빅데이터 관련 이른바 '비식별화' 입법에 대한 의견서

2015년 6월 8일

경실련 소비자정의센터
진보네트워크센터

1. 빅데이터 산업 활성화를 위해 정부가 행정입법으로 개인정보 보호법의 규범을 완화하려는 시도를 계속하고 있습니다.

정부는 "빅데이터로 창조경제 시동건다"(2013. 4. 21. 미래창조과학부 보도자료)는 기조 하에 빅데이터 산업 활성화를 추진해 왔습니다. 빅데이터 산업 활성화를 위해 방송통신위원회가 추진해온 <빅데이터 개인정보보호 가이드라인>은 대통령 직속 개인정보 보호위원회의 위법성 지적(2014. 7. 30. 의결) 등 논란 끝에 2014. 12. 23. 행정규칙으로 발표되었습니다.

방통위 가이드라인은 그 제목에 '개인정보보호'가 포함되어 있음에도 불구하고, 핵심 취지는 '비식별화'된 개인정보에 개인정보 보호 규범의 예외를 두는 것에 있습니다. 한국 정부가 창안한 '비식별화'라는 개념은, 해외에서 인정되고 있는 '익명화'에 비해 그 내용이 모호할 뿐 아니라, 핵심 취지가 기업으로 하여금 정보주체의 동의 없이 개인정보를 수집 및 처리할 수 있게끔 허용하겠다는 것입니다. 이는 결국 현행 개인정보보호 규범을 우회하거나 약화시키겠다는 것으로서, 빅데이터 산업 활성화의 명분으로 국민의 기본권에 중대한 제한을 가져올 것입니다.

국민들은 계속되는 개인정보 유출 사고('14 카드3사 1억 4백만 건 유출)와 개인정보 유상판매 사건('15 고객정보 약 2천4백만 건을 개당 1,980원 혹은 2,800원씩 받고 보험회사들에 유상판매한 혐의로 홈플러스 경영진 형사기소)을 겪으면서 개인정보 보호 문제를 민감하게 인식하고 있습니다. '비식별화' 라는 개념은 국가적인 개인정보 유출 사고 이후 국민 앞에 정부와 국회가 앞다투어 제시한 개인정보 보호 비전과도 역행하는 것입니다.

그러나 방통위의 가이드라인 발표 이후, 정부 다른 부처에서도 '비식별화' 개념을 무분별하게 도입하고 있어 시민사회로부터 깊은 우려를 사고 있습니다. 특히 지난 6월 3일 금융위원회는 빅데이터 산업 활성화를 위해 신용정보법 시행령을 개정하여 비식별정보를 개인신용정보에서 제외하겠다고 밝혔습니다. 개인정보를 동의받은 목적으로만 이용이 가능하도록 규정하고 있는 개인정보 보호법에도 불구하고, 개인정보를 비식별화할 경우 정보주체가 동의하지 않아도 빅데이터 처리 등에 사용할 수 있도록 허용하겠다는 의미입니다.

2. 현재 '비식별화' 개념은 방송통신위원회의 행정규칙인 <빅데이터 가이드라인>이나 금융위 시행령(예정)에 그치지 않고 다음 법률안에서 법정 개념으로 도입하고 있습니다.

- 개인정보 보호법 전부개정법률안(강은희의원 대표발의, 의안번호 19-13932)
- 정보통신망 이용촉진 및 정보보호 등에 관한 법률 일부개정법률안(강길부의원 대표발의, 의안

번호 19-14200)
- 개인정보 보호법 일부개정법률안(부좌현의원 대표발의, 의안번호 19-14166)

위 세 개 법안은 모두 빅데이터 산업 활성화 차원에서 정부가 창안한 '비식별화' 개념을 도입하고 있으며, 그 규율 내용도 방통위 가이드라인의 핵심내용과 다르지 않습니다. 상호간의 미세한 차이점에도 불구하고 이 법안들의 공통된 면모는, 현행 개인정보 보호법에서 '비식별화'라는 새로운 예외대상을 신설하는 것을 주요골자로 합니다.

먼저, 개인정보 보호법 전부개정법률안(강은희 의원안)의 관련 주요내용은 다음과 같습니다.
- 통계·연구, 시장조사, 마케팅 등의 목적을 위한 경우에는 개인정보 비식별화 조치를 통해 정보주체의 동의 없이 이를 처리할 수 있도록 하고 개인정보 파기 요건 및 이에 관한 예외 사유를 규정하여, 개인정보 처리 과정에서의 유연성을 부여함(안 제39조 및 제40조).
- 안전성 확보에 필요한 보호조치를 하지 아니하여 비식별화 처리한 개인정보를 분실·도난·유출·변조 또는 훼손당한 자는 2년 이하의 징역 또는 1천만원 이하의 벌금에 처함(안 제99조 제4항 제7호)

정보통신망 이용촉진 및 정보보호 등에 관한 법률(강길부 의원안)의 주요내용은 다음과 같습니다.
- "비식별화 방법"을 이용해 빅데이터의 활용을 증진하면서도 개인정보를 안전하게 보호하기 위하여, 정보통신서비스 제공자가 비식별화된 개인정보를 이용하는 과정에서 개인정보가 발생하는 경우에는 이를 파기하거나 다시 비식별화하는 의무를 부과함(안 제24조의3).
- 비식별화의 기술적 기준 및 개인정보의 파기 및 추가적인 비식별화에 관하여 필요한 사항은 대통령령으로 정함(안 동조 제3항)
- 비식별화 개인정보를 이용하는 과정에서 개인정보가 생성되었음에도 불구하고 이를 지체 없이 파기하거나 비식별화하지 아니한 자에게 3천만원 이하의 과태료를 부과함 (안 제76조제1항 제2호의4).

