가 최초로 계획한 기법과 기준에 따라 정확히 수행되었는가?
- ③ 개인식별요소가 제거된 데이터에 대한 이용, 제공 및 사후관리 체계는 마련되어 있는가?

## 02

### 용어정의

- ◎ 개인정보 : 살아있는 개인에 관한 정보<sup>1)</sup>로서 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 포함(개인정보보호법 제2조)
- ◎ 식별(identification) : 개인을 분별하여 알아보는 것
- ◎ 식별자(identifier) : 성명(유일한 경우), 주민등록번호, 이메일 주소, 전화번호 등과 같이 그 자체로 특정 개인을 직접 식별할 수 있는 데이터<sup>2)</sup>
- ◎ 준식별자(quasi-identifier) : 연령, 성별, 거주 지역, 국적 등과 같이 해당 데이터만으로는 직접적으로 특정 개인을 식별할 수는 없지만, 다른 정보와 결합하여 개인을 식별할 수 있는 데이터
- ◎ 비식별화(de-identification) : 정보의 일부 또는 전부를 삭제 · 대체 하거나 다른 정보와 쉽게 결합하지 못하도록 하여<sup>3)</sup> 특정 개인을 알아볼 수 없도록 하는 일련의 조치<sup>4)</sup>

1) 정보(information) : 관찰이나 측정을 통하여 수집한 자료를 실제 문제에 도움이 될 수 있도록 정리한 지식 또는 그 자료

2) 데이터(data) : 컴퓨터가 처리할 수 있는 문자, 숫자, 소리, 그림 따위의 형태로 된 정보

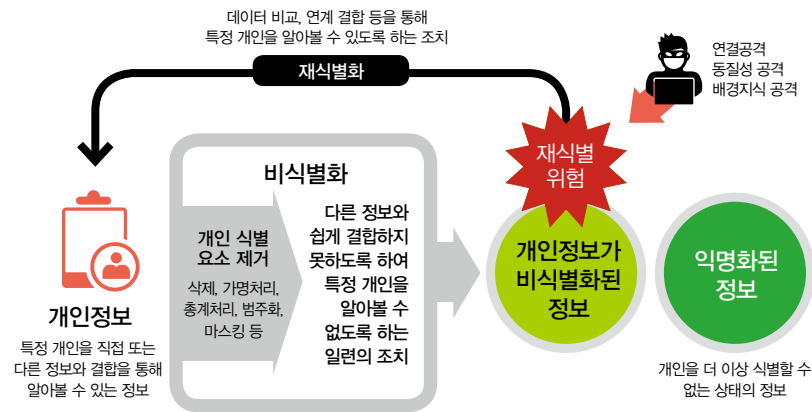
3) 개인식별요소 제거 방법(공공정보 개방 공유에 따른 개인정보 보호 지침, 행정자치부, 2013.9.)

- 가명처리 : 개인정보 중 주요 식별 요소를 다른 값으로 대체하여 개인 식별을 곤란하게 하는 것
- 총계처리 : 데이터의 총합 값을 보임으로서 개별 데이터의 값을 보여주지 않도록 하는 것
- 데이터 삭제 : 데이터 개방 · 공유 목적에 따라 데이터 셋에 구성된 값 중에 필요없는 값 또는 개인 식별에 중요한 값을 삭제하는 것
- 범주화 : 데이터의 값을 범주의 값으로 변환하여 명확한 값을 감추는 것
- 데이터 마스킹 : 공개된 정보 등과 결합하여 개인을 식별하는데 기여할 확률이 높은 주요 개인 식별자가 보이지 않도록 처리하여 개인을 식별하지 못하도록 하는 것



- ◎ 익명화(Anonymization) : 비식별화 조치의 궁극적인 상태로, 개인에 대한 재식별이 더 이상 불가능한 상태
- ◎ 재식별화(re-identification) : 비식별화한 개인정보를 다른 정보 또는 데이터와 비교, 연계, 결합 등을 통해 특정 개인을 알아볼 수 있도록 하는 일련의 조치

〈그림 1〉  
개인정보 비식별 및  
재식별 개념



〈표 1〉  
식별자 예 <sup>5)</sup>

- 성명(한자, 영문 성명 포함)
- 주소
- 고유식별정보(주민등록번호, 여권번호, 외국인등록번호, 운전면허번호)
- 연월일 : 생일(양/음력), 기념일(결혼, 돌, 환갑 등), 사망일, 자격증 취득일 등
- 전화번호(휴대폰번호, 집전화, 회사전화)
- 팩스번호
- 전자메일
- 의료기록번호
- 건강보험번호, 복지수급자번호
- 계좌번호, 카드번호
- 각종 자격증 및 면허번호
- 자동차번호
- 각종 기기의 등록번호 & 일련번호
- IP 주소, Mac 주소
- 홈페이지 URL
- 사진(정지, 동영상, 유사 이미지, CCTV 영상 등)
- 신체 식별 정보(지문, 음성, 홍채 등)
- 기타 유일 식별번호 : 군번, 사업자등록번호 특성(별명), 식별코드(아이디, 아이핀 값(cn, dn)) 등

4) ‘비식별화(De-identification)’라는 용어는 개인정보 식별요소를 제거하는 조치의 의미가 강하며 주로 북미 지역에서 사용되고 있으며, ‘익명화(Anonymization)’ 용어는 개인을 더 이상 식별할 수 없는 상태, 즉 개인에 대한 재식별이 불가능한 상태를 의미하며 학계와 EU에서 주로 사용하고 있음. 비식별화(De-identification) 및 익명화(Anonymization)라는 용어가 궁극적으로 추구하는 바는 개인을 식별할 수 없는 상태나 식별할 수 없도록 하는 것을 의미한다고 할 수 있음

- “HIPAA De-Identification Guidance”(USA, 2012.12.)
- “‘Best Practice’ Guidelines for Managing the Disclosure of De-Identification Health Information”(Canada, 2010.10.)
- “Opinion 05/2014 on Anonymization Techniques”(EU, 2014.5.)
- “Anonymisation : managing data protection risk code of practice”(UK, 2012.11.)

5) 미국의 Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule(OCR, 2012.11)에서 제안하고 있는 식별 요소의 내용을 고려하여 작성하였으며, 상황에 따라 식별자로 기능하지 않을 수 있음

〈표 2〉  
준식별자 예 <sup>6)</sup>

개인 특성	- 성별, 생년, 생일, 연령(나이), 국적, 고향 · 거주지 시군구명, 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, 상병코드, 투약코드, 진료내역 등
신용 특성	- 세금 납부액, 신용등급, 기부금 등 - 건강보험료 납부액, 소득분위, 의료급여자 등
경력 특성	- 학교명, 학과명, 학년, 성적, 학력 등 - 직업, 직종, (전·현)직장명, 부서명, 직급 - 자격증명, 경력 등
전자적 특성	- PC사양, 비밀번호, 비밀번호 질문/답변, 쿠키정보, 접속일시, 방문일시, 서비스 이용 기록, 위치정보, 접속로그 등 - IP 주소, MAC 주소, HDD serial 번호, CPU ID, 원격 접속 여부, Proxy설정 여부, VPN 설정 여부, USB serial 번호, Main Border serial 번호, UUID, OS 버전 - 사용 OS 버전, 기기 제조사, 모델명, 단말기 ID, 네트워크 국가 코드, SIM Card 정보 등
가족 특성	- 배우자, 자녀, 부모, 형제 여부 등 - 가족정보, 법정대리인 정보 등
위치 특성	- GPS 데이터, RFID 리더 접속기록, 특정 시점 센싱 기록 - 인터넷 접속, 핸드폰 사용 기록, 사진 등

6) 구체적 상황에 따라 ‘식별자’로도 기능할 수 있음

## II. 개인정보 재식별 및 위협 요소

### 01 개인정보 재식별 요소

개인정보가 비식별화된 데이터로부터 개인정보를 재식별 하기 위해서는 재식별 주체, 내부 및 외부 데이터 등의 요소가 필요함

#### 가. 재식별 주체

- ◎ 개인정보 재식별이 발생하는 원인의 하나로 범죄자, 스토키, 해커 등과 같이 개인정보처리자 외부의 제3자가 재식별 주체가 될 수 있음
  - 트위터 계정의 실제 이용자를 찾아내 해당 이용자의 행동양식을 알아내고, 이메일, 전화번호를 획득해 스팸문자를 보낸다든지, 특정 행동양식을 파악하려는 재식별 주체는 언제든 나타날 수 있음
  - 특히 스토키와 같은 부류의 악성 재식별 주체와 재식별 기술이 결합한다면 특정 개인의 프라이버시를 심각하게 침해할 수 있음
- ◎ 기업 등 개인정보처리자 자체의 관점에서 소비자의 다양한 행동 양식은 제품 판매와 마케팅 전략에 있어 매우 중요한 요소임
  - 기업은 보다 명확하고 정확하게 고객을 식별함으로써 보다 정밀한 경영 및 마케팅 전략을 수립하는 것이 가능하므로, 재식별 의지가 높은 주체중 하나로 꼽을 수 있음

- ◎ 해커·기업 등 이외에도 개인정보 재식별에 대해 관심이 높은 재식별 주체로는 정부주체와 작·간접적으로 연관이 있는 처리자 소속 직원 등이 있을 수 있으며,
  - 개인정보에 대한 재식별 행위도 언제, 어느 곳, 어떠한 데이터를 통해서도 발생할 수 있는 가능성이 있음

### 나. 내부 데이터

- ◎ 개인정보처리자 내부의 다양한 부서에서 다양한 목적과 다양한 항목의 개인정보를 수집 및 관리할 수 있음
- ◎ 조직내 정보의 결합 및 분석을 통한 기업 활동 활성화는 빅데이터 분석의 중요성이 강조되면서 점점 더 장려되는 상황임
- ◎ 조직내 부서간 보유하고 있는 정보의 결합은 개인정보 재식별을 가능하게 하는 요인으로 고려될 수 있음
- ◎ 조직 내부의 정보결합은 두 개의 데이터베이스 테이블에서 공유된 정보를 기반으로 하나의 정보 열과 또 다른 정보 열이 결합하는 것으로, 조직 내부 데이터 간 결합으로 개인을 재식별할 위험이 있는 2차 데이터를 생성이 가능함

### 다. 외부 데이터

- ◎ 재식별 주체는 개인정보가 비식별화된 데이터에서 발견한 재식별 가능 요소를 이용해서, 외부에서 수집 가능하거나 인터넷 등에 공개되어 있는 데이터와 비교, 연계, 결합 등을 통해 개인을 재식별할 수 있음
- ◎ 블로그, SNS 이용 등의 증거로 조성된 풍부한 외부 정보와 쉽고 빠른 검색 환경은 외부 데이터를 이용한 재식별 가능성을 크게 높임
  - ※ 2009년 시작된 페이스북의 프로필 작성 서비스는 자발적 개인정보 공개의 대표 사례
- ◎ 특히 악의적이고 목적의식이 뚜렷한 경우, 다양한 데이터를 결합해 자신들의 목적에 적합한 개인을 식별하고자 할 가능성이 높음
- ◎ 최근 정부의 공공정보 개방·공유 정책 등에 따라 공개적으로 수집할 수 있는 데이터의 증가로, 재식별 의도가 있는 이용자들은 이미 공개된 정보조각들을 보유하고 있을 가능성이 높음

## 02

### 개인정보 재식별 위협 요소

- ◎ 개인정보가 비식별화된 데이터의 재식별 가능성과 그 위험에 대비하기 위한 개인 식별요소 제거 기술, 프라이버시 보호 모델 등에 대한 연구는 지속적으로 이루어지고 있지만 개인정보 재식별의 위험을 완전히 제거하는 것은 매우 어려운 작업임
- ◎ 개인정보가 비식별화된 데이터에 대한 재식별 위험을 증가시키는 위협 요소로는 데이터 공개 증가에 따른 연결 가능 정보의 증가, 맞춤형 광고를 위한 데이터 수집 증가 및 집중화, 사물인터넷 환경과 비식별 정보의 증가 등 다양한 원인이 있음

### 가. 데이터 공개의 증가

- ◎ 개인정보가 비식별화된 데이터의 재식별 위험이 높다는 주장의 가장 큰 배경에는 공개 데이터의 증가가 있음
- ◎ 두 종류의 개인정보가 비식별화된 데이터를 보유한 연구자나 악의적인 해커가 새로운 세 번째 데이터를 확보한다면 이전에 식별하지 못했던 비식별 정보로부터 개인을 식별해낼 가능성은 더욱 증가함
- ◎ 이러한 현상은 의료, 웹서비스 등과 같이 특정 분야의 데이터에서만 존재하는 것이 아니라, 서로 전혀 다른 분야에서 공개된 데이터간에도 발생할 수 있음
- ◎ 공개 데이터의 증가는 그 목적이 순수하든 악의적이든 재식별의 가능성을 높일 수 있으며, 이렇게 재식별된 데이터는 개인에 대한 심각한 프라이버시 침해로 이어질 수 있음

