|  |  |  |
| --- | --- | --- |
| **KSKSKSKS**  **KSKSKSK**  **KSKSKS**  **KSKSK**  **KSKS**  **KSK**  **KS** | | KS X ITUTX1148 |
|  | **통신 서비스 제공자를 위한**  **비식별 프로세스 프레임워크**  KS X ITUTX 1148:2020 | |
| **방 송 통 신 표 준 심 의 회**  **2022년 12월 28일 제정** | | |

**심 의 : 정보보호 기술심의회**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | | 성명 | |  | | 근 무 처 | |  | | 직위 | |  | |
| (회 장) | |  | | 염흥열 | |  | | 순천향대학교 | |  | | 교수 | |  | |
| (위 원) | |  | | 권대성 | |  | | 국가보안기술연구소 | |  | | 책임연구원 | |  | |
|  | |  | | 김재성 | |  | | 한국인터넷진흥원 | |  | | 센터장 | |  | |
|  | |  | | 나재훈 | |  | | 한국전자통신연구원 | |  | | 전문위원 | |  | |
|  | |  | | 이창훈 | |  | | 서울과학기술대학교 | |  | | 교수 | |  | |
|  | |  | | 전명근 | |  | | 충북대학교 | |  | | 교수 | |  | |
|  | |  | | 진승헌 | |  | | 한국전자통신연구원 | |  | | 책임연구원 | |  | |
| (간 사) | |  | | 최영선 | |  | | 국립전파연구원 | |  | | 과장 | |  | |

**원안작성협력 : 정보보호 전문위원회**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 성명 |  | 근 무 처 |  | 직위 |  |
| (대표전문위원) |  | 권혁찬 |  | 한국전자통신연구원 |  | 책임 |  |
| (위 원) |  | 김종성 |  | 국민대학교 |  | 교수 |  |
|  |  | 김태경 |  | 명지전문대학교 |  | 교수 |  |
|  |  | 문호건 |  | 인제대학교 |  | 교수 |  |
|  |  | 윤석웅 |  | 한국인터넷진흥원 |  | 팀장 |  |
|  |  | 이창훈 |  | 서울과학기술대학교 |  | 교수 |  |
|  |  | 전동훈 |  | ㈜슈프리마비브이 |  | 수석 |  |
|  |  | 채수영 |  | 국가보안기술연구소 |  | 책임 |  |
|  |  | 홍득조 |  | 전북대학교 |  | 교수 |  |
| (간 사) |  | 박복룡 |  | 국립전파연구원 |  | 주무관 |  |

표준열람 : 국립전파연구원(http://www.rra.go.kr)

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

제 정 자：방송통신표준심의회 위원장 담당부처：과학기술정보통신부 국립전파연구원

제 정：2022년 12월 28일

심 의：방송통신표준심의회 정보보호 기술심의회

원안작성협력：정보보호 전문위원회

━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

이 표준에 대한 의견 또는 질문은 국립전파연구원 웹사이트를 이용하여 주십시오.

이 표준은 방송통신표준화지침 제18조의 규정에 따라 매 5년마다 방송통신표준심의회에서 심의되어 확인, 개정 또는 폐지됩니다.

목 차

[머 리 말 ii](#_Toc77594649)

[개 요 iii](#_Toc77594650)

[1 적용범위 1](#_Toc77594651)

[2 인용표준 1](#_Toc77594652)

[3 용어와 정의 1](#_Toc77594653)

[4 약어 5](#_Toc77594655)

[5 비식별 프로세스 개요 5](#_Toc77594656)

[5.1 데이터 생명주기 모델 및 비식별 단계 6](#_Toc77594657)

[5.2 비식별 고려 사항 7](#_Toc77594658)

[6 비식별 프로세스 프레임워크 8](#_Toc77594659)

[6.1 1단계 ― 사전 검토 9](#_Toc77594660)

[6.2 2단계 ― 비식별 적용 10](#_Toc77594661)

[6.3 3단계 ― 비식별 프로세스의 적정성 평가 11](#_Toc77594662)

[6.4 4단계 ― 사후 관리 11](#_Toc77594663)

[7 비식별 데이터의 유용성 12](#_Toc77594664)

[7.1 비식별 데이터 단계 12](#_Toc77594665)

[7.2 데이터 공개 모델 14](#_Toc77594666)

[7.3 데이터 공개 모델 및 데이터 단계 간 상관관계 16](#_Toc77594667)

[부속서 A (규정) 적정성 평가 절차 17](#_Toc77594668)

[부속서 B (규정) 비정형 데이터에 대한 비식별 접근 방식 20](#_Toc77594674)

[부속서 C (참조) 일반적인 비식별 기술의 예시 22](#_Toc77594675)

[부속서 D (참조) 비식별 프로세스 접근 방식 24](#_Toc77594683)

[참고문헌 27](#_Toc77594686)

[KS X ITUTX 1148:2020 해 설 28](#_Toc77594687)

머 리 말

이 표준은 방송통신발전기본법에 따라 방송통신표준심의회의 심의를 거쳐 제정한 방송통신표준이다.

이 표준은 저작권법의 보호 대상이 되고 있는 저작물이다.

이 표준의 일부가 기술적 성질을 가진 특허권, 출원공개 후의 특허출원, 실용신안권 또는 출원공개 후의 실용신안등록출원에 저촉될 가능성이 있다는 것에 주의를 환기한다. 관계 중앙행정기관의 장과 방송통신표준심의회는 이러한 기술적 성질을 가진 특허권, 출원공개 후의 특허출원, 실용신안권 또는 출원공개 후의 실용신안등록출원에 관계되는 확인에 대하여 책임을 지지 않는다.

개 요

이 표준은 2020년 제1판으로 발행된 ITU-T X.1148 Framework of de-identification process for telecommunication service providers를 기초로 기술적 내용 및 대응국제표준의 구성을 변경하지 않고 작성한 방송통신표준이다.

통신 조직은 개인정보를 포함하여 개인에 관한 데이터를 수집, 관리, 사용 및 공유한다. 결과적으로, 개인정보 보호를 위해 데이터 비식별 기술을 활용하고 있다. 본 표준은 운영 단계와 함께 비식별 프로세스 프레임워크를 설명하며, 데이터 생명주기 모델과 이해 당사자 역할을 기반으로 통신 서비스 제공자를 위한 비식별 프로세스의 데이터 공개 모델과 데이터 단계를 규정한다.

권고안 ITU-T X.1148은 데이터 생명주기 모델을 기반으로 비식별 프로세스에 대한 개요를 제공하며, 운영 단계의 비식별 프로세스 프레임워크 및 비식별 프로세스의 이해 당사자를 규정한다. 또한 비식별 프로세스의 데이터 공개 모델 및 데이터 단계를 설명하며, 부속서와 부록에서는 다양한 비식별 접근 방식과 보기를 제공한다.

**방송통신표준**

**KS X ITUTX1148:2020**

|  |
| --- |
| **통신 서비스 제공자를 위한**  **비식별 프로세스 프레임워크** |

|  |
| --- |
| Framework of de-identification process  for telecommunication service providers |

# 적용범위

이 표준은 데이터 생명주기 모델을 기반으로 비식별 프로세스에 대한 개요를 제공하며, 운영 단계의 비식별 프로세스 프레임워크 및 비식별 프로세스의 이해 당사자를 규정한다. 또한 비식별 프로세스의 데이터 공개 모델 및 데이터 단계를 설명하며, 부속서와 부록에서는 다양한 비식별 접근 방식과 보기를 제공한다.

이 표준에서 정의하고 있는 비식별 프로세스는 국내 개인정보보호법에서 정의하고 있는 가명 처리를 위한 기술적 수단으로 활용될 수 있다.

# 인용표준

해당 사항 없음.

# 용어와 정의

이 표준의 목적을 위하여 다음의 용어와 정의 및 약어를 적용한다.

## 총계 데이터(aggregated data) 해당 그룹의 통계 속성 모음과 같은 데이터 주체 그룹을 나타내는 데이터

[출처: ISO/IEC 20889]

**익명화**(anonymization)   
개인정보 처리자(PII controller) 단독으로 또는 다른 당사자와 협조해 개인정보(PII)를 직접 또는 간접적으로 개인정보 주체와 더 이상 식별할 수 없도록 비가역적으로 변경하는 과정

[출처: ISO/IEC 29100]

## 3.3 속성(attribute) 내재적 특징

[출처: ISO/IEC 20889]

## 3.4 데이터셋(dataset) 데이터 모음

[출처: ISO/IEC 20889]

## 3.5 비식별(de-identification) 식별 데이터와 데이터 주체 사이의 관계를 제거하는 절차를 표현하는 일반적인 용어

[출처: ISO 25237]

## 3.6 비식별 프로세스(de-identification process) 식별 속성 집합 및 데이터 주체 간의 연관성을 삭제하는 프로세스

[출처: ISO/IEC 20889]

## 3.7 비식별 기술(de-identification technique) 정보 및 개별 데이터 주체 간의 연관 범위를 줄이기 위해 데이터셋을 변환하는 방법

[출처: ISO/IEC 20889]

## 3.8

## 비식별 데이터셋(de-identified dataset) 비식별 프로세스를 적용한 데이터셋

[출처: ISO/IEC 20889]

## 3.9 비식별 정보(de-identified information) 잔여 정보로 개인을 식별할 수 없도록 개인정보가 삭제되었거나 가려졌기 때문에 개인을 식별하는 데 사용될 수 있다는 합리적 근거가 없는 레코드

[출처: NISTIR 8053]

## 3.10 차등 프라이버시(differential privacy) 특정 데이터 주체의 입력 데이터셋 포함 여부와 관계없이 통계 분석 결과에 따른 확률 분포에서 최대 지정 값의 차이를 생성하는 정규 프라이버시 보호 모델