개인정보 보호법 일부개정법률안(부좌현 의원안)의 주요내용은 다음과 같습니다.
- 빅데이터의 분석·활용 과정에서 개인정보가 유출되지 않도록 관련 규정을 마련한다는 취지 속에 개인정보처리자가 통계작성,학술연구,실태조사를 목적으로 개인정보를 처리하거나 이미 공개된 정보를 재가공하는 과정에서 개인정보가 유출되지 아니하도록 개인정보처리자에게 개인정보 비식별화 조치 의무를 부여함(안 제22조의2제1항).
- 그러한 한편으로 개인정보처리자가 개인정보를 비식별화하여 처리하는 경우에는 정보주체의 동의를 받지 아니할 수 있도록 하여 현행 법규범을 완화하고(안 제22조의2제2항).
- 개인정보처리자가 개인정보를 비식별화하여 처리하거나 비식별화된 개인정보를 처리하는 때에는 개인정보가 생성되지 않도록 하고, 이 과정에서 안전성 확보에 필요한 기술적·관리적 및 물리적 조치를 하도록 하고(안 제22조의2제3항 및 제4항), 안전성 확보에 필요한 보호조치를 하지 아니하여 개인정보를 분실·도난·유출·변조 또는 훼손당한 자는 2년 이하의 징역 또는 1천만원 이하의 벌금에 처함(안 제73조 제1호).

이하에서는 위 법안들과 방통위 가이드라인에 규정된 '비식별화' 개념을 보다 구체적으로 비판합니다.

3. ‘비식별화’ 개념은 개인정보 보호규범을 약화시키기 위해 정부가 창안한 근거없는 명분에 불과합니다.

- 현재 제안된 ‘비식별화’의 개념에 따르면 기업을 비롯한 개인정보 처리자들은 정보주체의 동의를 받지 않고 인터넷에서 개인정보를 수집하고, 저장하고, 분석하고, 심지어 제3자에게 판매하는 것이 가능해집니다. 반면 정보주체는 누가 언제 어떻게 자기의 정보를 수집하고, 분석하고, 조합하여 사고 파는지 알지 못하여 매우 불안한 상황에 처해질 것으로 보입니다. 이는 지금까지의 개인정보 보호 체계가 허물어 지는 것을 의미합니다.

◎ 예상되는 상황 (1)
현행 개인정보 보호법에서 규정하고 있는 제한에서 벗어나, 기업은 페이스북, 블로그, 트위터, 홈페이지, 카페, 직거래 사이트 등에 올려진 개인정보를 정보주체에게 동의 받지 않고 전부 수집할 수 있게 된다. 또 수집한 개인정보에 대해 비식별화 처리를 하면 제한없이 저장하고, 조합하고, 분석하고, 가공할 수 있고, 다른 기업에게 돈을 받고 팔거나 마찬가지로 사올 수도 있게 된다.

◎ 예상되는 상황 (2)
현행 개인정보 보호법에서 규정하고 있는 제한에서 벗어나, 기업은 개인의 위치정보(내비게이션 정보 포함), 콘텐츠 이용 내역, TV 시청 정보, 도서 구매 정보, 쇼핑 내역 정보, 카드 사용 정보 등 내밀한 정보도 정보주체의 동의 없이 비식별화한 후 마음대로 조합, 분석, 가공, 판매할 수 있게 된다.

◎ 예상되는 상황 (3)
현행 개인정보 보호법의 보호에서 벗어나, 정보주체는 자신의 개인정보에 대한 통제권을 상실하게 된다. 비식별화한 자신의 개인정보를 어느 기업이 어떻게 가지고 있는지, 누가 누구에게 판매했는지, 비식별화 조치는 어느 정도나 안전한 것인지 정보주체는 알지 못한다. 정보주체는 끊임없는 마케팅의 표적이 되고, 자신의 통제권 바깥에서 벌어지는 개인정보 유출사고가 터질 때마다 지금까지보다 더 큰 불안에 떨 수 밖에 없게 된다.

- 2005년 헌법재판소는 공개된 개인정보를 포함하여 자신의 개인정보의 공개와 이용에 관하여 정보주체가 스스로 결정할 수 있는 개인정보자기결정권이 우리 헌법에서 보호하는 기본권이라고 선언하였습니다. 2011년 이러한 개인정보자기결정권을 구체화하려는 취지에서 개인정보 보호법이 제정되었습니다. 헌법재판소의 결정에 따르면 공개된 개인정보에 대한 수집과 이용 역시 정보주체의 권리로서 헌법으로 보호 받고 있습니다.

개인정보자기결정권은 자신에 관한 정보가 언제 누구에게 어느 범위까지 알려지고 또 이용되도록 할 것인지를 그 정보주체가 스스로 결정할 수 있는 권리이다. 즉 정보주체가 개인정보의 공개와 이용에 관하여 스스로 결정할 권리를 말한다.
개인정보자기결정권의 보호대상이 되는 개인정보는 개인의 신체, 신념, 사회적 지위, 신분 등과 같이 개인의 인격주체성을 특징짓는 사항으로서 그 개인의 동일성을 식별할 수 있게 하는 일체의 정보라고 할 수 있고, 반드시 개인의 내밀한 영역이나 사사(私事)의 영역에 속하는 정보에 국한되지 않고 공적 생활에서 형성되었거나 이미 공개된 개인정보까지 포함한다. 또한 그러한 개인정보를 대상으로 한 조사·수집·보관·처리·이용 등

의 행위는 모두 원칙적으로 개인정보자기결정권에 대한 제한에 해당한다.
- 헌재 2005. 5. 26. 99헌마513 등

- 개인정보보호법, 정보통신망법은 이 법률들이 규율하는 개인정보에 대하여, 살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보 뿐 아니라 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함하여 정의하고 있습니다. 이 정의는 유럽연합의 개인정보보호지침의 개인정보의 정의 규정이나, 새로 제안된 유럽 GDPR의 개인정보의 정의 규정 혹은 각국 개인정보보호 관련 법률의 개인정보에 대한 정의와 크게 다르지 않습니다. 이러한 개인정보 정의에 따르면 '다른 정보와 결합하여 특정 개인을 알아볼 수 있다면', 즉, 재식별화(re-identification)가 가능하면 개인정보로서, 개인정보보호법의 규율대상입니다.

개인정보 보호법
제2조(정의)
1. "개인정보"란 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.