### 나. 맞춤형 광고 데이터의 증가와 데이터 집중의 심화

- ◎ 인터넷 산업계에서 광고 플랫폼 사업자를 통한 맞춤형 광고는 점점 더 중요한 비중을 차지하고 있을 뿐만 아니라, 다양한 업종에서 맞춤형 광고를 도입 및 활용하는 사례는 더욱 증가하고 있음
- ◎ 온라인상에서 맞춤형 광고를 위해 이용되는 개인정보가 비식별화된 데이터에는 개인의 행태, 성향, 위치 등의 개인정보가 있을 수 있음

- ◎ 광고사업자는 정보수집 애플리케이션을 통해 개인을 식별할 수 있는 식별자 뿐 아니라 IP 주소, 세션 정보, 검색 기록 등의 준식별자 관련 데이터를 무차별적으로 수집할 수 있음
- ◎ 맞춤형 광고를 제공하는 기업에서 더 많은 데이터를 보유하면 할수록, 이들이 보유하는 데이터는 개인을 식별할 수 있는 데이터 속성을 보유할 가능성이 더욱 높아질 것으로 예상됨
- ◎ 이렇게 집중된 데이터를 통해 개인이 재식별 될 가능성이 있으며, 이를 이용한 개인정보 침해가 발생할 가능성도 배제할 수 없음

#### 다. 사물통신 환경과 비식별 정보의 증가

- ◎ 최근 급속히 발전하고 있는 사물통신(IoT)<sup>7)</sup>, 모바일 환경 등에서의 기술은 개인에 관한 위치정보 등 새로운 형태의 데이터를 생성
- ◎ 성명, 주소, 휴대전화번호와 같은 전통적인 개인 식별정보 이외에도 센서, IP주소, 기기 등으로부터 수집 가능한 위치·생체 정보 등 광범위한 데이터는 개인을 식별할 수 있는 요소가 됨
- ◎ 실제로 더 많은 기기들이 IP 주소를 가지게 되는 초연결 사회에서는 IP주소를 통해 특정 기기는 물론 위치 등에 대한 식별이 이루어질 가능성이 커짐
- ◎ 사물통신과 결부된 정보와 위치정보는 개인의 행동양식과 생활패턴을 알려주는 민감한 비식별 정보로 기존의 식별정보 이상으로 개인을 재식별 할 수 있을 것으로 예상

7) 사물통신(IoT, Internet of Things) : 가전기기, 자동차, 생활용품 등의 모든 사물에 센서와 통신기능이 탑재되고, 이들 기기간 상호 통신을 통한 지능형 서비스를 제공 가능하게 하는 융합 기술

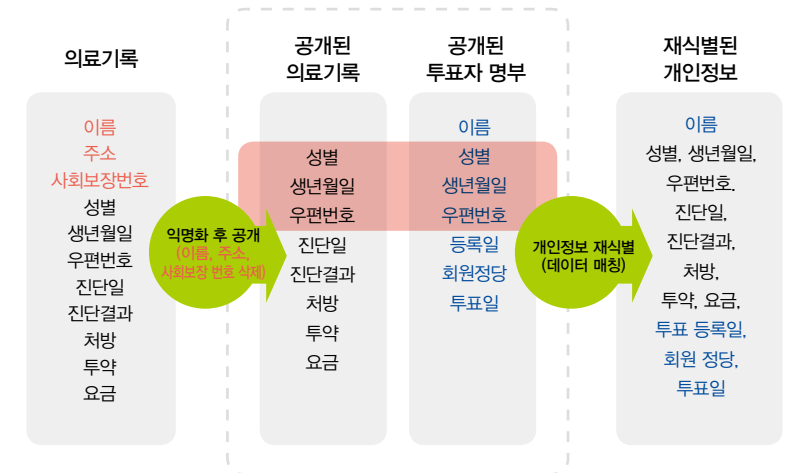
## 03

### 개인정보 재식별 사례

#### 가. 매사추세츠 주 사례 (1997, 미국)

- ◎ 미국 매사추세츠 주의 단체 보험위원회가 매사추세츠 주정부 소속 공무원의 의료 병원 출입 기록을 요약하여 공개
  - 공개 정보에서 이름, 주소, 사회보장번호와 그 외의 식별정보는 제거했으나, 환자 관련 100여개의 속성정보는 미삭제하여 공개함
  - ※ 준식별정보에 해당하는 우편번호 3자리, 생년월일, 성별 정보가 포함
- ◎ 또한 매사추세츠 주에서는 이름, 성별, 생년월일, 우편주소가 포함된 투표자 명부를 공개(판매)하고 있었음
- ◎ 일련의 연구자가 투표자 명부와 의료 데이터를 매칭해서 특정 개인에 대한 재식별을 시도하였으며, 그 결과로 주지사의 정보를 식별해냄
- ◎ 즉, 매사추세츠 주가 공개한 의료 데이터와 투표자 명부로부터 특정인(주지사)의 의료정보를 재식별 가능성이 밝혀짐

〈그림 2〉  
의료데이터와  
투표자 명부를 이용한  
개인정보 재식별



### 나. 아메리카 온라인 사례 (2006, 미국)

- 아메리카 온라인(AOL, America Online)은 학술연구를 위해 65만 명의 사용자가 3개월간 AOL의 검색엔진으로 검색한 이력 리스트 2,000만 건을 공개
- 사용자명과 IP 주소를 비식별화 하였지만, 사용자명은 유용성 확보를 위해 특정 번호의 식별자로 교환하여 공개
- 뉴욕 타임즈(New York Time)의 기자 2명이 검색 이력으로부터 이용자 '4417749'가 62세의 텔마 아놀드(Thelma Arnold)라는 미망인으로 조지아 주 릴번이라는 지역에 살고 있다는 사실을 밝힘
- 자료를 공개한 그 다음 주에 AOL은 데이터 공개를 중지하고 사과했으며, 관련 연구자와 그 상사를 해고하고, 최고기술책임자(CTO)가 사임함

### 다. 넷플릭스 사례 (2006, 미국)

- 온라인 영화 상영 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, 데이터에 대한 처리 내용을 연결하기 위해 독특한 식별자, 사용자에게 의한 영화의 평가, 평가한 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

〈그림 3〉  
뉴욕타임즈의  
재식별 관련 기사

The New York Times

August 8, 2009

## What Revealing Search Data Reveals

ACLS posted, but later removed, a list of the Web search inquiries of 636,000 unnamed users at a new Web site for academic researchers. An interview with one of those unnamed users, Theresa Arnold, confirmed that her data reveal what she was searching for, why and on which Web sites.

### A sample of Theresa Arnold's search data released by ACLS

0017700	swing sets	2009-05-24	10:39:30	4	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017701	swing sets	2009-05-24	10:39:30	4	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017702	swing sets	2009-05-24	10:39:30	10	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017703	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017704	swing sets	2009-05-24	10:39:30	4	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017705	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017706	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017707	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017708	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017709	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017710	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017711	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017712	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017713	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017714	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017715	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017716	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017717	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017718	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017719	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017720	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017721	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017722	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017723	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017724	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017725	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017726	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017727	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017728	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017729	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017730	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017731	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017732	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017733	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017734	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017735	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017736	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017737	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017738	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017739	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017740	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017741	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017742	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017743	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017744	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017745	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017746	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017747	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017748	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017749	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017750	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017751	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017752	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017753	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017754	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017755	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017756	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017757	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017758	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017759	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017760	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017761	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017762	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017763	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017764	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017765	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017766	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017767	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017768	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017769	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017770	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017771	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017772	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017773	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017774	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017775	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017776	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017777	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017778	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017779	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017780	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017781	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017782	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017783	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017784	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017785	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017786	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017787	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017788	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017789	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017790	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017791	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017792	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017793	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017794	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017795	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017796	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017797	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017798	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017799	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>
0017800	swing sets	2009-05-24	10:39:30	1	<a href="http://www.24magazine.com">http://www.24magazine.com</a>

### Why she search

"I was thinking about my grandchildren."

"I was looking for more."

"A woman was in the [public] bathroom crying. She was going through a divorce. I thought there was a place called 'Dances by Lori,' for singles."

"I wanted to find out what my husband was worth."

The New York Times

## 라. SNS에 노출된 개인정보를 이용한 개인 재식별 (2013, ETRI)

- ◎ 페이스북 667만개, 트위터 277만개의 한국인 이용자 계정에 업로드한 데이터를 이용해서 개인에 대한 재식별 가능성 분석
- ◎ 분석 결과 기존에 비식별 정보라고 생각되던 정보로 개인을 특정할 수 있는 경우가 3% 이상이고, 다른 정보와 조합을 통해 개인을 특정할 수 있는 경우가 최대 45%에 달하는 것으로 분석
- ◎ 페이스북과 트위터에 공개된 이용자들의 정보와 정보를 공개한 이용자는 알지 못했던 데이터의 조합을 통해 개인을 재식별 할 수 있는 가능성과 그 정도를 확인

## Ⅲ. 개인정보 비식별화에 대한 적정성 평가

### 01

#### 적정성 평가 개요

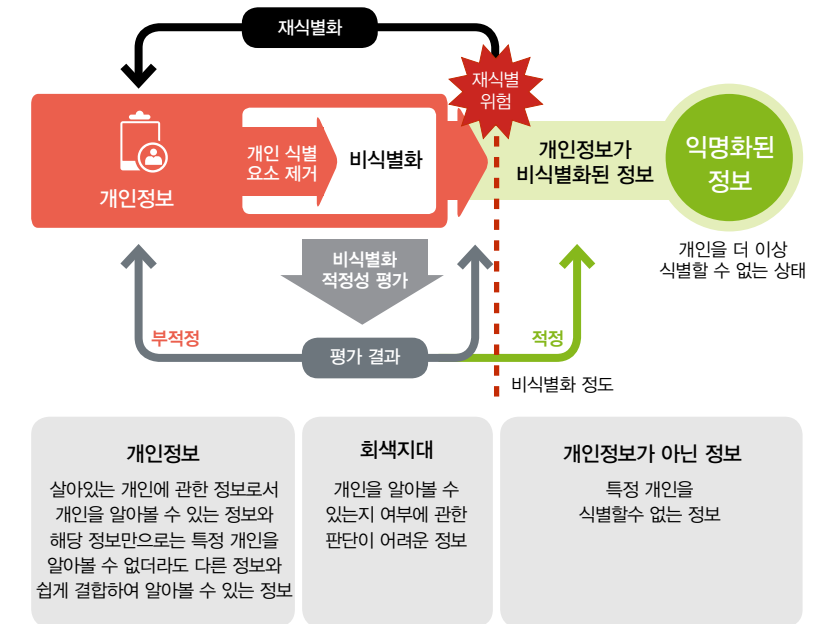
##### 가. 평가 배경

- ◎ 개인정보가 포함된 데이터는 개인정보 식별 요소 제거 기법을 통해 개인을 식별할 수 없는 형태로 데이터를 변경 및 이용 가능
  - 개인을 더 이상 식별할 수 없도록 변경한 데이터는 개인정보보호법상 개인정보에 해당하지 않기 때문에, 데이터 개방·공유 및 분석의 제한이 없음
- ◎ 그러나 개인식별요소의 제거 조치가 적절히 수행되지 않은 데이터는 다른 데이터와의 결합을 통해 특정 개인을 식별할 수 있는 위험을 내포함
- ◎ 따라서 정보주체의 개인정보를 보호하면서, 데이터 개방·공유 및 빅데이터 분석 활성화를 위해서는
  - 개인정보가 포함된 데이터에 대한 개인정보 식별 요소 제거 조치가 적정하게 이루어 졌는지를 확인하기 위한 체계적이고 객관적인 평가 방안이 요구됨

##### 나. 평가 목적

- ◎ 개인정보 식별 요소 제거 기법은 간단해 보이지만 실제 그 적용이 적절히 이루어졌는지를 판단하는 것은 매우 어려운 작업임
  - 특정 맥락에서는 비식별화 된 것으로 평가된 데이터라도 공격자의 다양한 시도, 새로운 데이터의 공개, 컴퓨팅 환경의 발전 등에 따라 재식별될 위험을 내포하고 있기 때문임

〈그림 4〉  
개인정보 비식별화에  
대한 적정성 평가 개념



- ◎ 본 평가는 개인정보가 포함된 데이터에 대한 개인식별요소 제거 조치가 적정하게 수행되었는지를 판단하기 위함
- ◎ 특히, 평가 대상 데이터에 대해 개인정보 비식별 조치를 추가적으로 수행해야 하는지, 현 상태로 이용·제공 및 공개 가능한지에 대한 판단 기준을 제공하기 위함

##### 다. 평가 대상

- ◎ 제한된 목적하에 개방·공유 및 분석 등의 활용을 위한 개인정보가 포함된 데이터

##### 라. 평가 추진 근거

- ◎ 개인정보보호법(제정 2011.3.29., 시행 2012.9.30.)

