TTA Standard

정보통신단체표준(국문표준)

TTAx.xx-xx.xxxx/R1 제정일: 20xx 년 xx 월 xx 일

비식별 처리를 위한 소프트웨어 프레임워크

Software Framework for De-identification

한국정보통신기술협회
Telecommunications Technology Association

표준초안 검토 위원회 응용보안 및 평가인증 프로젝트그룹(PG504)

표준안 심의 위원회 정보보호 기술위원회(TC5)

	성명	소 속	직위	위원회 및 직위	표준번호
표준(과제) 제안	박설하	금융보안원	사원	-	
표준 초안 작성자	남기우	금융보안 표준화협의회 (이지서티)	책임	-	
	김기태	금융보안 표준화협의회 (파수닷컴)	부장	-	
	김남식	금융보안 표준화협의회 (펜타시스템)	부장	-	
	임형진	금융보안 표준화협의회 (금융보안원)	팀장		
	박설하	금융보안 표준화협의회 (금융보안원)	사원		
사무국 담당	김재웅	TTA	단장	-	
	문서연	TTA	전임		

본 표준 발간 이전에 접수된 지식재산권 확약서 정보는 본 표준의 '부록(지식재산권 확약서 정보)'에 명시하고 있으며, 이후 접수된 지식재산권 확약서는 TTA 웹사이트에서 확인할 수 있습니다.
본 표준과 관련하여 접수된 확약서 외의 지식재산권이 존재할 수 있습니다.

발행인 : 한국정보통신기술협회 회장
발행처 : 한국정보통신기술협회
13591, 경기도 성남시 분당구 분당로 47
Tel : 031-724-0114, Fax : 031-724-0109
발행일 : 20xx.xx.

서 문

1 표준의 목적

이 표준의 목적은 비식별 처리 소프트웨어의 기능을 정의하여, 빅데이터 활용 시 개인정보를 보호하기 위해 비식별 처리 전용 소프트웨어를 도입할 경우 적합한 기능을 가진 소프트웨어를 식별하는데 참조할 수 있도록 하는 것이다.

2 주요 내용 요약

이 표준은 빅데이터 분석·활용 시, 빅데이터 내의 개인정보를 보호하기 위한 비식별 처리 소프트웨어의 구조와 계층별 요구사항 및 기능을 정의한다.

3 인용 표준과의 비교

3.1 인용 표준과의 관련성

- 해당 사항 없음

3.2 인용 표준과 본 표준의 비교표

- 해당 사항 없음

Preface

1 Purpose

The standard can be referred to adopt de-identification software for personal data protection when using big data.

2 Summary

The standard defines architecture, requirements and functions of de-identification software by layer.

3 Relationship to Reference Standards

None.

목 차

비식별 처리를 위한 소프트웨어 프레임워크
(Software Framework for De-identification)

1 적용 범위

이 표준은 비식별 처리 소프트웨어 제품들에 적용할 수 있다. 비식별 처리 소프트웨어란 개인을 식별할 수 있는 정보들을 식별이 불가능 하도록 처리하기 위한 소프트웨어를 말한다.

이 표준은 비식별 처리 소프트웨어의 구조, 요구사항, 기능에 대하여 정의한다. 비식별 처리 소프트웨어의 구조를 4개의 계층으로 구분하고, 각 계층별로 기능 수행을 위한 하위 모듈에 대해 명세한다.

2 인용 표준

해당 사항 없음

3 용어 정의

3.1 가명화(Pseudonymization)
개인식별가능정보를 다른 정보로 대체하기 위해 수행되는 처리 절차 및 방법
[출처] ISO/IEC CD 20889

3.2 익명화(Anonymization)
식별가능 데이터와 정보 주체 사이에 직접 또는 간접적으로 연관성을 제거하는 처리 절차 및 방법

3.3 비식별화(De-identification)
식별가능 데이터와 정보 주체 사이에 연관성을 제거하기 위해 이용 가능한 모든 처리 절차 및 방법
[출처(3.2~3.3)] ISO/IEC 29100

3.4 비식별화 처리(De-identification process)
정보주체와 데이터 식별 속성 사이의 연관을 제거하는 과정

3.5 총계화 데이터(Aggregated data)

정보 주체의 그룹을 나타내는 데이터로서 그 그룹의 통계적 속성 모음

3.6 재식별화(Re-identification)
식별되지 않은 데이터 집합과 데이터 원본의 주체와 연관시키는 처리 절차 및 방법
[출처(3.4~3.6)] ISO/IEC DIS 20889

4 약어

DMB Digital Multimedia Broadcasting
RDBMS Relation Data Base Management System
HDFS Hadoop Distributed File System
ID Identifier
QI Quasi-Identifiers
SA Sensitive Attribute
NSA Non-Sensitive Attribute

5 비식별 처리 소프트웨어의 구조

5.1 빅데이터의 비식별 처리

이 표준에서는 빅데이터 활용을 위한 데이터 처리 과정을 '데이터 수집'→'데이터 처리'→'데이터 제공'의 3가지 단계로 정의하고 각 단계에서 비식별 처리를 위한 차별적 요구사항을 정의한다.

• 데이터 수집 - 데이터 분석 과정에서 프라이버시 노출을 방지하고자 데이터 수집 시점에 식별자 제거 등의 조치가 필요할 수 있다.