**비고** 차등 프라이버시는 다음을 제공

a) 개인정보 보호로 간주되는 통계 분석의 결과에서, 원본 데이터셋 분석 결과와 데이터 셋의 데이터 주체가 추가 또는 삭제된 후 얻은 결과를 구별할 수 없다고 단정하는 개인정보 보호의 수학적 정의

b) 누적된 개인정보 손실을 모니터링하고 손실 한도에 대한 상한(또는 예산)을 설정할 수 있는 개인정보 보호 척도이며, 공식적인 정의는 다음과 같음. 은 양의 실수, M은 데이터셋을 입력으로 사용하는 무작위 알고리즘이며, 단일 요소(하나의 데이터 주체를 가지는 데이터)가 다른 모든 데이터셋 D1 및 D2에 대해 M, mml\_m1 범위의 모든 부분집합 S에서 알고리즘에 사용되는 임의성의 확률을 따를 경우, 알고리즘 M은 -차등 프라이버시를 보장

[출처: ISO/IEC 20889]

## 3.11 식별자(identifier) 특정 운영 상황에서 데이터 주체를 유일하게 식별할 수 있는 데이터셋의 속성 집합

[출처: ISO/IEC 20889]

## 3.12 식별 속성(identifying attribute) 특정 운영 상황에서 데이터 주체를 식별하는 데 유일하게 기여하는 데이터셋의 속성

[출처: ISO/IEC 20889]

## 3.13 프라이버시 이해 당사자(privacy stakeholder) 개인정보 처리와 연관된 활동 또는 결정에 영향을 주거나, 영향 받거나, 영향받는다고 스스로 인정하는 자연인 또는 법인, 공공 기관, 기관, 다른 기구

[출처: ISO/IEC 29100]

## 3.14 가명화(pseudonymization) 식별 정보를 가명으로 변환하여 개인정보에 적용하는 과정

[출처: ISO/IEC 20889]

## 3.15 준식별자(quasi-identifier) 데이터셋의 다른 속성과 결합해 데이터 주체를 개별화할 수 있는 데이터셋의 속성

[출처: ISO/IEC 20889]

## 3.16 레코드(record) 단일 데이터 주체와 관련된 속성 집합

[출처: ISO/IEC 20889]

## 3.17 재식별(re-identification) 비식별된 데이터셋의 데이터를 원본 데이터 주체와 연결하는 프로세스

**비고** 데이터셋에 특정 데이터 주체를 설정하는 프로세스를 포함

[출처: ISO/IEC 20889]

## 3.18 개별화(single out) 데이터 주체를 유일하게 식별하는 특성 집합을 식별해 데이터셋에서 데이터 주체에 속하는 레코드를분리하는 처리절차

[출처: ISO/IEC 20889]

## 3.19 제3자(third party) 개인정보 처리자 또는 개인정보 프로세서의 직접 통제하에 데이터를 처리하도록 인가받은 개인정보주체, 개인정보 처리자와 개인정보 프로세서 그리고 자연인 외 프라이버시 이해 당사자

[출처: ISO/IEC 29100]

## 3.20 신뢰할 수 있는 제3자(trusted third party) 보안 활동과 관련하여 다른 주체가 신뢰하는 보안 기관 또는 그 대리인

[출처: ISO/IEC 18014-1]

## 3.21 k-익명성(k-anonymity) 데이터셋에서 각 식별자에 대해 k개 이상의 레코드를 포함하는 등가 클래스가 있는 정규 프라이버시 보호 모델

[출처: ISO/IEC 20889]

## 3.22 l-다양성(l-diversity) 선택 속성에 대해 각 등가 클래스가 l개 이상의 값을 갖게 하는 정규 프라이버시 보호 모델

**비고** **l**-다양성은 선택 속성의 등가 클래스에서 공유된 값의 다양성을 보장하는 하한값 **l**을 제공하는 데이터 셋의 속성

[출처: ISO/IEC 20889]

## 3.23 t-근접성(t-closeness) 등가 클래스에서 선택 속성의 분포와 전체 데이터 집합에서 속성의 분포 간 간격이 임계값 t 이하가되게 하는 정규 프라이버시 보호 모델

**비고** 이 속성을 포함한 모든 등가 클래스가 **t**-근접성을 갖는 경우, 데이터 집합은 선택 속성과 관련하여  **t**-근접성을 보유

[출처: ISO/IEC 20889]

## 3.24 데이터 처리자(data controller) 개인의 목적으로 데이터를 사용하는 자연인 외에 데이터 처리 목적과 도구를 결정하는 이해 당사자(또는 프라이버시 이해 당사자)

## 3.25 데이터 프로세서(data processor) 데이터 처리자를 대신해서 데이터 처리자의 지시에 따라 데이터를 처리하는 이해 당사자

## 3.26 개인정보 보호 책임자(data protection officer) 개인정보 처리자(PII controller)가 개인정보 보호 법/규제 요구사항을 독립적인 방식으로 준수하기 위해 지정한 사람

**비고** 개인정보 처리자는 데이터 처리자와 동의어

## 3.27 데이터 주체(data subject) 데이터와 관련된 주체

**비고** 개인정보 주체(PII principal), 데이터 주체(data principal)의 동의어

## 3.28 프로세스(process) 정보 또는 데이터와 관련하여, 정보 또는 데이터를 취득, 기록 또는 보유하거나 정보 또는 데이터에 다음과 같은 작업을 수행하는 것을 의미

— 정보 또는 데이터의 구성, 조정 또는 변경

— 정보 또는 데이터의 검색, 검토 또는 사용

— 전송, 유포 또는 기타 방법에 따른 정보 또는 데이터 공개

— 정보 또는 데이터의 배열, 조합, 차단, 삭제 또는 파기

# 약어

|  |  |
| --- | --- |
| DP | 차등 프라이버시(Differential Privacy) |
| DPO | 개인정보 보호 책임자(Data Protection Officer) |
| PII | 개인정보(Personally Identifiable Information) |
| TTP | 신뢰할 수 있는 제3자(Trusted Third Party) |

# 비식별 프로세스 개요

비식별 프로세스의 목적은 데이터 주체의 기밀성을 보호하는 것이다. 따라서 데이터 분석가는 데이터에 개인정보(PII)가 포함되어 있을 수 있으므로 유의미한 정보 추출을 위한 데이터 분석 전과 후의 보안을 고려해야 한다.

본 절은 데이터 분석 환경, 데이터 생명주기 모델, 비식별 프로세스에서 주체의 역할 및 기타 비식별 프로세스의 고려 사항을 정의한다.

## 데이터 생명주기 모델 및 비식별 단계

일반적으로 조직은 개인정보 보호 및 보안 목표에 따른 비식별 목표를 설정한다. 본 절에서는 데이터 생명주기를 정의하며, 이 데이터 생명주기 모델을 기반으로 한 비식별 프로세스의 고려 시점을 설명한다.

데이터 생명주기의 개념은 재식별 가능성 분석을 기반으로 적절한 통제를 선택하는 데 사용된다. 본 표준의 **5.1.1** 에서 **5.1.5**에 서술된 바와 같이 데이터 생명주기를 정의한다.

**그림 1**은 데이터 생명주기 모델의 비식별 프로세스에 대한 개요이다.

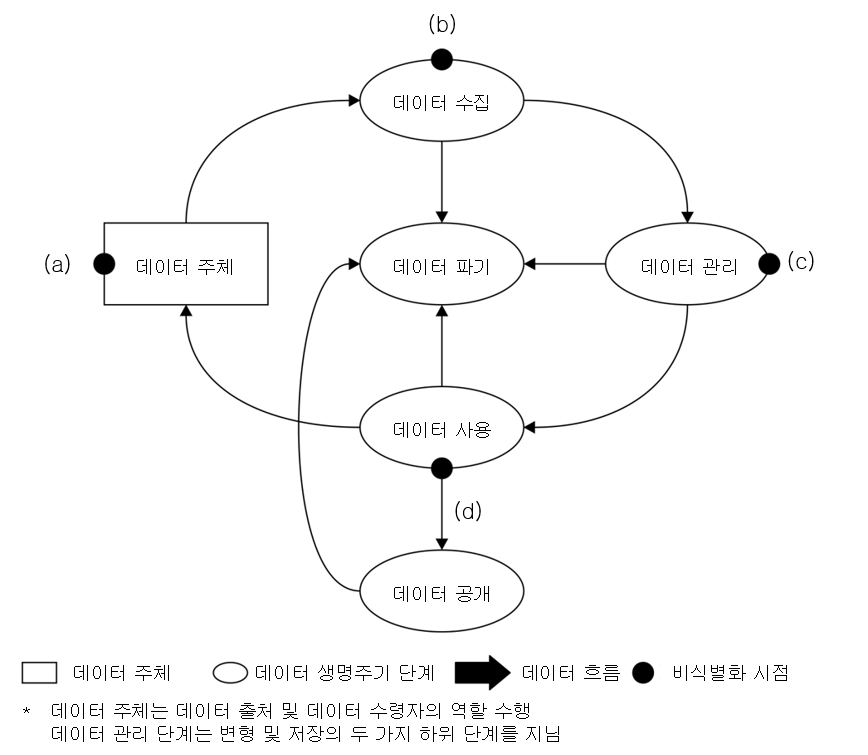


그림 1 — 데이터 생명주기 모델의 비식별 프로세스

### 데이터 수집 단계

데이터는 데이터가 참조하는 개인인 데이터 주체로부터 수집된다. 데이터 수집 결과로 생성된 데이터셋은 개인정보를 포함할 수 있다. 비식별 프로세스를 통해 모든 개인정보가 삭제된 새로운 데이터셋을 생성한다. 조직 내부에서는 원본 데이터셋 대신 비식별된 데이터셋을 사용하도록 권장한다.

이 모델을 사용한 비식별 프로세스로 다음 중 하나를 선택할 수 있다.

— 데이터 수집 과정(**그림 1**의 (b))

— 데이터를 수집했으나 식별자가 필요하지 않은 경우(**그림 1**의 (a))

데이터 관리(데이터 변환 및 데이터 저장)에 필요하지 않은 식별자는 수집하지 않아야 한다.

### 데이터 관리 단계

식별자 저장을 방지하려면 데이터 변환 후 데이터 저장 전에 비식별 프로세스를 적용해야 한다(**그림 1**의 (c)). 조직은 재식별 가능성을 고려하여 명확한 접근 통제, 최대 보존 기간, 그리고 비식별 데이터 간 연결 가능성을 최대한 줄이는 데이터 삭제 정책을 설정하도록 권장한다. 또한 의도된 사용 목적이 허용되는 범위 내에서 데이터 총계화와 같은 익명화 기술을 고려하도록 조직에 권장한다.