- (제3조) 개인정보처리자는 처리 목적에 필요한 범위에서 적합하게 개인정보를 처리하여야 하며, 그 목적 외의 용도로 활용하여서는 아니됨
- (제15조~제39조) 개인정보 수집·이용, 제공, 파기 등 처리 단계별 준수사항, 정보주체의 권리 보장, 개인정보의 안전성 확보조치 등

◎ 정부 3.0 기본계획(2013.6., 행정자치부)

※ '정보개방에 따른 보안 및 개인정보보호 대책'(46p)

- 정부3.0 대비 개인정보보호 대책 수립
  - 정보 개방·공유 등 처리 단계별 개인정보 보호지침 마련 및 이행 독려
    - ※ (수집·분석) 동의의무, 고지사항 등 준수 → (제공·개방) 민감정보 필터링, 비식별화 등 보호조치 → (저장·관리) 암호화 조치, 접근권한 관리 등 안전조치
  - 익명화 등 개인 비식별화 기법 보급 및 적용 지원
- 안전한 정보 개방·공유를 위한 개인정보보호 법령 적용 지원
  - 개인정보 보호법익을 고려한 정보 개방·공유의 범위·수준·방법 제시

## 마. 평가 시기

- ◎ 개인정보가 비식별화된 데이터를 인터넷 등에 공개하거나, 제 3의 기관 등에 제공하는 경우 관련 기관이 필요하다고 인정될 때 자율적으로 평가 수행 가능

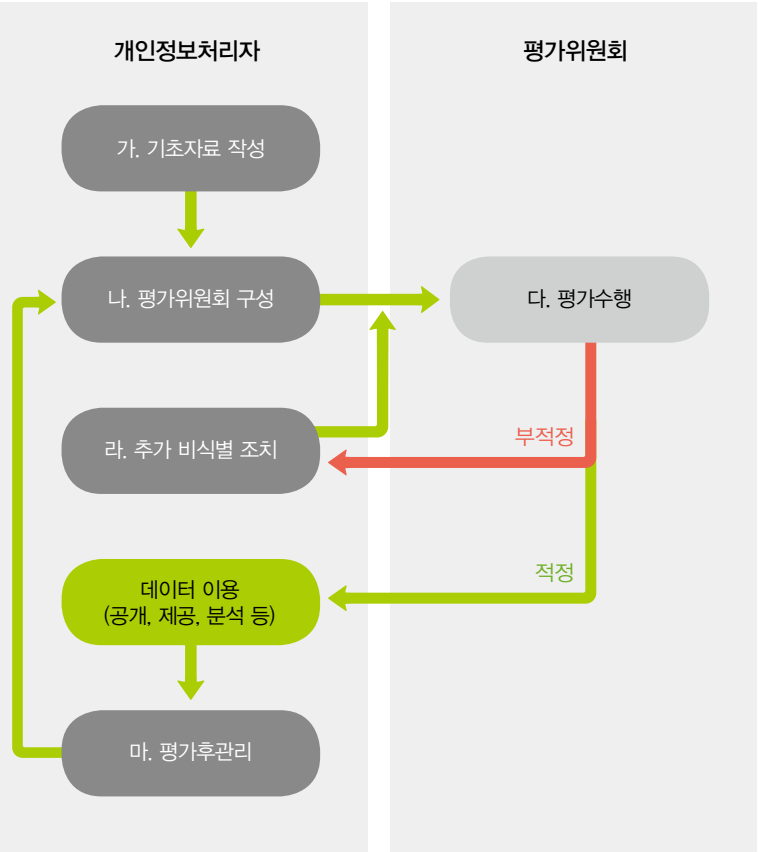
## 바. 안내서 활용을 위한 참고 사항

- ◎ 본 안내서는 데이터에 대한 비식별화 조치 뿐 아니라, 데이터 이용자, 데이터 제공 및 공개 유형, 개인정보 침해 가능성 등에 대해서 종합적으로 고려하여 검토할 수 있는 평가 방안을 제시하고자 하였음

- ◎ 본 안내서를 이용해서 개인정보 비식별화에 대한 적정성을 평가하고, 관련 데이터를 공개 또는 제공하는 것에 대한 책임은 해당 개인정보처리자에게 있으며, 평가위원회는 전문가로서의 판단을 제공하는 역할을 수행함
- ◎ 따라서 개인정보처리자는 평가를 위한 기초자료의 작성, 평가위원회의 구성 및 운영, 평가 시행, 평가 후 관리 등이 충실히 수행될 수 있도록 노력해야 함
- ◎ 본 안내서는 정형 데이터를 비식별화 하는 것에 대한 평가를 중심으로 개발되었으나, 비정형 데이터를 비식별화한 것에 대한 평가를 위하여 응용할 수 있음

02  
평가 절차

◎ 평가 대상 데이터에 대한 개인식별요소 제거 조치의 적정성을 평가하기 위한 절차는 ‘가. 평가 기초자료 작성’, ‘나. 평가위원회 구성’, ‘다. 평가 수행’, ‘라. 추가 비식별 조치’, ‘마. 평가 후 관리’ 등 다섯 단계로 <그림 5>와 같이 구성됨



<그림 5>  
개인정보 비식별화에 대한  
적정성 평가 절차

가. 기초자료 작성

- ◎ 개인정보처리자에서는 <표 3>에서 제시한 기초자료를 성실히 작성하고 준비해야 함
- ◎ 특히 평가위원회를 구성해서 평가를 수행하고자 하는 경우에는 <표 3>의 필수 사항은 반드시 작성하고, 평가위원회에 제출해야 함

<표 3>  
개인정보 비식별화에 대한  
적정성 평가를 위한 기초자료

구 분	기초자료	비 고
데이터	데이터 특성(크기, 생성 및 관리 환경 등), 세부 항목별 명세, 적용 코드 명세 및 원본 예시	필수사항
	개인식별요소가 제거된 평가 대상 데이터 및 세부 항목별 명세	필수사항
개인 식별요소 제거 현황	개인식별요소 제거 기법 적용 기준	필수사항
	평가 대상 데이터에 대한 $k$ -익명성, $l$ -다양성, $t$ -근접성 값 산출 결과	선택사항
이용기관의 관리수준	데이터 이용 기관의 데이터 이용 목적, 이용 방법, 이용기간, 데이터 접근 가능자 현황 등 데이터 활용에 관한 사항	필수사항
	데이터를 제공하는 경우 데이터를 제공받는 방법 및 데이터를 보호하기 위해 취하는 일련의 조치에 대한 현황	필수사항
	데이터 이용 및 제공과 관련이 있는 계약서 또는 협약서 사본 ※ 계약서가 없는 경우에는 없는 사유 제시	필수사항
	데이터 이용 기관에서 보유하거나, 보유할 수 있는 개인정보 관련 데이터(세부 내용 및 항목 등)에 관한 사항	선택사항
	데이터 이용기관의 PIPL, PMS 등 개인정보보호 관련 인증서 사본	선택사항

나. 평가위원회 구성

- ◎ 평가위원회는 개인정보처리자의 개인정보보호책임자(CPO)가 지정하는 3인 이상의 평가위원으로 구성하며, 평가위원은 홀수로 지정하여 평가위원의 의견이 1:1로 대립되지 않도록 구성



- ◎ 평가위원의 과반수 이상은 해당 평가 대상 기관 및 평가 대상 데이터의 이용과 관련이 없는 외부의 전문가로 지정
  - 관련 업무영역의 전문가 1인, 개인정보 비식별화 전문가 1인, 법률 전문가 1인은 필수적으로 포함되도록 평가위원회를 구성
- ◎ 평가위원장은 외부에서 위촉된 평가위원 중에서 호선으로 선출하며, 평가위원회 회의 개최, 심사, 종료, 평가결과의 발표 등 평가위원회의 운영과 관련된 전반적인 사항을 관장함
- ◎ 평가위원장은 다른 평가위원들과 협의하여 평가위원별 전문성을 고려한 업무 분장을 실시하고 평가업무를 수행할 수 있음
- ◎ 평가위원회는 착수회의를 포함해서 최소 2회 이상 운영하여야 하며, 추가적인 평가위원회의 개최가 필요할 경우에는 평가위원회의 협의를 거쳐 운영 회수 및 기간 등을 연장할 수 있음

#### 다. 평가 수행

- ◎ 평가위원회는 개인정보처리자에서 제공한 <표 3>의 ‘기초자료’와 ‘3. 세부 평가 방법’을 기반으로 평가 대상 데이터에 대한 개인식별요소 제거 기법 및 비식별 수준의 적정성에 대해 평가하고, 최종적으로 ‘적정’ 또는 ‘부적정’ 의견을 제시하며, 필요시 기초자료에 대한 보완을 요청할 수 있음
- ◎ 평가위원회는 평가 대상 데이터의 특성을 종합적으로 고려하여, ‘3. 세부 평가 방법’의 각 평가 단계별 평가지표 및 평가기준에 대해 평가위원회의 협의를 거쳐 조정 및 보완하여 사용할 수 있음
  - 단, 협의가 이루어지지 않는 경우에는 과반 수 이상의 평가위원이 선택한 평가 지표 및 기준을 채택해서 사용할 수 있음

- ◎ 평가위원회가 ‘적정’으로 평가한 경우에는 해당 정보만으로는 특정 개인을 알아볼 수 없을 뿐만 아니라, 다른 정보와 쉽게 결합하여도 특정 개인을 식별할 수 없는 상태라는 것을 의미함
- ◎ 평가위원회가 ‘부적정’으로 평가한 경우에는 관련 데이터의 개인식별요소 제거 등 추가조치가 필요하다는 것을 의미함
  - 평가위원회는 추가적인 개인식별요소 제거 조치에 대한 사항(적용기법 및 비식별 수준)을 구체적으로 제시해야 함
- ◎ 평가위원회는 컴퓨팅 환경의 발전, 평가 대상 데이터의 특성, 연계 가능한 데이터의 공개 가능성 등 다양한 환경변화를 고려하여, 평가 대상 데이터에 대해 일정 시간 경과 후 재평가를 받을 필요가 있는지에 대한 의견을 제시할 수 있음
  - 재평가가 필요하다고 인정된 경우에는 1년, 3년 등 구체적인 재평가 일정을 제시해야 함

#### 라. 추가 비식별 조치

- ◎ 개인정보처리자는 평가결과가 ‘부적정’인 경우, 평가위원회의 평가 의견을 고려하여 해당 데이터에 대한 개인식별요소 제거 조치를 추가적으로 수행<sup>8)</sup>
- ◎ 개인정보처리자는 평가위원회에서 제시한 개인식별요소 제거 기법 및 비식별 수준을 고려하여 개인정보가 식별되지 않도록 조치하고 관련 내용을 문서로 작성하여 관리해야 함
- ◎ 개인식별요소 추가 제거 작업이 완료된 경우에는 관련 ‘기초자료’를 보완하고, 평가위원회에 추가적인 개인식별요소 제거 작업이 적정히 수행되었는지에 대한 확인을 요청해야 함

8) 독일 뮌헨공대(Technische Universität München)에서 추진된 “ARX - Powerful Data Anonymization” (<http://arx.deidentifier.org/>) 프로젝트에서 프라이버시 모델( $k$ -익명성,  $l$ -다양성,  $t$ -근접성 등)을 이용하여 데이터의 익명화를 지원하는 GUI 프로그램 및 관련 소스 코드를 GNU 라이선스로 배포하고 있음

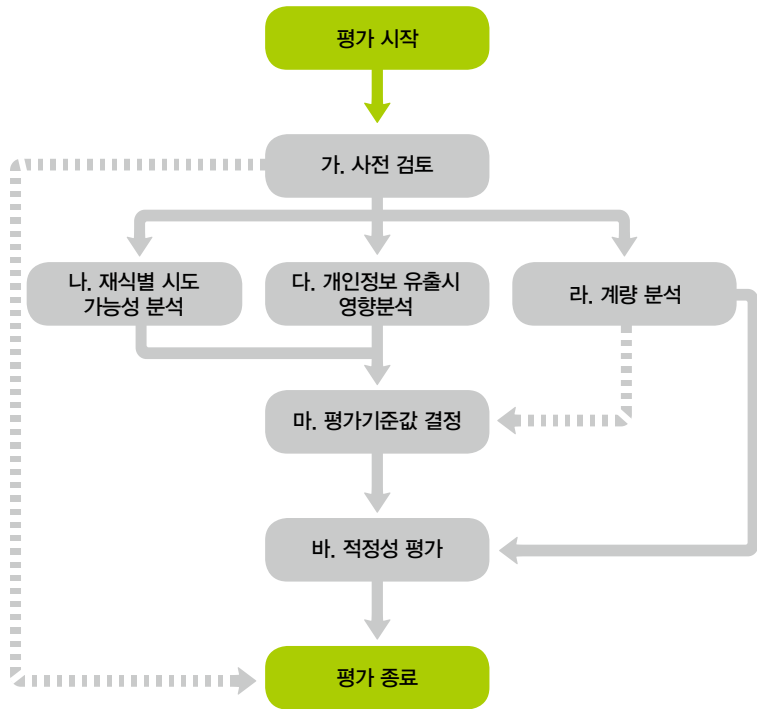
마. 평가 후 관리

- ◎ 평가위원회 또는 개인정보처리자가 인정한 경우에는 평가 대상 데이터에 대해 사후에 개인정보 비식별화 적정성 평가를 재 실시 할 수 있음
  - 개인정보처리자가 비식별 적정성에 대해 사후적으로 재평가를 고려해야 하는 경우는 <표 13> “개인정보 비식별화에 대한 적정성 재평가 검토 질문서”를 참고 하여 수행할 수 있음
- ◎ 개인정보 비식별화에 대한 적정성 평가를 재 실시 하는 경우에는 평가위원회를 재구성 할 수 있으며, 평가의 연속성 차원에서 기존 평가 위원 중 최소 1인 이상을 포함하여 구성할 수 있음

03

세부 평가 방법

세부 평가 방법은 <그림 6> 과 같이 사전검토, 재식별 시도 가능성 분석, 개인정보 유출시 영향 분석, 계량 분석, 평가기준값 결정 및 적정성 평가로 구성된다.