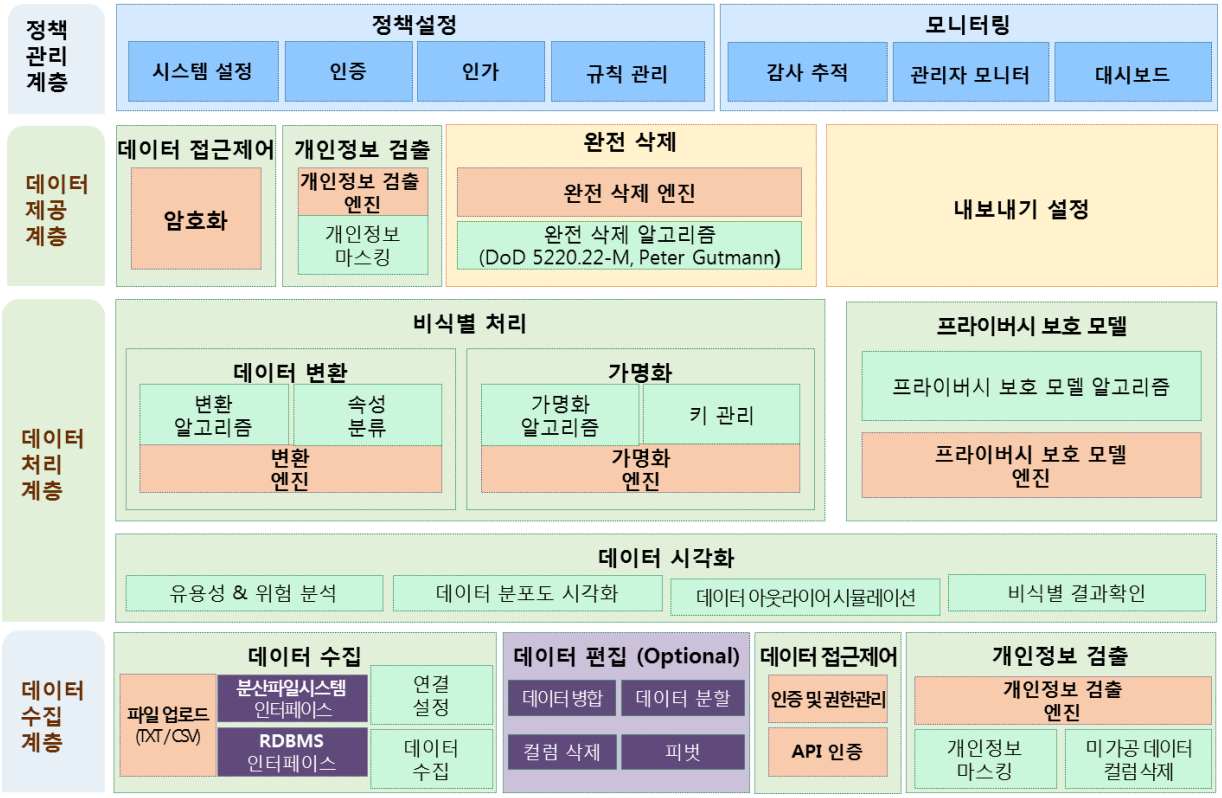
• 데이터 처리 - 데이터 이용 시 특정 개인을 추론하는 것을 방지하고자 비식별 기술을 적용할 수 있다. 데이터 분석 시 다양한 추론에 의한 재식별이 발생하지 않도록 k-익명화 및 확장 방식, 차분 프라이버시 모델 등을 적용할 수 있다.

• 데이터 제공 - 데이터 유출 시 프라이버시 보호를 위하여 식별자 이외의 데이터에 대하여 암호화를 적용할 수 있다.

5.2 소프트웨어 구조

개인정보가 포함된 데이터에 비식별 처리를 수행하는 소프트웨어는 데이터 수집, 데이터 처리, 데이터 제공 및 소프트웨어의 전체적인 관리와 조직의 비즈니스 프로세스를 연계 할 수 있는 정책 관리 역할이 필요하다.

(그림 5-1)에서는 빅데이터 활용을 위한 비식별 처리 소프트웨어의 구조를 나타내고 있다. 구조는 4개 계층으로 나누어져 있으며, 각 계층별로 기능 수행을 위한 하위 모듈이 존재한다.



(그림 5-1) 비식별 처리 소프트웨어의 구조

6 비식별 처리 소프트웨어의 계층별 요구사항 및 기능

6.1 정책 관리 계층

6.1.1 정책 설정(Policy Configuration)

• (PM-시스템설정-R1) 네트워크, 스토리지, 분산 처리 시스템 등의 시스템 설정을 할 수 있어야 한다.

- (PM-시스템설정-R1-C-1) 네트워크, 스토리지, 분산처리 시스템 등 비식별 처리 소프트웨어 시스템의 환경 설정 기능
- (PM-시스템설정-R1-C-2) 시스템 설정을 통한 원본 데이터 접근 기능 (Optional)

• (PM-시스템설정-R2) 분산 처리 기반의 솔루션의 경우, 시스템 설정에서 시스템의 확장 및 축소가 가능해야 한다.

- (PM-시스템설정-R2-C-1) 시스템 설정을 통한 시스템 확장 및 축소 기능

• (PM-사용자 관리-R1) 사용자 관리 기능을 지원해야 한다.
  - (PM-사용자 관리-R1-C-1) 비식별 처리 소프트웨어 사용자에 대한 등록 및 수정 등 관리 기능
  - (PM-사용자 관리-R1-C-2) 비식별 처리 소프트웨어 관리자에 대한 등록 및 수정 등 관리 기능

• (PM-사용자 인증-R1) 사용자 인증 기능을 지원해야 한다.
  - (PM-사용자 인증-R1-C-1) 비식별 처리 소프트웨어 사용자 및 관리자에 대한 비밀번호를 통한 인증 기능(전자금융감독규정의 비밀번호 규칙 준수)
  - (PM-사용자 인증-R1-C-2) 비식별 처리 소프트웨어 사용자 및 관리자에 대한 이중(2-Factor) 인증 기능(Optional)
  - (PM-사용자 인증-R1-C-3) 비식별 처리 소프트웨어 사용자 및 관리자에 대한 생체 인증 기능(Optional)