### 데이터 사용 단계

조직 내에서 데이터 관리용으로 개인정보(PII)가 필요한 경우, 데이터 공유용 데이터셋을 공개하기 전에 데이터를 비식별하도록 권장한다(**그림 1**의 (d)).

### 데이터 공개 단계

데이터는 ‘데이터 공유’ 계약과 같은 추가 관리 통제 항목에 따라 제3자와 공유할 수 있다. 또한 비식별 데이터셋도 공개할 수 있다. 비식별 데이터는 공개(Public), 반공개(Semi-public) 또는 비공개(Non-public)의 3가지 모델로 구분해 공개한다. 필요한 비식별 프로세스는 선택한 데이터 공개 모델에 따라 다를 수 있다.

### 데이터 파기 단계

데이터 파기는 데이터의 수집, 관리, 사용 및 공개와 같은 모든 단계에서 이루어질 수 있다. 데이터 복구를 방지하기 위해 검증된 방법으로 데이터를 파기해야 한다. 특히 재식별 가능성이 감지되면 데이터 파기를 고려해야 한다.

## 비식별 고려 사항

데이터 생명주기 동안 비식별 프로세스를 적용하면 효율성이 증가한다. 그러나 데이터 흐름에 참여하는 당사자 간의 관계에 따라 데이터 비식별의 시점은 데이터 수집 전, 수집 후, 저장 전 또는 저장 후 전달받는 당사자와 공유하기 직전으로 결정된다. 즉 수집 전은 **그림 1**의 (a), 수집 후는 그림 **그림 1**의 (b), 저장 전은 **그림 1**의 (c) 및 데이터 흐름에서 전달받는 당사자와 공유하기 직전인 **그림 1**의 (d)를 참조한다. 비식별 프로세스 적용 시점은 각 사용 사례에서 특정 비식별 기술의 효과를 강화하는 보안 및 기타 조직적 조치의 실행 가능성에 영향을 준다. 비식별 프로세스는 사용 목적이 익명화 기술을 지원하지 않는 경우 데이터 주체의 기밀성을 보호하는 데 유용한 기술이지만, 그 자체로는 데이터 주체를 보호하는 데 충분하지 않으며, 포괄적인 데이터 보호 프레임워크의 일환으로 고려해야 한다. 이번 절은 각 단계의 기능과 고려 사항을 설명한다.

### 데이터 수집

로컬 비식별 프로세스(또는 소스 비식별 프로세스)는 가장 보편적으로 사용되는 접근 방식으로, 분석용 데이터를 공개하기 전에 개인이(또는 개인의 데이터를 처리하는 처리자가) 모든 개인정보를 삭제하도록 허용하는 조치이다. 데이터 수집 단계와 직접적으로 관련된 비식별 프로세스 중 하나는 데이터 최소화이다. 데이터 주체를 수집하는 모든 데이터 처리자는 의도된 사용 목적에 따른 필수 데이터를 정확하게 정의하며, 정의된 매개변수로 데이터 수집을 제한해야 한다. 또한 데이터 필드를 줄이려면, 데이터 수집/전송에서 불필요한 개인정보를 제외하는 구체적인 프로세스를 마련해야 한다. 다른 비식별 프로세스로는 데이터 총계화가 있다. 데이터 처리자는 사용 목적에 따라 개별 데이터 주체를 반드시 선별할 필요가 없는 모든 경우에 데이터 총계화를 고려해야 한다.

### 데이터 관리

#### 데이터 변환

데이터 변환 단계에서는 총계화, 통계 공개 제한, 암호화 등과 같은 비식별 기술을 적용할 수 있다. 데이터 변환은 직접 수집 후 장기 저장 전, 상당한 저장 기간 후 접근 전, 또는 접근 통합을 포함한 하나 또는 여러 단계에서 적용할 수 있다. 데이터 치환 또는 총계 처리에 의한 변환은 수집 후 공개까지 어느 시점에서도 적용할 수 있다. 수집 직후 적용할 경우, 데이터를 치환 또는 총계 처리하면 데이터 유출 시 데이터 주체에 대한 잠재적 피해를 줄일 수 있지만, 데이터를 연결, 병합 또는 업데이트할 가능성이 줄어든다. 데이터 변환 방법은 데이터 주체가 노출될 경우의 잠재적 피해를 신중히 고려한 후 선택해야 한다. 노출 위험을 줄이는 데 사용되는 기술을 향후 잠재적 사용 및 분석에 영향을 미칠 수 있으므로, 변환 결정은 추후 데이터 사용 목적에 따라 지원되어야 하는 분석에 대하여 고려해야 한다.

#### 데이터 저장

데이터 저장은 데이터 처리자 또는 처리자의 지시를 따르는 당사자가 개인정보를 포함한 데이터를 모든 형태의 비휘발성 메모리에 저장하는 프로세스이다. 저장 단계에서는 정보 보안 및 프라이버시 통제에 중점을 두고 있으므로, 이번 절은 통제를 요약하되 세부 고려 사항은 열거하지 않는다(ISO/IEC 27001). 저장 단계에서 일반적인 정보 보안 및 프라이버시 통제의 예로는 접근 통제, 유지 관리, 보안 평가, 인증 절차, 사고 모니터링 및 대응, 감사 등이 있다. 특히 조직은 최대 데이터 보존 및 삭제 정책을 준수함으로써 사용 목적을 달성하는 데 필요한 최대 보존 기간이 경과하면 데이터를 저장하지 않고 완전히 파기해야 한다. 예를 들어, 데이터 공유 계약에는 데이터 수령자가 수령 후 1년 이내와 같이 일정 기간 내에 데이터를 파기해야 함을 명시하고 있으며, 이러한 계약 조항은 법적으로 의무화할 수 있다.

### 데이터 사용

비식별 데이터는 비식별 프로세스 이후 저장된 데이터 속성에 따라 여러 목적과 활용을 위해 수집, 저장 및 공유할 수 있다. 비식별 데이터셋을 공개하는 주된 이유 중 하나는 연구 목적으로 원시 데이터의 가치와 속성을 연구할 기회를 제공하는 것이다(ISO/IEC 20889). 따라서 비식별 프로세스는 개인정보를 보호하는 동시에, 정보의 유용성을 최대한 유지해야 한다. 또한 비식별 프로세스의 목적에 따라 데이터 공개 모델을 포함한 여러 상황의 활용을 고려해야 한다. 비식별 데이터를 공개할 경우, 조직은 광범위한 이해 당사자가 포함된 전문가 위원회에서 공개 관련 데이터 주체에 대한 잠재적 영향을 고려해 공개 여부를 결정해야 한다. 이 경우 위험 평가 및 점검 사항은 주로 평가 지침과 재식별 위험을 완화하는 데 적합한 공개 체계를 결정하기 위해 사용된다.

비식별 기술은 특정 사용 사례에 필요한 “적용성” 또는 “유용성” 수준에 따라 선택한다.

# 비식별 프로세스 프레임워크

이 절은 **그림 2**와 같이 4개의 단계에 걸쳐 비식별 개인정보를 제공하는 비식별 프로세스 프레임워크를 설명한다[Guidelines on De-identification Measures].

**1단계** ― **사전 검토**

1단계는 해당 데이터의 개인정보 포함 여부를 검증한다. 데이터에 개인정보가 포함되어 있으면 2단계 비식별이 필요하다.

**2단계** ― **비식별 적용**

2단계는 데이터 비식별을 포함해 해당 데이터셋에서 특정 개인정보의 추론을 방지한다. 이 단계에서는 전체적 또는 부분적으로 개인정보 요소를 삭제 또는 변환하는 방법을 적용한다. 개인정보 요소는 식별자, 준식별자 및 민감한 속성을 포함한다.

**3단계** ― **적정성 평가**

3단계는 개인정보 요소를 포함한 비식별 데이터셋의 적정성을 평가한다. 평가 시 해당 데이터셋의 개인정보 포함 여부, 재식별 가능성, 결과적으로 재식별이 가능한 연결 가능성을 고려한다.

**4단계** ― **사후 관리**

4단계는 재식별을 방지하는 관리적 및 기술적 안전 방안을 포함한다.

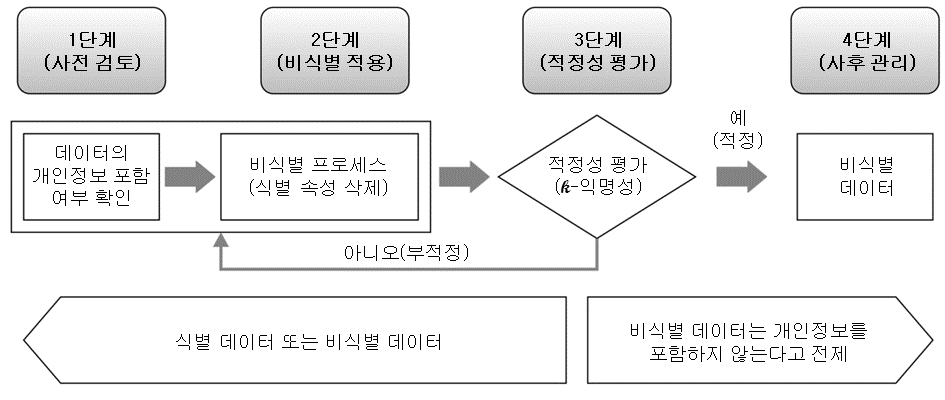


그림 2 — 비식별 프로세스

각 단계는 다음의 **6.1 – 6.4**에서 설명한다.

## 1단계 ― 사전 검토

다양한 목적으로 데이터를 사용 또는 제공하려는 조직은 먼저 정책과 표준을 결정해야 한다. 정책과 표준 결정 시 다음 사항을 고려하도록 권장한다.

— 정보의 비식별 처리 목적 및 용도는 무엇인가?

— 비식별 데이터는 어떤 종류의 데이터 속성으로 구성되는가?

— 데이터 비식별 프로세스에 어떤 기술이 사용되는가?

