

# 금융 빅데이터 처리를 위한 비식별 처리 기술의 국제 표준화 동향

황송이·임형진 (금융보안원)

목 차	1. 서 론	3. 국제 표준화 동향
	2. 비식별 처리 개념과 국내 추진 현황	4. 결 론

## 1. 서 론

비식별 처리는 식별 가능한 데이터를 수정하여 특정 개인을 식별 할 수 없도록 처리하는 프로세스이며, 비식별 분야에서는 비식별 처리를 수행할 때 데이터의 유용성은 손상시키지 않고 비식별 수준을 높이는 것에 대한 기술적 방법론 연구가 활발하게 진행되고 있다. 또한 비식별 처리 환경과 처리 목적 등에 따라 적절한 수준의 비식별 처리 여부를 판단하기 위해, 데이터 비식별 보장 요구 사항 개발의 필요성이 대두되고 있는 추세이다.

본 고에서는 국내 비식별 처리 추진 현황과 더불어, 현재 비식별 분야에서 표준화된 프레임워크와 데이터 비식별 평가 요구사항 제시의 필요성을 반영하기 위해 개발된 국제표준 2건(비식별 처리 프레임워크), 데이터 비식별 보장을 위한 요구사항2) 및 표준화 추진 동향을 살펴보고자 한다.

## 2. 비식별 처리 개념과 국내 추진 현황

### 2.1 비식별 처리의 개념

가명처리(Pseudonymisation)란 개인정보를 추가 정보 없이는 특정 개인을 알아볼 수 없도록 처리하는 것이다. 즉, 데이터 세트 내에 있는 직접 식별자를 다른 속성으로 교체하는 것으로 예시 기법으로 암호화, 해시가 있다. 가명정보는 가명조치된 정보로, 개인정보이기 때문에 개인정보보호법의 적용을 받으며 제한된 목적하에 기술적·관리적으로 안전조치하여 활용 가능하다. 추가 정보란 가명조치 과정에서 생성되는 정보로 이러한 추가 정보는 원본 정보와 별도로 분리해 보관해야 하고 이에 대한 기술적·관리적 안전조치가 필요하다.

익명처리(Anonymisation)란 합리적으로 예상되는 모든 수단(비용·시간·기술발전)을 고려해도 개인을 식별할 수 없도록 처리하는 것이다. 익명정보는 개인정보가 아니기 때문에 개인정보보호법 적용대상이 아니며 자유롭게 활용 가능하다.

익명처리와 가명처리는 개인정보보호법 제2조 제1호를 활용하기 위해 식별성을 제거하는 안전

1) ITU-T X.1148(Framework of de-identification process for telecommunication service providers)  
2) ITU-T X.1148 - Supplement on requirements for data de-identification assurance

조치라는 점에서는 동일하다. 다만 ‘익명처리’는 합의를 예상되는 모든 수단(비용, 시간, 기술 발전)을 고려할 때 원본으로 연결 및 복원이 불가능한 반면, ‘가명처리’는 추가 정보 없이는 특정 개인을 알아 볼 수 없지만, 추가 정보를 사용하면 원본으로 복원하고 연결할 수 있는 가능성이 존재한다. 따라서 이런 추가적 정보는 반드시 별도로 분리해 보관해야 한다.

## 2.2 국내 비식별 처리 관련 추진 경과

전 세계적으로 기존의 개인정보 보호를 우선했던 관점이 안전한 데이터의 활용 관점으로 변화하면서, 우리나라 정부에서도 개인정보보호와 활용을 동시에 모색하고자, 행정안전부를 포함한 6개 관계부처 합동으로 개인정보 비식별 조치 가이드라인<sup>3)</sup>(이하 가이드라인[1])을 마련하여 데이터 활용 활성화를 추진하였다. 그리고 비식별 조치를 지원하는 전문기관으로 한국인터넷진흥원, 금융보안원, 신용정보원, 한국정보화진흥원, 사회보장정보원 및 한국교육학술정보원을 지정한 바 있다. 가이드라인에 따르면 개인정보처리자는 개인정보를 비식별 조치하여 적정성 평가단으로부터 적정하다는 평가를 받은 뒤 이를 활용할 수 있고, 이종 산업간 비식별 정보 결합도 가능하다.

