빅데이터 기반의 개인정보 비식별화 동향

김 동 국[1], 이 혁[2]

◈ 목 차 ◈

1. 서 론

전 세계 기업의 30%가 이미 교통문제 해결을 위한 미래예측, 재난 및 안전관리, 신규세원 발굴, 세금징수 및 복지서비스의 효율적인 수행을 위해 빅데이터를 '구현'하고 '활용'하는 추세이다. 2013년 20%에 머물렀던 빅데이터의 도입은 2015년 3월 기준 30%로 증가하고 있다. (테크프로 리서치, 2015) 국내에서도 고객관리, 재난공공분야, 제조분야, 보건의료분야, e-Business에 이르기 까지 외국과 비교하여 다양성과 사업 단계 측면에서 유사한 수준으로 활성화되고 있으며 국내 시장 경우 2017년 4,260억 원(연평균 26.9%, NIA)으로 성장될 예정이다[1]. 이와 같이 빅데이터는 비 정형데이터와 데이터의 상관관계를 통해 새로운 성향을 예측함에 따라 그 상업적 가치를 인정받아 IT 환경의 새로운 화두로 급부상하게 되었으며 현재 새로운 형태의 마케팅 시장을 형성하는 원동력이 되고 있다. 하지만 해킹 등 비인가자의 위협으로 인하여 고객의 주민등록번호, 프라이버시 등 민감정보와 같은 중요 데이터가 유출될 경우 기업의 신뢰도 저하 및 국가 신인도 등에 심각한 위험을 초래할 수 있으므로 개인정보를 활용하여 주요 정보는 감추고 나머지 기타 정보로 소비자 성향을 분석하여 빅데이터를 활용하고자 하는 형태인 비식별화 방법을 이용하여 비즈니스에 활용하는 방법을 고려하게 되었다.

본 논고에서는 빅데이터에 환경에서 비식별화를 수행할 시 고려해야 하는 국내외 법률 및 개인 식별정보의 항목을 소개하고 활용단계에서 검토해야 할 비식별화 적용기준을 제시하며 기술 유형에 대해 소개하고자 한다. 또한 향후 식별화를 적용할 때 구현 시 발생할 수 있는 주요 시사점에 대해서도 논의하고자 한다.

2. 국내외 비식별화 법률 현황 및 개인 식별정보

2.1 국내외 비식별화 기반 법률

빅데이터 관련 정보의 개방과 활용의 요구가 증가하면서 방대하고 다양한 유형의 개인정보(SNS 메시징, 위치, 선호, 관심, 웹검색 이력, 성별, 나이 등)를 실시간으로 분석하여 비즈니스 활용의 근거를 제공하는 빅데이터의 특성상 분석과정에서의 개인정보의 노출 우려가 증가하는 상황에 따라 2013년 9월 안전행정부는 빅데이터 활용을 장려하고 효과적으로 개인정보를 보호하기 위한 목적으로 개인정보[3] 비식별화 기준을 발표하였다.

1) 한국IT컨설팅(주)
2) 한국IT컨설팅(주)
3) 비식별화는 데이터 값 삭제, 가명처리, 총계처리, 범주화, 데이터 마스킹 등을 통해 개인 정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 결합하여도 특정 개인을 식별할 수 없도록 하는 조치[개인정보 비식별화 기술활용 안내서]

국내의 경우 '13년 안전행정부, 공공정보 개방 공유에 따른 개인정보 보호 지침'을 필두로 [표1] 개인식별 요소 제거 처리기법 및 상세규칙을 고려하게 되었다.

국내는 미국, 영국과는 다르게 개인정보에 대한 식별 가능한 요소를 정하여 불필요한 요소는 삭제하고 비식별화 검토사항은 전문가 검증도 거치도록 함으로써 명확성을 높이고 있으며 추후 주기적인 모니터링을 통한 재식별 가능성을 완화하는데 주안점을 두고 있다. 또한 비식별화 해야 하는 개인정보의 범위 선정 시 그 자체로 개인식별이 가능한 정보를 열거하여 우선 삭제하도록 명시하고 있다.

