일본의 빅데이터 프라이버시 보호방안[1)]

Privacy Protection of Big Data in Japan

송태민 한국보건사회연구원 연구위원

1. 서론

최근 스마트폰, 스마트 TV, RFID, 센서 등의 급속한 보급과 모바일 인터넷과 소셜미디어의 확산으로 데이터량이 기하급수적으로 증가하고 데이터의 생산 · 유통 · 소비 체계에 변화를 주면서 데이터가 경제적 자산이 되는 빅 데이터 시대를 맞이하게 되었다[2)]. 세계 각국의 정부와 기업들은 빅데이터가 향후 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하며, 빅데이터의 분석과 활용을 통하여 사회전반에 걸쳐 새로운 부가가치를 창출하려고 노력하고 있다. 세계적으로 빅데이터의 활용에 대한 논의가 본격적으로 시작되면서, 세계경제포럼(WEF)에서는 빅데이터를 미래의 새로운 가능성을 여는 2012년의 가장 중요한 기술로 지목하고 있다[3)].

빅데이터의 등장과 함께 공공 및 민간기관에서는 개인정보를 단순히 한 개인을 식별하기 위한 목적이 아닌 다양한 목적으로 폭넓게 활용되면서 개인의 정보 유출과 프라이버시 침해에 대한 우려는 높아지고 있다. 개인에 관련된 정보를 정부나 공공기관 혹은 서비스 제공자가 실시간으로 감시하는 하는 행위는 사회 전체의 안전과 편의라는 공통의 가치를 위해 개인은 자신의 프라이버시를 부분적으로 포기한다고 암묵적으로 동의한 것으로 여겨지고 있으나, 실제적으로는 개인의 가치를 다양하게 반영한 선택적인 계약에 의한 것이 아니라 일률적인 점이 문제가 되고 있다. 빅데이터의 물결은 현대 정보사회에서 거부할 수 없는 거대한 흐름이다. 빅데이터의 존재를 긍정적으로 인식하고 이를 적극적으

1) 본 고의 일부 내용은 '송태민, 진달래, 이중순, 안지영, 박대순(2013). **인터넷 건강정보 게이트웨이 시스템 구축 및 운영-빅데이터 활용방안을 중심으로-**, 한국보건사회연구원' 을 분석 정리한 것임을 밝힘.

2) 송태민, 송주영(2013). **빅데이터 분석방법론**, 한나래아카데미.

3) 2012년 세계경제포럼, Rethinking Personal Data: Strengthening Trust

로 활용함으로써 빅데이터는 개개인을 위한 혹은 개개인이 새로운 비즈니스의 기회를 창출하는데 없어서는 안 되는 획기적인 자원으로서 그 가치를 더해 갈 것이다. 따라서 빅데이터의 긍정적 활용을 위해서는 우선 빅데이터의 이용으로 인해 발생할 수 있는 부작용과 이에 대한 대책을 검토하고 개개인의 존중을 기본으로 하는 사회전체의 공정한 규칙을 마련해야 할 것이다. 그 중에서 가장 시급한 문제 중의 하나가 빅데이터의 활용에 있어서의 프라이버시의 보호 방안이다. 본 연구에서는 우리나라와 사회 제도면에서 유사한 일본을 중심으로 빅데이터의 활용을 위한 프라이버시 보호방안에 대해 살펴보고자 한다.

2. 개인정보와 프라이버시의 개요

1) 개인정보 개요

개인정보의 법률상의 정의로는 '개인정보법(제2조 제1호)'과 '정보통신망 이용촉진 및 정보보호 등에 관한 법률(제2조 제6호)'에서 "살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다)를 말한다."로 규정되어 있다. OECD는 개인정보보호지침(Guidelines on the Protection of Privacy and Transborder Flows of Personal Data)에서 개인데이터(Personal Data)는 식별되거나 식별될 수 있는 개인에 관한 모든 정보를 지칭한다고 정의하고 있다. EU도 개인정보보호지침(Directive95/46EC)에서 개인 데이터(PersonalData)는 식별되거나 식별될 수 있는 자연인에 관한 모든 정보를 지칭한다고 규정하고, 식별 가능한 개인이란 직접 또는 간접적으로 신원확인 번호, 신체적, 생리적, 정신적, 경제적, 사회적 동일성(identity)을 나타내는 요소를 참조하여 그 신원이 확인될 수 있는 사람을 지칭한다고 정의하고 있다.