정보통신망 이용촉진 및 정보보호 등에 관한 법률
제2조(정의)
6. "개인정보"란 생존하는 개인에 관한 정보로서 성명·주민등록번호 등에 의하여 특정한 개인을 알아볼 수 있는 부호·문자·음성·음향 및 영상 등의 정보(해당 정보만으로는 특정 개인을 알아볼 수 없어도 다른 정보와 쉽게 결합하여 알아볼 수 있는 경우에는 그 정보를 포함한다)를 말한다.

- 현재 개인정보를 수집한 후 개인을 식별할 수 없게 하여 처리하는 것은 제한적으로 적법합니다. 현행 개인정보보호법은 개인정보 주체의 개인정보자기결정권을 실질적으로 보장하기 위하여 개인정보처리자가 '익명화'하더라도 정보주체의 동의 없이는 해당 개인정보를 통계 목적이나 연구 목적 등으로 제공하는 경우 외에는 제공할 수 없도록 규정하였습니다(이 법 제18조 제2항 제4호). 이처럼 개인정보보호법 규율의 예외가 되려면 '가명' 등으로 '비식별화'가 아니라 '익명화'가 되어 더 이상(no longer possible) 개인을 (재)식별할 가능성이 완전히 사라져야 합니다.

개인정보보호법
제18조(개인정보의 목적 외 이용·제공 제한)
① 개인정보처리자는 개인정보를 제15조제1항에 따른 범위를 초과하여 이용하거나 제17조제1항 및 제3항에 따른 범위를 초과하여 제3자에게 제공하여서는 아니 된다.
② 제1항에도 불구하고 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다(...)
4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우

- 반면 방송통신위원회의 <빅데이터 가이드라인>이나 현재 발의된 법안들은 '익명화'가 아닌 '비식별화'라는 개념을 규정하고 있습니다. '비식별화'는 익명화와 달리 '재식별화'의 가능성을 내포하고 그 (상업적) 활용성을 보장하고자 하는 개념입니다. 이는 헌법에 의해 보호받고 있는 개인정보 자기결정권의 행사를 침해합니다. 또한 현행 개인정보보호법에서 그 보호대상으로 규정하고 있는 개인정보의 전체 정의에 심각한 혼란을 야기합니다. 현재의 개인정보 관련 법률들에서는 직접적으로 식별되는 개인정보 뿐 아니라 다른 정보와 쉽게 결합하여 간접적으로 식별되는 개인정보도 차별없이 보호대상으로 정의하고 있기 때문입니다.

> 방통위 가이드라인
> 제2조(정의)
> 4. "비식별화"란 데이터 값 삭제, 가명처리, 총계처리, 범주화, 데이터 마스킹 등을 통해 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 쉽게 결합하여도 특정 개인을 식별할 수 없도록 하는 조치를 말한다.

> 강은희 의원안
> 제2조(정의)
> 9."비식별화"란 데이터 값 삭제,가명처리,총계처리,범주화 등을 통해 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 쉽게 결합하여도 개인을 식별할 수 없도록 하는 조치를 말한다.

> 강길부 의원안
> 제24조의3(비식별개인정보의 이용)
> ① 정보통신서비스 제공자는 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 쉽게 결합하여도 특정 개인을 알아볼 수 없도록 하는 조치(이하 "비식별화"라 한다)를 할 수 있다.

> 부좌현 의원안
> 제22조의2(개인정보 비식별화에 관한 특례)
> ① 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 개인정보를 비식별화(가명처리,범주화 등 대통령령으로 정하는 방식으로 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 결합하여도 개인을 알아볼 수 없도록 하는 조치를 말한다)하여 이를 처리할 수 있다.

- '비식별화'는 국제적으로 유례가 없는 개념입니다. 유럽연합이나 해외의 사례에서도 'de-identification'이라는 용어가 아닌 'anonymisation'라는 용어를 사용합니다. 예를 들어 영국의 ICO(Information Commissioner's Office)에서 발행한 비슷한 가이드라인의 경우 그 제목이 'Anonymisation: managing data protection risk code of practice'(익명화 : 데이터 보호 위험의 관리, 시행지침)입니다. 이를 '비식별화'에 대한 규범으로 해석하는 것은 잘못입니다.
- 무엇보다 비식별화 법안 등은, 비식별화만 한다면 개인정보주체의 동의를 받지 않고도 공개된 개인정보와 이용내역 정보를 포함한 개인정보를 수집, 저장, 조합, 분석 및 제공 등 처리할 수

있도록 하였습니다. 이는 현행 개인정보보호법이 보호하는 개인정보(직접적인 식별 뿐 아니라 다른 정보와 쉽게 결합하여 식별되는 경우도 해당)의 경우에 원칙적으로 정보주체의 동의를 받도록 한 이 법 제정 취지를 몰각하고 헌법 및 국제인권규범에서 보장하는 개인정보 자기결정권을 침해하고 있습니다.

> 방통위 가이드라인
> 제4조(공개된 정보의 수집·이용)
> ① 정보통신서비스 제공자가 개인정보가 포함된 공개된 정보를 비식별화 조치한 경우에는 이용자의 동의 없이 수집·이용할 수 있다. 다만, 이용자의 동의를 받거나 법령상 허용하는 경우에는 비식별화 조치를 취하지 아니하고 수집·이용할 수 있다.
> (...)
> 제10조(제3자 제공) 정보통신서비스 제공자는 개인정보가 포함된 공개된 정보, 이용내역정보, 생성 정보의 경우, 이용자의 동의를 얻어 제3자에게 제공할 수 있다. 다만, 비식별화 처리된 공개된 정보, 이용내역정보, 생성 정보는 이용자 동의 없이 제3자 제공이 가능하다.

> 강은희 의원안
> 제39조(개인정보의 비식별화)
> ① 개인정보처리자는 통계·연구·분석,공공정책의 수립,시장조사,마케팅 등의 목적을 위하여 필요한 경우 정보주체의 동의 없이 개인정보를 비식별화하여 처리할 수 있다.