<그림 6>  
개인정보 비식별  
적정성 세부 평가 방법

가. 사전 검토

- ◎ 제출된 기초자료와 담당자 인터뷰 등을 기반으로 개인정보 비식별화에 대한 적정성 평가 수행 여부를 검토
- ◎ 첫째, 개인정보처리자가 작성 및 제출한 기초자료가 필수사항을 모두 포함하고 있고, 적절히 작성되었는지 검토
  - <표 3>에 따른 기초자료가 제출되지 않은 경우, 개인정보처리자에 추가적인 자료의 제출 및 보완을 요구할 수 있음

- ◎ 둘째, 평가 대상 데이터의 특성에 대해 확인하고, 개인을 식별할 수 있는 식별 요소를 포함하고 있는지 검토
  - 평가대상 데이터가 생성 및 관리되는 환경, 데이터의 규모, 시간 흐름에 따른 축적 여부 등 데이터의 특성에 대해 확인함
  - 평가 대상 데이터가 <표 1> 과 같은 식별자 또는 <표 2>와 같은 준식별자를 포함하고 있는지 여부 등 검토
  - 평가 대상 데이터가 개인 식별자를 포함하고 있는 경우, 개인식별요소 제거 조치가 ‘부적정’ 한 것으로 판단하고 평가 종료 가능
- ◎ 셋째, 기초자료로 제출된 ‘개인식별요소 제거 조치 기법 및 기준’에 따라 관련 조치가 적절히 수행되었는지 검토
  - <표 3>의 ‘데이터 원본 예시 및 세부 항목별 명세’, ‘개인식별요소가 제거된 평가 대상 데이터 및 세부 항목별 명세’ 및 ‘개인식별요소 제거 기법 적용 기준’을 검토
  - ‘개인식별요소 제거 기법 적용 기준’에 따라 개인식별요소 제거 조치가 적용되지 않은 경우, ‘부적정’ 한 것으로 판단하고 평가를 종료할 수 있음

나. 재식별 시도 가능성 분석

- ◎ 평가 대상 데이터를 이용하거나 제공받으려는 자의 개인정보 재식별 시도 가능성에 대한 분석은 ‘1) 재식별 의도 및 능력’, ‘2) 정보보호 능력’에 대해 우선 분석을 실시하고,
  - 그 결과를 종합하여 최종적으로 ‘3) 재식별 시도 가능성’을 분석함

1) 재식별 의도 및 능력 분석

- ◎ 데이터 이용자 또는 요청자의 재식별 의도 및 능력에 대하여 검토하고 평가를 실시함
- ◎ 개별 평가위원은 <표 4> 평가지표의 세부 질문에 대해 검토하고 개별 평가 지표별로 ‘예’ 또는 ‘아니오’로 평가를 실시함

<표 4>  
재식별 의도 및  
능력 분석 평가 지표

구 분	세부 지표	평 가
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 금지 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 자원(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 연계 관련	인터넷, SNS 등에 평가대상 데이터와 결합 가능한 데이터가 존재할 수 있음	예/아니오

- ◎ 평가위원별 평가점수는 개별 평가위원이 각 평가지표에 대해 ‘예’로 평가한 지표의 개수를 합산해서 산출함
- ◎ 개별 평가위원의 점수를 합산한 후 평가위원의 수로 나누어 ‘재식별 의도 및 능력 분석’의 평균 점수를 구하고, <표 5> 평가 기준에 따라 ‘높음’, ‘중간’, ‘낮음’으로 1차 평가결과를 도출함

〈표 5〉  
재식별 의도 및  
능력 분석 평가 기준

- ◎ 1차 평가결과를 기초로 평가위원회의 토의를 거쳐 최종 확정하며, 1차 평가 결과보다 하위의 기준을 적용하는 경우에는 이에 대한 사유를 명확히 문서로 남겨야 함
- ◎ 비식별 정보를 인터넷 등 일반에 공개하는 경우에는 점수와 무관하게 ‘재식별 의도 및 능력 분석’의 평가 결과를 ‘높음’으로 평가

구 분	평가 기준
높 음	평균 점수가 5점 이상인 경우 또는 데이터를 일반에 공개 하는 경우
중 간	평균 점수가 3점 이상, 5점 미만인 경우
낮 음	평균 점수가 3점 미만인 경우

2) 정보보호 능력 분석

- ◎ 데이터 이용자 또는 요청자가 제공받은 데이터에 대한 정보보호 능력이 어떠한 수준인지 검토하고 평가를 실시함
- ◎ 개별 평가위원은 〈표 6〉 평가지표의 세부 질문에 대해 검토하고 개별 평가 지표별로 ‘예’ 또는 ‘아니오’로 평가를 실시함
- ◎ 평가위원별 평가점수는 개별 평가위원이 각 평가지표에 대해 ‘예’로 평가한 지표의 개수를 합산해서 산출함
- ◎ 개별 평가위원의 점수를 합산한 후 평가위원의 수로 나누어 ‘정보보호 능력 분석’의 평균 점수를 구하고, 〈표 7〉 평가 기준에 따라 ‘높음’, ‘중간’, ‘낮음’, ‘없음’으로 1차 평가결과를 도출
- ◎ 1차 평가결과를 기초로 평가위원회의 토의를 거쳐 최종 확정하며, 1차 평가 결과보다 하위의 기준을 적용하는 경우에는 이에 대한 사유를 명확히 문서로 남겨야 함
- ◎ 비식별 정보를 인터넷 등 일반에 공개하는 경우에는 점수와 무관하게 ‘정보보호 능력 분석’의 평가 결과를 ‘없음’으로 평가

〈표 6〉  
정보보호 능력 분석 평가 지표

구 분	세부 지표	평 가
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 PIPL, PIMS 등의 인증을 받음	예/아니오

〈표 7〉  
정보보호 능력 분석 평가 기준

구 분	평가 기준
높 음	평균 점수가 6점 이상인 경우
중 간	평균 점수가 4점 이상, 6점 미만인 경우
낮 음	평균 점수가 4점 미만인 경우
없 음	인터넷 등 일반에 공개하는 경우

3) 재식별 시도 가능성 분석

- ② 1) 재식별 의도 및 능력 분석<sup>9)</sup>, 2) 정보보호 능력 분석<sup>9)</sup>의 결과를 이용해서 비식별화된 데이터의 재식별 시도 가능성 분석을 최종적으로 실시
- ② 재식별 시도 가능성에 대한 평가는 ‘빈번한’, ‘가능한’, ‘가끔’, ‘거의 없는’ 등 4단계로 평가
- ② <그림 7>와 같이 1) 재식별 의도 및 능력 분석<sup>9)</sup>의 결과값과, 2) 정보보호 능력 분석<sup>9)</sup>의 결과값이 교차하는 지점의 평가값으로 재식별 시도 가능성 분석

<그림 7>  
재식별 시도 가능성 분석표

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

다. 개인정보 유출시 영향 분석

- ② 데이터가 의도적 또는 비의도적으로 노출 및 유출되어, 재식별 되었을 때 정보주체에게 미치는 영향에 대해 분석
  - 특히 경제적 또는 비경제적(개인정보 또는 프라이버시 침해)인 피해를 줄 수 있는 가능성에 대해 평가를 실시
- ② 개별 평가위원은 <표 8> 평가지표의 세부 질문에 대해 검토하고 개별 평가 지표별로 ‘예’ 또는 ‘아니오’로 평가를 실시함
- ② 평가위원별 평가점수는 개별 평가위원이 각 평가지표에 대해 ‘예’로 평가한 지표의 개수를 합산해서 산출함

<표 8>  
개인정보 유출시  
영향 분석 평가 지표

- ② 개별 평가위원의 점수를 합산한 후 평가위원의 수로 나누어 ‘개인정보 유출 위험성 분석’의 평균 점수를 구하고, <표 9> 평가 기준에 따라 ‘높음’, ‘중간’, ‘낮음’으로 1차 평가결과를 도출함
- ② 1차 평가결과를 기초로 평가위원회의 토의를 거쳐 최종 확정하며, 1차 평가 결과보다 낮은 기준을 적용하는 경우에는 이에 대한 사유를 명확히 문서로 남겨야 함

구 분	세부 지표	평 가
공개 또는 유출시 영향 관련	데이터가 공개 또는 유출 <sup>9)</sup> 되었을 때 법적, 도덕적, 기술적 이슈로 사회적인 혼란을 가져올 가능성이 있음	예/아니오
	데이터가 공개 또는 유출되었을 때 관련 정보주체의 개인정보 또는 프라이버시를 침해할 수 있음	예/아니오
	데이터가 공개 또는 유출되었을 때 관련 정보주체에게 경제적 또는 비경제적 손실을 발생시킬 수 있음	예/아니오
	데이터가 공개 또는 유출되었을 때 데이터 이용자 또는 요청자에게 경제적 또는 비경제적 손실을 발생시킬 수 있음	예/아니오

9) 의도하지 않게 사용되거나, 권한이 없이 사용된 것 포함

〈표 9〉  
개인정보 유출시  
영향 분석 평가 기준

구 분	평가 기준
높 음	평균 점수가 2점 이상인 경우
중 간	평균 점수가 1점 이상, 2점 미만인 경우
낮 음	평균 점수가 1점 미만인 경우

라. 계량 분석

- ◎ 평가 대상 데이터의 특성을 고려하여 평가 대상 데이터에 대한 비식별 수준을 분석할 수 있는 분석 기법을 선정하고 분석함
  - ※ 평가위원회에서 데이터의 특성, 비식별화 정도 등을 고려해서 분석기법 선정
- 비식별 정도를 분석하기 위한 기법에는  $k$ -익명성( $k$ -anonymity),  $l$ -다양성( $l$ -diversity),  $t$ -근접성( $t$ -closeness) 등의 프라이버시 모델 이용 가능
  - ※ 각 기법의 세부 내용은 '[부록2]. 프라이버시 보호 모델' 참조
- ◎ 분석결과는 '마. 평가기준값' 결정시 참고할 수 있으며, 필요시 재분석 할 수 있음
- ◎ 평가 대상 데이터에 대한 비식별 정도에 대한 계량 분석은 개인정보처리자가 직접 수행하거나, 외부의 공신력 있는 전문기관(전문기술지원기관, 연구소, 학교 등)에 의뢰하여 수행할 수 있음

마. 평가 기준 값 결정

- ◎ 평가위원회는 평가 대상 데이터에 대한 개인식별요소 제거 조치의 적정성을 계량적으로 평가하는데 필요한 평가 기준 값을 〈그림 8〉의 사례와 같이 결정하여 사용할 수 있음
  - 평가 기준 값으로는 ' $k$ -익명성', ' $l$ -다양성', ' $t$ -근접성' 값 등이 단독 또는 복수 개 이상으로 설정될 수 있음
- ◎ 평가 기준 값 결정시 고려 사항
  - 평가 대상 데이터의 준식별자 항목 수, 규모, 시간 흐름에 따른 누적 데이터 존재 여부 등의 데이터 특징

〈그림 8〉  
평가 기준 값 예시<sup>10)</sup>

다. 개인정보 유출시 영향					
침해위험 높음	$k = 5$ $l = 3$	$k = 10$ $l = 3$	$k = 15$ $l = 4$	$k = 30$ $l = 5$ $t < 0.3$	
침해위험 중간	$k = 3$ $l = 2$	$k = 5$ $l = 2$	$k = 10$ $l = 3$	$k = 15$ $l = 4$	
침해위험 낮음	$k = 3$ $l = 2$	$k = 5$ $l = 2$	$k = 5$ $l = 2$	$k = 15$ $l = 3$	
	거의 없는	가끔	가능한	빈번한	나. 재식별 시도 가능성

10) 〈그림 8〉의 “세부 평가 기준 값”은 단순 사례이며, 실제 적용시 일반적인 기준 값으로 이용하는 것은 적정하지 않을 수 있음. 기준값에 대한 결정은 평가위원회의 검토 및 논의에 따라 적용 프라이버시 모델 및 기준을 정하여 사용해야 함



바. 적정성 평가

- ◎ 평가위원회는 ‘마. 평가 기준 값 결정’에서 도출된 평가 기준 값과 ‘라. 계량 분석’에서 계산된 분석 값을 비교하여 개인정보 비식별화에 대한 1차 평가 결과인 ‘적정’ 또는 ‘부적정’을 도출
- ◎ 최종적인 평가는 1차 평가 결과를 기초로 평가위원회의 토의를 거쳐 최종 확정하며, 1차 평가 결과와 반대의 결과를 도출한 경우에는 이에 대한 사유를 명확히 문서로 남겨야 함