한편, 미국의 프라이버시법(Privacy Act, 1974)에서의 개인정보는 행정기관이 보유하는 개인기록(Record)에서 개인에 관한 정보(information about an individual)의 개개 항목 또는 그 집합을 의미하며, 여기에는 개인의 이름, 식별번호, 부호, 지문, 성문 등이 있고 특정 개인과 연결 지을 수 있는(Linkable) 정보가 포함되어 있다. 최근에는 스마트폰의 ID와 같이 해당 개인의 고유한 식별자(identifying particular)까지 확대되어 개인정보로 취급되고 있다. 일본의 '개인정보의보호에 관한 법률'상에 정의 된 개인정보는 '생존하는 개인에 관한 정보로서, 해당 정보에 포함된 성명, 생년월일과 그 외 개인을 식별할 수 있는 것(타의 정보와 쉽게 결합하여 그에 의해 특정의 개인을 식별할 수 있는 것을 포함한다.)'이라고 규정되어 있어 우리나라의 개인정보 법에서와 표현이 유사하다.

위에서 보는 바와 같이 각국에서 개인정보의 법적인 정의는 비슷하며, 우리나라의 것을 기준으로 개인정보를 전체적으로 열거하면 '성명, 주민등록번호를 비롯하여 주소, 성별, 본적, 가족관계와 여권번호, 운전면허증 번호 등 신상에 관한 기본정보와, 신장, 체중, 장애정도, 생

체인식 정보, DNA, 혈액형, 지병 등과 같이 신체의 특징을 나타내거나 건강 의료에 관련된 정보, 소득, 재산, 보험가입 현황, 신용정보, 채권 채무 등의 경제관계 정보, 병역, 직업 경력, 사회활동 경력, 전과기록 및 법률위반 기록 등의 사회경력 정보, 친구, 선후배, 애인 등의 인적 관계 정보, 종교, 취미, 사상, 신조, 가치관, 정치적 성향 등의 내면 정보를 비롯하여 기타 통신 내역, 위치정보, 음주 흡연량' 등의 정보가 포함된다.

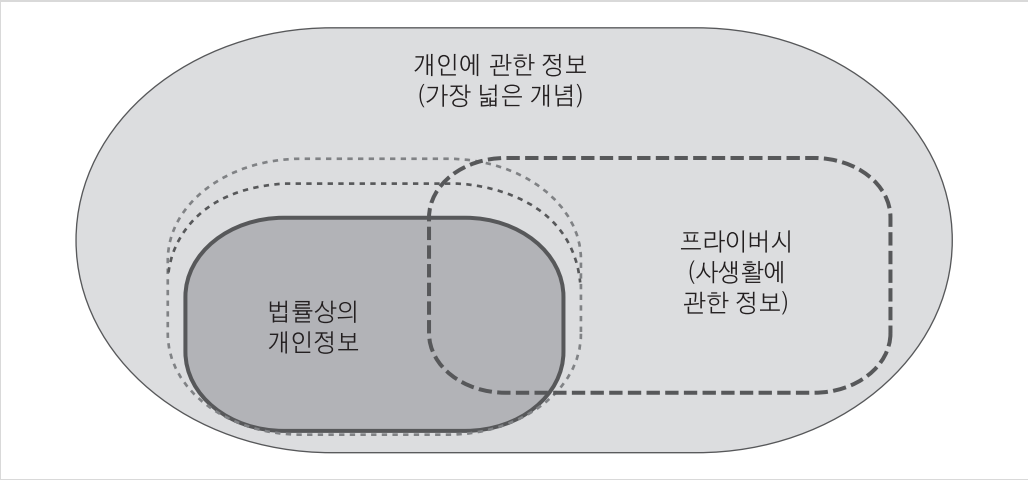
개인정보는 본인확인 정보로서의 식별정보와 본인에 부속된 속성정보로 대별할 수 있다. 식별정보에는 성명, 주민등록번호, 여권번호와 얼굴사진, 지문, 성문, 홍체, 유전자 등의 생체정보가 있다. 이는 그 자체로서 개인을 식별하거나 특정 지을 수 있는 정보이다. 성명의 경우 동명이인이 존재하는 경우도 있지만 사회 통념상 독립적으로 개인을 특정 하는데 사용되고 있음으로 식별정보로서 분류된다. 한편 주소, 생년월일, 성별, 인종, 국적 등은 단독으로 특정 개인을 식별할 수는 없지만 조합에 의해 본인을 식별할 수 있는 준식별 정보이다. 속성정보에는 건강의료 정보, 경제상황 정보, 개인의 신용정보, 학력과 경력정보 등이 있다. 반면, 신용정보와 학력, 경력정보에 포함된 신용카드번호나 학번, 사번과 같은 개인ID 정보는 일반적으로 기본 식별정보와 준식별 정보를 근거로 발행되어 개인을 특정 하는데 사용됨으로 식별정보로 분류할 수 있다.