> 강길부 의원안
> 제24조의3(비식별개인정보의 이용)
> ① 정보통신서비스 제공자는 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 쉽게 결합하여도 특정 개인을 알아볼 수 없도록 하는 조치(이하 “비식별화”라 한다)를 할 수 있다.
> ② 정보통신서비스 제공자는 비식별화를 통하여 다른 정보와 결합하여도 특정 개인을 쉽게 알아볼 수 없도록 가공된 정보(이하 “비식별화개인정보”라 한다)를 이용하는 과정에서 개인정보가 생성되는 경우 이를 지체 없이 파기하거나 추가적인 비식별화를 하여야 한다.

> 부좌현 의원안
> 제22조의2(개인정보 비식별화에 관한 특례)
> ② 개인정보처리자가 개인정보를 비식별화하여 처리하거나 비식별화된 개인정보를 처리하는 경우에는 제15조제1항제1호, 제17조제1항제1호 및 제18조제2항제1호에도 불구하고 정보주체의 동의 없이 이를 처리할 수 있다.

- 특히 방통위 가이드라인은 비식별화 처리 후 재식별화되더라도 ‘처리중단’이 아닌 ‘재비식별화’를 하여 ‘계속 이용’하도록 하고 있다는 점을 주목해야 합니다. 이는 ‘비식별화’란 개념이 명백히 ‘익명화’와 달리, 개인정보의 지속적인 이용을 보장한다는 사실을 보여주고 있습니다. 현행 개인정보 보호법의 정의에서 뿐 아니라 국제 개인정보보호 규범에 따르면, 재식별이 가능한 개인정보도 개인정보로서, 동의 없이 처리하는 것은 위법합니다. 개인정보 처리에 대해 정보주체의 동의가 없다면, 개인정보 처리자는 그 처리를 즉각 중단하고 해당 개인정보를 회수하거나 폐기해야 마땅합니다.

- 방통위 가이드라인에서 비식별화하면 동의 없이 사용할 수 있다고 본 '이용내역 정보'(이용자가 정보통신서비스를 이용하는 과정에서 자동으로 발생하는 서비스 이용기록, 인터넷 접속정보, 거래기록 등의 정보)는 매우 민감한 개인정보입니다. 통신사실에 대한 자료는 현행 통신비밀보호법에 따라 수사기관에 제공될 때 법원의 허가가 필요하기도 합니다. 쇼핑 내역, 검색 내역, 통신 내역, 의료 등의 이용내역 정보 또한 개인의 사상, 종교, 성적 취향, 정치적 신조, 노동조합 가입여부, 인종, 건강, 성생활에 대한 정보 등 매우 민감한 정보를 포함할 수 있다는 점에서 이에 대해 동의 없이 제공하는 것은 중대한 기본권 침해로 이어질 수 있습니다.
- 특히 우리나라 정보 환경에서 비식별화란 개인정보 보호에 아무런 효과가 없습니다. 통신사, 인터넷 포털, 유통, 신용카드사, 은행 등 개인정보가 대기업으로 집중되는 정도가 심하고, 그 동안 주민등록번호나 휴대전화번호 등 개인을 식별할 수 있는 정보가 광범위하게 활용되어 왔기 때문입니다. 그 동안 대규모 개인정보 유출 사례도 많아서 비식별화나 익명처리를 해도 개인정보를 안전하게 익명화하기는 거의 불가능한 상황입니다.
- 외국은 빅데이터 문제를 연구하며 더 이상 개인정보가 아니라고 볼 수 있는 '익명화'에서도 재식별될 가능성을 염두에 두고 매우 신중한 접근을 하고 있습니다. 개인정보를 설령 익명화하였더라도 기술 발전에 따라 개인정보 처리비용이 계속 낮아지고 가용 정보가 증가하기 때문에 추후 개인을 재식별할 가능성이 높기 때문입니다. 그런데 그간 수많은 사고들로 개인정보 보호의 토대가 취약해진 우리나라에서 '익명화'도 아니고 '비식별화' 법안을 추진하여 개인정보 보호의 예외를 넓히는 것은 국민의 눈을 속이고 기업의 무분별한 개인정보 처리에 포괄적인 허가장을 내어주는 것입니다.

4. <u>빅데이터 시대 필요한 개인정보 보호는 '비식별화'가 아니라 '프로파일링 규제'입니다.</u>

- 최근 국제사회는 빅데이터가 개인정보에 끼치는 위협으로 프로파일링(profiling) 문제에 주목하고 있습니다(별첨자료 참조). 프로파일링이란 개인을 평가하거나 개인의 업무실적, 경제상태, 위치, 건강, 선호, 행동을 분석 예측하기 위해 이루어지는 개인정보의 자동화된 처리를 말합니다(유럽 GDPR). 한마디로 개인별 평가, 분석, 예측을 자동적으로 처리하는 것으로 빅데이터 시대 개인정보 보호의 국제 이슈로 떠오르고 있습니다.
- 그러나 방통위 가이드라인은 프로파일링의 개념을 '정보처리 시스템'이라는 모호한 말 속에 암시적으로 담았으며, 보호하는 것이 아니라 그 활용을 독려하고 있습니다.

> 방통위 가이드라인
> 제2조(정의)
> 3. "정보 처리시스템"이란 공개된 개인정보 또는 이용내역정보 등을 전자적으로 설정된 체계에 의해 조합·분석 등 처리하여 새로운 정보를 생성하는 시스템을 말한다.

- '정보처리 시스템'이라는 모호한 개념을 통하여 정보주체의 동의 없이 개인에 대한 새로운 정보를 생성하여 그를 자동적으로 평가, 분석, 예측하는 것은 현행 법률에 위배될 뿐 아니라 헌법에서 보장하고 있는 개인정보자기결정권을 중대하게 침해합니다.
- 현재 발의된 법안들 중에서 프로파일링 처리를 규제하는 등 빅데이터 시대 개인정보를 보호할 수 있는 대책은 보이지 않습니다. 유럽 GDPR에서 프로파일링 처리에 대한 동의, 프로파일링을 거부할 권리, 프로파일링의 제한 등 정보주체의 권리를 규정하고 있는 점과 대조적입니다.