해외 경우는 주로 미국, 영국, 일본에서 시행하고 있는 비식별화 지침 및 가이드의 주요 특징에 대해서 다루고자 한다.

우선 미국은 '12년3월 개인정보의 비식별화 가이드라인(FTC, '12.3)를 수립하여 특정한 소비자, 컴퓨터 및 기타 개인을 식별할 수 있는 장치들과 연관될 수 있는 것(reasonable linkability)은 어떤 정보라도 보호의 대상이 되어야 한다고 규정하고 있다.

- FTC는 특정데이터가 비식별화 기준을 아래의 3가지 내용으로 제시하고 있으며 업계가 이를 이행한다면 특정 데이터가 식별가능하지(reasonable linkable) 않은 것으로 판단하고 있다

① 개인, 컴퓨터 및 디바이스에 대한 정보를 추론(식별)할 수 있는 데이터의 삭제, 수정, 'noise' 추가, 통계적으로 샘플링, 총계처리 등의 적절한 방법을 자율적으로 판단하여 반드시 비식별 조치를 취함

② 데이터를 공개할때 해당 데이터를 비식별화하여 이용한다는 점, 향후 비식별화 상태를 유지할 것, 재식별하지 않을 것을 개인정보주체에게 공개적으로 약속함

③ 데이터를 제3자에 제공할 때 해당 데이터를 제공받은 어떠한 이용자도 재식별하지 않도록 계약상 반드시 요구함

〔표1〕 개인식별요소 제거 처리기법 및 상세규칙 〔4〕

구분	공공정보 개방·공유에 따른 개인정보 보호지침	빅데이터 활용을 위한 개인정보 비식별화 사례집	빅데이터 개인정보보호 가이드라인	개인정보 비식별화에 대한 적정성 자율평가 안내서	빅데이터 활용을 위한 개인정보 비식별화 기술활용안내서
발표시기	2013.09	2014.05	2014.12	2014.12	2015.05
주관부처 및 기관	-안전행정부	-미래창조과학부/한국정보화진흥원	-방송통신위원회	-행정자치부/한국정보화진흥원	-미래창조과학부/한국정보화진흥원
활용분야	-공공	-공공/민간	-공공/민간(주로 정보통신서비스제공자)	-공공/민간	-공공/민간
내용	-지침개요 -기본방침 -개인정보처리단계별 준수사항 -비식별화 조치방법	-개요 -빅데이터 활용 단계별 개인정보 비식별화 처리 -비식별화 처리 사례	-목적 -정의 -개인정보의보호 -공개된 정보의 수집·이용 -이용내역정보의 수집·이용 -새로운 정보의 생성 -민감정보 생성의 금지 -통신내용의 조합 · 분석 또는 처리금지 -공개된 정보 및 이용내역정보의 이용 -제3자제공 -적용범위	-개요 -개인정보재식별 및 위협요소 -개인정보비식별화에 대한 적정성평가 -재식별위험관리 방안	-개요 -분야별 개인정보 참고법령 및 조치사항 -비식별화기술 실무 활용방법

- 즉, 비식별화 해야 하는 데이터의 범위는 개인 뿐 아니라 개인의 각종 디바이스의 식별 가능성이 최소화되는 수준으로 업계가 방법을 자율적으로 선택하여 반드시 조치하되, 재식별에 대해서는 개인정보 주체에게 재식별하지 않을 것을 공개적으로 약속하도록 하고, 제3자에게는 비식별화 데이터 제공 시에도 계약상에 재식별 방지를 요구하도록 권고하고 있다[9].

영국의 경우 '12년 12월 정보 감독청(ICO), '데이터 비식별화 실행 규칙에 따라 지침을 수립하고 있다. 데이터 비식별화 실행 규칙은 비식별화 해야 하는 개인정보의 범위는 개인 식별 가능성이 최소화되는 수준으로 업계가 자율적으로 판단하여 조치하도록 하고 있으며 비식별화 방법에 대한 주요 특징은 다음과 같다.