법률상에서 정의된 '개인정보'와 이를 기술하는데 사용된 '개인에 관한 정보'를 비교해 보면 후자가 전자를 포함하는 개념임을 내포하고 있다(그림 1). 즉 개인에 관한 정보가 보다 넓은 개념의 집합으로 그 중 일부가 개인정보인 것이다. 개인에 관한 정보 중에는 개인의 식별로 연결되기가 어려워 법률적으로 개인정보에 포함되지 않는 것들이 있다. 따라서 이러한 정보는 개인에 관한 정보로서의 그 중요성이나 의미와 가치가 적은 것으로 취급되어 왔다. 경우에 따라서는 '개인에 관한 정보'에서 제외되었다. 그러나 정보통신 기술의 발전과 인터넷의 등장으로 지금까지 비개인정보로 취급되었던 사항들이 다른 정보와 결합하여 개인을 특정 지을 수 있게 사회가 변화해 감에 따라 개인정보가 차지하는 영역이 점차 확대 되어 가고 있는 상황이다. 과거에는 혼잡한 길거리의 사진에 촬영된 사람들의 정보는 '개인에 관한 정보'에서 제외되었고 도서관 자료의 열람정보와 같은 행동 이력 등은 사회적인 상식의 범위에서 개인정보로 간주하지 않았던 사항들이었다(그림 2). 최근 들어 인터넷 상의 프라이버시 보호에 대한 논의와 함께 그러한 정보 들이 개인정보의 범주에 포함되고 있다. 특히 데이터의 분석 기술의 발달과 함께 그러한 경향은 더욱 가속되고 있으며, 그동안 의미나 존재 가치가 희박했던 비개인 정보들이 활용을 전제로 새로운 가치를 창출하는 빅데이터로서 재평가 받고 있는 것이다.

2) 프라이버시 개요

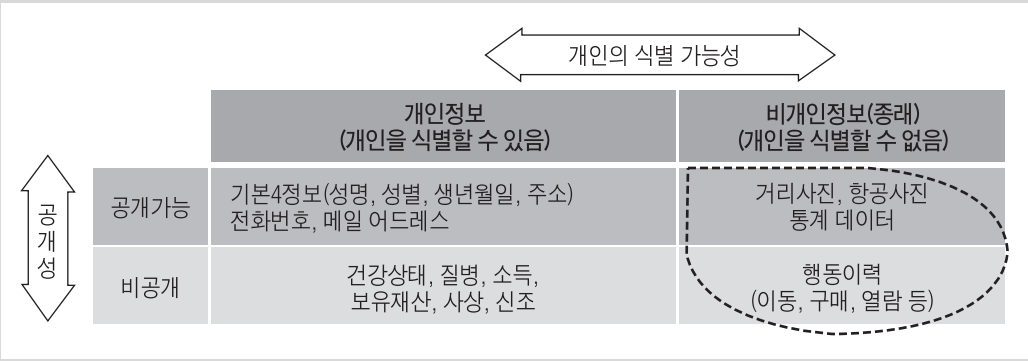
개인에 관한 정보 중에는 사생활에 관계되는 프라이버시(Privacy)의 부분이 있다(그림 3). 프라이버시의 사전적인 의미로서는 개인의 사생활이나 사적인 일, 또는 그것을 남에게 알려지

그림 1. 개인에 관한 정보와 개인정보 및 프라이버시의 관계



자료: 노무라연구소 小林慎太郎 ビッグデータ社会におけるプライバシー ～「個人情報」から「プライバシー」の保護へ～' 第176回NRIメディアフォーラム資料 2012년11월26일

그림 2. 개인정보 식별성과 공개성



자료: 노무라연구소 小林慎太郎 ビッグデータ社会におけるプライバシー ～「個人情報」から「プライバシー」の保護へ～' 第176回NRIメディアフォーラム資料 2012년11월26일

지 않거나 간섭받지 않을 권리를 말하며, 사생활(私生活)로 번역하기도 하지만 프라이버시는 권리를 포함한 더 큰 범주에 속한다. 프라이버시를 내용에 따라 분류하면, 결정프라이버시(Decisional privacy), 공간프라이버시(Spatial privacy 혹은 Locational privacy), 의도프라이버시(Intentional privacy), 정보프라이버시(Informational privacy), 통신프라이버시

(Communicational privacy), 물리적정신적 프라이버시(Physical and psychological privacy)[4] 등으로 나눌 수 있다.

프라이버시는 초기에 '혼자 있을 권리(the right to be left alone)'로서 자기 자신에 관한 정보가 남에게 함부로 공개되거나 침해되지 않는 권리의 소극적인 개념에서 정보화의 물결과 함께 '자신의 정보를 통제할 수 있는 권리(the right to control information about oneself)'로 발전하게 되었다[5]. 즉 자신의 정보를 수집, 가공, 유통 및 제공하는 데에 있어 접근권 및 통제권을 가지고, 언제, 무슨 정보를, 어느 범위까지, 누구에게 유통시키느냐를 스스로 결정하는 '정보의 자기결정권'을 의미하게 된 것이다. 이처럼 프라이버시 권리는 당초의 소극적인 개념에서 시작하여 정보사회에서는 '정보의 자기결정권'과 같은 적극적이고 능동적인 권리의 개념으로 바뀐 것이다.