1) k-익명성 값을 이용한 비식별 적정성 평가

- ◎ ‘라. 계량 분석’에서 분석된 평가 대상 데이터의 k-익명성 분석값이 평가위원회에서 결정한 ‘평가 기준 값’ 보다 작은 경우에는 <표 10>과 같이 개인식별요소 제거 조치가 ‘부적정’한 것으로 평가
- ◎ ‘라. 계량 분석’에서 분석된 평가 대상 데이터의 k-익명성 분석값이 평가위원회에서 결정한 ‘평가 기준 값’ 보다 크거나 같은 경우에는 <표 10>과 같이 개인식별요소 제거 조치가 ‘적정’한 것으로 평가



<표 10>  
k-익명성 기반  
적정성 평가 사례

k-익명성 값을 이용한 개인정보 비식별화에 대한 적정성 평가			
계량분석의 k-익명성 값	<	평가 기준 값 (k-익명성 값)	계량분석의 k-익명성 값 >= 평가 기준 값 (k-익명성 값)
			
‘부적정’ (개인식별요소 제거 조치 필요)		‘적정’ (개인식별요소 제거 조치 불필요)	

2) l-다양성 값을 이용한 비식별 적정성 평가

- ◎ ‘라. 계량 분석’에서 분석된 평가 대상 데이터의 l-다양성 분석 값이 평가위원회에서 결정한 ‘평가기준값(l-다양성)’ 보다 작은 경우에는 <표 11>과 같이 개인식별요소 제거 조치가 ‘부적정’한 것으로 평가
- ◎ ‘라. 계량 분석’에서 분석된 평가 대상 데이터의 l-다양성 분석 값이 평가위원회에서 결정한 ‘평가기준값(l-다양성)’ 보다 크거나 같은 경우에는 <표 11>과 같이 개인식별요소 제거 조치가 ‘적정’한 것으로 평가



<표 11>  
l-다양성 기반  
적정성 평가 사례

l-다양성 값을 이용한 개인정보 비식별화에 대한 적정성 평가			
계량분석의 l-다양성 값	<	평가 기준 값 (l-다양성)	계량분석의 l-다양성 값 >= 평가 기준 값 (l-다양성)
			
‘부적정’ (개인식별요소 제거 조치 필요)		‘적정’ (개인식별요소 제거 조치 불필요)	

3) *t*-근접성 값을 이용한 비식별 적정성 평가

- ‘라. 계량 분석’에서 분석된 평가 대상 데이터의 *t*-근접성 분석 값이 평가위원회에서 결정한 ‘평가기준값(*t*-근접성)’ 보다 작은 경우에는 <표 12>와 같이 개인식별요소 제거 조치가 ‘적정’한 것으로 평가
  - 통상 *t*-근접성 값의 범위는 0에서 1사이의 소수이며, 0에 가까울수록 개인을 식별할 가능성이 적다는 것을 의미함
- ‘라. 계량 분석’에서 분석된 평가 대상 데이터의 *t*-근접성 분석 값이 평가위원회에서 결정한 ‘평가기준값(*t*-근접성)’ 보다 크거나 같은 경우에는 <표 12>과 같이 개인식별요소 제거 조치가 ‘부적정’한 것으로 평가

<표 12>  
*t*-근접성 기반  
적정성 평가 사례

<i>t</i> -근접성 값을 이용한 개인정보 비식별화에 대한 적정성 평가	
계량분석의 <i>t</i> -근접성 값 <b>&gt;=</b> 평가 기준 값 ( <i>t</i> -근접성 값)	계량분석의 <i>t</i> -근접성 값 <b>&lt;</b> 평가 기준 값 ( <i>t</i> -근접성 값)
	
‘부적정’ (개인식별요소 제거 조치 필요)	‘적정’ (개인식별요소 제거 조치 불필요)

01

재식별  
위험에 대한  
관리적 조치

가. 데이터 제공 및 위탁 계약시 재식별 금지 관련 조항 반영

- 개인정보가 비식별화된 데이터를 제3의 기관에 제공하거나, 처리를 위탁하는 경우, 개인정보가 재식별될 위험을 관리하기 위한 내용을 계약서에 반드시 반영

< 계약서 특수조건 반영 내용 예시 >

- 제00조(개인정보 재식별 금지)
- ① 을은 갑으로부터 제공받은 데이터를 비식별화하여 이용 및 유지하여야 하고, 이를 이용해서 개인을 재식별하기 위한 행위를 수행해서는 안된다.
  - ② 갑으로부터 제공받은 데이터를 을이 제3자에게 제공하거나 처리를 위탁하는 경우, 을은 관련 데이터를 이용해서 개인을 재식별하기 위한 행위를 금지하는 것을 반드시 문서로서 명확히 하여야 한다.
  - ③ 을은 갑으로부터 제공받은 데이터에 대한 재식별이 이루어지거나 재식별의 가능성이 현저하게 높아지는 상황이 발생하면, 즉시 데이터 처리 작업을 중지하고 관련 사항을 갑에게 알리며, 필요한 협조를 제공해야 한다.
  - ④ 을은 제1항에서 제3항까지의 사항을 이행하지 않아 발생하는 모든 결과에 대한 모든 책임을 진다.
- ※ 개인정보가 비식별화된 데이터를 제공한 기업(이하 “갑”), 제공받은 기업(이하 “을”)



- (데이터 처리시 재식별 금지) 데이터를 제공 또는 위탁 받은 기관에서는 데이터를 비식별화된 상태로 처리하고, 데이터를 이용한 재식별 시도를 금지해야 한다는 것을 명확히 함
- (재 제공 및 위탁시 재식별 금지 명문화) 데이터를 제공 또는 위탁 받은 기관에서 관련 데이터를 제3의 기관에 재제공하거나, 처리를 재위탁하는 경우에도 재식별 시도를 금지한다는 것을 명확히 함
- (위험 발견시 통지) 재식별이 이루어지거나, 재식별의 가능성이 높아지는 상황 발생시, 관련 사항에 대해 데이터 이용자가 데이터 제공자에게 통지하는 책임을 명확히 함

나. 데이터 공개 시 재식별 금지 관련 조항 게시

- ◎ 개인정보 비식별화 데이터를 인터넷 등에 공개하는 경우, 관련 데이터를 이용하는 개인 또는 단체가 개인정보를 재식별하기 위한 행위를 금지하는 내용을 이용약관 등의 형태로 게시

〈 이용약관 또는 저작권 내용 사례 〉

제00조(개인정보 재식별 금지)

- ① 데이터 이용자는 데이터를 비식별화하여 이용 및 유지해야 하며, 이를 이용해서 개인을 재식별하기 위한 행위를 수행해서는 안된다.
- ② 데이터 이용자는 데이터에 대한 재식별이 이루어지거나 재식별의 가능성이 현저하게 높아지는 상황이 발생하면, 즉시 데이터 처리 작업을 중지하고 관련 사항을 데이터 제공자에게 알려야 한다.
- ③ 데이터 이용자는 제1항 및 제2항의 사항을 이행하지 않아 발생하는 모든 결과에 대해 형사 및 민사상 책임을 진다.

※ 데이터 이용자는 개인정보가 비식별화된 데이터를 이용하는 자

- (데이터 처리시 재식별 금지) 공개된 데이터를 이용하는 기관에서는 데이터를 비식별화된 상태로 처리하고, 데이터에 대한 재식별 시도를 금지해야 한다는 것을 명확히 함

- (위험 발견시 통지) 재식별이 이루어지거나, 재식별의 가능성이 높아지는 상황 발생시, 데이터 이용자는 데이터 제공자에게 관련 사항을 통지해야 한다는 점을 명확히 함

다. 재식별 가능성 모니터링

- ◎ 개인정보를 비식별화한 데이터를 이용 및 처리하는 기관에서는 개인정보 비식별화에 대한 적정성 평가를 재 실시 해야 하는 경우가 발생하는지를 상시 또는 정기적으로 모니터링 및 관리할 필요가 있음

〈표 13〉  
개인정보 비식별화에 대한  
적정성 재평가 검토 질문서

구 분	세부 지표	비 고
내부 환경의 변화	평가 대상 데이터와 연계 가능한 추가적인 개인정보의 수집 · 보유 · 이용 · 제공 및 공개 범위를 변경하는 경우	
	평가 대상 데이터에 대한 비식별 수준을 변경 요청이 있는 경우	
	서비스 이용 과정에서 생성되는 정보를 평가 대상 데이터와 결합해서 기존에 존재하지 않던 2차 정보가 생성되는 경우	
	신규 또는 추가로 구축되는 시스템이 평가 대상 데이터에 대한 접근을 관리 또는 통제 하는 보안체계에 중대한 변화를 초래하는 경우	
외부 환경의 변화	평가 대상 데이터를 이용한 재식별 사례가 알려진 경우	
	평가 대상 데이터를 이용해서 개인을 식별할 수 있는 방안 또는 기술이 공개된 경우	
	평가 대상 데이터와 새롭게 연계 가능한 데이터가 출현하거나 공개된 것으로 알려진 경우	

- ◎ <표 13>의 질문서 항목 중 어느 하나에 해당하는지에 대해 정기적으로 점검하고, 관련 사항이 발생하는 경우 개인정보 비식별화에 대한 적정성 평가를 다시 실시할 필요가 있음

#### 라. 개인정보 재식별시의 대응 매뉴얼 마련 및 시행

- ◎ 기관별 데이터 공개, 제공, 위탁 특성을 고려하여 개인정보 재식별시 대응 매뉴얼을 마련하고, 관련 개인정보취급자에 대한 교육을 정기적으로 실시함

- ◎ 제3자에게 제공 및 위탁하거나 일반에 공개한 데이터가 개인정보를 재식별하거나, 재식별 할 가능성이 높은 것으로 확인된 경우, 개인정보처리자는(정보를 제공받은 자에게 귀책사유가 있는 경우, 제공받은 자는) 해당 정보주체의 개인정보를 보호하고, 2차적으로 발생 가능한 개인정보침해를 최소화하기 위해 신속하게 다음과 같이 조치하여야 함

#### 가. 데이터 공개 중단 및 회수

- ◎ 신속하게 홈페이지 등에 공개되어 있는 관련 데이터에 대한 공개를 중단하며, 온라인으로 조회 서비스 등을 제공 하는 경우에도 해당 데이터에 대한 접근을 통제함
- ◎ 기록매체 등에 저장되어 있는 경우에는 신속하게 관련 매체의 배포처를 확인하고 회수 처리함

#### 나. 데이터 제공 및 처리 위탁 중단

- ◎ 재식별 위험이 발생한 데이터에 대한 제3자 제공 및 처리 위탁 등 일체의 제공 행위를 즉시 중단함
- ◎ 추가적인 비식별 조치가 완료되기 전까지 관련 데이터의 이용 및 제공을 금지함

- ◎ 계약 등의 이슈가 제기되는 경우에는 관련 데이터의 재식별을 통한 경제적, 비경제적 피해를 고려하여 해당 기관과 신속히 협의 및 제공을 중단하기 위한 절차를 마련할 필요가 있음

#### 다. 데이터 재식별 위험 통지

- ◎ 개인정보가 재식별된 데이터를 이용 및 처리하고 있는 기관에 개인정보 재식별 위험에 대해 안내하고, 관련 데이터 및 관련 데이터로부터 파생된 데이터에 대한 처리의 중단을 요청
- ◎ 필요시 관련 데이터 및 관련 데이터로부터 파생된 데이터를 삭제해 줄 것을 요청

#### 라. 개인정보 유출통지 및 유출신고

- ◎ 재식별된 개인정보를 이용한 2차 피해를 최소화하기 위해 관련 정보주체에게 개인정보 유출 통지 조치를 <표 14>와 같이 실시
  - 개인정보처리자는 개인정보 유출 사실을 알게된 때 지체 없이 정보주체에게 유출관련 내용을 서면, 전자우편, 팩스, 전화 등의 방법으로 통지
    - ※ 유출 통지 시 포함 사항 : 1. 유출된 개인정보의 항목, 2. 유출된 시점과 그 경위, 3. 유출피해를 최소화하기 위하여 정보주체가 할 수 있는 방법 등에 대한 정보, 4. 개인정보처리자의 대응조치 및 피해구제절차, 5. 정보주체에게 피해가 발생한 경우 신고 등을 접수할 수 있는 담당부서 및 연락처
- ◎ 유출된 개인정보 건수가 1만명 이상인 경우에는 <표 15>와 같이 행정자치부에 개인정보 유출 신고를 실시하고 홈페이지를 통해 7일 이상 게재

#### 마. 추가적 비식별화 조치

- ◎ 재식별이 가능하다고 인지된 데이터에 대해 추가적인 개인식별요소 제거 조치를 수행
- ◎ 추가적인 개인식별요소 제거 조치가 수행된 데이터를 대상으로 개인정보 비식별화에 대한 적정성을 재평가