— 재식별의 위험 수준과 이에 따른 부작용은 무엇인가?

— 특정 개인의 정보가 재식별될 경우, 어떤 대응책을 적용할 수 있는가?

— 재식별 처리 수준은 어떻게 평가하는가?

— 비식별 프로세스에 소요되는 인력과 비용은 어떻게 결정하는가?

사전 검토를 구성하는 구체적인 고려 사항은 데이터의 종류와 사용 목적에 따라 달라질 수 있지만 일련의 표준 설정을 권장한다.

다양한 목적으로 데이터를 처리하려는 조직은 특정 데이터의 개인정보 포함 여부를 확인하기 위해 적합한 표준을 참조해야 한다. 데이터가 개인정보를 포함하지 않는다고 판단하더라도 조직은 가용 데이터 간 연결 가능성을 고려해 위험을 최소화하도록 적절한 조치를 취해야 한다. 또한 데이터가 개인정보를 포함하는 경우 비식별 단계가 필요하다.

개인정보 포함 여부에 대한 판단 기준은 다음을 참조한다.

— 데이터의 종류, 형태, 특성 및 형식에 관한 특별한 제한이 없다.

— 데이터 프로세서가 데이터를 사용해 개인을 식별할 수 있는 경우, 해당 데이터는 개인정보로 간주한다.

— 데이터는 개인에 관한 것이어야 하므로, 복수의 개인으로 구성된 그룹의 통계값은 개인정보로 간주하지 않는다.

— 추가 정보와 결합하여 개인을 식별할 수 있는 데이터는 개인정보로 간주한다. 여기서 추가 정보는 보통 공개/가용 정보를 의미한다.

## 2단계 ― 비식별 적용

### 식별자에 대한 비식별

“식별자”는 개인 또는 개인과 관련된 사물에 고유하게 부여된 값 또는 이름과 같은 데이터이다. 일반적으로 “식별자”의 수집을 최소화하며, 데이터셋에 포함된 식별자를 삭제한다.

다만, 의도된 목적상 반드시 필요한 식별자는 다음 데이터를 포함할 수 있다.

— 고유 식별 번호(주민등록번호, 사회보장번호(SSN), 여권번호, 외국인 등록번호, 운전면허 번호 등)

— 이름(한자, 영문 성명 등)

— 상세 주소(번지, 도로명 등)

— 날짜 정보(생일, 기념일(결혼 등), 자격증 취득일 등)

— 전화번호(휴대전화, 집 전화, 사무실, 팩스 등)

— 의료 기록 번호, 건강 보험 번호, 복지 수급자 번호 등

— 은행 계좌 번호, 신용카드 번호 등

— 사진(일반 사진, 동영상 CCTV 영상 등)

— 생체인식 데이터(지문, 음성, 홍채 등)

— 이메일 주소, IP주소, MAC 주소, 홈페이지 URL 등

— 식별 코드(사원 번호, 고객 번호 등)

— 기타 고유 식별 번호(군번, 사업자 등록 번호 등)

### 준식별자 및 식별 속성에 대한 비식별

일반적으로 데이터셋에 포함된 준식별자는 데이터 사용 목적과 관련이 없으면 삭제해야 한다. 데이터 사용과 관련된 준식별자에 식별 요소가 있는 경우, 가명화 및 총계화와 같은 비식별 기술을 사용해야 한다.

행위 정보와 같이 식별 가능성이 큰 데이터는 비식별 프로세스 및 익명화 기술을 적용해야 한다.

### 비식별 기술

가명화, 총계화, 데이터 범주화 및 데이터 마스킹을 포함한 다양한 기술을 개별적으로 또는 조합해 사용할 수 있다. 가명화 기술만 적용하는 것은 비식별 기술로 충분하지 않을 수 있다.

각 기술을 구현하기 위해 사용할 수 있는 여러 유형의 기술들이 있다. 따라서 데이터의 사용 목적과 특정 기술의 장단점을 기반으로 가장 적합한 기법을 선택해 활용해야 한다. 비식별을 완료한 후에는 다음 단계로 진행한다.

## 3단계 ― 비식별 프로세스의 적정성 평가

비식별 프로세스가 충분하지 않은 경우, 다른 데이터를 결합하거나 다양한 추론 기법을 사용해 개인을 식별할 수 있다.

재식별 위험을 줄이려면, 사용하기 전 비식별 데이터의 적정성 평가를 수행해야 한다. 적정성 평가 시 다음 사항을 고려한다.

— 비식별 프로세스의 목적

— 비식별 프로세스에 관련된 데이터 속성의 유형(식별자 포함 여부)

— 적정한 비식별 수준

적정성 평가는 개인정보 보호 책임자(DPO), 위임받은 신뢰할 수 있는 제3자(TTP) 또는 외부 평가단이 수행할 수 있다.

적정성 평가는 프라이버시 보호 모델 중에서 k-익명성 모델을 사용한다. k-익명성은 기본적인 평가 수단이다. 필요시 추가 평가 모델(l-다양성, t-근접성, 차등 프라이버시(DP) 등)을 적용할 수 있다.

적정성 평가에 대한 자세한 내용은 **부속서 A**를 참조한다.

## 4단계 ― 사후 관리

### 비식별 데이터의 보호 조치

비식별 데이터가 유출되거나 다른 데이터와 결합할 경우, 재식별 가능성을 방지하기 위한 보호 조치는 다음과 같다.

— 관리적 보호 조치: 비식별 데이터 파일의 관리자를 지정하고, 비식별 데이터의 공유 여부를 결정하며, 사용 목적을 달성하면 데이터를 즉시 파기한다.

— 기술적 보호 조치: 비식별 데이터 파일에 대한 접근을 제한하고, 접근 기록을 관리하며, 보안 프로그램을 설치 및 운영한다.

또한, 보안 조치는 비식별 데이터가 유출될 경우 취해야 할 보호 조치를 포함하는데, 이에 해당하는 보호 조치는 다음과 같다.

— 유출 원인을 분석하고 추가 유출을 방지하기 위해 관리적, 기술적 안전 조치를 모두 이행한다.

— 유출된 비식별 데이터를 회수하고 파기한다.

### 재식별 가능성 모니터링

비식별 데이터를 사용하거나 제3자에게 제공할 의도가 있는 데이터 프로세서는 재식별 가능성을 정기적으로 모니터링해야 한다.

재식별 가능성이 감지될 경우, 비식별 데이터를 제공받은 데이터 프로세서에게 데이터의 처리 중단, 회수 및 파기를 요청해야 한다.

### 제3자 계약에 대한 요구사항

비식별 데이터를 제3자에게 제공 또는 위탁할 경우, 계약에 재식별 위험 관리를 포함해야 한다. 재식별 위험 관리는 다음과 같다.

— 제3자에 대한 데이터 공개를 데이터 주체에게 고지한다.

— 가능한 경우, 제3자에게 익명화 데이터를 제공한다.

— 재식별 금지: 비식별 데이터를 제공받거나 처리를 위탁받은 데이터 프로세서가 다른 데이터와 결합해 재식별하는 것을 금지하도록 규정한다.

— 재제공 및 재위탁 제한: 비식별 데이터를 제공하거나 처리를 위탁받은 계약에 재제공 또는 재위탁의 허용 범위를 규정한다.

— 재식별 위험 고지: 데이터가 재식별되었거나 재식별 가능성이 높아진 경우, 데이터 처리를 중지하고 이를 제공자와 위탁자에게 고지해야 하는 의무를 규정한다.

### 재식별에 대한 조치

비식별 데이터가 재식별될 경우, 데이터 처리를 중단하고 개인정보의 유출 방지를 위해 필요한 조치를 취해야한다.

재식별 데이터는 즉시 파기한다.

# 비식별 데이터의 유용성

## 비식별 데이터 단계

이 절은 데이터에 의해 개인이 식별되는 정도, 데이터의 특성(속성)과 개인의 연관 방식을 설명하기 위해 데이터 종류로 표현 가능한 데이터 비식별 프로세스의 단계를 정의한다. 데이터의 사용 또는 처리 맥락에서, 데이터 명세에는 데이터 종류를 비롯해 데이터에 의해 개인이 식별되는 정도 또는 데이터의 특성 집합과 개인을 연결할 수 있는 정도에 대한 설명도 포함되어야 한다.

**그림 3**은 비식별 프로세스에서 식별 데이터부터 비식별 데이터에 이르는 데이터 단계들을 보여준다. 각 단계는 연속적으로 다른 수준의 재식별 위험 가능성이 있다. 데이터 종류는 비식별 단계에 따라 데이터셋이 거치는 특정 단계의 특성을 갖는다.

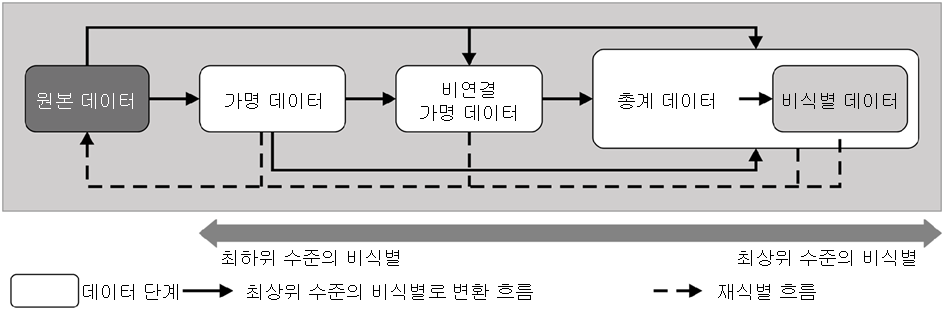


그림 3 — 비식별 데이터 단계

**그림 3**과 같이, 모든 데이터는 비식별 단계를 거친다. 우측(최상위 수준의 비식별)으로 갈수록 과거의 날씨 정보와 같이 개인과 관련되지 않는 데이터이므로 개인정보에 대한 위험이 낮아진다. 좌측(최하위 수준의 비식별) 끝은 특정 개인과 직접 연결된 식별 데이터라고 볼 수 있다. 두 종류의 데이터 단계 사이에서 재식별 시도를 통해 특정 개인을 연결하거나 특정 그룹의 개인을 연결할 수 있는 데이터도 있고, 개인에 기반을 두지만, 다시 연결할 수 없는 데이터도 있다. 일반적으로 비식별 프로세스는 요구되는 유용성을 유지하면서 비식별 데이터 공개로 인한 위험을 낮추도록 우측 방향의 비식별 데이터로 변환될 수 있게 설계되어야 한다.