프라이버시권의 법적 지위는 유엔에 있어서의 세계인권 선언[6]을 비롯하여 각국은 헌법과 법률에 국민의 의해 기본권으로 보호하고 있다. 세계 인권선언(Universal Declaration of Human Rights)12조에는 '어느 누구도 자신의 사생활, 가족, 가정 또는 통신에 대하여 자의적인 간섭을 받지 않으며, 자신의 명예와 신용에 대하여 공격을 받지 아니한다. 모든 사람은 그러한 간섭과 공격에 대하여 법률의 보호를 받을 권리를 가진다.'라고 규정하고 있다. 우리나라 헌법에서는 제16조 공간에 대한 프라이버시, 제17조 사생활의 비밀과 자유로서의 프라이버시, 제18조 통신에 대한 프라이버시를 국민의 기본권으로 보장하고 있다. 이를 침해하였을 때는 민법에서 불법적 행위에 의한 침해로서 제750조(불법행위의 내용)에 의해 대한 손해를 배상할 책임이 있고, 제751조(재산 이외의 손해의 배상)에서 자유 또는 명예를 해하거나 기타 정신상의 고통에 대해 배상할 책임이 발생한다.

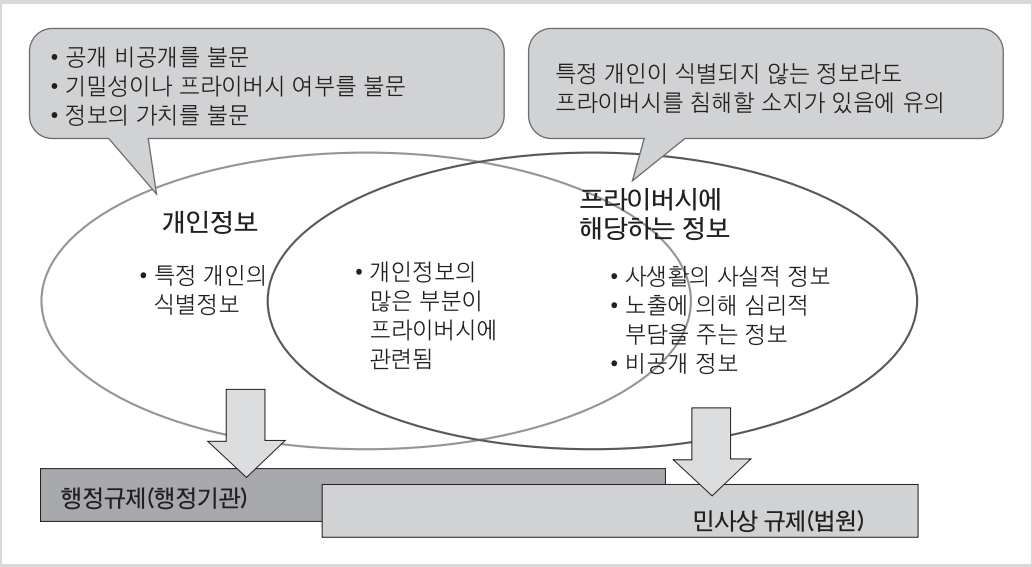
3) 익명성의 개요

익명의 사전적 의미는 '자신의 본래 이름 혹은 아이덴티티를 밝혀 드러내지 않고 숨기는 것'을 말한다. 익명화(Anomymize)된 정보는 개인정보보호법 상의 개인정보에는 해당하지 않는다는 인식이 일반적이다. 그러나 현재까지 익명화가 어떻게 처리되어야 하는 지에 대한 명확한 설명이나 구체적인 기준이 없는 실정이다. 익명화란 식별요소를 '1대 다'의 관계를 유지하면서 본인에의 1대 1'의 도달 가능성을 없애는 작업이라고 볼 수 있다[7].

개인정보에 대한 처리의 한 형태로 '가명화'(Pseudonymity)가 있다. 가명화란 일반적인 식별요소를 상대적인 식별요소로 치환한 것이다. 예를 들어 <이름+상품구입 이력>로 구성된 정보를 <회원ID+상품구입 이력>으로 바꾸는 것이다. 이는 해당 회원 ID에 상대적인 본인 도달

4) http://itlaw.wikia.com/wiki/Right_of_privacy. 2013/09/13.
5) http://en.wikipedia.org/wiki/Right_to_privacy. 2013/09/13.
6) http://www.humanrights.com/what-are-human-rights/universal-declaration-of-human-rights.html. 2013/09/13.
7) 中田響(2007). 個人情報性の判断講造, 慶藤義塾大学メディア・コミュニケーション研究所紀要No.57, pp.145~161.