2018년에는 부처별로 나뉘어 있는 개인정보보호 중복규제를 없애고, 개인정보를 활용할 수 있는 데이터 3법(개인정보보호법, 정보통신망법, 신용정보법)이 발의됐다. 데이터 3법은 2020년 1월 9일 국회 본회의를 통과해 그해 8월 5일 본격 시행됐다.

3) 국무조정실, 행정자치부(現행정안전부), 방송통신위원회, 금융위원회, 미래창조과학부(現과학기술정보통신부), 보건복지부 등 관계부처가 합동으로 개인정보보호법령의 틀 내에서 빅데이터를 안전하게 활용될 수 있도록 개인정보 비식별 조치 가이드라인을 마련, 발표함(2016년 6월 30일)

개인정보보호위원회를 주축으로 정부는 데이터 경제 활성화를 위해 2020년 8월 개인정보보호법 개정으로 가명정보 제도가 도입됐다. 이러한 몇 번의 개정을 거쳐 2022년 5월 ‘가명정보 처리 가이드라인’이 발표됐다.

가명정보 조치 가이드라인에 따라 국내 개인정보보호 전담기관인 한국인터넷진흥원에 개인정보 비식별 지원센터를 설치하였고, 통신, 금융, 보건, 공공 등 관련 5개 부처는 분야별 개인정보 비식별 조치 전문기관을 지정하였다.

분야별 개인정보 비식별 조치 전문기관의 컨트롤타워로서, 전문기관이 원활한 업무 수행을 할 수 있도록 비식별 조치 전문기관 협의회 운영 및 전문기관 운영에 필요한 관련 규정 마련을 지원한다. 또한 가이드라인의 준수여부에 대한 점검, 국내외 가명정보 처리 관련 정책과 제도, 기술을 연구하고 가이드라인을 올바르게 활용할 수 있도록 지원하는 등 관련 정책의 전반적인 점검과 조율을 수행한다.

## 3. 국제 표준화 동향

### 3.1 (ISO/IEC 20889:2018) 비식별 기술 표준

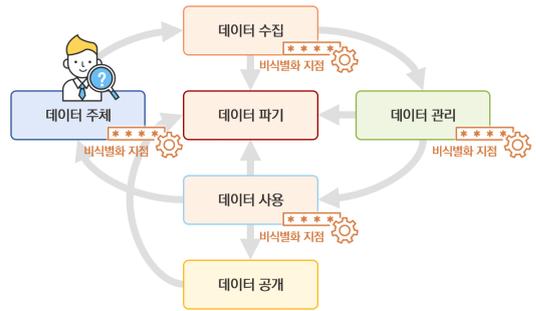
ISO/IEC 20889 표준은 ISO/IEC JTC1에서 개발한 비식별 분야의 최초 국제표준<sup>2)</sup>이며, 표준화된 용어를 사용하여 기존의 비식별화 기술을 분류하고 재식별화의 위험을 줄이기 위한 기반 기술 및 각 기술의 적용 가능성 등의 특성에 대해 설명하는 표준이다. 표준의 주요 내용으로는 재식별 공격의 종류, 비식별 처리 기술의 종류, 일반 프라이버시 측정 모델, 비식별화 기술 적용을 위한 일반적인 원칙, 비식별화 도구/기술/모델의 특성 소개가 있다.

한편, 20889에서 제시된 기술들은 테이블 형식으로 변환할 수 있는 데이터셋에 적용 가능한 내

용이며, 자유 형식의 텍스트나 이미지, 오디오, 비디오를 포함하는 복잡한 데이터셋에는 적용이 불가능하다. 본 표준의 가장 큰 특징은 여러 기술과 정량적인 관점의 위험도 등을 고려했을 때 동일한 공격 위험에 상충된 도움을 줄 수 있는 비식별 기술이 있다면, 유용성 관점에서 최적의 기술을 선택할 것을 제시하는 것이다. 유사한 흐름으로 본 표준에서는 안전하면서도 유용성을 함께 보장하는 공동의 비식별 처리 방법론이 존재하지 않음을 제시하고 있다.