### 원본 데이터 단계

식별 데이터의 원본 데이터 단계에서는 해당 정보로부터 개인을 식별할 수 있으므로 데이터는 특정 개인과 명확하게 연관될 수 있다. 식별자에 관한 지침은 ISO/IEC 29100의 **4.4.1**을 참조한다.

### 가명 데이터 단계

가명 데이터 단계는 모든 식별자를 가명으로 대체하기 때문에, 가명을 지정한 당사자 외에는 데이터를 복구할 수 없다. 그러나 가명 데이터는 다른 데이터와의 연결 가능성을 통해 재식별될 수 있다.

이는 본 표준 **3.14**에서 “가명화”로 정의한 데이터에 해당한다.

### 비연결 가명 데이터 단계

비연결 가명 데이터 단계에서는 모든 식별자를 삭제하거나 가명으로 대체하며 지정 함수를 삭제 또는 복구할 수 없으므로, 처리한 당사자를 포함한 어떤 당사자도 연결을 복구할 수 없다. 그러나 비연결 가명 데이터는 다른 데이터와의 연결 가능성을 통해 재식별 될 수 있다.

### 총계 데이터 단계

총계 데이터 단계에서 개인 수준의 속성이 개인 수준의 항목을 포함하지 않고 결합된 통계 데이터로 변환되므로, 데이터는 추론할 수 없는 다른 사람에 대한 정보를 형성한다. 특정 개인을 식별할 수 있는 일부 변수의 조합이 교차할 경우, 총계화 기술을 적용하면 모든 총계 데이터는 임계값 이하의 식별 가능성에 도달할 수 없다.

이는 이 표준 **3.1**에서 “총계 데이터”로 정의한 데이터에 해당한다.

### 비식별 데이터 단계

비식별 데이터 단계에서 데이터 단독 또는 기타 데이터와 결합하여 개인을 직간접적으로 식별할 수 없는 합리적 신뢰 수준의 방식으로 데이터 연결을 해제하고 속성을 변경(속성자를 무작위화 또는 일반화하여)한다.

## 데이터 공개 모델

비식별 데이터 공개 모델은 데이터 분석 상황에 따라 3가지 모델로 분류한다(The anonymisation decision-making framework).

즉, 비식별 데이터는 3가지 데이터 공개 모델인 공개(Public), 반공개(Semi-public) 또는 비공개 모델(Non-public)로 전달할 수 있다.

데이터 공개 모델은 허용하는 가용성 및 정보보호 수준이 다르다. 또한 데이터의 공개 목적 및/또는 법적 요구사항에 따라 모델의 적합성도 다를 수 있다. 데이터 공개 모델은 선택한 모델에 따라 필요한 비식별 프로세스가 다를 수 있어, 비식별 프로세스에서 중요한 역할을 한다.

이 3가지 데이터 공개 모델은 **7.2.1** 부터 **7.2.3**에서 다룬다.

### 공개(Public) 모델

일반적인 공개 모델로 모든 사람이 등록이나 조건 없이 데이터에 접근할 수 있다. 해당 모델의 예로는 조직에서 공개하는 가용 데이터 및 웹 포털과 같은 오픈 액세스 데이터 저장소에 게시된 데이터를 들 수 있다. 조직은 데이터셋을 공개해 모든 사람이 자유롭게 사용 및 재게시할 수 있도록 한다.

데이터 공개 시 접근자 및 접근 방식을 비롯하여 정보에 제한을 두지 않는 것이 일반적이다. 따라서 다운로드한 데이터셋의 개인을 식별할 수 없는 경우는 공개 모델로 취급한다.

**7.2.2**의 정보 요청에 대한 접근의 경우도 정보 요청자가 정보의 처리, 개인정보 또는 보안에 관한 약관에 동의할 필요가 없다면 공개 모델로 취급한다.

### 반공개(Semi-public) 모델

반공개 모델은 공개 모델보다 더 제한적이며, 데이터 접근 권한을 취득하기 위한 공식적인 요청 및 승인 프로세스가 있다. 해당 사례의 경우, 데이터 수령자는 이용 약관에 동의하거나 온라인 선택을 통해 체결되는 계약에 서명할 수 있다. 온라인 선택으로 체결되는 계약은 데이터로 수행할 수 있는 작업과 데이터 처리 방법에 제한을 둘 수 있는 온라인 이용 약관이다. 그렇지만 누구든 해당 데이터를 다운로드할 수 있다.

비식별은 데이터셋에 대한 정보 요청 접근에 대응하는 경우에도 유용할 수 있다. 조직은 비식별 프로세스를 적용하여 정보의 유용성을 유지하면서 개인정보를 보호하는 방식으로 요청에 대응할 수 있다. 다음과 같은 정보 시스템으로 데이터를 공유할 경우, 조직은 일부를 제한하는 접근 통제를 적용할 수 있다.

— 데이터에 접근하는 모든 사용자에게 접근 전 등록 및 연락처 정보를 요구한다.

— 개인 신원 확인용 인증 프로토콜을 채택한다.

— 개인의 소속 또는 자격 증명을 기반으로 계층화된 접근 시스템을 사용하여 다른 당사자에게 차등화된 수준의 접근 권한을 부여한다.

또한 분석가가 데이터 처리자에게 분석을 요청하여 데이터 처리자가 분석을 수행하는 경우와 같이 데이터 공유가 필요하지 않은 사례도 데이터 접근에 해당된다. 이러한 경우, 이와 같은 사례는 조직의 데이터 공유에 해당되지 않을 수 있다.

### 비공개(Non-public) 모델

개인정보가 포함된 데이터셋은 국가 규제 지침에 따라 공개를 허용한 경우에 한 해 조직 내, 조직 간에 공유할 수 있다. 따라서 기관은 공개 허용이 되지 않는 데이터셋을 공유하려면 모든 개인정보를 삭제해야 한다. 비공개 모델은 최소 가용성을 제공하지만, 보호 수준이 높기 때문에 상대적으로 낮은 수준의 비식별 프로세스를 적용한다.

조직 간에 정보를 공유할 경우 데이터셋에 대한 접근은 해당 조직에 국한되므로 데이터 공유 계약을 통해 정보의 개인정보보호 및 보안에 대한 요구사항을 설정 및 시행할 수 있다. 비공개 모델로 취급할 경우, 당사자 간의 데이터 공유 계약을 체결해야 한다. 데이터 공유 계약은 공개에 따른 위험 완화 전략의 주요 부분으로, 다음과 같은 이용 약관을 포함한다.

— 허용 접근 명세(수령자 통제)

— 데이터 보안 요구사항(인프라 통제)

— 사용 제한, 특히 다른 파일과의 연결 및 의도적 재식별 금지(기타 데이터 및 거버넌스 통제)

— 사용 완료 즉시 데이터 파기에 대한 요구사항(거버넌스 통제)

데이터 공유 계약의 목적은 다음과 같은 세 가지 측면이 있다.

— 데이터 처리자가 신뢰하는 개인 또는 조직과 그렇지 않은 개인 또는 조직을 명확하게 구별한다.

— 접근 조건을 규정할 수 있는 프레임워크를 제공한다.

— 개인/조직이 해당 접근 조건을 위반하는 경우 제재 또는 처벌을 규정할 수 있다.

### 데이터 공개 모델 비교

데이터 흐름 환경에서 재식별 가능성을 제한하는 방법 중 하나는 데이터의 취득 및 사용 방식을 통제하는 것이다. 데이터 공개 모델에 따라 통제 항목을 분류할 수 있는데, 이 경우 각 모델은 서로 다른 장단점을 갖는다. 이에 조직은 여러 사용 사례 및 이에 따른 개인정보 위협을 해결하기 위해 이러한 모델을 결합한 계층적 접근 통제 방식도 적용할 수 있다.

추가적으로 데이터 공개 모델은 다수의 또는 주기적 공개 가능성을 고려한다. 아래 제시된 모델들은 제한이 없는 모델부터 엄격하게 제한하는 모델까지 서로 다른 접근 범위를 지닌다. **표 1**은 데이터 공개 모델 간 차이점을 비교해 제시한다.

**표 1** — **데이터 공개 모델 비교**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **공개(Public) 모델** | **반공개(Semi-public) 모델** | **비공개(Non-public) 모델** |
| 접근 권한 | * 공개된 데이터에 누구나 자유롭게 접근 가능 | * 공개된 데이터에 승인받은 개인 또는 조직과 이에 속하는 당사자만 접근 가능 | * 공개된 데이터에 특정 개인 또는 조직과 이에 속하는 당사자만 접근 가능 |
| 사용 사례 | * 웹포털을 통한 무제한 데이터 접근처럼 누구나 자유롭게 사용 가능 | * 온사이트 안전 조치 설정 * 접근 권한 제공 * 원격 가상 접근 * 분석 서버를 통한 접근 | * 조직 내 및 조직 간 공유 |
| 사용 권한 | * 데이터 재사용 및 재배포에 대한 무제한 권한 | * 승인받은 개인 또는 조직만 사용 가능 | * 데이터 재사용, 재게 시 또는 배포 금지 |
| 재식별 공격 | * 선전용 시범적 공격 | * 내부자의 의도적 공격 * 친분을 이용한 데이터셋의 부주의한 개인 인식 * 데이터 유출 | |

## 데이터 공개 모델 및 데이터 단계 간 상관관계

### 비공개(Non-public) 모델

데이터 출처로부터 데이터를 비공개 모델로 공유할 경우, 데이터 비식별 프로세스를 적용해야 한다. 일반적인 상황에서는 비공개 모델이더라도 비연결 가명 데이터와 높은 수준의 비식별 데이터를 사용한다. 이 경우에는 가명화, 암호화, 합성, 삭제 등의 비식별 도구를 사용할 수 있다.