5. 이상과 같은 이유에서 우리는 빅데이터 산업 활성화를 위해 방통위 가이드라인이나 금융위원회에서 도입하고 있는 '비식별화'개념에 반대합니다. 더불어 '비식별화' 개념을 법적으로 도입한 강은희, 강길부, 부좌현의원안에도 반대합니다.

전세계적으로 빅데이터 산업이 개인정보 보호에 미칠 영향을 둘러싼 논의가 진지하게 이루어지고 있습니다. 특히 계속된 개인정보 유출 사고로 개인정보 보호 토대가 취약해진 것으로 지적받는 우리나라에서 지금 필요한 것은 성급한 입법이 아닙니다. 정부와 국회는 빅데이터 시대 예상되는 기업의 무분별한 개인정보 처리로부터 소비자의 개인정보를 보호하기 위하여 프로파일링을 규제하는 등 개인정보를 보호하기 위한 조치 마련에 우선적으로 나서야 할 것입니다.

끝.

<별첨자료> 빅데이터 (EU 자료 번역)

==========

빅데이터 Big Data

◎ 옮긴이주 ◎

유럽에서는 빅데이터에 대해 개인정보를 보호할 수 있는 방안을 진지하게 검토해 왔습니다. 유럽연합에서 개인정보를 주무하고 있는 개인정보보호 작업반(ARTICLE 29 DATA PROTECTION WORKING PARTY)에서는 지난 2013년 4월 2일 목적제한에 대한 의견(Opinion 03/2013 on purpose limitation)을 채택하였는데, 이 의견의 부록으로 개인정보 보호 측면에서 빅데이터가 가지고 있는 우려점에 대해 잘 설명하고 있습니다.
인터넷에서 정보를 수집하여 처리하는 빅데이터는 원칙적으로 개인정보가 처음 수집되었을 때의 '목적'과 다른 목적으로 처리될 가능성이 있다는 점에서 원칙적으로 개인정보 보호법의 '목적 구속의 원칙'을 위배할 우려가 있습니다.
이 의견서는 그런 상황에서 빅데이터 처리의 '두 가지 시나리오'를 제시하고 있습니다. (1) 처리자가 트렌드와 정보의 연관성을 추적하는 경우와 (2) 처리자가 개인들에게 맞춤하여 추적하는 경우입니다. 빅데이터 처리에서 개인이 식별될 가능성이 없도록 익명화되어 (1)과 같이 처리되는 경우에는 큰 문제가 없지만 (2)와 같이 처리되는 경우에는 처리되는 정보주체들이 동의를 받아야 합법적이라고 이 의견서는 지적하고 있습니다.
결국 빅데이터 시대에서도 정보주체의 헌법적 기본권인 개인정보 자기결정권은 여전히 중요한 문제인 것입니다.

* http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

'빅데이터'와 '빅데이터 분석'이란 무엇인가?

앞서 3장 2.5절에서 간략히 강조하였듯이, '빅데이터'는 정보의 유용성과 자동화된 처리 측면에서 기

하급수적인 성장을 나타내고 있다. 빅데이터는 기업, 정부, 거대조직에서 보유한 방대한 대용량 디지털 데이터를 말하는데, 이 데이터들이 컴퓨터 알고리즘을 통해 대규모로 분석되고 있다. 빅데이터는 향상된 기술력을 필요로 한다. 많은 양의 데이터를 수집하고 저장할 수 있어야 할 뿐 아니라 (분석 애플리케이션을 이용하여) 정보가 가지고 있는 가치 전체를 분석하고 이해하고 이점을 취할 수 있어야 하기 때문이다. 빅데이터는 궁극적으로 더 많은 정보에 기반한 더 나은 의사결정을 이끌 것으로 기대된다.

건강, 모바일 통신, 스마트 그리드, 교통 관리, 부정행위 적발, 마케팅과 소매업 등 온라인과 오프라인의 다양한 영역에 빅데이터를 다루는 여러 애플리케이션이 존재한다. 빅데이터는 일반적인 트렌드와 연관성을 파악하는 데 사용될 수 있지만, 그 처리 결과가 개인들에게 직접적인 영향을 미칠 수도 있다. 예를 들어, 마케팅과 광고 영역에서 빅데이터는 소비자의 개인적인 취향, 행동, 태도를 분석하고 예측하는 데 사용될 수 있고, 나중에는 그 소비자의 프로파일에 근거한 맞춤 할인, 특가 판매, 맞춤 광고 등 그 소비자에 대해 취해질 '조치와 결정'에 영향을 미칠 수 있다.

빅데이터로 인해 개인정보보호와 프라이버시권에 야기될 위험성과 어려움은 어떤 것들이 있는가?

그 혁신가능성에도 불구하고 빅데이터는 개인정보보호와 프라이버시권에 중대한 위협을 야기할 수 있다. 특히 빅데이터에 대해 다음과 같은 우려가 제기된다.

- 데이터 수집, 추적, 프로파일링의 가파른 증가 규모에 대한 우려. 수집된 데이터들의 다양성과 상세함, 데이터들이 종종 다른 많은 출처의 데이터들과 결합된다는 사실을 고려함.
- 정보 보안에 대한 우려. 양적 팽창에 비해 뒤쳐지는 것으로 보여지는 보호 수준에 기인함.
- 투명성에 대한 우려. 충분한 정보가 제공되지 않는다면 개인(소비자나 이용자)은 자신이 이해하지 못하고 통제할 수 없는 결정에 종속될 수 밖에 없음.
- 부정확성, 차별, 배제, 경제적 불균형에 대한 우려. (하단에서 논의)
- 정부 감시의 강화 가능성

사용된 분석 애플리케이션의 유형에 따라 부정확하거나 차별적이거나 불법적인 결과로 이어질수 있다. 특히, 알고리즘은 연관성에 주목하여 통계적인 추론을 산출하는데, 이것이 마케팅이나 다른 의사결정에 부당하고 차별적인 영향을 미칠수 있다. 이는 현존하는 편견이나 스테레오 타입을 영속시키고, 사회적 배제와 계층화 문제를 악화시킬 수 있다.