• (PM-권한부여-R1) 사용자와 관리자에 대한 RBAC(Role-based access control)이 가능해야 한다.
  - (PM-권한부여-R1-C-1) 비식별 처리 소프트웨어 관리자 및 부 관리자에 대한 권한 부여 기능
  - (PM-권한부여-R1-C-2) 비식별 처리 소프트웨어 사용자에 대한 권한 부여 기능
  - (PM-권한부여-R1-C-3) 사용자의 역할 및 권한에 따른 사용자 그룹 관리 기능

• (PM-프로젝트관리-R1) 비식별 처리 프로젝트의 관리가 가능해야 한다.
  - (PM-프로젝트관리-R1-C-1) 단일 비식별화 작업에 대한 프로젝트 관리 기능
  - (PM-프로젝트관리-R1-C-2) 다수의 비식별화 작업에 대한 프로젝트 관리 기능

• (PM-규칙관리-R1) 다양한 규칙 관리가 가능해야 한다.
  - (PM-규칙관리-R1-C-1) 동일한 스키마일 경우 기존의 비식별 규칙을 복제하여 사용할 수 있는 기능
  - (PM-규칙관리-R1-C-2) 개인정보 검출에 필요한 정규 표현식 관리 기능
  - (PM-규칙관리-R1-C-3) 개인정보 검출에 필요한 규칙 관리 기능(예: 예외처리)
  - (PM-규칙관리-R1-C-4) 준식별자로 지정된 컬럼에 적용할 기능의 레벨 관리 기능(예: 주소, 나이 등의 범주화 레벨)
  - (PM-규칙관리-R1-C-5) 자주 사용하는 비식별 기법에 대한 관리 기능

6.1.2 모니터링(Monitoring)

• (PM-사용자감사추적-R1) 사용자의 모든 행위에 대해 로그를 통한 감사 및 추적이 가능해야 한다.
- (PM-사용자감사추적-R1-C-1) 모든 행위에 대한 로그 생성(로그 생성 시 데이터에 대한 내용은 포함하지 않음)
- (PM-사용자감사추적-R1-C-2) 생성된 로그를 통한 사용자 행위 감사 기능

• (PM-사용자감사추적-R2) 로그 생성 시 변조 가능성을 제거해야 하며, 생성된 로그에 대해 변조 여부를 확인해야 한다.
- (PM-사용자감사추적-R2-C-1) 로그 생성 시 변조 가능성 제거 기능(예: 해시값 생성)

• (PM-관리자모니터-R1) 관리자의 모든 행위에 대해 로그를 통한 감사 및 모니터가 가능해야 한다.
- (PM-관리자모니터-R1-C-1) 모든 행위에 대한 로그 생성(로그 생성 시 데이터에 대한 내용은 포함하지 않음)
- (PM-관리자모니터-R1-C-2) 생성된 로그 기반 관리자의 모든 행위에 대한 감사 기능
- (PM-관리자모니터-R1-C-3) 로그 생성 시 해시를 포함하는 등의 방법으로 변조 가능성 제거

• (PM-대시보드-R1) 사용자가 시스템의 상태 등을 쉽게 확인할 수 있도록 대시보드를 제공해야 한다.
- (PM-대시보드-R1-C-1) 사용자 역할별 비식별 처리 소프트웨어 작업 이력 확인 기능
- (PM-대시보드-R1-C-2) 작업 화면 바로가기 기능
- (PM-대시보드-R1-C-3) 비식별 처리 결과에 대한 통계적 시각화 기능
- (PM-대시보드-R1-C-4) 실시간 시스템 리소스 모니터링 기능

6.2 데이터 수집 계층

6.2.1 데이터 수집(Data Import)

• (DC-데이터수집-R1) 파일을 처리하여 데이터를 가지고 오는 기능을 제공해야 한다.
- (DC-데이터수집-R1-C-1) 파일 업로드를 통한 데이터 수집 기능(CSV, TXT 파일 업로드 기능 제공)
- (DC-데이터수집-R1-C-2) 데이터 일부를 샘플링 하여 보여주는 기능

• (DC-데이터수집-R2) RDBMS와 연계하여 데이터를 가지고 오는 기능을 제공해야 한다.(Optional)

- (DC-데이터수집-R2-C-1) RDBMS 연결에 필요한 설정 정보 저장 기능
- (DC-데이터수집-R2-C-2) RDBMS 내 메타 정보 확인 기능
- (DC-데이터수집-R2-C-3) 사용자가 선택한 정보 요소(컬럼)에 대해 SQL 코드 자동 생성을 통한 비식별 대상 데이터 생성 및 수집 기능
- (DC-데이터수집-R2-C-4) 데이터 일부를 샘플링 하여 보여주는 기능

• (DC-데이터수집-R3) 분산 파일시스템(예: HDFS)과 연계하여 데이터를 가지고 오는 기능을 제공해야 한다.(Optional)

- (DC-데이터수집-R3-C-1) 분산 파일 시스템(예: HDFS) 연결에 필요한 설정 정보 저장 기능
- (DC-데이터수집-R3-C-2) 분산 파일 시스템(예: HDFS) 내 메타 정보 확인 기능
- (DC-데이터수집-R3-C-3) 사용자가 선택한 정보 요소에 대해 SQL 코드 자동 생성을 통한 비식별 대상 데이터 생성 및 수집 기능
- (DC-데이터수집-R3-C-4) 비식별 대상 데이터 일부를 샘플링 하여 보여주는 기능

6.2.2 데이터 편집(Data Editing)

• (DC-데이터편집-R1) 비식별 대상 원본 데이터에 대한 편집 기능을 제공해야 한다.