그러나 양자 간 특별 계약 또는 관련법이 적용될 경우, 데이터의 분석 및 저장 단계에서 가명 데이터를 사용할 수 있다.

### 반공개(Semi-public) 모델

데이터 출처로부터 데이터를 반공개 모델로 공유할 경우, 비공개 모델보다 높은 수준의 비식별 프로세스를 적용해야 한다. 재식별을 방지하기 위해서 통계 처리를 수행한다. 총계 데이터와 상대적으로 높은 수준의 비식별 데이터는 반공개 모델로 공개할 수 있다. 구체적으로 통계 처리, 무작위화 등의 비식별 도구를 사용할 수 있다.

**표 1**과 같이, 데이터에는 승인받은 개인 또는 조직만 접근할 수 있으므로, 공개 모델보다 상대적으로 낮은 수준의 비식별 프로세스를 허용할 수 있다.

### 공개(Public) 모델

데이터 출처로부터 데이터를 공개 모델로 공유하는 경우, 반공개 모델보다 높은 수준의 비식별 프로세스를 적용해야 한다. 비식별 데이터를 취득하는 프로세스를 수행하여 **표 1**과 같이 프로세스 결과를 공개 모델로 사용할 수 있다.

1. (규정)  
     
   적정성 평가 절차

이 **부속서**는 적정성 평가 절차의 수행 모델을 제공한다(Guidelines on De-identification Measures). (**그림 A.1** 참고)

적정성 평가 절차 단계는 다음과 같다.

— (기초자료 작성) 데이터 처리자는 적정성 평가에 필요한 데이터 명세, 비식별 프로세스 현황, 이용 기관의 관리 수준 등의 기초자료를 작성한다. “이용 기관”은 비식별 프로세스 이후 비식별 데이터를 활용하고자 하는 기관을 의미한다.

— (평가단 구성) 개인정보 보호 책임자는 평가단을 구성하거나, DPO 또는 TTP에게 요청하여 평가를 수행할 수 있다.

— (평가 수행) 평가단은 개인정보 관리자(PII manager)가 작성한 기초자료를 활용해 비식별 프로세스 수준의 적정성을 평가한다.

— (추가 비식별 프로세스) 데이터 처리자는 평가 결과가 적정하지 않은 경우, 평가단의 의견을 반영해 추가 비식별 프로세스를 수행한다.

— (데이터 활용) 비식별 프로세스가 적정하다고 평가받은 경우, 빅데이터 분석 등의 목적으로 데이터를 이용하거나 제공할 수 있다.

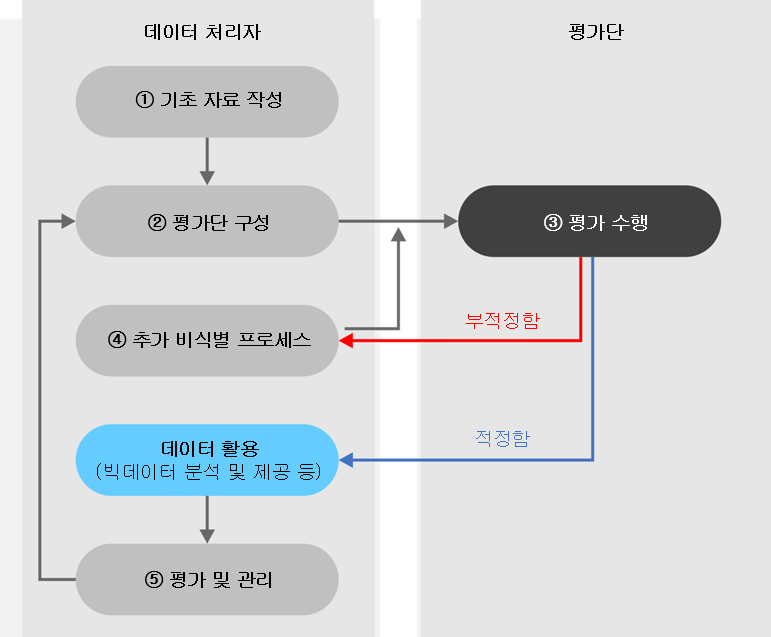


그림 A.1 — 비식별 프로세스에 대한 적정성 평가 절차

* 1. 기초 자료 작성

데이터 처리자는 적정성 평가에 필요한 평가 주체인 데이터 명세, 비식별 프로세스 현황, 이용 기관의 관리 수준 등의 기초 자료를 작성한다.

* 1. 평가단 구성

개인정보 보호 책임자는 평가단을 구성한다. 외부 전문가를 위촉할 경우, 분야별 전문 기관이 운영하는 전문가 집단에서 법무 전문가 및 비식별 전문가를 1명 이상 임명한다.

또한 데이터의 활용 목적과 직접적인 이해관계가 없는 자로 평가단을 구성한다.

* 1. 평가 수행

평가단은 기초자료와 k-익명성 모델을 사용해 비식별 프로세스의 적정성을 평가한다.

— (사전 검토) 데이터 처리자는 준비된 기초 자료 및 인터뷰를 검토하고 데이터셋에 개인 식별 요소의 포함 여부, 사용 목적 및 비식별 기술의 정합성을 확인한다.

— (재식별 시도 가능성 분석) 데이터를 사용 또는 수령하는 데이터 처리자의 의도, 개인정보 보호 수준 및 역량을 포함해 재식별 시도 가능성을 분석한다.

— (재식별 시 영향 분석) 데이터가 의도적 또는 비의도적으로 재식별될 경우, 데이터 주체에 미칠 수 있는 영향을 평가한다.

— (계량 분석) 데이터 처리자가 제공한 k값의 정확성을 검증한다.

— (평가 기준값 결정) 평가단은 재식별 가능성, 재식별시 영향, 계량 분석 결과, 데이터 사용 목적 등을 종합적으로 고려하여 평가 기준값을 결정한다.

— (적정성 평가) 평균 기준값과 계량 전 분석 결과의 산출값을 비교하여 비식별 프로세스의 적정성을 결정한다.

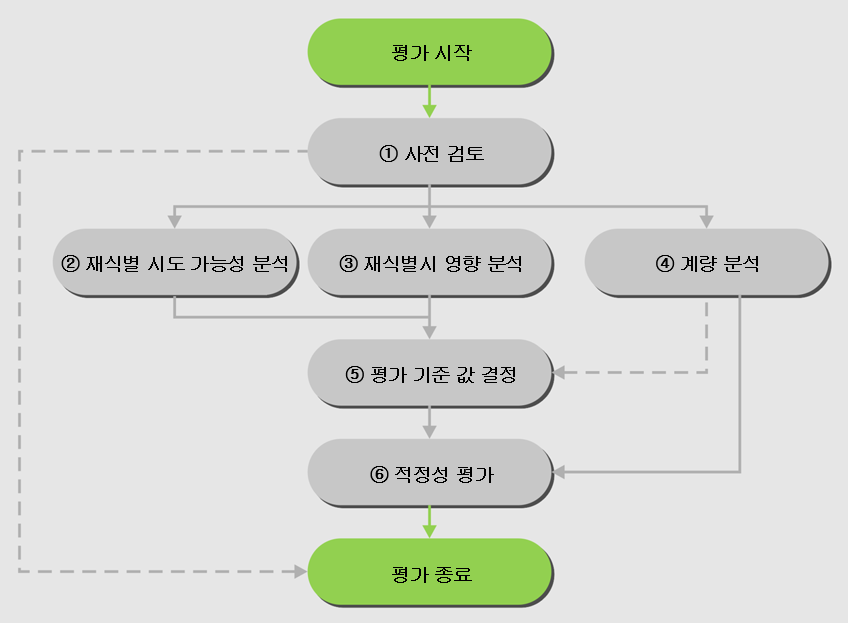


그림 A.2 — 적정성 평가 절차

* 1. 추가 비식별 프로세스

— 데이터 처리자는 평가 결과가 부적정한 경우, 평가단 의견을 기반으로 추가 비식별 프로세스를 이행해야 한다.

— 평가단은 데이터 처리자가 추가 비식별 프로세스를 완료한 후 재평가를 수행해야 한다.

* 1. 데이터 활용

— 비식별 프로세스의 평가(재평가) 결과가 적정한 경우, 비식별 데이터를 빅데이터에 활용하거나 제3자에게 제공하는 것을 허용한다.

— 재식별 위험이 높은 관계로 데이터 공개 모델에 따른 위험 완화 전략이 없을 경우, 일반 대중 또는 승인되지 않은 데이터 사용자에게 데이터를 제공하거나 공개하는 것을 원칙적으로 금지한다.

— 데이터 사용 목적을 달성하거나 데이터가 불필요한 경우, 데이터를 즉시 파기한다.

— 데이터 활용 과정에서 사후 관리 단계에 따른 조치를 준수해야만 비식별 데이터 형태로서 유효하게 데이터를 활용할 수 있다.

1. (규정)  
     
   비정형 데이터에 대한 비식별 접근 방식

비정형 데이터의 비식별 체계는 정형 데이터의 경우와 달리 정형 데이터 필드 대신에 원시 데이터에 적용된다. 다음 사진의 경우, 비식별은 얼굴 이미지를 삭제하거나, **그림 B.1** 과 같이 다른 얼굴로 교체하는 것을 의미한다.

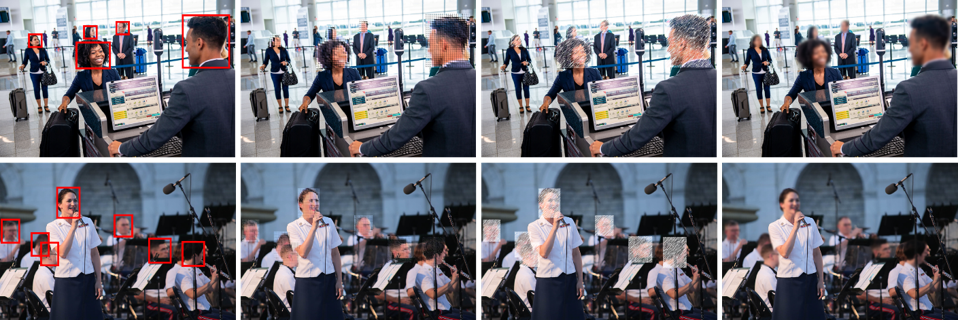


그림 B.1 — 안면 비식별 프로세스 예시

비정형 데이터는 다음과 같은 4개 유형으로 구분한다.

