빅데이터 환경에서 개인정보 비식별 처리 방법 표준 개발

임형진(금융보안원 책임연구원, hjlim@fsec.or.kr)

본고에서는 비식별 처리에 대한 국제표준[1] 개발 내용을 기반으로 빅데이터 환경에서 비식별 처리를 위한 고려사항과 비식별 처리를 위해 요구되는 방법을 소개하고자 한다.

정보 통신 기술 및 서비스의 급속한 발전과 폭발적인 성장으로 데이터의 양이 늘어나고 복잡해짐에 따라 기존의 데이터 처리 방법과 도구를 사용하여 주어진 시간 내에 데이터를 효율적으로 분석하는 것이 매우 어려워졌다. 이러한 문제를 해결하기 위해 개발된 패러다임을 빅데이터라고 하며, 기존 데이터 특성에 비해 양, 다양성 및 처리속도 측면에서 더 광범위한 데이터 특성을 가진다. 빅데이터의 중요한 활용성은 다양한 출처로부터 데이터를 결합하여 더 많은 새로운 정보를 도출할 수 있다는 것이다. 일반적으로 기업 간에는 대규모 데이터로부터 새로운 가치를 창출하기 위해 상호 간 데이터 교환을 필요로 한다. 하지만 이러한 데이터 교환을 통해 데이터를 결합할 경우 특정 개인과 관련된 패턴이 나타날 수 있는 위험이 야기된다.

대용량 데이터 기술 및 서비스는 데이터 보존 정책을 유지함과 동시에 전체 데이터 처리 과정에서 개인 식별 정보 및 민감한 데이터의 보호를 요구한다. 실제로 기업은 비즈니스 요구사항과 관련 법규 및 규정에 따라 개인정보보호 및 보안 조치를 수행하게 되며, 결과적으로 기업들은 데이터 교환 과정에서 데이터에 대해 비식별 처리가 필요하게 된다. 또한 기업은 데이터를 수집하기 전·후 또는 저장 전에 비식별 처리를 수행해야 하는지, 혹은 상대 기업과 데이터를 공유하기 전에 비식별 처리를 수행해야하는지에 대한 결정이 필요하다.

비식별 처리란 식별 가능한 데이터를 수정하여 특정 개인을 식별 할 수 없도록 처리하는 프로세스를 의미한다. 데이터의 유용성을 손상시키지 않으면서 높은 수준의 비식별 처리는 불가능하다. 일반적으로 단순히 식별 정보만 삭제하는 낮은 수준의 비식별 처리는 재식별 가능성을 차단하기에 충분하지 않다. 반면에 매우 높은 수준의 비식별 처리는 다른 데이터에서 동일한 혹은 유사한 개인의 데이터를 연결하지 못하게 하여 빅데이터의 많은 잠재적 이점을 저해할 수 있다. 또 다른 쟁점으로, 데이터 처리자가 수집한 데이터를 분석한 후 어떤 목적으로 사용하는지 알 수 있어야 한다는 것이다. 즉 개인 정보의 원래 수집 목적에 한해 사용되는 것이 보장되어야 한다.

---

1) 금융보안원에서는 2016년 ITU-T SG17(정보보호) 정기회의에서 비식별 처리 프레임워크에 대한 표준 개발 필요성을 제안하였고, 회의 참여 국가들과 논의를 거쳐 2019년까지 해당 표준을 개발하는 과제가 채택됨

금융보안원이 주도적으로 추진하고 있는 ITU-T 국제표준에는 기술 중심의 비식별화를 언급하던 기존의 국내·외 가이드라인 및 표준에는 포함되지 않은 빅데이터 활용을 위한 효율적인 비식별 처리 방안을 위한 요소들을 제시하고 있다. 예를 들면, 데이터 모델(생명주기), 데이터 공유모델, 비식별 처리 수준의 개념, 비식별 처리 결과 절차 및 방법 등의 표준 내용을 개발할 계획이다.

<표 1> 가이드/표준 간 비교

항목/가이드 (출간연도)	ISO WD 20889 (2017.6)	ENISA guideline (2015.12)	NISTTR 8053 (2015.10)	UKAN Guideline (2016.9)	NIST 800-188 (2016.12)
비식별 용어	De-identification	Anonymization	De-identification	Anonymization	De-identification
데이터 모델 (생명주기)	Data flow scenario	Data value chain	Data flow model	×	Data lifecycle models
데이터 공유모델	×	×	○	○	○
비식별 처리 수준의 개념	×	×	○	×	○
비식별 처리 기술 분류	○	○	○	○	×
비식별 처리 절차 및 방법	×	×	○	○	○
비식별 처리 결과 평가 방법	×	×	○	○	○
비식별 처리 SW 기능요구사항	×	×	×	×	○

※ 자료: ITU-T draft Recommendation X.fdip(2017.8)

향후 정보의 유용성을 살리면서도 비식별 수준을 적정히 유지할 수 있는 협력형 비식별 처리 방안의 개발이 요구된다. 또한 빅데이터를 처리하는 과정에서 여러 참여자들 간에 다양한 형태의 정보 교환 상황이 발생할 수 있으며, 이 경우 개인정보보호에 대한 책임 소재 이슈가 발생할 수 있다. 빅데이터 활용 과정에서 요구되는 책임 소재를 명확히 규정하고 각 참여자는 부여된 역할 범위 내에서 데이터 처리를 수행하는 것이 필요할 것이다.

<표 2> 비식별 처리 데이터 활용 모델

항목	공개적 데이터 활용 모델	반공개적 데이터 활용 모델	비공개적 데이터 활용 모델
접근 용이성	모든 사람이 자유롭게 데이터 접근 가능	허가된 사람 혹은 기관은 누구나 데이터에 접근 가능	특정 개인 혹은 기관만이 데이터에 접근 가능
활용 사례	웹 포털을 통한 데이터 공개	• 제한된 공간에서의 데이터 접근 • 요청에 의해 정보를 전달 • 원격 접근을 통한 데이터 열람 • 분석 결과만을 전달	기업·기관 간에 데이터 제공
재식별 위험 수준	매우 높은 위험	매우 높은 위험	엄격한 제한 하에 보통 위험
이용 권한	무제한 사용과 재사용	접근 권한이 있는 개인과 기업에게만 제공	데이터의 재활용 금지
가능한 재식별 공격	가능한 모든 공격	• 재식별을 위한 내부자의 의도적인 공격 • 부주의한 데이터 설정에 의한 개인정보 노출 • 데이터 유출	

※ 자료: ITU-T draft Recommendation X.fdip(2017.8)

향후 ITU-T 국제표준회의에서는 이러한 고려사항에 기반하여 우리나라 비식별 처리 가이드라인의 내용을 구조화하고, 타 국가 및 기관들의 기준들을 수용하여 표준안에 반영할 예정이다. 또한 리스크 평가와 관리에 관한 부분과 비식별 처리 절차 등도 포함하여 2019년까지 비식별 국제표준을 개발 완료할 예정이며, 비식별 기술을 중점적으로 다루고 있는 ISO/IEC에서 개발 중인 비식별 처리 기술 전문가들과도 표준개발상황을 공유하여 협력할 예정이다.