### 3.2 (ITU-T SG17 X.1148) 비식별 처리 프레임워크

금융보안원과 KISA가 ITU-T SG17(정보보호)에서 개발한 ‘비식별 처리 프레임워크 표준[3]’은 2016년 9월 SG17 정기회의에서 금융보안원과 한국전자통신연구원, 순천향대학교, 중국 차이나 모바일이 공동 에디터로 참여하여 2016년 9월에 새롭게 진행가능 과제로 채택된 신규과제이다. 한편, 2016년 8월에 한국인터넷진흥원이 비식별 조치 지원센터로 정식 지정되면서 12월부터 본 표준의 공동 에디터로서 참여하게 되었고 현재는 두 기관에서 주도적으로 최종 표준제정을 위한 절차를 진행 중에 있다.



(그림 1) 데이터 흐름에 따른 비식별 처리 필요지점

본 표준은 비식별 처리절차에 대한 세계 최초 표준으로 전 산업분야에 관계없이 적용이 가능하다. 단, 비식별 처리 세부기술은 ISO 20889와 범위가 중복되므로 별도로 다루지 않고 있다.

비식별 처리절차 표준의 주요 내용은 ①데이터 흐름(라이프사이클)에 따른 비식별 처리가 필요한 지점 정의[그림 1], ②국내 가이드라인에서 제시한 4단계의 비식별 처리 절차[그림 2], ③비식별 정보 제공 모델별 특성과 비식별 처리 수준에 따른 데이터 형태[그림 3]이다.

본 표준은 기존 20889 국제표준에서 각 기술에 대한 소개를 주로 나열했던 것과 달리, 전체 비식별 처리 절차를 프레임워크로 구성·개발하였다는 것과 국내 가이드라인의 비식별 처리 절차를 포함하여 한국이 비식별 조치 분야의 국제표준을 선도



(그림 2) 비식별 처리 절차





(그림 4) 데이터 비식별화 보증을 위한 9가지 요구사항

주체가 적정성 평가단일 경우, 평가시 이용된 K-익명성 모델(사용했을 경우에 한함)에 대한 계량 분석이 포함된다. 데이터 상황에 대한 위험도를 측정해 처리수준을 결정한다. 그리고 처리수준 평가 결과에 따라 익명처리 안전성 평가 기준을 적용한 것에 대한 만족 여부 평가를 한다. 이것은 처리 기준의 적용 여부에 대해 적정성 평가단이 판단해 처리 구준을 충족할 경우 ‘적정’, 그렇지 않을 경우 ‘부적정’으로 판단할 수 있다.

데이터 상황 기반의 비식별 적용 시 비식별 정보가 활용되는 환경 관리는 데이터의 익명화 수준을 결정하는 중요한 요소다. 또한 활용되는 시점에 이전의 데이터에 대한 흔적 역시 매우 중요한 요소다. 데이터 이용 환경은 완전공개 환경, 데이터 이용 합의서를 기반으로 하는 환경, 안전한 고립 환경에서 데이터를 활용하는 밀실 환경(샌드박스)의 세 가지 환경을 다루고 있다. 이에 따라 사후관리 방안도 이 세 가지 환경을 기반으로 기술한다.

데이터 이용 합의서 기반 환경에서의 기술적 관리 방안은 데이터의 제공 형태가 가명정보인지 익명정보인지에 따라 차이가 많다. 가명처리 과정에서 생성된 추가정보(암호키, 매핑표 등)가 유출되

거나 불법적인 재식별 조치에 악용되지 않도록 별도 보관 및 추가적인 기술적·관리적 보호조치 적용이 필요하다. 또한 원본 데이터 및 가명정보와 물리적으로 분리해 저장·관리해야 한다. 다만 불가피한 사유로 물리적 분리가 어려운 경우 테이블 분리처럼 논리적으로 분리하되 엄격한 접근 통제를 적용한다.