그림 3. 개인정보와 프라이버시권에 해당하는 정보의 관계



자료: 鈴木 正朝「プライバシーの權利と個人情報保護法」総務省マイナンバーシンポジウム資料 2012년11월

가능성이 인정되는 한 익명화는 아니다. 익명화된 정보는 식별 요소성이 부족해서 타 정보와의 용이한 조합 가능성이 인정되지 않음으로 해서 개인정보에서 벗어 날 수 있다. 이에 비해 가명화된 정보는 특정의 사업자와의 관계에서는 여전히 개인정보인 것이다.

한편, 웹서비스를 이용하는 데 있어 프라이버시를 보호하는 차원에서 가명을 사용하기도 한다. 그러나 사업자의 입장에서는 이용자에 대한 계속적이고 부가적인 서비스 등에 있어서의 편의성을 위해 혹은 이용자를 식별하고 추적하거나 부정행위를 막을 목적으로 ID등록을 요구하고 있다. 또한 연락수단으로서 메일주소를 요구하고 있기도 하다. 이런 관계에서 일단 ID를 취득하게 되면 비록 가명에 의한 회원등록이라고 할지라도 본인 도달가능성이 어느 정도 실현된다고 볼 수 있다. 현재 공학적인 입장으로부터 프라이버시 보호를 위한 여러 가지 기술 들이 활발히 연구되고 있는 가운데 익명성이나 프라이버시 등의 정의가 각각의 연구 내에서 이루어지고 있다. 그러나 복수의 연구 성과를 비교해 익명성의 강도를 비교한다든가, 여러 기술 중에서 필요로 하는 기술을 찾고자 할 때 용어 등이 달라 상호간의 관계가 명확하지 않은 경우가 많다[8]. 이러한 용어 통일의 실현을 목표로 한 활동으로서 Pfitzmann(2010)[9]의 용어집이 잘 알려져

8) 真野健(2011). '匿名性 · プライバシーの工学的定式化とその学際的応用', 電子情報通信学会誌, 94(9), p.788~794.
9) http://dud.inf.tu-dresden.de/Anon_Terminology.shtml. 2013/09/13.

있다. Pfitzmann 용어집에서는 프라이버시 보호에 관한 개념 어구 등을 망라해 일관성이 있게 정의하고 이를 연구자들 사이에서 공유하는 것을 목적으로, 2000년 이래 거듭되는 개정판이 공개되고 있다. 예를 들면, 익명성(Anonymity)은 자신의 속성을 공개하지 않고 서비스나 자원을 이용할 수 있는 상태라고 정의하며, 익명성을 결정하는 것은 정보 제공자와 수신자의 관계인 관찰불능성(Unobservability), 연결불능성(Unlinkability), 가명성(Pseudonymity)의 요소로 설명된다[10]. 또한, 관찰불능성(Unobservability)란, 정보의 제공자와 수신자, 혹은 커뮤니케이션 자체의 관찰이 불가능한 상태를 나타낸다. 관찰불능성을 만족하면 해당 존재가 발견되지 않는 상태(undetectable)에 놓여진다. 연결불능성(Unlinkability)이란 복수의 세션이 동일 인물에 의한 것인지가 판정 되지 않는 것을 의미하고 세션끼리의 관련성을 맺기가 불가능한 상태를 나타낸다. 복수의 세션이 동일인에 의한 것인지가 분명하면, 세션 간을 관련(링크)지을 수 있고, 관련지어진 대상에 대해서는 식별자를 가명(Pseudonym)의 형태로 발행할 수 있다. 관련지을 수 있는 정보가 많아질수록 가명으로 연결된 정보가 축적되어 간다. 물론 복수의 세션을 관련지을 수 없는 상태(링크 불능)를 유지할 수 있으면 세션에 대응하는 가명에 의한 본인 도달 가능성은 생기지 않는다. 가명에 의해 실명을 은닉 하더라도 세션이 어느 행위자에 의한 것일까를 판정할 수 있는 경우가 있다.