① 재식별에 대해서는 최소한의 검증(motivated intruder test)을 거친 후 비식별 데이터의 사용제한 및 접근 통제를 통해 철저한 사후관리를 강조
② 정해진 수신자만 동의된 목적에 따라서만 이용 가능, 보안 등 훈련 받은 수신자만 접근, 재식별 시도 금지, 직원 비밀 준수 의무 부과, 데이터 접근 암호화, 복사 제한 등[9]

일본에서는 빅데이터의 방대한 정보를 활용하여 신산업 기회를 창출하고 경제재생계획을 실현하기 위해 개인정보보호에 관한 법률('14.06)에 빅데이터 관련 규칙을 넣어 개정하였으며 주요 특징은 다음과 같다.

① 일본총리 산하 IT종합전략본부는 전략본부 내 '개인데이터 검토회'에서 쟁점이 되었던 개인정보 구분 및 보호 요건 완화함
② 상품 구매이력, 웹사이트 이용 현황, 스마트폰 위치정보 등의 데이터는 본인 동의가 없어도 개인이 특정되지 않는 비식별화를 조건으로 제3자에게 제공하는 것을 인정함
③ 빅데이터 환경에서 다루는 대부분의 정보가 누구의 것인지 특정할 수 없도록 비식별화 처리를 하는 것을 권고함
④ 빅데이터를 활용한 신산업 창출 및 확대를 위해 규제완화 등의 조치가 필요하며, 이와 병행하여 사생활 침해 등을 방지하기 위한 개인정보 보호 장치의 면밀한 검토가 필요함[10]

2.2 비식별화를 위한 국내 개인식별정보

국내에서는 현행 법률을 기반으로 개인식별정보에 대하여 침해사고 발생 시 그 피해를 최소화 하기 위하여 법규에서 제시하는 비식별화 과정을 수행하는 것을 고려하고 있다. 국내 개인정보 비식별화는 일반, 공공, 민간으로 구분되며 민간경우는 정보통신, 상거래, 금융·신용, 보건·의료로 4부분으로 구분한다. 비식별화된 개인정보는 기술발전에 따라 다른 개인정보와의 조합으로 인하여 식별 개인정보로 인식될 수 있으므로 정기적으로 비식별 정보를 구분하여 분류해야 한다. 가령, 주민등록번호, 운전면허, 주소, 연락처 등은 개인을 식별할 수 있는 개인정보이지만 행태정보, 성향정보, 위치 정보 등은 비식별 정보로서 기술발전에 따른 상관 분석을 통해 개인 식별이 가능할 수 있다. [표 2]는 국내 비식별화를 위한 법률 기반에 따른 개인식별정보에 대한 법적 근거와 항목에 대한 예를 제시하고 있다.

〔표2〕 법률기반 개인 식별정보

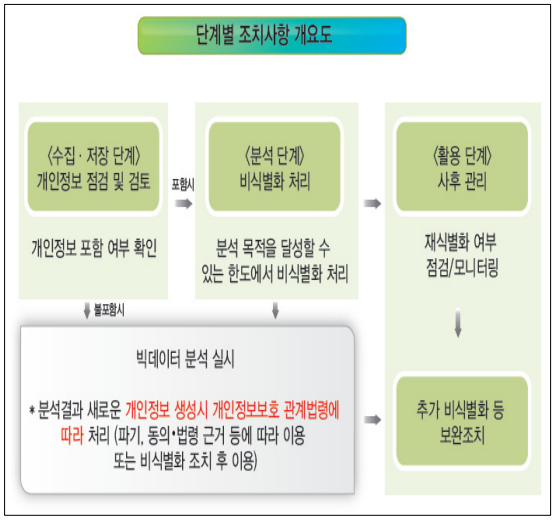
구분	근거	개인 식별 정보 항목
일반	• 개인정보보호법 제18조, 제23조, 제24조 제1항, 제24조 제2항, 제24조 제3항	• 주체자의 사생활을 침해할 수 있는 식별정보 (ex. 의료정보, 정신적 성향 등) • 주체자의 신분 확인을 위한 일반 식별정보 (ex. 이름, 주민등록번호, 주소 등)
공공 부문	• 전자정부법 제42조	• 정당한 사용자임을 인증하는 식별정보 (ex. 인증서 일련 번호, 유효기간 등)