- (DC -데이터편집-R1-C-1) 서로 다른 테이블로 구성된 데이터를 하나의 테이블로 병합하는 기능
- (DC -데이터편집-R1-C-2) 하나의 테이블로 구성된 데이터를 여러 개의 테이블 로 분할하는 기능
- (DC -데이터편집-R1-C-3) 필요 없는 정보 요소를 지우는 기능
- (DC -데이터편집-R1-C-4) 데이터의 행과 열을 변환하는 기능

6.2.3 데이터 접근제어(Data Access Control)

• (DC-데이터접근제어-R1) 별도의 데이터 수집 시스템이나 데이터 전송 시스템이 있는 경우, 데이터 전달 시 인증 및 권한 관리를 통해 안전한 제공이 가능해야 한다.(예: 인가된 사용자만 데이터 암·복호화 가능)

- (DC-데이터접근제어-R1-C-1) 분산 파일시스템(예: HDFS) 및 RDBMS에서 개인정보가 포함된 원본 데이터를 비식별 처리 소프트웨어로 전달 시 적용할 인증 및 관리 기능

• (DC-데이터접근제어-R2) 비식별 처리 시스템은 API 인증을 통해 지정된 시스템

의 정상적인 요청에만 동작하도록 해야 한다.

- (DC-데이터접근제어-R2-C-1) 비식별 처리 소프트웨어가 데이터를 처리하기 전 적용할 API에 대한 인증 기능

6.2.4 개인정보 검출(Privacy Finding)

• (DC-개인정보검출-R1) 원본 데이터에 대하여 개인정보 검출 기능을 제공해야 한다.

- (DC-개인정보검출-R1-C-1) 다양한 개인정보 패턴(정규식)에 의거한 원본 데이터에 포함된 개인정보 검출 기능
- (DC-개인정보검출-R1-C-2) 대용량 원본 데이터에 개인정보 검출을 위한 고속 병렬 처리 기술 적용 기능(예: 인메모리(In-memory) 기반 병렬 처리 기술)

• (DC-개인정보검출-R2) 원본 데이터에서 검출된 개인정보에 대해 마스킹(Masking) 또는 삭제 기능을 제공해야 한다.

- (DC-개인정보검출-R2-C-1) 검출된 개인정보에 대한 마스킹 기능
- (DC-개인정보검출-R2-C-2) 개인정보가 검출된 컬럼 삭제 기능
- (DC-개인정보검출-R2-C-3) 마스킹 및 데이터 컬럼 삭제 기능

6.3 데이터 처리 계층

6.3.1 비식별 처리(De-identification Processing)

• (DP-비식별처리-R1) 개인정보 특성(ID, QI, SA, NSA)을 구분하여 설정할 수 있어야 한다.

- (DP-비식별처리-R1-C-1) 사용자 조작을 통한 수동 속성 분류 기능
- (DP-비식별처리-R1-C-2) 개인정보 정규 표현식 패턴에 의거한 자동 속성 분류 기능(Optional)

• (DP-비식별처리-R2) 다양한 비식별화 기능을 제공해야 한다. (가이드라인에 명기된 17가지 비식별 조치 기법)

- (DP-비식별처리-R2-C-1) 식별자를 정해진 규칙에 의해 대체하거나 사람의 판단에 따라 가공하여 자세한 개인정보를 숨기는 기능
- (DP-비식별처리-R2-C-2) 정보 가공 시 일정한 규칙의 알고리즘을 적용하여 암호화함으로써 개인정보를 대체하는 기능
- (DP-비식별처리-R2-C-3) 기존의 정보 요소를 사전에 정해진 외부 변수 값과 연계하여 교환하는 기능
- (DP-비식별처리-R2-C-4) 데이터 전체 또는 부분을 집계(총합, 평균 등) 처리하는 기능

- (DP-비식별처리-R2-C-5) 데이터 셋 내 일정 부분 레코드만 집계 처리하는 기능(예: 다른 값에 비하여 오차 범위가 큰 항목을 평균으로 변환하는 기능)
- (DP-비식별처리-R2-C-6) 집계 처리된 값에 대하여 라운딩(올림, 내림, 반올림)을 적용하여 최종 집계 처리하는 기능
- (DP-비식별처리-R2-C-7) 개인정보를 타인의 정보와 섞어서 전체 정보에 대한 손상 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 기능(예: 개인이 식별되지 않도록 하기 위해 데이터를 삭제하지 않고 재배열)
- (DP-비식별처리-R2-C-8) 원본 데이터에서 식별자를 단순 삭제하는 기능
- (DP-비식별처리-R2-C-9) 식별자의 일부만 삭제하는 기능
- (DP-비식별처리-R2-C-10) 다른 정보와 뚜렷하게 구별되는 레코드를 전부 삭제하는 기능
- (DP-비식별처리-R2-C-11) 잠재적으로 개인을 식별할 수 있는 속성자까지 전부 삭제하는 기능
- (DP-비식별처리-R2-C-12) 데이터를 평균값으로 변환 또는 범주화하는 기능
- (DP-비식별처리-R2-C-13) 수치 데이터를 임의의 수 기준으로 올림 또는 내림 하는 기능(랜덤 라운딩)
- (DP-비식별처리-R2-C-14) 수치 데이터를 해당하는 값의 범위로 표현하는 기능
- (DP-비식별처리-R2-C-15) 랜덤 라운딩에서 특정 값 변경 시 행과 열의 합이 일치하지 않는 경우, 이를 제어하여 일치시키는 기능(제어 라운딩)
- (DP-비식별처리-R2-C-16) 개인 식별 가능 정보에 임의의 숫자 등 잡음을 추가하는 기능(임의 잡음 추가)
- (DP-비식별처리-R2-C-17) 특정 항목의 일부 또는 전부를 공백 또는 대체 문자로 바꾸는 기능(공백과 대체)