1) 비정형 텍스트 데이터: 웹데이터, 보고서, 블로그, 언론 기사 등

2) 비정형 비디오 데이터: 모든 비디오 데이터는 비정형이며 일부 태그 정보에서 정규화 데이터를 제공한다.

3) 비정형 오디오 데이터: 모든 오디오 데이터는 비정형이며 일부 태그 정보 또는 인식된 오디오는 텍스트 데이터로 변환된다.

4) 비정형 로그 데이터: 기계 생성 로그 데이터는 비정형이지만 일반적으로 패턴이 있기 때문에 정형 형태로 변환할 수 있다.

텍스트, 음성, 이미지 및 비디오를 포함한 비정형 데이터의 구문 정보를 표현하려면, 비식별 프로세스 시스템에 다음 3가지 장치가 필요하다.

1) 멀티미디어 데이터 입력에서 텍스트 메타 정보를 감지하는 멀티미디어 정보 검출기는 다음을 포함한다.

— 음성 입력에 포함된 객체 또는 활동을 추적하기 위해 음성 입력을 텍스트로 변환하는 음성 검출기

— 이미지 입력에서 문자로 변환하는 광학 문자 인식 검출기

— 이미지 입력에 포함된 객체 또는 활동을 변환하거나, 이미지 또는 동영상 입력에서 사진을 삭제하는 시각 자료 검출기

— 이미지 또는 동영상 입력에서 텍스트 문장으로 변환하는 시각 자료 문장화 검출기

2) 지식기반 형성기는 다음과 같이 텍스트 메타 정보 및 컨텍스트 정보를 외부 구성의 구문 및 내부 정보의 의미로 나누며, 나눠진 각 정보는 다음을 포함한다.

— 구문 정보는 멀티미디어 데이터를 생성하는 소스 정보, 소스에서 생성된 멀티미디어 데이터 정보 및 의미 영역에서 추출된 객체 검출 정보를 포함한다.

— 의미 정보는 멀티미디어 데이터를 구성하는 의미 영역에 포함된 이벤트 정보 및 상황 정보를 포함한다.

3) 비식별 장치는 지식기반 및 텍스트 메타 정보에서 식별 가능한 개인정보를 삭제한다.

비정형 데이터의 비식별 체계는 관련 요구사항 및 보안 강화 수준을 다음과 같이 규정한다.

— 비식별 대상: 활용 또는 온라인 서비스를 위해 보호해야 하는 대상 개체는 무엇인가?

— 비식별 방법: 어떤 비식별 체계를 적용하는가? 비식별은 어떤 수준(예시: 블랙박스, 픽실레이션(pixilation), 블러링(blurring))으로 적용되는가?

— 비식별 및 재식별: 복구 또는 재식별 요구를 판단해야 한다. 정책에 따라 범죄 수사에 원본 사진이 필요한 경우, 비식별 처리된 사진을 복구할 수 있는가?

1. (참조)  
     
   일반적인 비식별 기술의 예시

이 **부속서**는 일반적인 비식별 기술에 대한 예시와 설명을 제공한다.

* 1. 비식별 기술의 통계화 도구

— 표본추출: 전체 데이터셋을 공개하는 대신, 전체 데이터셋의 표본을 공개하는 처리 방법이다. 하위 샘플을 공개하면 재식별 가능성을 낮출 수 있다.

— 총계: 전체 데이터셋의 대푯값을 생성하는 통계 함수의 집합이다.

* 1. 비식별 기술의 암호화 도구

— 결정적 암호화: 암호화 알고리즘의 개별 실행에서 주어진 평문과 키에 대해 항상 동일한 암호문을 생성하는 암호화 체계이다.

— 순서 보존 암호화: 평문값의 순서를 유지하는 암호화 체계이다.

— 동형 암호화(ISO/IEC 18033-6): 암호문을 연산할 수 있으며, 암호화 결과는 복호화한 평문의 작업 결과와 동일한 암호화 체계이다.

— 형태 보존 암호화(NIST 800-38G): 암호문이 평문과 동일한 형태인 암호화 체계이다.

* 1. 삭제 기술

— 마스킹: 필드 값을 대체 또는 삭제하는 처리 방법이다. 삭제 기술은 전화번호를 별표 또는 무작위로 생성된 가명으로 대체하는 것을 포함한다.

— 로컬 삭제: 선택한 레코드에서 특성 속성자를 삭제하는 처리 방법이다. 데이터를 삭제하면 개인정보 보호는 강화되지만, 데이터셋의 유용성은 낮아질 수 있다.

— 레코드 삭제: 데이터셋에서 전체 레코드를 삭제하는 처리 방법이다.

* 1. 가명화 기술

데이터 주체와의 연결을 삭제하고, 데이터 주체와 관련된 특정 특성 집합 및 하나 이상의 가명 간의 연결을 추가하는 처리 방법이다. 일반적으로 가명화는 직접 식별자를 무작위로 생성된 값과 같은 가명으로 대체한다. 직접 식별자의 예로는 이름, 이메일 주소 및 정부 발행 번호가 있다. 모든 직접 식별자 및 잠재적으로 추가되거나 잔존하는 모든 식별 속성을 가명으로 대체한다.

* 1. 일반화 기술

— 라운딩: 간결하고 단순하며 명시적으로 표현할 수 있는 근삿값으로, 수칫값을 대체하는 처리 방법이다.

— 상하단 코딩: 값이 상한 또는 하한보다 높은 속성을 최댓값 또는 최솟값의 임계값으로 설정하는 처리 방법이다.

* 1. 무작위화 기술

— 잡음 추가: 데이터셋의 선택 속성에 예측할 수 없는 임의 잡음을 추가하는 처리방법이다.

— 순열: 데이터셋의 레코드에서 선택 속성자를 수정하지 않고 교환하는 처리 방법이다.

— 부분 총계: 연속 속성의 모든 값이 특정 알고리즘 방식으로 산출된 평균값으로 대체되는 처리 방법이다.

* 1. 합성 데이터

합성 데이터는 사전 정의한 통계 데이터 모델의 마이크로 데이터를 인위적으로 생성하는 접근 방식이다. 정의에 따라, 합성 데이터셋은 기존 데이터 주체에서 수집한 데이터를 포함하지 않지만, 의도된 목적에 따라 실제 데이터처럼 보인다.

1. (참조)  
     
   비식별 프로세스 접근 방식

이 **부속서**는 일반적인 비식별 프로세스 접근 방식에 대한 예시와 설명을 제공한다.

* 1. 데이터 중심 비식별 접근 방식

개인정보 노출을 방지하기 위해 비식별 기술로 원본 데이터를 수정한다는 점을 감안하면, 유용성과 개인정보 보호 간에 상대적 역학 관계가 발생한다. 따라서 정확성 손실을 최소화하면서 개인정보를 보호하는 것이 과제이다. 이상적으로, 데이터 사용자는 원본 데이터를 실행한 분석 결과의 정확성을 비식별 데이터 분석에서도 유지할 수 있어야 한다. 실제로 데이터 유용성의 손상 없이 완벽한 비식별 프로세스를 적용하기는 어렵다. 빅데이터의 경우, 데이터의 양과 다양성으로 인해 이러한 문제가 증가하고 있다. 한편, 식별자를 삭제만 하는 낮은 수준의 비식별 프로세스는 다른 출처의 동일한 개인 또는 유사한 개인의 데이터 연결을 방지할 수 있는데, 이는 결국 빅데이터의 잠재적 이점을 저해할 수 있다. 이번 절은 유용성과 개인정보 보호 간의 역학 관계를 다루는 데이터 중심 비식별 프로세스의 2가지 접근 방식을 설명한다. 데이터 사용에 대한 특정 및 일반 유용성 측정은 비식별 공개 데이터셋의 유용성 측정 방법으로 사용할 수 있다.

**유용성 우선 비식별 접근 방식**

빅데이터의 개인정보는 대체로 독립적인 여러 출처로부터 수집한다. 따라서 동일한(또는 동일 종류/유사한) 개인에 속하는 레코드를 연결하는 기능은 빅데이터 생성의 핵심이다. 유용성 우선 비식별 접근 방식은 휴리스틱 매개변수를 선택하고 적합한 유용성 특성을 유지하는 비식별 기술을 마이크로 데이터셋에서 실행한 후 노출 위험을 측정한다. 따라서 유용성 우선 비식별 접근 방식은 느리고 공식적인 개인정보 보호 보장이 결여된다. 예를 들어, 원본과 비식별 데이터셋 간의 레코드 연결을 시도해 경험적으로 재식별 위험을 추정할 수 있다. 현존 위험이 지나치게 높다고 판단되면, 엄격한 개인정보보호 매개변수로 유용성 손실을 감수하는 비식별 기술을 다시 실행하고, 반복적으로 매개변수를 변경해 통상적인 공식 통계 수준까지 경험적 노출 위험이 감소되어야 한다. 물론 연결 가능성은 유용성 관점에서 바람직하지만, 이는 개인정보 보호에 대한 위협도 된다. 연결의 정확성은 원본 데이터셋보다 비식별 데이터셋에서 상당히 감소해야 한다. 비식별 기술 또는 비식별 프라이버시 보호 모델과 호환되는 연결 가능성에 따라(해당 기술/모델에서) 동일한 개인에 대한 비식별 데이터의 연결 가능성과 연결 방법이 결정된다.

**프라이버시 우선 비식별 접근 방식**

프라이버시 보호 모델은 재식별 노출 위험 및 속성 노출 위험에 대한 상한을 보장하는 매개변수를 적용한다. 이는 모델 매개변수에서 파생된 매개변수와 함께 모델별 비식별 기술을 사용하는 모델이다. k-익명성 및 확장자, -차등 프라이버시는 잘 알려진 프라이버시 모델이며, 대체로 데이터의 유용성/연결 가능성을 저해한다. 프라이버시 우선 비식별 접근 방식을 적용해 생성된 비식별 데이터의 유용성이 지나치게 낮으면, 해당 프라이버시 보호 모델에 유용성 손상이 적은 비식별 기술을 적용하거나, 완화된 프라이버시 매개변수를 선택하거나 다른 비식별 프라이버시 보호 모델을 선택해야 한다.