구분		근거	개인 식별 정보 항목
		• 주민등록법 제10조	• 신분 확인정보와 가족구성원 정보를 통해 확인될 수 있는 식별정보 (ex. 성명, 성별, 세대주와의 관계 등)
		• 공공기관의 정보공개에 관한 법률 제18조 • 공공기록물 관리에 관한 법률 제37조	• 주체자의 신분 확인을 위한 일반 식별정보 (ex. 이름, 주민등록번호, 연락처 등)
		• 민원사무처리에 관한 법률 제26조 • 국가정보화 기본법 제39조	• 본인·대리인 확인을 위한 식별정보 (ex. 주민등록번호, 대리인 신분증 등)
민간 부문	정보통신	• 정보통신망 이용촉진 및 정보보호 등에 관한법률 • 전자서명법 제24조	• 회원 관리를 위한 사용자 식별 정보 (ex. 이름, ID, PW 등) • 정당한 사용자임을 인증하는 식별정보 (ex. I-PIN인증, 단말정보, 휴대폰정보 등)
		• 전자금융거래법 제25조	• 휴대폰 결제 서비스 수행을 위한 식별정보 (ex. 결제수단별 개인정보, 카드번호, 비밀번호 등)
		• 전기통신사업법 제83조	• 주체자의 신분 정보 및 통신상의 사용자 정보에 대한 식별정보 (ex. 이름, ID, 주민등록번호 등)
		• 위치정보보호법 • 통신비밀보호법	• 업무 수행 및 처리를 위한 통신상의 식별정보 (ex. 접속 IP정보, GPS 정보 등)
		• 청소년보호법 제29조, 제16조	• 제한된 연령 확인에 대한 식별정보 (ex. 법정 생년월일, 법정 대리인 정보 등)
	상거래	• 전자문서 및 전자거래기본법 제12조 • 정보통신망 이용촉진 및 정보보호 등에 관한 법률 제23조, 제24조 • 전자상거래 등에서의 소비자보호에 관한 법률 제12조	• 전자문서 서비스를 위한 식별정보 (ex, 공인전자주소, 송신자, 수신자 등) • 통신의 안전한 조치를 위해 확인할 수 있는 식별정보 (ex. 비밀번호, 계좌번호, 주민등록번호 등) • 거래 기록 및 배송을 확인하기 위한 식별정보 (ex. 배송 주소지, 수령인 연락처 등)
		• 전자서명법 제24조	• 정당한 사용자임을 인증하는 식별정보 (ex. 가입자 이름, 전자서명검증정보, 인증서 일련번호)
	금융신용	• 신용정보의 이용 및 보호에 관한 법률 제33조 • 금융실명거래 및 비밀 보장에 관한 법률 제4조	• 신용정보 및 거래능력을 판단할 수 있는 식별정보 (ex. 재산, 소득, 대출 보증 등) • 금융기관의 거래내역을 판단할 수 있는 정보 (ex. 주민등록번호, 계좌번호, 거래실적 자료 등) • 이용자 및 거래내용의 정확성을 확인하기 위한 식별정보 (ex. 전자금융업자에 등록된 이용자번호, 이용자의 생체 정보, 등)
		• 전자금융거래법 제26조 • 전자금융 감독규정 제5조의 3	• 신용정보 및 거래능력을 판단할 수 있는 식별정보 (ex. 재산, 소득, 대출 보증 등) • 금융기관의 거래내역을 판단할 수 있는 정보 (ex. 주민등록번호, 계좌번호, 거래실적 자료 등) • 이용자 및 거래내용의 정확성을 확인하기 위한 식별정보 (ex. 전자금융업자에 등록된 이용자번호, 이용자의 생체정보, 등)