〈표 14〉  
개인정보 유출시 통지 방법

구 분	내 용	비 고
통지 방법	◎ 개인정보처리자는 개인정보 유출을 알게된 때 지체 없이 정보주체에게 개인정보 유출 관련 사항 통지  1. 서면, 전자우편, 모사전송, 전화, 문자전송 또는 이와 유사한 방법(필수사항) 2. 1번 통지방법과 동시에 홈페이지 등을 통하여 7일 이상 공개 ※ 1번 통지방법으로 연락 불가시 홈페이지 게시	개인정보보호법 시행령 제40조 지침 제28조 지침 제29조
통지 내용	1. 유출된 개인정보의 항목 2. 유출된 시점과 그 경위 3. 유출피해 최소화하기 위하여 정보주체가 할 수 있는 방법 등에 대한 정보 4. 개인정보처리자의 대응조치 및 피해구제 절차 5. 정보주체에게 피해가 발생한 경우 신고 등을 접수할 수 있는 담당부서 및 연락처	개인정보보호법 제34조
통지 시기	◎ 개인정보 유출사고를 안 때부터 지체없이 (5일 이내) 통보 ◎ 단, 개인정보 구체적인 유출내용 확인 불가한 경우 다음 사항을 알리고 추후 확인된 사항 추가안내 가능  1. 정보주체에게 유출이 발생한 사실 2. 통지내용 중 확인된 사항	개인정보보호법 시행령 제40조
통지 결과 신고	◎ 통지결과* 등 지체없이 신고기관에 신고 * (예시) 통지내용·결과, 홈페이지에 게시한 경우 게시기간 및 내용 화면캡처 등 입증자료 첨부	개인정보보호법 제34조
기타 관련 시스템 구축	◎ 홈페이지 개인별 유출 확인 시스템 구축 － 방법 : 이름, 생년월일, 휴대폰번호로 유출여부 확인 － 안내문구 : ‘개인정보 유출된 고객입니다.’ ‘개인정보 유출된 고객이 아닙니다’ 등으로 확인	－

〈표 15〉  
개인정보 유출시 신고 방법

구 분	내 용
신고기관	◎ 행정자치부 또는한국인터넷진흥원에 신고
신고대상	◎ 1만명 이상 정보주체의 개인정보 유출시
신고시기	◎ 1만건 이상 ◎ 지체없이(5일이내) 신고
신고방법	◎ 전자우편, 팩스, 인터넷 사이트를 통해 유출사고 신고 및 신고서 제출 ◎ 시간적 여유가 없거나 특별한 사정이 있는 경우 : 전화를 통해 통지 내용 신고 후 유출신고서 제출 가능
신고내용	◎ 기관명, 통지여부, 유출 개인정보 항목·규모, 유출시점·경위 ◎ 유출피해 최소화 대책·조치 및 결과 ◎ 정보주체가 할 수 있는 피해 최소화 방법 및 구제절차 ◎ 담당부서·담당자 연락처 등
신고접수기록	◎ 행정자치부장관 또는 전문기관은 신고 접수 사실 확인 (신고자 전자우편, 팩스)
과태료	◎ 통지하지 아니한 자, 신고하지 아니한 자 3천만원 이하의 과태료 부과
홈페이지 게시	◎ 1만명 이상 개인정보 유출시 개별 통지와 함께 반드시 7일 이상 게시
기타 매체관리	◎ 홈페이지가 없는 경우 등에는 서면·사업장 등 보기 쉬운 장소에 법 제34조제1항 각호의 사항 7일 이상 게시
관련 법령	◎ 법 제34조(개인정보 유출 통지 등) ◎ 법 제75조(과태료) ◎ 시행령 제39조(개인정보 유출신고의 범위 및 기관), 제40조(개인정보 유출통지의 방법 및 절차)

부록

- 01 / 비식별화 조치 방법
- 02 / 프라이버시 보호 모델
- 03 / 참고자료

01 / 비식별화 조치 방법<sup>1)</sup>

비식별화  
처리 원칙

⦿ 원칙적으로 그 자체로 개인을 식별할 수 있는 정보는 삭제

〈 그 자체로 개인을 식별할 수 있는 정보 (예시) 〉

- ① 쉽게 개인을 식별할 수 있는 정보(이름, 전화번호, 주소, 생년월일, 사진 등)
- ② 고유식별정보(주민등록번호, 운전면허번호, 외국인등록번호, 여권번호)
- ③ 생체정보(지문, 홍채, DNA 정보 등)
- ④ 기관, 단체 등의 이용자 계정(등록번호, 계좌번호, 이메일 주소 등)

– 또는 개인을 식별할 수 있는 정보의 삭제처리 대신 (일부)속성 정보로 대체처리 가능

\* 단, 공개하려는 속성정보는 함께 공개되는 정보 및 인터넷 등에 공개되어 있는 정보와 결합하여 개인을 식별할 수 없어야 함

〈개인 식별 정보의  
비식별화 처리 예시〉

주민등록번호	연금 수급액	속성정보	연금 수급액
550001-1000001	100만원	남성(또는 50대)	100만원
450002-1000002	150만원	남성(또는 60대)	150만원
350003-2000003	200만원	여성(또는 70대)	200만원
250004-2000004	250만원	여성(또는 80대)	250만원

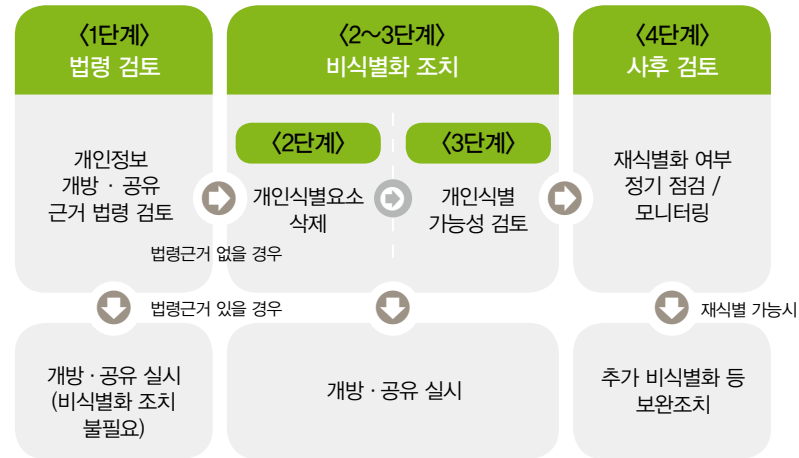
⦿ 삭제 후 남아있는 정보의 추가 가공(삭제 · 변환 등) 등을 통해 제공받는 자가 보유한 정보 및 인터넷 · 언론 등에 공개되어 있는 정보와 쉽게 결합하여 개인을 식별할 수 없도록 조치

※ ‘[참고] 개인식별요소 제거 요령(예시)’ 참조

1) 공공정보 개방 공유에 따른 개인정보 보호 지침(행정자치부, 2013.9.) 제4장 비식별화 조치 방법

## 단계별 조치사항

〈 비식별화를 위한  
단계별 조치 사항 〉



### ◎ 1단계 : 관계법령 검토

관계 법령 또는 정보주체의 동의 등을 통하여 개인정보가 포함된 공공정보의  
개방 · 공유 허용 근거 확인

☞ 합법적 개방 · 공유 처리 근거가 있다면 그에 따라 개인정보가 포함된 공공정  
보의 개방 · 공유 실시

\* 민감정보 또는 고유식별정보의 개방 · 공유를 위해서는 해당 내용에 관하여 구체적 법  
령 근거 또는 (별도) 동의 필요

### ◎ 2단계 : 개인식별요소 제거

개방 · 공유가 허용된 특정 개인정보를 제외한 모든 개인식별요소를 제거

☞ 개인식별요소의 제거는 개인정보 자체를 삭제하거나 개체의 개별 값을 임의의  
참고 값 또는 집단의 값으로 대체하는 등의 방법 활용(p.24 참고)

### ◎ 3단계 : 개인 식별 가능성 검토

제공받는 자가 보유하고 있거나 공개되어 있는 정보와의 결합을 통한 개인 식  
별 가능성 검토

– 통계 · 수학 · 과학적 지식을 보유한 다양한 분야의 전문가들로 하여금 간접 개  
인식별 정보 및 기타 데이터를 통한 재식별 가능성 검토

### 〈참고 : 재식별 가능성이 높은 정보(예시)〉

- 소수 집단에 관한 정보(90대 이상 연령자, 도서산간 거주자, 희귀질병 감염자 정보 등)
- 연속하여 공개되는 패널 데이터 등(분기별 공개하는 환자진료 및 처방에 따른  
회복 관련 정보 등)
- 링크 정보를 가지고 있는 집단에게 정보를 개방 · 공유하는 경우(자동차 번호별  
소유자를 알고 있는 처리자에게 자동차 번호를 제공하는 경우 등)
- 집단과 그 구성원이 알려져 있는 경우로서 동일 속성을 가진 집단에 관한 정보

– 비식별화 처리를 한 공공기관이 비식별화된 정보를 활용하는 경우 비식별화 처  
리 전에 보유한 개인 관련 정보를 활용 · 연계하여 개인을 식별할 수 없도록 내  
부 규정 등을 보완

\* 비식별화된 정보를 비식별화 처리 전에 습득한 개인관련 정보와 매칭하여 사용하는  
경우 개인정보의 목적 외 이용에 해당될 수 있음

### 〈참고 : 재식별 가능성이 낮도록 규정을 보완하는 방법(예시)〉

- 해당 정보에 접근 가능한 사람을 제한
- 데이터의 이용과 관련 데이터의 추가적 공개 등에 관하여 부가적 제한을 둠
- 비식별화된 정보의 적절한 이용 통제를 위한 접근 제한과 같은 절차를 도입
- 가능한 신속하게 데이터를 파기하도록 절차와 수단을 도입

### ◎ 4단계 : 사후검토

시간의 경과에 따라 데이터 분석기술의 진화 및 관련 공개정보가 누적되어 재  
식별 위험이 증가할 수 있으므로 비식별화 기법 및 재식별 가능성에 관한 주기  
적 모니터링 실시

– 재식별이 되는 경우 추가 비식별화 등의 보완 조치 및 향후의 비식별화 처리 기  
법 개선시 반영

※ 비식별화된 정보를 재식별하는 경우 재식별 하는 자가 정보주체의 동의 또는 법령 근  
거 마련 등 필요

[참고]  
개인식별요소 제거 요령  
(예시)

처리기법	주요내용
가명처리 (Pseudonymisation)	개인정보 중 주요 식별요소를 다른 값으로 대체하여 개인식별을 곤란하게 함  (예) 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대 서울 거주, 국제대 재학  * 다른 값으로 대체하는 일정한 규칙이 노출되어 역으로 개인을 쉽게 식별할 수 있어서는 안된다
총계처리 (Aggregation) 또는 평균값 대체 (Replacement)	데이터의 총합 값을 보임으로서 개별 데이터의 값을 보이지 않도록 함  (예) 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팔쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm  * 단, 특정 속성을 지닌 개인으로 구성된 단체의 속성 정보를 공개하는 것은 그 집단에 속한 개인의 정보를 공개하는 것과 마찬가지로의 결과가 나타나므로 그러한 정보는 비식별화 처리로 볼 수 없음 (예) 에이즈 환자 집단 임을 공개하면서 특정인물 '갑'이 그 집단에 속함을 알 수 있도록 표시하는 것은 '갑'이 에이즈 환자임을 공개하는 것과 마찬가지임)
데이터 값 (가치) 삭제 (Data Reduction)	데이터 공유·개방 목적에 따라 데이터 셋에 구성된 값 중에 필요 없는 값 또는 개인식별에 중요한 값을 삭제  (예) 홍길동, 35세, 서울 거주, 한국대 졸업 → 35세, 서울 거주 (예) 주민등록번호 901206-1234567 → 90년대 생, 남자 (예) 개인과 관련된 날짜 정보(자격 취득일자, 합격일 등)는 연 단위로 처리 (예) 연예인·정치인 등의 가족 정보(관계정보), 판례 및 보도 등에 따라 공개되어 있는 사건과 관련되어 있음을 알 수 있는 정보
범주화 (Data Suppression)	데이터의 값을 범주의 값으로 변환하여 명확한 값을 감춤  (예) 홍길동, 35세 → 홍씨, 30-40세
데이터 마스킹 (data masking)	공개된 정보 등과 결합하여 개인을 식별하는데 기여할 확률이 높은 주요 개인식별자가 보이지 않도록 처리하여 개인을 식별하지 못하도록 함  (예) 홍길동, 35세, 서울 거주, 한국대 재학 → 홍**, 35세, 서울 거주, **대학 재학  * 남아 있는 정보 그 자체로 개인을 식별할 수 없어야 하며 인터넷 등에 공개되어 있는 정보 등과 결합하였을 경우에도 개인을 식별할 수 없어야 한다

## 02 / 프라이버시 보호 모델

### 가. 배경

- ◎ 공개된 데이터에 대한 연결공격(linkage attack)을 방어하기 위해 제안된 프라이버시 보호 모델로 2002년 L. Sweeney가 제안한 모델

#### 〈 공개 데이터의 취약점 〉

- ◎ 공개 데이터
  - 공공기관, 의료기관, 민간기업 등에서 공개 또는 배포하는 데이터
  - 데이터 분석가는 이러한 데이터를 분석하여 새로운 부가가치를 창출함
- ◎ 개인정보를 포함한 공개 데이터
  - 많은 공개 데이터는 개인정보를 포함함 (예: <표 1>)
  - 일반적으로 공개 데이터에서 이름, 주민등록번호 등과 같이 개인을 직접적으로 식별하게 하는 속성은 삭제되어 배포됨
  - 그러나 준식별자에 의한 연결 공격이 가능하며, 이로 인해 민감한 정보가 노출되는 프라이버시 문제 발생 가능
- ◎ 연결공격(linkage attack)
  - 공개 데이터간의 결합으로 인해 개인의 민감한 정보가 노출될 수 있음
  - 실제 미국 매사추세츠 주에서 '선거인명부'와 '공개 의료데이터'가 결합하여 개인의 병명이 노출된 사례가 있음
  - 예를 들어, <표 2>의 선거인명부와 <표 1>의 의료데이터가 지역 코드, 연령, 성별에 의해 결합되면, 개인의 민감한 정보인 병명이 드러날 수 있음  
예 : 김민준 (13053, 28, 남자) → 환자 레코드 1번 → 전립선염