* 1. 역할 중심 비식별 접근 방식

이 절은 비식별 프로세스의 역할과 책임에 따라 3가지 유형의 접근 방식을 설명한다. 역할 중심 접근 방식은 개괄적으로 ‘누가’, ‘무엇을’ 및 ‘어디에서 어떻게’ 라는 질문에 대한 답변으로 특징지을 수 있다.

— 누가 데이터에 접근하는가?

— 어떤 분석이 수행되거나 수행되지 않는가?

— 데이터 접근/분석은 어디에서 수행되며 접근 권한은 어떻게 취득하는가?

**중앙집중형 비식별 프로세스**

통계 노출 통제 프로세스는 전체 원본 데이터셋에 접근할 수 있는 데이터 처리자가 수행하는 중앙집중형 비식별 프로세스에 중점을 둔다. 이 중앙집중형 접근 방식은 다음과 같은 장단점이 있으며 **표 D.1**에 제시되어 있다.

**표 D.1** — **중앙집중형 비식별 프로세스의 특성**

|  |  |
| --- | --- |
|  | **설명** |
| 장점 | * 개인은 제공하는 데이터 기록을 비식별할 필요가 없다. 전산 처리 능력과 비식별 프로세스에 대한 전문성을 보유한 데이터 처리자가 전체 데이터셋에 적정한 비식별 프로세스를 적용할 수 있다. * 데이터 처리자는 원본 데이터셋에 대한 통합처리 능력을 보유하고 있으므로, 데이터 유용성 및 현존하는 노출 위험 간의 균형을 최적화할 수 있다. |
| 단점 | * 데이터 처리자는 모든 원본 데이터에 접근할 수 있으므로, 원본 데이터를 제공하는 모든 당사자가 데이터 처리자를 신뢰할 수 있어야 한다. 데이터 처리자가 국가 통계 기관인 공식 통계에서는 문제가 되지 않지만, 여러 데이터 출처에서 수집하는 데이터 처리자가 민간 기업(예: 데이터 브로커)인 일반적인 빅데이터 환경에서는 주요 장애물이 될 수 있다. * 특히 빅데이터의 경우, 단일 관리자가 비식별 프로세스를 적용하려면 전산 처리 부담이 지나치게 커질 수 있다. * 관리자 대부분이 단일 빅데이터 처리 환경에 해당하므로 중앙집중형 접근 방식을 관리할 수 없다. |

로컬 비식별 프로세스 및 협력형 비식별 프로세스는 상기의 장단점을 보완한다.

**로컬 비식별 프로세스**

로컬 비식별 프로세스는 개인(데이터 주체)이 데이터를 수집하는 데이터 처리자를 신뢰하지 않는(또는 부분적으로 신뢰하는) 환경(빅데이터 포함)에 적합한 공개 제한 접근 방식이다. 데이터 처리자에게 데이터를 제공하기 전에 각 주체는 비식별 프로세스를 먼저 적용한다. 개인정보 보호에 유의하여 주어진 출처에 의해 수집된 데이터는 제공 전 수집 출처에서 비식별을 적용해야 한다. 그러나 다른 주체의 데이터가 없는 주체가 본인의 데이터에 비식별 프로세스를 적용하므로, 이는 출처 별 독립적인 비식별 프로세스로 인해 중앙집중형 비식별 프로세스보다 정보 손실이 많아진다. 다시 말해, 해당 주체는 데이터셋에 대한 통합 처리 능력이 없어, 노출 위험 제한 및 정보 손실 간의 균형을 찾기가 어렵다.

**협력형 비식별 프로세스**

협력형 프로세스의 높은 개인정보 보호 수준을 결합한 방식이다. 중앙집중형 비식별 프로세스의 문제는 데이터 처리자가 적합하게 데이터를 사용하거나 비식별 프로세스를 적용한다는 것을 신뢰하지 않으면, 데이터 주체는 거짓 데이터를 제공하거나 데이터를 제공하지 않을 수 있는데, 이로 인해 응답 편향 또는 무응답 편향이 발생할 수 있다. 따라서 데이터 주체는 보유한 데이터와 연관된 노출 위험을 결정한 다음, 분산 및 협력 방식으로 적정 보호 수준을 로컬로 적용하고자 협력할 수 있다. 이러한 접근 방식의 2가지 주요 특징은 다음과 같다.

— 동일한 개인정보보호 수준을 적용하는 중앙집중형 접근 방식으로 취득한 데이터셋보다 정보 손실이 크지 않다. 로컬 접근 방식보다 정보 손실이 적다.

— 데이터 주체와 데이터 처리자 모두 기타 특정 데이터의 기밀 속성에 관해 최종 비식별 데이터셋에 포함된 지식 이상을 얻을 수 없다. 데이터 수집기보다 높은 수준의 개인정보보호를 제공하는 중앙집중형 접근 방식보다 신뢰할 수 있다.

또한 협력형 접근 방식은 외부 강제 메커니즘 없이 원활하게 작동하는 프로토콜로 이어질 수 있다. 마이크로 데이터 비식별 프로세스에서 주체가 취득한 개인정보 보호는 다른 사람의 개인정보 보호에 영향을 미친다. 협력형 접근 방식에서 안전한 다자간 변환을 위해서는 당사자가 다른 당사자에게 데이터셋을 제공할 필요 없이 둘 이상의 당사자가 보유한 모든 데이터셋을 변화할 수 있는 전자 프로토콜이 필요하다. 안전한 다자간 변환은 모든 데이터를 중앙집중형으로 저장하지 않고도 쿼리를 변환할 수 있어 데이터 유출로 인한 피해를 줄여 주며, 서로를 완전히 신뢰하지 않는 다자간 계산을 허용한다. 다자간 계산은 특정 상황에서 개인정보보호와 유용성 모두를 개선할 수 있다.

참고문헌

[1] ISO/IEC 11770(all parts), Information technology – Security techniques – Key management

[2] ISO/IEC 18033-6, Information technology security techniques – Encryption algorithms – Part6: Homomorphic encryption.

[3] ISO/IEC 20889 (2018), Privacy enhancing data de-identification terminology and classification of techniques

[4] ISO/IEC 27001 (2018), Information technology – Security technique – Information security management systems

[5] ISO/IEC 29100 (2011), Information technology – Security technique – Privacy framework.

[6] NIST Special Publication 800-38G (2016), Recommendation for Block Cipher Modes of Operation: Methods for Format-Preserving Encryption

[7] NIST 8053(2015), De-Identification of Personal Information

[8] Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. (2004), Order preserving encryption for numeric data, SIGMOD ‘04 Proceedings of the 2004 ACM SIGMOD international conference on Management of data, Paris, France, June, pp564-574

[9] Korean Ministry of the Interior, Guidelines on De-identification Measures. June 2016. <http://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE\_000000000821178&fileSn=2&nttId=7187&toolVer=&toolCntKey>(Last accessed 26 July 2019), <https://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE\_000000000827161&fileSn=0>(English, last accessed 12 December 2019)

[10] UK Anonymization Network, The anonymisation decision-making framework, 2016. <https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>

KS X ITUTX 1148:2020  
해 설

이 해설은 이 표준과 관련된 사항을 설명하는 것으로 표준의 일부는 아니다.

# 1 개요

## 1.1 제정의 취지

이 표준은 2020년 제1판으로 발행된 Framework of de-identification process for telecommunication service providers를 기초로 기술적 내용 및 대응 국제표준의 구성을 변경하지 않고 작성한 방송통신표준으로, 데이터 생명주기 모델을 기반으로 비식별 프로세스에 대한 개요를 제공하며, 운영 단계의 비식별 프로세스 프레임워크 및 비식별 프로세스의 이해 당사자를 규정하기 위해 제정되었다.

또한 비식별 프로세스의 데이터 공개 모델 및 데이터 단계를 설명하며, 부속서와 부록에서는 다양한 비식별 접근 방식과 보기를 제공한다.

## 1.2 주요 제정 내용

통신 조직은 개인정보를 포함하여 개인에 관한 데이터를 수집, 관리, 사용, 및 공유한다. 결과적으로 개인정보 보호를 위해 데이터 비식별 기술을 활용하고 있다.

이 표준은 운영 단계와 함께 비식별 프로세스 데이터 프레임워크를 설명하며, 데이터 생명주기 모델과 이해 당사자 역할을 기반으로 통신 서비스 제공자를 위한 비식별 프로세스의 데이터 공개 모델과 데이터 단계를 규정한다.

이 표준은 다음과 같은 내용으로 구성되어 있다.

* 비식별 프로세스 개요

재식별 가능성을 기반으로 적절한 통제를 선택하기 위하여 데이터 생명주기 단계(수집, 관리, 사용, 공개, 파기)를 정의하고, 이를 기반으로 비식별 프로세스의 고려 시점별 특징과 보안 고려 사항을 설명함

* 비식별 프로세스 프레임워크

비식별 프로세스 프레임워크의 4개의 단계(사전 검토, 비식별, 적정성 평가, 사후 관리)에 대하여 정의하고, 단계별로 수행해야하는 사항과 수행 과정에서의 고려 사항을 설명함

* 비식별 데이터의 유용성

데이터 특성과 개인의 연관을 설명하기 위한 데이터 단계(원본, 가명, 비연결 가명, 총계, 비식별)를 정의하고 단계별 특징에 관해 설명하며, 데이터 분석 상황에 기반한 데이터 공개 모델(공개, 반공개, 비공개)을 정의하고 비교하였으며 데이터 모델과 데이터 단계 간 상관관계를 설명함

**KS X ITUTX 1148:2020**

|  |
| --- |
| **KSKSKS**  **KSKSK**  **KSKS**  **KSK**  **KS**  **KSK**  **KSKS**  **KSKSK**  **KSKSKS** |

|  |
| --- |
| **Framework of de-identification process**  **for telecommunication service providers** |
| **ICS 00.000.00** |