구분		근거	개인 식별 정보 항목
		• 특정 금융거래 정보의 보고 및 이용 등에 관한 법률 제5조의 3	• 자금이체 수행을 위한 식별정보 (ex. 송금인 성명, 계좌번호, 수취인의 정보)
	보건 의료	• 의료법 제21조 응급의료에 관한 법률 제22조의 2항 • 산업안전보건법	• 정확한 환자의 진료를 위해 확인가능 한 식별정보 (ex. 주민등록번호, 의료기록, 가족력 등) • 신체의 질병정보를 통해 인지될 수 있는 식별정보 (ex. 감염병명, 혈액정보, 조직정보 등)
		• 후천성면역결핍증예방법 • 감염법의 예방 및 관리에 관한 법률 제74조	• 정확한 환자의 진료를 위해 확인가능 한 식별정보 (ex. 주민등록번호, 의료기록, 가족력 등) • 신체의 질병정보를 통해 인지될 수 있는 식별정보 (ex. 감염병명, 혈액정보, 조직정보 등)
		• 장애인 차별금지 및 권리규제 등에 관한 법률 제22조	• 신체 장애정보를 통해 확인 가능한 식별정보 (ex. 주민등록번호, 신체장애, 장애등급 등)
		• 국민건강보험법 제5조	• 가족구성원의 정보를 통해 확인할 수 있는 식별정보 (ex. 가족구성원의 이름, 출생지, 소득 등)

2.3 빅데이터 활용 단계별 비식별화 적용

개인정보 비식별화 단계별 조치사항인 [그림1]은 비식별정보, 생성정보, 공개정보등을 통해 수집되어지는 정보를 단순정보와 개인정보로 구별하고 분석단계에서는 목적을 달성할 수 있는 한도 내에서 비식별화 처리하는 일련의 과정을 개요를 통해 제시하고 있다.



(그림1) 비식별화 절차 개요도 [5]

분석결과 개인정보보호 관계법령에 따라 새로운 개인정보가 생성 시 정보주체의 동의를 받은 후 법령근거에 따라 이용 또는 비식별화 조치 후 이용한다. 그 이후 활용단계에서는 재식별화 여부를 점검 모니터링하고 추가되는 비식별화 등에 대한 보완조치를 이행하도록 한다. 예를 들어, 서해 또는 남해 등 외딴 섬에 거주하는 소수집단의 80대 이상 연령 등은 간단한 정보만 확인이 되더라도 재식별이 가능하기 때문에 추가적인 조치가 필요하다.

또한, 여타 개인정보와 같이 빅데이터에서 이용되는 정보도 수집, 저장, 분석, 이용·제공, 파기에 따른 개인정보 생명주기에 의거한 처리 기준을 정의할 수 있다.

수집단계에서는 데이터 수집 대상이 특정 개인인 경우 법률의 허용 규정이 있거나 사전에 정보주체의 동의를 얻어야 하며 개인정보가 포함된 공개된 데이터를 수집하는 경우 법률의 허용 규정이나 정보주체의 사전 동의가 없으면 개인정보를 비식별화해야 한다. 또한 주민등록번호, 민감정보 등 수집 제한 데이터를 수집하지 않도록 유의해야 한다.

저장단계는 개인정보가 저장, 처리되는 빅데이터 처리 시스템(정보 조합, 분석, 처리 시스템)에 대하여는 관계 법령에 따른 안전 조치 또는 보호조치를 하여야 한다. 수집단계에서 필터링 되지 않거나 저장된 데이

터를 유형화하는 과정에서 추출된 고유식별정보는 암호화 또는 다시 비식별화 하는 등의 보호조치를 해야 한다. 그리고 분석 단계는 개인정보가 포함된 공개정보, 이용내역 정보는 비식별화 조치를 한 다음 조합, 분석 또는 처리를 해야 하며 목적을 달성한 후 지체없이 파기 또는 비식별화 조치를 하여야 한다.