나아가, 보다 큰 틀에서 보자면, 대용량 데이터와 이러한 데이터를 검사하는데 사용된 정교한 분석도구의 유용성은 큰 기업과 소비자 간에 경제적인 불균형도 증가시킬 수 있다. 이러한 경제적 불균형은 소비자에게 제공되는 생산품과 서비스 관련해서 부당한 가격 차별을 불러올 수 있을 뿐 아니라, 상당히 침해적이고 생활에 지장을 주고, 개인에게 맞춤한 타겟 광고와 제공물들로 이어질수 있다.

이는 개인에게 또다른 중요한 부정적인 결과를 낳을 수 있다. 구직 기회, 은행 대출, 건강보험 선택사항과 관련한 경우들과 같은 예에서 말이다.

개인정보를 [목적]부합적 분석에 추가이용하려면 어떤 보호수단이 필요한가?

[목적]부합성 평가에 있어서는, 수집 목적과 맥락, 정보주체의 합리적 기대, 개인정보의 속성과 정보주체에 미치는 영향 간의 관계 등 3장 2.2절에 서술된 모든 관련 요소들이 고려되어야 한다. 공정한 처리를 보장하고 부당한 영향을 방지하기 위해 채택된 보호수단에 대한 평가 또한 중요하다. 덧붙여, '역사적, 통계적, 과학적 목적'과 관련한 특별 조항이 또한 관련이 있다.

어떤 보호수단이 필요한지 알아보기 위해, 두 가지 다른 시나리오를 구분하는 것이 도움이 될 수 있다. 첫째로는, 데이터를 처리하는 기관이 트렌드와 정보의 연관성을 추적하는 것을 원하는 경우이다. 두번째는, 기관이 개인에게 관심을 갖는 경우이다.

첫번째 시나리오에서는 기능적 분리 개념이 핵심적이다. 이 데이터를 (마케팅이나 다른) 연구에 추가적으로 이용할 때 [목적]부합적인 것으로 볼수 있는지 아닌지 판단하려면, 그 한도가 어디까지인지가 중요한 근거가 된다. 이런 사례들에서는 개인정보처리자가 정보의 기밀성과 보안을 보장할 필요가 있으며, 기능적 분리를 보장하기 위해 모든 필요한 기술적, 조직적 조치를 취해야 한다.

두번째 가능한 시나리오는 어떤 조직이 개인 소비자의 취향, 행동, 태도를 구체적으로 분석하고 예측하기를 원할 때인데, 이는 나중에 이 소비자와 관련하여 취해지는 '조치나 결정'에 영향을 미칠 것이다.

이런 경우에는, 자유롭고, 구체적이고, 충분한 정보에 입각하고 명확한 '옵트인' 동의가 거의 대부분 요구되어야 하고, 그렇지 않으면 추가 이용이 부합적이라고 볼수 없다. 어떤 경우에는 그런 동의가 요구되어야만 한다는 사실이 중요한데, 직접 광고(direct marketing), 행태 광고(behavioural advertisement), 데이터 판매, 위치기반 광고(location-based advertising)나 추적기반 디지털 시장조사(tracking-based digital market research)와 같은 경우가 그렇다.

자신의 동의에 충분한 정보를 제공받고 투명성을 보장받기 위해, 정보주체/소비자는 자신의 '프로파일'에 접근할 수 있어야 할 뿐 아니라, 프로파일을 생성하는 의사결정 로직(알고리즘)에도 접근할 수 있어야 한다. 달리 말하면, 기관은 의사결정 기준을 공개해야 한다. 이는 결정적이며, 빅데이터 세계에서 그 어느 때보다 중요한 보호수단이다. 대개 민감한 것은 수집된 개인정보 자체가 아니다. 오히려 개인정보로부터 야기되는 사생활 침해와 그런 사생활 침해가 이루어지는 방식이 민감한 것이다. 나아가 프로파일의 생성으로 이어지는 데이터의 출처는 공개되어야 한다.

특히 부적절한 사생활 침해 위험성을 고려한다면, 정보주체/소비자들이 원할 때 자신의 프로파일을 수정하거나 갱신할 수 있어야 한다는 점도 중요하다. 개인정보처리자들이 좀더 정확한 정보에 기반하여 마케팅 등에서 의사결정을 하고자 할 때에도 득이 될 것이다.

나아가, 많은 경우 소비자/정보주체들로 하여금 자신의 데이터에 대해 이전가능하고, 이용자 친화적이며 기계판독가능한 형식으로 직접 접근할 수 있도록 허용하는 등의 보호수단은, 소비자/정보주체들이 권능을 발휘하게끔 하고 거대기업과 정보주체/소비자 간의 경제적인 불균형을 시정하는데 도움이 될 것이다. 이는 또한 개인들이 빅데이터가 창출한 '부를 공유'할 수 있도록 하고, 개발자들이 이용자들에

게 추가적인 기능과 애플리케이션을 제공하도록 장려할 것이다.

예를 들어, 에너지 소비자에게 이용자 친화적인 형태로 정보에 접근할 수 있도록 하면 주택소유자들이 좀더 쉽게 요금제를 바꾸거나 가스/전기 효율을 최대화할 수 있을 것이다. 또한 이들에게 자신의 에너지 소비를 모니터하고 자신의 생활양식을 바꿔 환경적 영향 뿐 아니라 청구금액을 감소시키도록 할 수도 있다.

데이터 이전성을 보장하는 것은 산업과 정보주체/소비자들이 빅데이터의 이점을 보다 조화롭고 투명한 방식으로 극대화할 수 있도록 한다. 이는 또한 부당하고 차별적인 관행을 최소화하고 의사결정 목적으로 부적절한 데이터를 사용하는 데 따른 위험성을 감소시킬 수 있으며, 산업과 정보주체/소비자 모두에게 득이 될 것이다.