〈표 1〉  
공개 의료데이터 사례

	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

	이름	지역 코드	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〉  
4-익명성 모델에 의해  
익명화된 의료데이터 사례  
( $k=4$ )

나. 정의

⦿ 주어진 데이터 집합에서 준식별자 속성값들이 동일한 레코드가 적어도  $k$ 개 존재해야 함

다. 의미

⦿ 데이터 집합의 일부를 수정하여, 모든 레코드가 자기 자신과 동일한(구별되지 않는)  $k-1$ 개 이상의 레코드를 가짐

⦿ 예를 들어, 〈표 1〉의 의료 데이터가 익명화 된 〈표 3〉에서 1~4, 5~8, 9~12 레코드는 서로 구별되지 않음<sup>12)</sup>

	준식별자			민감한 정보
	지역 코드	연령	성별	질병
1	130**	< 30	*	전립선염
2	130**	< 30	*	전립선염
3	130**	< 30	*	고혈압
4	130**	< 30	*	고혈압
5	1485*	> 40	*	위암
6	1485*	> 40	*	전립선염
7	1485*	> 40	*	고혈압
8	1485*	> 40	*	고혈압
9	130**	3*	*	위암
10	130**	3*	*	위암
11	130**	3*	*	위암
12	130**	3*	*	위암

⦿ 따라서, 익명화된 데이터 집합에서는 공격자가 정확히 어떤 레코드가 공격 대상인지 알아낼 수 없도록 하여 “프라이버시 보호”

예) 김민준 → 레코드 1~4 → 전립선염 또는 고혈압

12) ‘\*’ 표시는 임의의 글자를 나타낸다. 가령, 지역코드 ‘130\*\*’은 ‘13000’ ~ ‘13099’ 범위 안에 존재하는 하나의 지역코드 값을 의미한다

- 여기서, 같은 준식별자 속성 값들로 식명화된 레코드들의 모임을 ‘동일 준식별자 속성 값 집합(equivalent class, 이하 동질 집합)’이라고 함
- 예) 레코드1~4, 5~6, 9~12

라. *k*-익명성 검증 방법

- k*-익명성 수준의 판별을 위해서는 우선적으로 <알고리즘 1> 동질집합을 산출 알고리즘을 이용하여 동질집합을 산출해야 함

<알고리즘 1>  
동질 집합 산출을 위한 알고리즘

1:	<i>R</i> 은 익명성을 판별하고자 하는 데이터 레코드들의 집합
2:	<i>EC</i> 는 ‘동질 집합’ <i>ec</i> 들의 집합 (공집합으로 초기화)
3:	for each 레코드 <i>r</i> ∈ <i>R</i> 에 대하여
4:	if <i>r</i> 의 준식별자가 어떤 <i>ec</i> <sub><i>i</i></sub> ( <i>EC</i> 의 원소)의 준식별자와 같다
5:	<i>r</i> 을 <i>ec</i> <sub><i>i</i></sub> 에 삽입
6:	else
7:	<i>EC</i> 에 새로운 <i>ec</i> <sub><i>i</i></sub> 를 생성
8:	<i>r</i> 을 <i>ec</i> <sub><i>i</i></sub> 에 삽입
9:	end if
10:	end for
11:	<i>EC</i> 를 반환

- 이후 <알고리즘 2> *k*-익명성 수준 판별을 위한 알고리즘을 이용하여 각 동질집합이 갖는 레코드의 개수인 *k*값을 계산함
- 동일한 레코드가 적어도 *k*개 존재해야 한다는 정의에 따라, 가장 작은 *k*값이 전체 데이터의 *k*-익명성을 대표함

<알고리즘 2>  
*k*-익명성 수준 판별 알고리즘

1:	<i>EC</i> 는 알고리즘 1을 통해 찾은 모든 ‘동질 집합’들의 집합
2:	<i>k</i> 는 정수, 무한대로 초기화 ( <i>k</i> =∞)
3:	<i>ec</i>  는 <i>ec</i> ( <i>EC</i> 의 원소)의 레코드의 개수
4:	for each <i>ec</i> ∈ <i>EC</i> 에 대하여
5:	if   <i>ec</i>   < <i>k</i>
6:	<i>k</i> =   <i>ec</i>
7:	end if
8:	end for
9:	<i>k</i> 를 반환



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

가. 배경

- ◎ k-익명성에 대한 두 가지 공격, 즉 동질성 공격 및 배경지식에 의한 공격을 방어하기 위해 2006년 A. Machanavajjhala 등이 제안한 프라이버시 보호 모델

〈k-익명성의 취약점〉

- ◎ 데이터가 k-익명화 되었더라도 민감한 정보가 충분히 다양하지 않으면 프라이버시 문제 발생 가능
- ◎ 취약점 1. 동질성 공격 (Homogeneity attack)
  - 데이터 집합에서 동일한 민감한 정보를 이용하여 공격 대상의 민감한 정보를 알아내는 공격
  - 〈표 3〉에서, 레코드 9~12의 민감한 정보는 모두 ‘위암’이므로, k-익명성 모델이 적용되었음에도 불구하고 민감한 정보가 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 할 때, 준식별자 조합(13068, 29, 여)에 따라 ‘이지민’은 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 추정 가능함
- ◎ k-익명성의 취약점의 원인
  - 다양성의 부족 (lack of diversity)
    - 익명화할 때 민감한 정보의 다양성을 고려하지 않음
    - 동일한 민감한 정보를 가진 (다양하지 않은) 레코드가 익명화되어 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 취약함

나. 정의

- ◎ l-다양성 (l-diversity)의 정의

- 주어진 데이터 집합에서 함께 익명화되는 레코드들은 (동질 집합에서) 적어도 l개의 서로 다른 민감한 정보를 가져야 함

다. 의미

- ◎ 익명화 과정에서, 충분히 다양한(l개 이상) 서로 다른 민감한 정보를 갖도록 동질 집합을 구성
- ◎ 민감한 정보가 충분한 다양성을 가지므로, 다양성의 부족으로 인한 공격에 방어가 가능하고, 배경지식으로 인한 공격에도 일정 수준의 방어력을 가짐
- ◎ 예를 들어, 〈표 4〉에서 모든 동질 집합은 3-다양성을 통해 익명화 되어 3개 이상의 서로 다른 민감한 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〉  
3-다양성 모델에 의해  
익명화된 의료데이터의 예  
(l = 3)

	준식별자			민감한 정보
	지역 코드	연령	성별	질병
1	1305*	≤ 40	*	전립선염
4	1305*	≤ 40	*	고혈압
9	1305*	≤ 40	*	위암
10	1305*	≤ 40	*	위암
5	1485*	> 40	*	위암
6	1485*	> 40	*	전립선염
7	1485*	> 40	*	고혈압
8	1485*	> 40	*	고혈압
2	1306*	≤ 40	*	전립선염
3	1306*	≤ 40	*	고혈압
11	1306*	≤ 40	*	위암
12	1306*	≤ 40	*	위암

〈알고리즘 3〉  
*l*-다양성 수준 판별 알고리즘

1:	<i>EC</i> 는 알고리즘 1을 통해 찾은 모든 ‘동질 집합’들의 집합
2:	<i>l</i> 은 정수, 무한대로 초기화 ( $l = \infty$ )
3:	$ e_{CS} $ 는 <i>ec</i> ( <i>EC</i> 의 원소) 의 서로 다른 민감한 정보의 개수
4:	for each $ec \in EC$ 에 대하여
5:	if $ e_{CS}  < l$
6:	$l =  e_{CS} $
7:	end if
8:	end for
9:	<i>l</i> 을 반환

라. *l*-다양성 검증 방법

- ◎ 각 동질 집합이 갖는 서로 다른 민감한 정보의 개수(*l*-다양성)를 탐색하기 위해 〈알고리즘 3〉을 이용하여 *l*-다양성 값을 산정
- ◎ 서로 다른 민감한 정보가 적어도 *l*개 존재해야 한다는 정의에 따라, 가장 작은 *l*값이 전체 데이터의 *l*-다양성을 대표함

마. 정보 유용성 (data utility)

- ◎ 정보 손실 (information loss)
  - 데이터를 익명화하게 되면, 프라이버시를 보호하는 대가로 일정량의 정보가 필연적으로 손실됨
  - 〈표 3〉, 〈표 4〉의 ‘\*’ 나 ‘)’ 등과 같이 일부 정보를 지우거나, 원본 값을 구간 값 또는 더 상위 개념의 값으로 일반화(generalization)하는 과정에서 원본 데이터의 정보가 일부 손실됨
  - 예: 연령 ‘23’을 구간 값 (20 ~ 25)으로 익명화, 성별 ‘여성’을 ‘\*’ (남성/여성을 구분 없이 모두 지칭함)로 익명화
- ◎ 프라이버시 보호-정보 손실 간의 관계
  - *k*-익명성과 *l*-다양성 모델에서 *k*, *l*값은 곧 프라이버시 보호 수준을 의미
  - *k*, *l*값이 증가할수록 프라이버시 보호 수준은 증가하지만, 이에 따라 많은 정보가 손실되므로 정보 유용성은 감소하는 경향을 보임

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

가. 배경

- ◎ *l*-다양성의 취약점(쏠림 공격, 유사성 공격)을 보완하기 위해 2007년 N. Li 등이 제안한 프라이버시 보호 모델

〈*l*-다양성의 취약점〉

- ◎ 쏠림 공격 (skewness attack)
  - 민감한 정보가 특정한 값에 쏠려 있을 경우, *l*-다양성 모델이 프라이버시를 보호하지 못함
  - 쏠림 공격의 예
    - 임의의 ‘동질 집합’이 99개의 ‘AIDS 양성’ 레코드와 1개의 ‘AIDS 음성’ 레코드로 구성되어 있다고 가정
    - 민감한 정보는 2-다양성을 만족하지만, 공격자는 공격 대상이 99%의 확률로 ‘AIDS 양성’이라는 것을 알 수 있음
- ◎ 유사성 공격 (similarity attack)
  - 익명화된 레코드의 민감한 정보가 서로 비슷하다면, *l*-다양성 모델을 통해 익명화된다고 할지라도 프라이버시가 노출될 수 있음
  - 유사성 공격의 예
    - 〈표 5〉는 3-다양성 모델을 통해 익명화된 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 ~ 50 백만원)

〈표 5〉  
3-다양성 모델에 의해  
익명화되었지만 유사성 공격에  
취약한 사례  
( $l=3$ )

	준식별자		민감한 정보	
	지역 코드	연령	급여 (백만원)	질병
1	476**	2*	30	위궤양
2	476**	2*	40	급성 위염
3	476**	2*	50	만성 위염
4	4790**	$\geq 40$	60	급성 위염
5	4790**	$\geq 40$	110	감기
6	4790**	$\geq 40$	80	기관지염
7	476**	3*	70	기관지염
8	476**	3*	90	폐렴
9	476**	3*	100	만성 위염

나. 정의

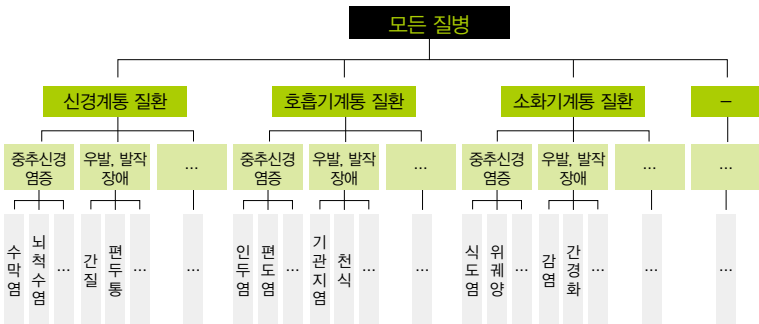
◎ 동질 집합에서 민감한 정보의 분포와, 전체 데이터 집합에서 민감한 정보의 분포가 이하의 차이를 보여야 하는 것

다. 의미

- ◎ 민감한 정보의 분포를 고려
  - 각 동질 집합에서 ‘민감한 정보의 분포’가 전체 데이터 집합의 그것과 비교하여 너무 특이하지 않도록 함
  - 〈표 5〉에서, 전체적인 급여 값의 분포는 30 ~ 110
  - 이 때, 레코드 1, 2, 3이 속한 동질 집합에서 급여의 분포는 30 ~ 50으로, 이는 전체 급여 값의 분포(30 ~ 110)와 비교할 때 극히 일부 → 공격자는 근사적인 급여 값을 알 수 있음
  - $t$ -근접성 모델은 이러한 동질 집합과 전체 데이터 집합 사이 분포의 과도한 차이를  $t$ -다양성 모델의 취약점으로 규정함