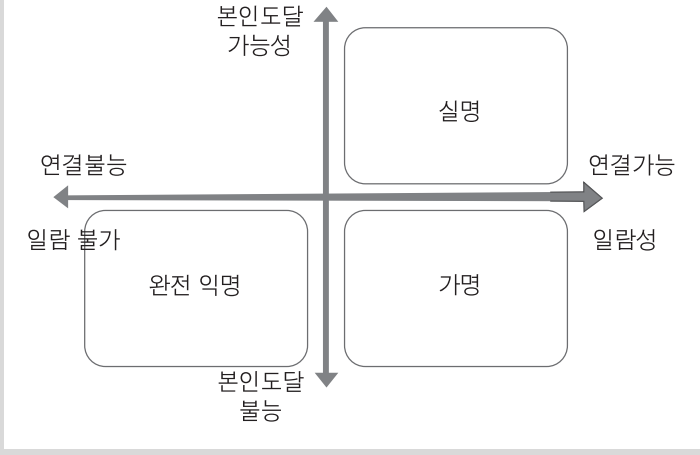
익명성의 강도 면에서 본다면 대상이 관찰불능(Unobservable)인 상태가 가장 익명성이 강하다고 할 수 있다. 관찰불능(Unobservable)이면 링크불능의 상태가 되며 가명도 발생하지 않는다. 관찰가능한 상태가 되면 세션의 행위자에 대해 가명이 발생한다. 그러나 연결불능(Unlinkable)이면 여전히 익명성이 높다고 할 수 있다. 세션의 행위자가 가명을 사용하는 경우 이미 언급한 바와 같이 인위적인 의미를 지닌 가명으로 인해 연결불능성(Unlinkability)이 충족되지 못하고 익명성은 약하게 된다. 복수의 세션이 연결 가능한 경우, 제삼자에 의해서 연결 가능한 세션이 일람 할 수 있는 상태가 될 때 이를 '일람성'이라고 정의한다. 가명과 같이 인위적인 ID에 의해 복수의 세션을 관련지을 수 있다고 해도 이를 제삼자가 일람 할 수 없으면 외관상 링크 불능성이 성립한다고 볼 수있다 그러나 가명에 의해 행한 댓글의 이력이나 구매 이력 등이 검색에 의해서 링크 가능한 형태로 모아지게 되면 일람성이 성립될 수 있고 익명성을 잃게 된다(그림 4).

3. 빅데이터 활용에 있어서의 프라이버시 보호 방안

빅데이터의 활용에 있어서의 가장 논란이 되고 있는 것은 개인의 각종 기록 정보를 수집하여 분석함으로 해서 생길 수 있는 프라이버시의

10) 新田明子, 江木啓訓(2007.9.21). 'リンク不能性および一 覧性の観点による匿名性の分類', 日本情報処理学会 SIG Technical Report, pp.153~158.

그림 4. 연결가능성, 일람성과 익명성의 강도



자료: 新田明子, 江木啓訓(2007.9.21). ‘リンク不能性および一覧性の観点による匿名性の分類’, 日本情報処理学会 SIG Technical Report, pp.153~158.

침해에 대한 위험성의 문제이다. 이는 외부로부터의 침입에 의한 데이터의 유출의 위험성과 데이터 취급자의 부적절한 업무처리로 인한 무의식적인 공개와 같은 잠재적 위험성, 그리고 내부자의 불법적인 열람 행위 등에서 비롯되는 위험성을 내포하고 있다. 프라이버시가 노출되거나 개인 정보가 악용된 경우 사회적으로 큰 파장이 일어남은 물론 경제적인 면에서도 손해 배상과 대책 비용의 발생, 사업 활동의 자숙 혹은 신용 저하에 따른 매출 감소 등의 비용이 발생한다. 실제적인 예로 일본의 경우 일본네트워크보안 협회의 보고서에 따르면 2011년에 연간 약 1,551건의 개인 정보 유출과 1,900억 엔(약 2조원) 이상의 손해 배상이 있다고 한다[11].

이러한 문제에 대해 적절히 대처하기 위해 사회적인 큰 틀에 있어서의 법제도적인 대책에서부터 비롯하여 조직적인 대책, 기술적인 대책 등이 검토되고 있다. 법제도의 측면에서는 한국과 일본에서는 개인정보법이 제정되어 있으며, 미국에는 옵트아웃을 기반으로 한 ‘소비자 프라이버시 권리장전(Consumer Privacy Bill of Rights)’, 유럽의 옵트아웃을 기반으로 한 ‘EU 데이터 보호 규정(General Data Protection Regulation)’ 등이 정비 되어 있다. 조직적인 대책으로는 법에서 규정한 안전관리 조치를 준수하는 차원에서 정보보호관리체계(ISMS)와 프라이버시 영향평가(PIA) 등의 시스템 운용과 관리상의 각종 기법이 도입되고 있다. 또한 기술적인 면에서도 프라이버시 보호 데이터마이닝(PPDM: Privacy-Preserving Data Mining) 등 여러 가지 기법이 연구 개발되어 적용되고 있다. 그러나 개인정보 보호를 과도하게 중시하면, 빅데이터의 활용을 저해하게 될 우려가 있다. 이는 서비스 공급자와 이용자의 양자가 상호간에 다양한 효용성을 얻을 수 있는 기회를 놓치는 것이 된다. 이러한 점에서 법제도적, 기술적 대책을 적절히 운용하여 정보 주체가 안심하고 자신의 데이터를 제공

11) 日本ネットワークセキュリティ協会, 2012年情報セキュリティインシデントに関する調査報告書 –個人情報漏えい編– 2012.12.7.

하게 하고 빅데이터가 적극적으로 활용되게 하는 일이 중요하다.