• (DP-비식별처리-R3) 가명화와 관련된 다양한 기법을 사용할 수 있어야 한다.
- (DP-비식별처리-R3-C-1) 가역적 가명화(키 관리 기능 포함)
- (DP-비식별처리-R3-C-2) 비가역적 가명화
- (DP-비식별처리-R3-C-3) 설정에 따라 별도의 시스템 또는 별도의 저장 영역에 보관하는 기능(예: 키 관리 기능)

6.3.2 프라이버시 보호 모델(Privacy Protection Model)

• (DP-프라이버시보호모델-R1) 원본 데이터에 대하여 프라이버시 보호 모델 기능을 제공해야 한다.
- (DP-프라이버시보호모델-R1-C-1) K-익명성(anonymity)
- (DP-프라이버시보호모델-R1-C-2) L-다양성(diversity)
- (DP-프라이버시보호모델-R1-C-3) T-근접성(closeness)(Optional)
- (DP-프라이버시보호모델-R1-C-4) 다른 프라이버시 보호 모델(Optional)

- (DP-프라이버시보호모델-R1-C-5) 기준값에 미 충족하는 결과는 기준값에 충족하도록 조치하거나 범주화하는 기능(솔루션의 특성에 따라 선택 가능)

6.3.3 데이터 시각화(Data Visualization)

• (DP-데이터시각화-R1) 비식별 처리 전·후의 변화를 비교할 수 있는 시각화 기능을 지원해야 한다.

- (DP-데이터시각화-R1-C-1) 비식별 전·후 동질 집합의 분포에 대한 시각화 기능

• (DP-데이터시각화-R2) 각 컬럼별로 비식별 처리 전·후 분포에 대한 시각화 기능을 제공해야 한다.

- (DP-데이터시각화-R2-C-1) 분포도를 조회할 비식별 대상 원본 데이터의 정보 요소에 대한 사전 선택 및 시각화 기능
- (DP-데이터시각화-R2-C-2) 비식별 대상 원본 데이터의 정보 요소별 분포도 시각화 기능
- (DP-데이터시각화-R2-C-3) 비식별 전·후 분포도 비교에 대한 시각화 기능
- (DP-데이터시각화-R2-C-4) 범주화 적용 시 범주화에 따른 분포를 비교하여 보여주는 확인 기능

• (DP-데이터시각화-R3) 원본 데이터의 각 컬럼에서 이상점(Outlier)에 대한 시각화 기능을 제공해야 한다.

- (DP-데이터시각화-R3-C-1) 이상점을 조회할 비식별 대상 원본 데이터의 정보 요소에 대한 사전 선택 및 시각화 기능(모든 정보 요소에 대하여 적용 가능)
- (DP-데이터시각화-R3-C-2) 이상점 기준값 설정 기능
- (DP-데이터시각화-R3-C-3) 사용자가 설정한 기준값에 의거한 비식별 대상 원본 데이터의 정보 요소별 이상점 시각화 기능
- (DP-데이터시각화-R3-C-4) 각 정보 요소에 대해 사용자가 설정한 기준값별 이상점 시뮬레이션 기능

• (DP-데이터시각화-R4) 비식별 처리 후 데이터에 대한 유용성 및 위험성(재식별 가능성) 분석 기능을 제공해야 한다.

- (DP-데이터시각화-R4-C-1) 프라이버시 보호 모델에 따른 유용성 및 위험성(재식별 가능성) 지표 제공 기능
- (DP-데이터시각화-R4-C-2) 비식별 전·후 데이터에 대한 활용도 지표 및 위험도 지표에 의거한 분석 및 결과 제공 기능
- (DP-데이터시각화-R4-C-3) 프라이버시 보호 모델에 따라 적용된 지표의 분석 결과를 제공하는 기능

6.4 데이터 제공 계층

6.4.1 데이터 접근제어(Data Access Control)

• (DP-데이터접근제어-R1) 비식별 처리된 데이터에 대한 안전한 제공을 위해 인증 및 권한 관리 기능을 제공해야 한다.

- (DP-데이터접근제어-R1-C-1) 비식별 후, 데이터 용량을 고려한 암호화 키 길이 및 알고리즘 설정 기능
- (DP-데이터접근제어-R1-C-2) 키 전달을 위한 공개키 암호화 적용 기능
- (DP-데이터접근제어-R1-C-3) 두 개의 경로를 통한 키 전달 기능

6.4.2 개인정보 검출(Privacy Finding)

• (DP-개인정보검출-R1) 비식별 처리된 데이터를 제공하기 전, 개인정보 검출 기능을 수행해야 한다.

- (DP-개인정보검출-R1-C-1) 모든 정보 요소에 대한 전체 검사 기능
- (DP-개인정보검출-R1-C-2) 특정 정보 요소에 대한 검사 기능
- (DP-개인정보검출-R1-C-3) 설정을 통한 강제 적용 기능

6.4.3 완전 삭제(Data Shredding)

• (DP-완전삭제-R1) 원본 데이터 및 비식별 처리 데이터를 필요 시 완전 삭제할 수 있는 기능을 제공해야 한다.