익명정보를 기술적·관리적으로 보호하는 방안의 경우 우선적으로 익명정보라 하더라도 이를 다른 개인정보와 연결해 재식별을 시도하는 행위는 금지된다. 익명처리를 통해 익명화된 정보를 활용하는 경우에는 기술의 발전에 따라 재식별화의 가능성이 있는지 모니터링할 필요가 있다.

익명정보의 익명화 수준(익명화 Level 1~4)은 데이터 사용 상황에 따라 평가한 결과, 평가 대상 기관보다 낮은 수준의 보호 수준을 가진 곳으로의 데이터 제공은 엄격히 제한해야 한다. 단 보호 수준이 높은 곳으로의 데이터 제공은 최초 데이터 이용 합의서의 내용을 기준으로 판단할 수 있으나 권장하지 않는다. 이는 기존 20889 국제표준이 비식별 조치와 관련된 기술에 대해 나열했던 것과 유사하게 비식별 조치된 정보의 안전성 보장을 위

해 고려해야 할 요구사항을 ①데이터 상황과 ②데이터 위험도로 분류하여 제시하는 내용으로 개발하였다. 본 표준안건은 2019년 1월 정식으로 채택하여 금융보안원과 한국인터넷진흥원은 2023년 9월 정기회의에서 최종 표준으로 확정되었다.

#### 4. 결 론

최근 챗GPT 열풍으로 북미, 유럽을 중심으로 개인정보 침해 우려가 제기됐다. 챗GPT가 직접 개인정보를 수집하지 않지만 학습을 위해 수집된 데이터에 개인정보가 포함될 수 있다는 이유에서다.

우리나라의 경우 2021년 AI챗봇 서비스 ‘이루다’의 개인정보 유출 사고를 계기로 ‘이루다2.0’에는 가명처리, 개인정보 필터링 기술이 적용되고 있다. 특히 챗GPT, AI 기술, 마이데이터 등 데이터 경제가 빠르게 성장하기 위해서는 개인정보 활용이 필요한 만큼 안전한 활용을 위해 비식별화, 개인정보보호 기술에 대한 관심이 높아질 것으로 전망한다.

특히, 이번 9월 ITU-T에서 최종 채택된 국제표준을 통해 데이터 환경요소와 데이터 기술요소에 따른 비식별 처리 시 보증 수준을 높이기 위한 요구사항을 제시함으로써 전 산업분야에 걸쳐 기업과 기관이 비식별 정보를 올바르게 활용할 수 있도록 함으로서 데이터 산업 경제 활성화에 기여할 수 있을 것으로 예상된다.

fication of techniques, 2018-11.

- [3] 최지선, 이예원, 오용석, 임형진. 비식별 처리 분야의 국제 표준화 동향. 정보보호학회지, 29(4), 13-18, 2019.
- [4] ITU-T Recommendation X.1148 : Framework of de-identification processing service for telecommunication service providers, 2020.
- [5] ITU-T Recommendation X.1148 - Supplement 39 on requirements for data de-identification assurance, 2023.

#### 저 자 약 력



황 송 이

이메일 : songyih@fsec.or.kr

- 2022년 과학기술연합대학원대학교(UST) 정보보호공학과 (석사)
- 2022년~현재 금융보안원 보안연구부 책임
- 관심분야: 금융보안, 차세대 보안 등

#### 참 고 문 헌

- [1] 행정자치부 외, 개인정보 비식별 조치 가이드라인-비식별 조치 기준 및 지원-관리체계 안내, 2016.
- [2] ISO/IEC 20889 - Privacy enhancing data de-identification technology and classi-



### 임 형 진

이메일 : hylim@fsec.or.kr

- 2006년 성균관대학교 컴퓨터공학 (박사)
- 2015년~현재 금융보안원 보안연구부 팀장
- 2009년~현재 과학기술정보통신부 위촉 ICT 국제표준전문가 (정보보호)
- 2019년~현재 여신금융협회 신용카드 단말기 시험 인증 위원회 위원
- 2009년~현재 ITU-T SG17(정보보호) 전문위원 / X.1153 (OTP), X.1157(FDS), X.1148(비식별 처리 프레임워크), X.Supl.39(데이터 비식별화 보증요건) 에디터
- 관심분야: 비식별 처리 기술, 금융보안, 차세대 보안 등