이용·제공 단계는 비식별화된 공개된 개인정보 등을 서비스 제공을 위하여 내부에서 이용할 때에는 정보주체가 이를 쉽게 확인할 수 있도록 공개해야 하며 공개된 개인정보 등의 분석 결과를 사생활 침해, 사회적 차별 조장 기타 사회 질서에 반하는 목적으로 이용해서는 안 된다.

파기단계는 개인 식별 정보가 포함된 분석 결과의 이용 목적이 달성되거나 보유기간이 경과한 경우 해당 개인식별정보를 즉시(통상 5일이내) 파기하거나 비식별화 조치를 취해야 한다.

3. 비식별화 적용 기준 및 기술유형

3.1 비식별화 적용 기준

비식별화에 대한 적용 기준은 세 가지로 요약할 수 있다. 첫째, 그 자체로 개인 식별이 가능한 정보는 삭제해야 한다. 단 수집 시 개인정보에 대한 자체이용, 제 3자 제공 등 활용에 대한 이용자 동의를 받았을 경우에는 비식별화 없이 활용이 가능하다.

둘째, 다른 정보와 결합에 따른 재식별 위험을 최소화해야 한다. 비식별화 기술은 재식별에 대한 일정한 한계를 가지므로 비식별화 목적을 명확히 하여 재식별화에 대한 위험 요소를 최소화 하도록 한다.

셋째, 정보가 식별될 수 있는 리스크를 고려하여 사후관리는 철저해야 한다. 특히 개인정보에 대한 처리 기술이 발전함에 따라 재식별 사례를 분석하여 주기적으로 비식별화 처리 기법개선에 반영해야 하며 빅데이터 분석과정에서 불필요한 개인정보가 새로 생성되거나 비식별화 처리된 정보가 재식별화 된 경우에는 지체없이 삭제하거나 비식별화 처리를 해야 한다[6].

3.2 비식별화 기술 유형

비식별화를 위한 적용기술은 5개 처리기법(가명처리, 총계처리, 데이터 값 삭제, 범주화, 데이터 마스킹)에 속한 총 18개 세부기술로 나눌 수 있다. [표3]은 비식별화 적용을 위한 세부기술유형에 대한 처리 예시이다.

〔표 3〕 비식별화 주요 기술 유형

처리기법	세부 기술	주요 내용 및 처리 예
가명처리 (Pseudonymisation)	① 휴리스틱 익명화 ② K-익명화 ③ 암호화 ④ 교환 방법	개인정보 중 주요 식별 요소를 다른 값으로 대체하여 개인 식별을 곤란하게 함 예) 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대 서울 거주, 국제대 재학
총계처리 (Aggregation)	⑤ 총계처리 ⑥ 부분집계 ⑦ 라운딩 ⑧ 데이터 재배열	데이터의 총합 값을 보임으로써 개별 데이터의 값을 보이지 않도록 함 예) 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm
데이터 값 삭제 (Data Reduction)	⑨ 속성값 삭제 ⑩ 속성값 부분 삭제 ⑪ 데이터 행 삭제 ⑫ 식별자 제거를 통한 단순 익명화	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
범주화 (Data Suppression)	⑬ 범주화 ⑭ 랜덤 올림 방식 ⑮ 범위 방법 ⑯ 제어 올림	데이터의 값을 범주의 값으로 변환하여 명확한 값을 감춤 예) 홍길동, 35세 → 홍씨, 30-40세
데이터 마스킹 (Data Masking)	⑰ 임의 잡음 추가 ⑱ 공백과 대체	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

출처 :개인정보 비식별화 기술활용 안내서

4. 결 론

본 기고에서는 현재 빅데이타 환경에서 향후 비식별화를 하기 위한 준거성 관점에서 국내외 법률 기반의 개인식별정보를 파악하고 빅데이터의 활용단계인 수집, 저장, 분석, 이용·제공, 파기 단계에서 검토해야 할 비식별화를 위한 검토사항과 비식별화 적용 기준에 대해서 살펴보았으며 비식별화 기술 18가지에 대한 유형에 대해서도 알아보았다.