- (DP-완전삭제-R1-C-1) DoD 5220.22-M(3path, 7path) 삭제 알고리즘
- (DP-완전삭제-R1-C-2) Peter Gutmann(35path) 삭제 알고리즘

6.4.4 내보내기 설정 지정(Export)

• (DP-내보내기설정지정-R1) 비식별 처리된 데이터에 대해 다양한 내보내기 옵션을 지정할 수 있어야 한다.

- (DP-내보내기설정지정-R1-C-1) 데이터 순서 변경 기능
- (DP-내보내기설정지정-R1-C-2) 범주화 레코드 표시 여부 설정 기능

6.4.5 임시 대체키 생성(Optional)

• (DP-임시대체키생성-R1) 비식별 처리 데이터 간의 결합을 위한 임시 대체키 생성을 지원해야 한다.

- (DP-임시대체키생성-R1-C-1) 단방향 암호화 알고리즘 지원 기능
- (DP-임시대체키생성-R1-C-2) 솔트(Salt)를 이용한 레인보우 테이블 공격(Rainbow table attack) 방어 기능

부 록 I

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

비식별 처리 소프트웨어 기능 체크리스트

1. 정책 관리 계층(PM)

색인번호	요구사항	색인	기능
시스템설정-R1	네트워크, 스토리지, 분산 처리 시스템 등의 시스템 설정을 할 수 있어야 한다.	C-1	네트워크, 스토리지, 분산처리 시스템 등 비식별 처리 소프트웨어 시스템의 환경 설정 기능
		C-2	시스템 설정을 통한 원본 데이터 접근 기능(Optional)
시스템설정-R2	분산 처리 기반의 솔루션의 경우, 시스템 설정에서 시스템의 확장 및 축소가 가능해야 한다.	C-1	시스템 설정을 통한 시스템 확장 및 축소 기능
사용자 관리-R1	사용자에 대한 관리 기능을 지원해야 한다.	C-1	비식별 처리 소프트웨어 사용자에 대한 등록 및 수정 등 관리 기능
		C-2	비식별 처리 소프트웨어 관리자에 대한 등록 및 수정 등 관리 기능
사용자 인증-R1	사용자에 대한 인증을 지원해야 한다.	C-1	비식별 처리 소프트웨어 사용자 및 관리자에 대한 비밀번호를 통한 인증 기능(전자금융감독규정의 비밀번호 규칙 준수)
		C-2	비식별 처리 소프트웨어 사용자 및 관리자에 대한 이중(2-factor) 인증 기능(Optional)
		C-3	비식별 처리 소프트웨어 사용자 및 관리자에 대한 생체 인증 기능(Optional)
권한부여-R1	사용자와 관리자에 대한 RBAC(Role-based access control)이 가능해야 한다	C-1	비식별 처리 소프트웨어 관리자 및 부 관리자에 대한 권한 부여 기능
		C-2	비식별 처리 소프트웨어 사용자에 대한 권한 부여 기능
		C-3	개별 사용자에 대한 역할 및 권한에 따른 사용자 그룹 관리 기능
프로젝트관리-R1	비식별 처리 프로젝트의 관리가 가능해야 한다.	C-1	단일 비식별화 작업에 대한 프로젝트 관리 기능
		C-2	다수의 비식별화 작업에 대한 프로젝트 관리 기능
규칙관리-R1	다양한 규칙 관리가 가능해야 한다.	C-1	동일한 스키마일 경우 기존의 비식별 규칙을 복제하여 사용할 수 있는 기능
		C-2	개인정보 검출에 필요한 정규 표현식 관리 기능
		C-3	개인정보 검출에 필요한 규칙 관리 기능(예: 예외 처리)
		C-4	준식별자로 지정된 컬럼에 적용할 기능의 레벨 관리 기능(예: 주소, 나이 등의 범주화 레벨)
		C-5	자주 사용하는 비식별 기법에 대한 관리 기능

사용자감사추적-R1	사용자의 모든 행위에 대해 로그를 통한 감사 및 추적이 가능해야 한다.	C-1	모든 행위에 대한 로그 생성(로그 생성 시 데이터에 대한 내용은 포함하지 않음)
		C-2	생성된 로그를 통한 사용자 행위 감사 기능
사용자감사추적-R2	로그 생성 시 변조 가능성을 제거해야 하며, 생성된 로그에 대해 변조 여부를 확인해야 한다..	C-1	로그 생성 시 변조 가능성 제거 기능(예: 해시값 생성)
관리자모니터-R1	관리자의 모든 행위에 대해 로그를 통한 감사 및 모니터가 가능해야 한다.	C-1	모든 행위에 대한 로그 생성(로그 생성 시 데이터에 대한 내용은 포함하지 않음)
		C-2	생성된 로그 기반 관리자의 모든 행위에 대한 감사 기능
		C-3	로그 생성 시 해시를 포함하는 등의 방법으로 변조 가능성 제거
대시보드-R1	사용자가 시스템의 상태 등을 쉽게 확인할 수 있도록 대시보드를 제공해야 한다.	C-1	사용자 역할별 비식별 처리 소프트웨어 작업 이력 확인 기능
		C-2	작업 화면 바로가기 기능
		C-3	비식별 처리 결과에 대한 통계적 시각화 기능
		C-4	실시간 시스템 리소스 모니터링 기능

2. 데이터 수집 계층(DC)