1) 정보보호관리체계 인증을 통한 프라이버시 보호

정보보호관리체계(ISMS: Information Security Management System)는 기업이나 조직이 정보보호 활동을 체계적이고 지속적으로 수행하기 위해, 보안정책을 수립하고 이에 근거한 계획의 작성, 실시, 운영, 그리고 일정 기간 후의 보안 방침과 계획의 재검토 등을 포함한 전체적인 위험관리를 지속적으로 수행하는 체계를 말한다[12]. 기업이나 조직이 ISMS를 보유하고 유지하고 있는 지에 대해서는 'ISMS 인증제도'를 통하여 제도적으로 보증 받는다. 이는 제삼자 기관에 의한 인증 심사에 의해 정보보안 관리체계의 국제규격인 ISO/IEC17799:2000 및 BS7799-2:1999에 입각한 평가를 받는 것이다. ISMS에 요구되는 범위는, ISO/IEC 15408 등이 정하는 기술적인 정보보안 대책의 레벨이 아니라 조직 전체에 있어서 보안 관리 체제를 구축 · 감시하고 위기관리를 실시하는 것이다. ISMS는 개별적인 문제에 대한 기술 대책 외에도 조직 관리의 일환으로서 스스로의 위험을 평가하여 필요한 보안 레벨을 정하고 계획에 따라 자원을 배분해서 시스템을 운용하는 것이다. 조직이 보호해야 할 정보 자산에 대해서, 기밀성, 완전성, 가용성을 균형 있게 유지하고 개선하는 것이 ISMS의 기본 개념이다.

2) 프라이버시 영향평가를 통한 프라이버시 보호

프라이버시 영향평가(PIA: Privacy Impact Assessment)는 개인정보를 수집하여 취급하는 정보 시스템의 기획, 구축, 보수 유지 과정에 있어서 개인정보 제공자의 프라이버시에의 영향을 '사전'에 평가하는 일련의 프로세스를 말한다. 이를 통하여 잠재적인 프라이버시 침해 위험성을 분석하고 대안적인 방법이나 보호방안을 검증하는 등 정보 시스템의 구축 운용을 적정하게 실시할 수 있도록 하는 과정을 포함한다. 설계 단계에서부터 프라이버시 보호 대책을 검토함으로써, 정보시스템 가동 후의 프라이버시 리스크를 최소한으로 억제할 수 있어 향후 시스템의 개보수에 따르는 추가 비용의 발생을 막을 수 있다. PIA는 2008년 4월 ISO22307 (Financial services Privacy impact assessment)로서 표준화 되었다. 부제의 타이틀에서 의미하는 바와 같이 금융 서비스를 제공하는 기업이 고객과 거래처 등의 재무 데이터의 처리와 관련된 프라이버시 보호와 리스크에 대처하기 위한 방법론이 정의되어 있다. 그러나 그 내용이 금융 관련 서비스에 특화한 것이 아니고 민간부문 및 공공부문을 불문하고 다양한 분야에 적용 가능한 것이다. 또한, 요구 사양이 각국의 법체계나 사회제도에 의존하지 않는 최대공약수적인 내용이다.

실제적으로 법령상의 해석이나 익명화와 같은 기술적인 대책만으로는 개인정보의 이용을

12) http://isms.kisa.or.kr/kor/intro/intro01.jsp. 2013/09/13.

정당화 하는데 충분한 근거를 마련할 수 없는 상황이 생길 때가 많다. 이런 상황을 해소하는 데 있어 PIA의 프로세스로서 활용하는 것이 유효할 수 있다. 즉 개인정보의 보호와 이용의 균형을 도모하기 위해 개인정보의 이용에 수반되는 프라이버시에의 영향(리스크)을 평가해서 정보 이용에 의해 초래되는 편익과 비교한다. 이를 토대로 정보 주체자를 비롯한 이해관계자들의 의향을 파악하여 이를 반영한 적절한 조치를 강구함으로써 정보 이용에 대한 정당성을 확보하는 것이다.

3) 프라이버시 보호기술을 통한 프라이버시 보호

개인과 관계되는 정보를 유통함에 있어서 법제도적인 대책과 조직적인 대책을 완수하기 위해서는 기술적인 대책이 뒷받침되어야 한다. 프라이버시 보호를 위해 정보 유통에 강한 제약조건을 두게 되면 정보의 가치를 잃게 되거나 손상되어 사회적으로는 물론 정보 주체자가 본인에게도 손실을 가져 오게 된다. 서비스를 개인화하는데 있어 개인 정보는 불가결한 요소이기 때문이다. 이러한 문제를 기술적으로 보완하기 위하여 최근 개인 정보의 이용과 프라이버시 보호의 밸런스를 취하는 방안이 연구되고 있다. 그 중에서 프라이버시 보호 데이터 마이닝에 관한 연구가 대표적으로 관심을 받고 있는 연구이다. PPDM는 개인정보나 기밀정보의 안전성을 유지하고 개인의 프라이버시를 보호하면서 대규모의 데이터로부터 특징이나 규칙성 등을 추출하고 새로운 지식을 발견하는 데 활용하기 위한 기술을 총칭하는 것이다.