비식별화는 공공, 민간기업의 마케터가 수많은 고객 데이터를 활용하여 빅데이터 기술 관점에서 수집, 가공, 분석의 결과를 활용할 때 사용되는 방법으로 국내의 비식별화는 “방송통신위원회의 <빅데이터 가이드라인>이나 현재 발의된 법안들은 ‘익명화’가 아닌 ‘비식별화’라는 개념으로 규정하고 있다. ‘비식별화(de-identification)’는 익명화(anonymisation)와 달리 재식별화(re-identification)의 가능성을 내포하고 상업적 활용성을 보장하고자 하는 개념으로서 핵심 취지가 기업으로 하여금 정보주체의 동의 없이 개인정보를 수집 및 처리할 수 있게끔 허용하겠다는 것이다[7][8]. 현재 국내 개인정보보호법 제18조 제2항 제4호에 따르면 현행 개인정보보호법은 개인정보 주체의 개인정보자기결정권을 실질적으로 보장하기 위하여 개인정보처리자가 개인정보를 ‘익명화’하더라도 정보주체의 동의 없이는 해당 정보를 통계 목적이나 연구 목적 등으로 제공하는 경우 외에는 제공할 수 없도록 규정하고 있다. 하지만 기업이나 공공기관에서 개인정보가 포함된 빅데이터를 통해서 자사의 마케팅 등의 목적에 이를 이용하려고 하고 있다는 점이며 그에 따른 비식별화를 위한 현실적인 주요 쟁점은 ‘[4]구글 데이터 등 개인정보를 식별할 수 있는 개인정보의 범위가 넓기 때문에 비식별화 해야 하는 데이터의 범위도 명확치 않다는 문제가 있다는 점이다. 또한 비식별 개인정보는 공개 시점에 외부 데이터와의 조합을 통해 개인이 식별되거나, 향후 분석 기술이 진화하고 공개정보가 많아질수록 식별가능성이 커진다는 문제가 발생한다.[1] 해외 사례를 예로 들면 ‘라타냐 스위니’(Latanya Sweeney) 교수가 매사추세츠주의 단체보험위원회가 ‘이름, 주소, 사회보장번호와 그 외의 식별정보’는 제거하였으나 ‘환자관련 100여개의 속성정보’는 미삭제하여 공개한 주정부 소속 공무원의 무료병원 출입기록 데이터와 ‘이름, 성별, 생년월일, 우편주소’가 포함된 판매용 투표명부를 매칭하여 확인한 매사추세츠 주지사인 ‘윌리엄 웰드’(William Weld)의 의료정보, 거주지 정보, 우편번호를 알아낸 사건은 개인에 대한 단순 정보를 조합하여 개인정보를 식별한 하나의 사례로 들 수 있다. 또한 국내 사례 경우 2013년 ERTI 보고서에 따르면 페이스북 667만개, 트위터 277만개의 한국인 이용자 계정에 업로드한 데이터를 이용해서 개인에 대한 재식별 가능성 분석했는데 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 분석되었다[6]. 결국 이와 같은 사례를 통하여 언제 어떠한 방식으로 정보가 결합하여 개인 식별 가능성을 가질지 알 수 없는 모든 정보들에 대해 사전에 모든 이용자들의 동의를 구하는 것은 현실적으로 불가능하다는 것을 알 수 있다.

따라서 정보의 공유와 개방요구가 점차 증대되고 기술발전이 진행되고 있는 빅데이타 환경에서는 필연적으로 완벽한 개인정보보호의 보장은 어렵지만 개인정보의 철저한 비식별처리와 추후 재식별 방지를 위한 지속적인 모니터링 등의 사후관리를 해야 한다. 또한 식별 가능성에 대한 일정한 기준과 이용자 사전동의를 통한 합리적 적용 및 업계의 자율적인 식별방지 노력이 필요하며 이에 대해 다양한 정보주체간에 합의가 선행되어야 할 것이다.