색인번호	요구사항	색인	기능
데이터수집-R1	파일을 처리하여 데이터를 가지고 오는 기능을 제공해야 한다.	C-1	파일 업로드를 통한 데이터 수집 기능(CSV, TXT 파일 업로드 기능 제공)
		C-2	데이터 일부를 샘플링 하여 보여주는 기능
데이터수집-R2	RDBMS와 연계하여 데이터를 가지고 오는 기능을 제공해야 한다.(Optional)	C-1	RDBMS 연결에 필요한 설정 정보 저장 기능
		C-2	RDBMS 내 메타 정보 확인 기능
		C-3	사용자가 선택한 정보 요소(컬럼)에 대해 SQL 코드 자동 생성을 통한 비식별 대상 데이터 생성 및 수집 기능
		C-4	비식별 대상 데이터 일부를 샘플링 하여 보여주는 기능
데이터수집-R3	분산 파일시스템(예: HDFS)과 연계하여 데이터를 가지고 오는 기능을 제공해야 한다.(Optional)	C-1	분산 파일 시스템(예: HDFS) 연결에 필요한 설정 정보 저장 기능
		C-2	분산 파일 시스템(예: HDFS) 내 메타 정보 확인 기능
		C-3	사용자가 선택한 정보 요소에 대해 SQL 코드 자동 생성을 통한 비식별 대상 데이터 생성 및 수집 기능
		C-4	데이터 일부를 샘플링 하여 보여주는 기능
데이터편집-R1	비식별 대상 원본 데이터에 대한 편집 기능을 제공해야 한다.	C-1	서로 다른 테이블로 구성된 데이터를 하나의 테이블로 병합하는 기능
		C-2	하나의 테이블로 구성된 데이터를 여러 개의 테이블 로 분할하는 기능
		C-3	필요 없는 정보 요소를 지우는 기능
		C-4	데이터의 행과 열을 변환하는 기능

데이터접근제어-R1	별도의 데이터 수집 시스템이나 데이터 전송 시스템이 있는 경우, 데이터 전달 시 인증 및 권한 관리를 통해 안전한 제공이 가능해야 한다.(예: 인가된 사용자만 데이터 암·복호화 가능)	C-1	분산 파일시스템(예: HDFS) 및 RDBMS에서 개인정보가 포함된 원본 데이터를 비식별 처리 소프트웨어로 전달 시 적용할 인증 및 관리 기능
데이터접근제어-R2	비식별 처리 시스템은 API 인증을 통해 지정된 시스템의 정상적인 요청에만 동작하도록 해야 한다.	C-1	비식별 처리 소프트웨어가 데이터를 처리하기 전 적용할 API에 대한 인증 기능
개인정보검출-R1	원본 데이터에 대하여 개인정보 검출 기능을 제공해야 한다.	C-1	다양한 개인정보 패턴(정규식)에 의거한 원본 데이터에 포함된 개인정보 검출 기능
		C-2	대용량 원본 데이터에 개인정보 검출을 위한 고속 병렬 처리 기술 적용 기능(예: 인메모리(In-memory) 기반 병렬 처리 기술)
개인정보검출-R2	원본 데이터에서 검출된 개인정보에 대해 마스킹(Masking) 또는 삭제 기능을 제공해야 한다.	C-1	검출된 개인정보에 대한 마스킹 기능
		C-2	개인정보가 검출된 컬럼 삭제 기능
		C-3	마스킹 및 데이터 컬럼 삭제 기능

3. 데이터 처리 계층(DP)

색인번호	요구사항	색인	기능
비식별처리-R1	개인정보 특성(ID, QI, SA, NSA)을 구분하여 설정할 수 있어야 한다.	C-1	사용자 조작을 통한 수동 속성 분류
		C-2	개인정보 정규 표현식 패턴에 의거한 자동 속성 분류(Optional)
비식별처리-R2	다양한 비식별화 기능을 제공해야 한다. (가이드라인에 명기된 17가지 비식별 조치 기법)	C-1	식별자를 정해진 규칙에 의해 대체하거나 사람의 판단에 따라 가공하여 자세한 개인정보를 숨기는 기능
		C-2	정보 가공 시 일정한 규칙의 알고리즘을 적용하여 암호화함으로써 개인정보를 대체하는 기능
		C-3	기존의 정보 요소를 사전에 정해진 외부의 변수 값과 연계하여 교환하는 기능
		C-4	데이터 전체 또는 부분을 집계(총합, 평균 등) 처리하는 기능
		C-5	데이터 셋 내 일정 부분 레코드만 집계 처리하는 기능(예: 다른 값에 비하여 오차 범위가 큰 항목을 평균으로 변환하는 기능)
		C-6	집계 처리된 값에 대하여 라운딩(올림, 내림, 반올림) 기준을 적용하여 최종 집계 처리하는 기능