(주석) 관련 사례

◎ 휴대전화 위치가 도로안전정비 정책을 지원한다 ◎

교통부는 다양한 경로로 움직이고 있는 휴대전화 - 종국적으로는 이들을 탑재한 교통수단 - 의 속도를 계산하기 위해서 휴대전화 위치정보를 사용할 수 있는지 통신사에 문의해 왔다. 휴대전화 데이터는 특정 도로 구간에서 속도가 보편적이라는 사실을 드러낸다. 따라서 이 정보들은 도로안전 정책을 수립하는 데 사용될 수 있는데, 이런 정책은 나중에 해당 지역에서 도로교통 사망사고 발생을 유의미하게 감소시키는 결과를 낳을 것으로 보인다. 정보주체가 재식별화될 위험성을 최소화하기 위해, 휴대전화 데이터는 교통부에 제공되기 전 효과적으로 익명화시킨다. 세심한 영향평가가 이루어지고, 침투테스트가 수행되고, 이해당사자들이 자문한다. 이런 시나리오에 대해 우리는 모든 요소들이 재식별화의 위험성을 매우 낮추거나 최소화시킬 것이고 정보주체에게 영향이 있더라도 비교적 낮은 영향을 미칠 것이 확실하다고 추정한다.

이 시나리오는 세부적인 부합성 평가를 요구한다. 처음에 특정한 목적으로 수집되는 통신 데이터는 이제 (도로교통 관련) 다른 목적으로 사용된다. 대부분의 사람들은 자신의 데이터가 다른 방식으로 사용될 것이라고 일반적으로 예상하지 않는다. 이는 [개인정보 수집의] 목적에 부합할 수 없다는 강력한 초기 표지가 될 수 있다. 수집된 휴대전화의 위치정보에 대한 관계적 감수성 또한 이런 평가를 지지할 수 있다.

그러나 이 경우, 이차적 목적으로 사용/제공되기에 앞서, 데이터는 효과적으로 익명화된다는 가정이 있다. 그러므로 두 가지 목적이 다르다 하더라도, 익명화가 완벽하게 적절하다는 가정에서라면(그래서 그 정보가 더이상 개인정보로 간주되거나, 재식별화될 위험성이 매우 낮은 회색지대에 떨어진다면) 이는 [목적에] 부합하지 않는 처리에 대한 우려를 감소시킬 것이다. 그럼에도 불구하고, 처리의 완전한 투명성과 같은 추가적인 보호수단이 여전히 권장된다. 특히, 완벽한 익명화가 보장될 수 없거나 [재식별화] 위험성이 남아 있다면, 이런 문제점을 공개해야 한다. [유럽 개인정보보호 디렉티브] 13조의 예외가 적용될 수 없다면, 그에 대한 규칙으로서 충분한 정보에 입각한 동의를 받아야 할 것이다.

◎ 구매습관을 통해 소비자의 임신을 예측하는 비밀 알고리즘 ◎

한 백화점이 고객들의 구매습관을 분석하고 새로운 마케팅 트렌드를 알아보고 고객들에게 특가판매와 할인쿠폰을 제공하기 위해 포인트적립카드 데이터를 사용한다. 백화점에서 사용한 혁신적인 분석 소프트웨어는 여성 고객이 임신했을 가능성과 몇개월인지를 높은 확률로 예측한다. 이 정보는 고객들의 프로파일에 맞춰 마케팅을 조정하기 위해 사용된다. 고객들이 포인트적립카드에 가입할 때는 [이런 상황에 대한] 구체적인 정보가 제공되지 않는다. 상세 계약조건(백화점 웹사이트에서 볼 수 있는)에는 단지 '포인트적립카드 데이터는 고객들에게 특가판매나 할인쿠폰을 제공하는 등 마케팅 목적으로 사용될 수 있습니다'라고만 언급하고 있다. 이 백화점은 한 십대소녀 아버지로부터 항의를 받는다. 이 소녀는 집 우편함으로 다량의 임신 관련 광고가 도착된 사실에 대해 추궁받았고 결국 임신 3개월이라는 사실이 발각되었다.

위 시나리오는 바로 명백한 프라이버시 문제를 제기하고 있다. 어떤 임산부들, 특히 임신 초기의 임산부들은 임신 소식을 본인만 알고 있거나 아주 밀접한 가족친지들에게만 알리고 싶어할 수 있다. 프로파일링이 (임신을 예측하기 위한 비밀 알고리즘을) 수행한 방식은 분명 다수 고객들이 기대하지 않았고, 부적절했고 무례한 것이었다. 문제는 (그 자체로는 침해성이 적은) 본래 수집된 데이터의 속성으로 인한 것이 아니다. 은밀하고 불쾌한 알고리즘을 사용하여 전반적인 프로파일(임신이나 그 개월수)을 예측하기 위해 데이터를 결합하고 추가적으로 처리하고 이용한 방식에서 발생한 것이다.

이 사례가 제기하는 다른 모든 이슈들을 차치하고 위 사실들을 종합해 보면, 우선적으로 데이터가 처리되는 방식과 보호수단(투명성 뿐 아니라 진실하고 충분한 정보에 기반한 동의 등)의 부족 때문에 [목적에] 부합할 수 없다는 강력한 지표가 존재한다. 이 사례는 다음 사례와 대비되는데, 이 사례 역시 고객 프로파일링에 대한 것이지만 보다 사회적으로 수용가능한 방식이다.

◎ 잔디깎는 기계에 대한 특가판매 ◎

원예용품과 DIY 장비를 판매하는 전국적인 대형매장이 고객들에게 보통 수준의 연간회비를 받고 포인트적립카드를 제공하며, 이 카드를 사용한 모든 구매액의 10% 에 대해 할인을 제공한다. 회사 웹사이트는 유익한 프라이버시 고지를 게시하고 있으며, 포인트적립카드에 가입하는 고객을 위해 선택사항을 명시한 단축본도 제공한다.