4) 정보주체의 자기결정권의 보호를 위해 주로 ‘동의’ 방식을 활용해왔으나 최근 빅데이타 환경에서는 데이터의 수집, 분석, 결과를 활용한 시점에 따라 데이터의 범위를 특정해내는 것이 불가능하다. 가령 2016년 가동 예정인 칠레의 LSST(large synoptic telescope) 망원경은 5일에 140테라바이트18)의 데이터를 수집하게 되며, 미국의 주식시장에서는 매일 진행되는 약 70억 주의 거래 가운데 ‘2/3’가 컴퓨터 알고리즘에 의해 진행되는 데이터 처리의 산물이다. 이처럼 각종 데이터의 수집으로 발생되는 데이터에 대해서 매번 센서들에 대해 수집하는 무궁무진한 정보들도 있다. 산업현장에서 사고예방을 목적으로, 또는 자연재해의 극복이나 보안의 목적으로 설치되어 끊임없이 신호를 보내오고 있는 수많은 센싱 데이터(sensing data)에 대해서 정보주체에 대한 기준을 가지고 매번 동의를 받을 수 있는 기준의 정립이 향후 필요하다[2][3].

참 고 문 헌

[1] 한국정보화진흥원, “정보화 통계집”, 2014
[2] Paul Ohm, “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization”, 57 UCLA Law Review, 1719~1720쪽, 2010
[3] 오길영, “데이터 상업화 과정으로서의 개인정보 비식별화”, 민주법학 제58호, 8쪽, 2015.07
[4] 정영철, “의료분야 빅데이터 활용을 위한 개인정보 비식별화 규정 현황과 과제”, 한국보건사회, 6쪽, 2015. 9
[5] NIA, 빅데이터 활용 단계별 개인정보 비식별화 처리
[6] KISA, “빅데이터_비식별화_기술_활용_안내서, 14 ~ 44쪽, 2015. 06
[7] 고학수 “빅데이터 산업 활성화를 위한 법적과제”, 한국정보법학회 정기학술세미나, 34쪽, 2015. 6
[8] 경실련 & 진보네트워크 보도자료, 2015. 6
[9] DIGIECO, “개인정보 비식별화 동향 및 시사점”, issue & trend, 6 ~ 7쪽, 2013.11
[10] 과학기술정책연구원 국외정책 동향, 2014. 07.21 http://www.stepi.re.kr/app/snt/view.jsp

◐ 저 자 소 개 ◑

김 동 국

1991년 서울과학기술대학교 전자계산학과 이학사
1998년 홍익대학교 국제경영대학원 회계학 전공 경영학 석사
2009년 서울과학기술대학교 IT정책전문대학원 글로벌산업융학 전공 공학 박사
2006년 ~ 2013년 에이쓰리시큐리티(주) 지식사업본부 컨설팅 이사(그룹장)
2015년 ~ 현재 한국IT컨설팅 보안사업본부 수석컨설턴트
2015년 ~ 현재 한국정보보호심사원협회(KISCA) 이사
2015년 ~ 현재 한국인터넷정보학회(KSII) 협동이사
관심분야 : 개인정보보호, 클라우드 컴퓨팅, IoT, 빅데이타

이 혁

2004년 부산대학교 생물학과 생물학 학사
2013년 ~ 2014년 인포섹 침해사고대응팀 선임
2015년 ~ 현재 한국정보보호심사원협회(KISCA) 정회원
2015년 ~ 현재 한국인터넷정보보호학회(KSII) 정보보호연구회 정회원
2014년 ~ 현재 한국IT컨설팅 보안사업본부 책임컨설턴트
관심분야 : 모의해킹, 기술적취약점 진단, 개인정보보호