		C-7	개인정보를 타인의 정보와 섞어서 전체 정보에 대한 손상 없이 특정 정보가 해당 개인과 연결되지 않도록 하는 기능(예: 개인이 식별되지 않도록 하기 위해 데이터를 삭제하지 않고 재배열)
		C-8	원본 데이터에서 식별자를 단순 삭제하는 기능
		C-9	식별자의 일부만 삭제하는 기능
		C-10	다른 정보와 뚜렷하게 구별되는 레코드를 전부 삭제하는 기능
		C-11	잠재적으로 개인을 식별할 수 있는 속성자까지 전부 삭제하는 기능
		C-12	데이터를 평균값으로 변환 또는 범주화하는 기능
		C-13	수치 데이터를 임의의 수 기준으로 올림 또는 내림 하는 기능(랜덤 라운딩)
		C-14	수치 데이터를 해당하는 값의 범위로 표현하는 기능
		C-15	랜덤 라운딩에서 특정 값 변경 시 행과 열의 합이 일치하지 않는 경우, 이를 제어하여 일치시키는 기능(제어 라운딩)
		C-16	개인 식별 가능 정보에 임의의 숫자 등 잡음을 추가하는 기능(임의 잡음 추가)
		C-17	특정 항목의 일부 또는 전부를 공백 또는 대체 문자로 바꾸는 기능(공백과 대체)
비식별처리-R3	가명화와 관련된 다양한 기법을 사용할 수 있어야 한다.	C-1	가역적 가명화(키 관리 기능 포함)
		C-2	비가역적 가명화
		C-3	설정에 따라 별도의 시스템 또는 별도의 저장 영역에 보관하는 기능(키 관리 기능)
프라이버시보호모델-R1	원본 데이터에 대하여 프라이버시 보호 모델 기능을 제공해야 한다.	C-1	K-익명성(Anonymity)
		C-2	L-다양성(Diversity)
		C-3	T-근접성(Closeness)(Optional)
		C-4	다른 프라이버시 보호 모델(Optional)
		C-5	기준값에 미 충족하는 결과는 기준값에 충족하도록 조치하거나 범주화하는 기능(솔루션의 특성에 따라 선택 가능)
데이터시각화-R1	비식별 처리 전·후의 변화를 비교할 수 있는 시각화 기능을 지원해야 한다.	C-1	비식별 전·후 동질 집합의 분포에 대한 시각화 기능
데이터시각화-R2	각 컬럼별로 비식별 처리 전·후 분포에 대한 시각화 기능을 제공해야 한다.	C-1	분포도를 조회할 비식별 대상 원본 데이터의 정보 요소에 대한 사전 선택 및 시각화 기능
		C-2	비식별 대상 원본 데이터의 정보 요소별 분포도 시각화 기능
		C-3	비식별 전·후 분포도 비교에 대한 시각화 기능

		C-4	범주화 적용 시 범주화에 따른 분포를 비교하여 보여주는 확인 기능
데이터시각화-R3	원본 데이터의 각 컬럼에서 이상점(Outlier)에 대한 시각화 기능을 제공해야 한다.	C-1	이상점을 조회할 비식별 대상 원본 데이터의 정보 요소에 대한 사전 선택 및 시각화 기능(모든 정보 요소에 대하여 적용 가능)
		C-2	아웃라이어 기준값 지정 기능
		C-3	사용자가 설정한 기준값에 의거한 비식별 대상 원본 데이터의 정보 요소별 이상점 시각화 기능
		C-4	각 정보 요소에 대한 사용자가 설정한 기준값 별 아웃라이어 시뮬레이션 기능
데이터시각화-R4	비식별 처리 후 데이터에 대한 유용성 및 위험성(재식별 가능성) 분석 기능을 제공해야 한다.	C-1	적용된 프라이버시 보호 모델에 따른 유용성 및 위험성(재식별 가능성) 지표 기능
		C-2	비식별 전·후 데이터에 대한 활용도 지표 및 위험도 지표에 의거한 분석 기능 및 결과 제공 기능
		C-3	활용도 지표를 프라이버시 보호 모델에 따라 변경 후 적용된 기법에 따라 필요한 지표를 제공하는 기능

4. 데이터 제공 계층(DU)

색인번호	요구사항	색인	기능
데이터접근제어-R1	비식별 처리된 데이터에 대한 안전한 제공을 위해 인증 및 권한 관리 기능을 제공해야 한다.	C-1	비식별 후, 데이터 용량을 고려한 암호화 키 길이 및 알고리즘 설정 기능
		C-2	키 전달을 위한 공개키 암호화 적용 기능
		C-3	두 개의 경로를 통한 키 전달 기능
개인정보검출-R1	비식별 처리된 데이터를 제공하기 전, 개인정보 검출 기능을 수행해야 한다.	C-1	모든 정보 요소에 대한 전체 검사 기능
		C-2	특정 정보 요소에 대한 검사 기능
		C-3	설정을 통한 강제 적용 기능
완전삭제-R1	원본 데이터 및 비식별 처리 데이터를 완전 삭제할 수 있는 기능을 제공해야 한다.	C-1	DoD 5220.22-M(3path, 7path) 삭제 알고리즘
		C-2	Peter Gutmann(35path) 삭제 알고리즘
내보내기설정지정-R1	비식별 처리된 데이터에 대해 다양한 내보내기 옵션을 지정할 수 있어야 한다.	C-1	데이터 순서 변경 기능
		C-2	범주화 레코드 표시 여부 설정 기능
임시대체키생성-R1	비식별 처리 데이터 간의 결합을 위한 임시 대체키 생성을 지원해야 한다.	C-1	단방향 암호화 알고리즘 지원 기능
		C-2	솔트(Salt)를 이용한 레인보우 테이블 공격(Rainbow table attack) 방어 기능

부 록 II-1

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

지식재산권 확약서 정보

II-1.1 지식재산권 확약서(1)

- 해당 사항 없음

II-1.2 지식재산권 확약서(2)

- 해당 사항 없음

※ 상기 기재된 지식재산권 확약서 이외에도 본 표준이 발간된 후 접수된 확약서가 있을 수 있으니, TTA 웹사이트에서 확인하시기 바랍니다.

부 록 II-2

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

시험인증 관련 사항

II-2.1 시험인증 대상 여부

- 해당 사항 없음

II-2.2 시험표준 제정 현황

- 해당 사항 없음

부 록 II-3

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

본 표준의 연계(family) 표준

- 해당 사항 없음

부 록 II-4

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

참고 문헌

[1] ISO/IEC CD 20889: Information technology - Security techniques - Privacy enhancing data de-identification techniques,

[2] ISO/IEC 29100 - Information technology — Security techniques — Privacy framework

[3] ISO/IEC 29151 - Information technology — Security techniques — Code of practice for personally identifiable information protection

부 록 II-5

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

영문표준 해설서

- 해당 사항 없음

부 록 II-6

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

표준의 이력

판수	채택일	표준번호	내용	담당 위원회
제1판			비식별 처리를 위한 소프트웨어 프레임워크	PG504