고지사항은 무엇보다도 명확히 서술 및 언급하기를, 고객이 '맞춤 할인을 제공받을 수 있도록 본인의 구매이력을 온라인으로 저장하고' 이 구매이력이 '구매양식을 분석하여 단골고객을 위한 맞춤 특가판매를 제공'하는 데 사용될 수 있다는 선택사항(옵션A)을 옵트인으로 선택할 수 있다. 혹은, 고객들이 자신의 포인트적립카드를 스스로 보관하고 여전히 10% 할인(또는 다른 일반적인 할인)을 제공받을 수 있다(옵션B)고 고지문은 설명한다. 즉 '나는 나의 세부적인 프라이버시가 지켜지기 바라며 일반적인 할인만을 제공받겠다'는 옵션을 선택함으로써 고객은 프로파일링되지 않고 맞춤 제공이나 할인을 받지 않을 수 있다. 보다 상세한 내용은 온라인과 오프라인으로 찾아볼 수 있다.

어느 봄날 단골 고객이자 열정적인 정원사가 맞춤 할인을 선택하였고, 우편으로 특가할인을 안내받았다. 자신의 낡은 잔디기계가 막 말썽을 피우기 시작할 때 보다 저소음이고 에너지효율이 높은 신상품을 30% 할인한다는 소식이다.

이 고객은 흥미가 생겨 보다 자세한 사항을 알고자 온라인을 방문한다. 각각의 카드 소지자는 맞춤 추

천상품과 특가에 대한 정보를 받을 수 있을 뿐 아니라 지난 5년간의 구매이력, 즉 해당 상점이 기본 설정에 따라 보유하고 있는 정보에 접근할 수 있다. 그 사이트는 구매를 분석하고 고객이 좋아할 만한 다른 상품을 추가적으로 추천하기 위해 이용자친화적인 많은 기능들을 가지고 있다. 또 원예 상점에서 이용하는 분석 소프트웨어가 작동되는 방식에 대해 매우 특징적인 정보도 게시하고 있는데, 이는 해당 산업의 공통된 관행에 초점을 둔 것이다. 예컨대 고객이 과거 구매했던 상품에 대한 특가판매는 자신의 구모델을 대체할 생각을 하기 시작할 때쯤 제공된다고 설명하고 있다.

또 이 게시글은 할인 적용이 다양한 요소들에 기반하여 최적화될 것이라고도 설명한다. 상점에서 고객이 월평균 지출한 금액(더 많이 지출할수록 할인폭이 커진다), 과거 특가구매 경력, 그밖의 여러 유사한 지표들이 그런 요소들로, 투명하고 상세하게 설명되어 있다. 진작부터 이런 투명성은 웹사이트의 '자유게시판' 에서 잔디깎는 기계들이 고장나는 평균시간에 대한 농담, 그리고 어떻게 시스템을 '속여' 더 많은 할인을 받을 것인가에 대한 공유 전략과 팁으로 이어져 왔다. 예를 들어, 최근에는 많은 소비자들이 웹사이트에서 자신이 쇼핑하고 있음을 드러내기 위해 일부러 할인 상품을 클릭하고, 더 높은 할인율에 더 잘 반응하겠다는 의사를 내비치곤 한다.

이 사이트는 또한 고객의 구매 이력을 평균 양식으로 다운로드할 수 있도록 허용한다. 예컨대 일부 고객들은 자신의 개인 재정을 계획하고 분석하기 위해 이 정보들을 자신이 사용하는 (별도의) 소프트웨어에 통합시켜 버릴 수도 있다.

임신 예측과 관련하여 위에 거론된 사례처럼, 이 경우에도 세부사항에 대한 복잡한 분석을 요구하며, 간략한 요지만으로는 물론 설명될 수 없다. 그럼에도 불구하고 두 가지 사례를 비교하는 것은 가치가 있다. 많은 유사성이 있지만 많이 다르기도 하다. 두 사례 모두 마케팅 목적으로 고객 프로파일링을 포함하였지만, 상식적으로 볼 때 첫번째가 대다수에게 불쾌함을 줄 것이 자명한 반면 두번째는 훨씬 덜 문제시될 것이다.

결국 첫번째 사례에서 가장 중요한 요소는, 겉보기에는 무해한 구매 데이터에서 임신을 예측하는 알고리즘의 이상한 능력이 [정보주체/소비자의] 기대에서 어긋났다는 사실이다. 반대로, 원예상점은 고객을 훨씬 더 예측가능하고 (심지어 편리하고) 합리적인 방식으로 프로파일한 것으로 나타났다. 구모델을 교체해야 할 때쯤 새로운 잔디깎는 기계를 할인해주는 방식으로 말이다. 특가판매를 제공하거나 회사가 제공시기를 계산하는 방식은 놀랍거나 불쾌한 일이 아니다. 결정적인 차이는 알고리즘이 설계되는 방식에 있다. 일반적으로 합리적인 대중의 기대에 부합하는지 여부 혹은 불쾌하거나 부당한 일이 있는지 여부 말이다.

이러한 관점에서, 마케팅 목적으로 추적하거나 프로파일링하는 것은 보통, 진실하고 명확하고 자유롭고 충분한 정보에 입각한 동의 등 법적 근거가 있을 때에만 목적부합적 사용으로 인정된다는 사실을 강조하고자 한다. 두번째 사례에서 원예상점은 고객들에게 투명성을 보장하고 선택권을 제공하기 위해 중요한 노력을 취한 것으로 보인다. 이런 보호수단은, 결국 예측성에 기여하고 합리적인 기대성을 확실히 할 수 있다. 전반적으로 공정함을 보장하고 정보주체가 예측못한 불쾌한 영향을 최소화할 수도 있다. 참으로, 회사가 그 의사결정 기준 – 프로파일링 알고리즘 – 을 공개한다면 부당하거나 불쾌한 방법을 사용하는 일이 적을 것이다.

마지막으로, 데이터의 속성도 [목적부합성] 평가에서 고려사항이 될 수 있다. 원예도구와 원예용품 구매의 세부 양식이 개인에 대한 중요한 정보로 드러났음에도 불구하고, 전반적으로 볼 때 이 정보들은 사람들이 방문하는 웹사이트, 대여/구입하는 도서나 영화, 또는 약국에서 구매하는 약품을 알아내는 것처럼 사생활을 침해하는 유형의 민감한 정보는 아닐 것이다. <끝>