PPDM에는 데이터가 가진 정보의 일부를 누락시키거나 개인을 특정할 수 있는 요소를 삭제 은폐하는 등의 익명화 수법과, 통계학적인 처리에 의해 DB에 잡음을 첨가하여 통계적인 성질을 유지하면서 데이터의 누설을 막는 교란 수법, 데이터가 가진 정보에 손상을 입히지 않고 당사자 간에 정보를 주고받는 암호화 수법 등이 있다. 현재는 정보 공개에 일반적으로 활용될 수 있는 익명화 기술이 많이 연구되고 있다. 그러나 데이터의 익명화는 정보의 손실을 초래하여 데이터 마이닝(data mining)의 결과에 영향을 미치게 됨으로 활용 용도에 따라 프라이버시 보호수준을 고려하여 익명화 처리에 특히 주의를 기울일 필요가 있다. 암호화 수법 중에는 데이터를 복수의 그룹에 나누어서 분리 보관하고, 그룹사이의 공개키 암호로 데이터를 암호화한 채로 필요에 따라 데이터 마이닝(data mining)을 위한 계산을 실시하여 그 결과만을 전체 그룹 사이에 공유하는 비밀 계산 기술도 최근 연구되고 있다.

4. 맺음말

현재 어느 나라를 막론하여 빅데이터의 활용에 있어 가장 큰 과제는 개인의 사생활 비밀보호 및 개인정보보호이다. 앞에서 이미 언급한 바와 같이 개인정보보호에 중점을 두면 빅데이터의 활용을 저해하게 될 우려가 있다. 개인정보보호법 등 관련 법률을 자의적으로 해석하여 수집된 개인 데이터를 공공의 목적으로 활용하

기 위해 제삼자에게 제공하는데 소극적인 기관이 적지 않다. 개인정보 보호법의 목적이 '개인정보의 수집 · 유출 · 오용 · 남용으로부터 사생활의 비밀 등을 보호함으로써 국민의 권리와 이익을 증진하고'로 되어 있지만 개인정보와 비개인정보를 명확히 구분하기가 곤란하고 비즈니스에 있어 자동적으로 수집되는 데이터가 비개인정보라고 할지라도 프라이버시를 침해할 가능성이 있다. 특히, 소셜미디어에 공개된 개인정보는 위변조, 오남용이 쉽고 상업적 이용을 위한 정보수집 등에 노출이 될 수 있기 때문에 프러이버시 침해 등의 문제가 발생할 가능성이 매우 높다.

개인 정보의 흐름이 이미 국경을 넘어 선 지가 오래이다. 구글, 트위트, 유튜브 등의 글로벌 기업의 서비스에 국내법 상에 문제가 있는 부분이 있더라도 이를 국내에서 국내법의 규정에 따라 제재를 가할 수는 없다. 따라서 이제는 국제협력을 도모하면서 선진국 수준에 부응하는 프라이버시 보호와 개인 정보 보호 제도를 정착시켜야 할 시점에 와 있다. 해외에서 이미 법적인 관점에서의 개인정보 보호보다는 사회적인 관점에서의 프라이버시 보호가 중시되는 경향으로 흐르고 있다. 최근 국회 입법조사처에서 검토한 우리나라의 빅데이터 활용에 있어 개인정보보호 법제의 대응방안)[13]을 살펴보면 다음과 같다. 첫째, 현행 개인정보보호 법제가 빅데이터의 활용을 저해한다는 의견이 강하게 제기되면서 '개인식별가능성' 요건을 완화하는 문제에 대한 단계적 접근을 통해 공감대를 형성해 나가야할 필요가 있다. 둘째, 분산된 개별 정보보호 법률들의 소관 부처들의 감독 및 규제 기능을 통합 또는 일원화 할 수 있는 감독기구와 관련한 체계정비가 요구된다. 셋째, 장기적인 관점에서 민간영역의 자율규제를 촉진하고 이 과정에서 형성된 법적 판단기준들을 검토하여 관련 법령을 개정할 수 있을 것으로 본다.

빅데이터로부터 개인을 보호하기 위해 가장 중요한 것은 특정 개인을 식별하지 못하도록 하는 익명화와 정보접근 및 정보처리에 대한 통제다. 그러나 정보접근 및 정보처리에 대한 통제를 강하게 하면 정보활용을 활성화할 수 없기 때문에 빅데이터의 '활용과 보호의 균형'에 대한 효과적인 정책이 우선적으로 마련되어져야 할 것이다[14]. 보건복지

13) 심우민(2013. 10. 11). 빅데이터 활용과 개인정보보호, **이슈와 논점**, 국회입법조사처.
14) 송태민 (2013. 9). 우리나라 보건복지 빅데이터 동향 및 활용방안, **과학기술정책, 192**, 과학기술정책연구원.