- ◎ ‘민감한 정보의 분포’를 조정하여 프라이버시를 보호
  - 민감한 정보가 특정 값으로 쏠리거나, 유사한 값들이 뭉치는 경우를 방지
  - 〈표 6〉에서  $t$ -근접성 모델에 따라 레코드 1,3,8이 하나의 동질 집합으로 익명화됨
  - 이 때, 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, 폐렴)으로 병명이 서로 다르면서, 질병이 ‘위’와 관련된 것 이외에 ‘폐’와 관계된 것이라서, 특정 부위의 질병임을 유추하기 어려움
  - 따라서 〈표 5〉의 경우와 비교하여, 공격자가 공격 대상의 민감한 정보를 추측 하기가 더욱 어려워짐
- ◎ 민감한 정보의 의미(semantics)까지 파악하는 프라이버시 모델
  - 민감한 정보의 의미를 고려하여 값의 분포를 계산함
  - 연속 속성(continuous attribute)의 경우 숫자 값을 통해 의미가 유사한 정도를 파악 (예: 급여)
  - 범주 속성(categorical attribute)의 경우 분류 트리(taxonomy tree, 〈그림 1〉 참고)를 이용해 의미가 유사한 정도를 파악 (예: 질병)

〈그림 1〉  
질병에 대한 분류 트리



〈표 6〉  
*t*-근접성 모델에 의해  
익명화된 데이터 사례

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

라. *t*-근접성 검증 방법

- ◎ 각 동질 집합의 ‘민감한 정보의 분포’와, 전체 데이터 집합에서의 ‘민감한 정보의 분포’간의 거리(차이)를 계산하기 위해 〈알고리즘 4〉를 이용해서 *t*-근접성 값 산출
  - 임의의 동질 집합에서 민감한 정보의 분포 *P<sub>ec</sub>*, 전체 데이터에서 민감한 정보의 분포 *Q*
  - 모든 동질 집합에 대하여, *P<sub>ec</sub>*와 *Q*의 차이(*D[P<sub>ec</sub>, Q]*)<sup>13)</sup> *t*를 계산
  - 분포의 차이가 *t* 이하여야 한다는 정의에 따라, 가장 큰 분포의 차이 값이 전체 데이터의 *t*-근접성을 대표함

〈알고리즘 4〉  
*t*-근접성 수준 판별  
알고리즘

1:	<i>EC</i> 는 알고리즘 1을 통해 찾은 모든 ‘동질 집합’들의 집합
2:	<i>t</i> 는 실수, 0으로 초기화 ( <i>t</i> =0)
3:	<i>P<sub>ec</sub></i> 는 임의의 <i>ec</i> 에서 민감한 정보의 분포
4:	<i>Q</i> 는 입력 데이터 집합 전체에서 민감한 정보의 분포
5:	for each <i>ec</i> ∈ <i>EC</i> 에 대하여
6:	if <i>t</i> < <i>D[P<sub>ec</sub>, Q]</i> / * <i>D[A, B]</i> 는 분포 A, B간의 거리*/
7:	<i>t</i> = <i>D[P<sub>ec</sub>, Q]</i>
8:	end if
9:	end for
10:	<i>t</i> 를 반환

13) 분포의 차이 *D[P<sub>ec</sub>, Q]*는 EMD (Earth Mover Distance) 식을 통해 계산. *t*-근접성 논문인 “*t*-Closeness : Privacy Beyond -Anonymity and -Diversity”(2007.4.) 참조

**ε-차분프라이버시**  
**(ε-differential**  
**privacy) :**  
*k*-익명성과  
*l*-다양성의  
취약점을 보완한  
프라이버시  
보호 모델

가. 배경

◎ *k*-익명성과 *l*-다양성의 취약점(최소성 공격)을 보완하기 위해 Dwork, C. 가 제  
안한 프라이버시 보호 모델

〈 *k*-익명성, *l*-다양성의 취약점

- ◎ 최소성 공격 (Minimality Attack)
  - *k*-익명성, *l*-다양성 프라이버시 모델을 이해하고 있는 공격자에 의해 민감한 정보가 유출될 수 있음
  - 최소성 공격의 예
    - 공격자는 2-익명성과 2-다양성을 만족하는 공개된 데이터베이스 테이블 〈표7〉에서 Ben의 준식별자는 q1이고, 모든 가능한 준식별자는 q1 또는 q2 뿐이라는 사실을 인지한다고 가정
    - 공격자는 2-익명성과 2-다양성을 만족하는 공개된 데이터베이스 테이블을 통해 Ben이 HIV에 걸렸다는 것을 알아낼 수 있음
    - 공격자는 〈표 7〉의 하단 2개 레코드는 (q2, NO)라고 공개되어 있기 때문에 상단 4개의 레코드만 변화하면서, 〈표 8〉과 같이 모든 가능한 경우의 테이블을 도출할 수 있음
    - 공격자는 〈표 8〉에서 중복이 없는 9개 테이블(1, 2[3], 4, 5[9], 6[7,10,11], 8[12], 13, 14[15], 16)을 도출 가능
    - 〈표 8〉의 9개 테이블 중 2-익명성과 2-다양성을 만족하는 7개 테이블은 그대로 공개 가능하지만, 2-익명성과 2-다양성을 만족하지 못하는 2개 테이블(4, 8[12])에서 q1은 HIV가 있는 것을 발견할 수 있어, 공격자는 〈표 9〉와 같이 Ben이 HIV에 감염된 것을 확인할 수 있음

〈표 7〉

원본테이블		익명화 테이블	
QID	질병	QID	질병
q1	HIV	Q	HIV
q1	HIV	Q	HIV
q2	NO	Q	NO
q2	NO	Q	NO
q2	NO	q2	NO
q2	NO	q2	NO
		2-다양성 만족	

〈가정〉

1. 공격자는 QID의 q1이 Ben인 것을 알고 있음
2. Q는 q1 또는 q2 인 것을 알고 있음
3. 익명화 테이블은 2-익명성과 2-다양성을 만족하는 것을 알고 있음

〈표 8〉

추측테이블1		추측테이블2		추측테이블3		추측테이블4	
QID	질병	QID	질병	QID	질병	QID	질병
q1	HIV	q1	HIV	q1	HIV	q1	HIV
q1	HIV	q1	HIV	q1	HIV	q1	HIV
q1	NO	q1	NO	q2	NO	q2	NO
q1	NO	q2	NO	q1	NO	q2	NO
q2	NO	q2	NO	q2	NO	q2	NO
q2	NO	q2	NO	q2	NO	q2	NO
II		II		II		II	
추측테이블5		추측테이블6		추측테이블7		추측테이블8	
QID	질병	QID	질병	QID	질병	QID	질병
q1	HIV	q1	HIV	q1	HIV	q1	HIV
q2	HIV	q2	HIV	q2	HIV	q2	HIV
q1	NO	q1	NO	q2	NO	q2	NO
q1	NO	q2	NO	q1	NO	q2	NO
q2	NO	q2	NO	q2	NO	q2	NO
q2	NO	q2	NO	q2	NO	q2	NO
II		II		II		II	
추측테이블9		추측테이블10		추측테이블11		추측테이블12	
QID	질병	QID	질병	QID	질병	QID	질병
q2	HIV	q2	HIV	q2	HIV	q2	HIV
q1	HIV	q1	HIV	q1	HIV	q1	HIV
q1	NO	q1	NO	q2	NO	q2	NO
q1	NO	q2	NO	q1	NO	q2	NO
q2	NO	q2	NO	q2	NO	q2	NO
q2	NO	q2	NO	q2	NO	q2	NO
II		II		II		II	
추측테이블13		추측테이블14		추측테이블15		추측테이블16	
QID	질병	QID	질병	QID	질병	QID	질병
q2	HIV	q2	HIV	q2	HIV	q2	HIV
q2	HIV	q2	HIV	q2	HIV	q2	HIV
q1	NO	q1	NO	q2	NO	q2	NO
q1	NO	q2	NO	q1	NO	q2	NO
q2	NO	q2	NO	q2	NO	q2	NO
q2	NO	q2	NO	q2	NO	q2	NO

〈표 9〉

원본테이블		익명화 테이블		⇒	추측테이블4		추측테이블8	
QID	질병	QID	질병		QID	질병	QID	질병
q1	HIV	Q	HIV		q1	HIV	q1	HIV
q1	HIV	Q	HIV		q1	HIV	q2	HIV
q2	NO	Q	NO		q2	NO	q2	NO
q2	NO	Q	NO		q2	NO	q2	NO
q2	NO	q2	NO		q2	NO	q2	NO
q2	NO	q2	NO		q2	NO	q2	NO

2-다양성 만족

\*두 테이블을 비교해서 q1(Ben)이 HIV에 걸렸다는 것을 도출 가능

나. 정의

- ◎ 데이터베이스 레코드들에 확률적으로 변조를 가하여 레코드에 대한 식별 가능성을 제한해 민감한 정보를 높은 확률로 추측하지 못하도록 막기 위한 모델

다. 의미

- ◎ 공개되는 테이블에 민감정보가 나올 확률을 고려
  - 각각의 모든 데이터베이스 레코드에 대해서 그 레코드가 원본 테이블에 존재 하든, 존재하지 않든 공개된 테이블에 나올 확률을 거의 같도록 조정해서
  - 각각의 레코드에 대한 민감한 정보를 찾더라도 100% 확신하지 못하도록 하여 민감한 정보를 보호하는 모델
- ◎ 데이터의 정확도는 어느 정도 희생
  - 데이터의 정확도는 어느 정도 희생해야 하며, 민감한 정보를 100% 확신하는 것 만을 막을 수 있을 뿐 민감한 정보를 가질 확률이 높다는 것까지 공개되는 것 을 막을 수 없는 한계가 있음

03 / 참고자료

행정자치부, “공공정보 개방 공유에 따른 개인정보 보호지침”, 2013.9.

행정자치부, “정부 3.0 기본계획”, 2013.6.

일본 개인데이터 위원회<sup>14)</sup> 기술검토 워킹그룹, “기술검토 워킹그룹 보고서”, 2013.12.10.

최대선 등, “소셜네트워크서비스 개인정보 노출 실태 분석”, 정보보호학회논문지, 2013.10.

Cynthia Dwork, “Differential Privacy”, ICALP, 2006.

Data Protection Working Party, “Opinion 05/2014 on Anonymization Techniques”, EU, 2014.4.10.

Department of Health & Human Services, “Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule”, USA, 2012.11.26.

EU, “Opinion 05/2014 on Anonymization Techniques”, EU Article 29 Data Protection Working Party, 2014.5.

Federal Trade Commission, “Data Brokers : A Call for Transparency and Account ability”, USA, 2014.5.

Florian Kohlmayer, Fabian Prasser, “ARX – Powerful Data Anonymization”, <http://arx.deidentifier.org/>, 2013.

14) 일본 총리산하 IT종합전략본부가 설립한 “개인데이터 위원회”는 일본의 개인정보보호 관련 연구계, 법조계, 산업계, 소비자 등의 전문가로 구성, 소비자청, 내각 비서실 산하 ICT 국가전략실, 총무성, 경제산업성이 지원

Health System Use Technical Advisory Committee, “‘Best Practice’ Guidelines for Managing the Disclosure of De-Identification Health Information”, CANADA, 2010.10.

ICO, “Anonymisation : managing data protection risk code of practice”, UK, 2012.11.

L. Sweeney, “ $k$ -anonymity: A model for protecting privacy”, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002.

Machanavajjhala, Ashwin, et al., “ $l$ -diversity: Privacy beyond  $k$ -anonymity”, ACM Transactions on Knowledge Discovery from Data 1.1, 2007.

Ninghui Le, et al., “ $t$ -Closeness : Privacy Beyond  $k$ -Anonymity and  $l$ -Diversity”, ICDE. Vol.7, 2007.4.

Paul Ohm, “Broken Promises Of Privacy : Responding To The Surprising Failure Of Anonymization”, UCLA Law Review 1701, 2010.

Wong, Raymond Chi-Wing, et al., “Minimality attack in privacy preserving data publishing”, Proceedings of the 33rd international conference on Very large data bases(VLDB) Endowment, 2007.

## 04 / 참여 전문가

한국정보화진흥원	개인정보보호단	단장	강종관
	개인정보보호기획부	부장	황성욱
	개인정보보호기획부	수석연구원	이재근
	개인정보보호기획부	수석연구원	김현진
	개인정보보호기획부	책임연구원	정영수
	개인정보보호기획부	책임연구원	김나현
외부전문가	서울대학교	교수	심규석
	서울대학교	교수	고학수
	고려대학교	교수	정연돈

## 개인정보 비식별화에 대한 적정성 자율평가 안내서

---

인 쇄 일 | 2014년 12월

발 행 일 | 2014년 12월

---

발 행 처 | 한국정보화진흥원

발 행 인 | 한국정보화진흥원

기획편집 | 한국정보화진흥원

주 소 | 서울특별시 중구 청계천로 14(무교동 77번지)  
NIA빌딩 10층

전 화 | 02-2131-0111~2

---

디 자 인 | 컬러커뮤니케이션즈(02-333-6555)

---