오픈넷 포럼

# 빅데이터와 사물인터넷 시대,
# 비식별화 정보는 개인정보인가?

## 빅데이터와 사물인터넷 시대에
## 개인정보는 어떻게 보호되어야 할까요?

**2016년 3월 21일(월) 오후 7시 30분**

**스타트업 얼라이언스 &스페이스**

**[오픈넷 포럼]**

# 빅데이터와 사물인터넷 시대, 비식별화 정보는 개인정보인가?

일시: 2016년 3월 21일(월) 오후 7시 30분 ~ 9시 30분
장소: 스타트업 얼라이언스 &스페이스
주최: 사단법인 오픈넷

<사회>
**박지환** | 오픈넷 변호사

<발제>
**"개인정보 비식별화 또는 익명화 쟁점"**
**심우민** | 국회입법조사처 입법조사관

<토론>
**박경신** | 고려대 법학전문대학원 교수, 오픈넷 이사
**전응준** | 유미 법무법인 변호사
**이영환** | 건국대 정보통신대학원 교수

발제

# 개인정보 비식별화 또는 익명화 쟁점

심우민

국회입법조사처 입법조사관

# 개인정보 비식별화 또는 익명화 쟁점*

심 우 민**

## 1. 개인정보 보호입법을 둘러싼 논란

□ 세계 각국은 네트워크 초연결 사회에서의 개인정보 보호 문제를 해소하기 위하여 개인정보 보호입법에 대한 진중한 논의를 이어가고 있음

○ EU의 경우에는 비교적 최근에 발의되어 최종 합의에 이른 「General Data Protection Regulation(이하, GDPR)」이 가장 대표적인 규제대응 사례임

○ 미국의 경우도 개인정보 보호에 관한 지속적인 논의가 이어져 왔으며, 2014년 5월 「BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES」라는 백악관 주도의 보고서가 발표된 바 있음

□ 즉 이러한 개인정보 보호입법의 개선논의는 비단 우리나라의 경우에만 문제시 되는 것이 아니라, 급격한 정보화 또는 데이터화의 국면에 처해 있는 대부분 국가들의 문제라고 할 수 있음

○ 개인정보의 개념정의와 관련하여, 다른 나라에 비하여 우리나라의 개인정보 개념정의 규정이 매우 포괄적이어서 문제가 있다는 지적이 산업계를 중심으로 제시되고 있는데, 실제 각국의 개인정보 보호법제의 개념정의 규정들과 비교해 볼 때, 그 근거가 다소 모호하다고 판단됨[1](cf. 형사제재 등)

- 좀 더 구체적으로 우리나라 현행 법제상의 개인정보 개념정의의 문제점에 관한 지적은 입법의 문제가 아니라, 법원 판결에서의 문제임을 확인해 둘 필요가 있음

### [표 1] 각국의 개인정보 개념정의 규정

| 국가 | 개인정보 개념정의 규정 |
|---|---|
| EU<br>(Data Protection Directive) | 개인정보는 '식별되거나 식별가능한 자연인'(이하, '정보주체'라 한다)에 관한 정보를 말한다; 식별가능한 개인이란 신분증 번호를 통하여 또는 신체적, 생리적, 정신적, 경제적, 문화적, 사회적 요소에 관한 하나 또는 그 이상의 정보를 통하여 직·간접적으로 알아볼 수 있는 사람을 의미한다. |
| 영국<br>(Data Protection Act) | 개인정보란 다음으로부터 식별될 수 있는 살아있는 개인에 관한 정보를 의미한다.<br>(a) 살아있는 개인을 식별할 수 있는 정보<br>(b) 살아있는 개인을 식별할 수 있는 정보 및 개인정보처리자가 보유하고 있거나 보유할 가능성이 있는 기타 정보<br>그리고 개인에 관한 의견, 개인정보처리자 또는 그 밖의 사람들의 개인에 관한 의사표현을 포함한다. |

---

* 이 발제문은 필자가 개인 식별정보 비식별화 논의와 관련하여 기존의 각종 토론회에서 발표하였던 내용들을 발표의 용이성을 위해 수정·보완한 것이며, 입법적 차원에서의 비식별화 또는 익명화 문제에 대해서는 별도의 연구결과를 제시할 것임을 밝혀둡니다.

** 국회입법조사처 입법조사관; 법학박사(legislation21@gmail.com)

1) 물론 국내에서의 개인정보 보호 규정 위반행위가 반드시 형사규제와 연계될 수 있어 문제가 있다는 지적에 대해서는 공감하는 측면이 있음

| | |
|---|---|
| 프랑스<br>(Loi n° 78-17 du 6 janvier 1978 relative a l'informatique, aux fichiers et aux libertes L'Assemblee nationale et le Senat ont adopte) | 신분증 번호 또는 자연인에 관한 하나 또는 그 이상의 요소로 직 간접적으로 식별되거나 식별가능한 자연인에 관한 모든 정보<br>개인이 식별가능한지 여부를 결정하기 위해서는 개인정보처리자 또는 그 밖의 사람들이 이용 또는 접근가능한 모든 수단을 고려해야 한다. |
| 독일<br>(Bundesdatens chutzgesetz) | 개인정보란 식별되거나 식별가능한 개인에 관한 인적 물리적 환경에 관한 모든 정보를 말한다. |
| 스웨덴<br>(Personal Data Act) | 이 법에서 사용되는 용어의 뜻은 다음과 같다.<br>개인정보 – 살아있는 자연인과 직 간접적으로 관련이 있을 수 있는 모든 유형의 정보 |
| 일본<br>(개인정보보호법) | 이 법률에서 개인정보라 함은 생존하는 개인에 관한 정보로서, 해당 정보에 포함되는 성명, 생년월일 기타 기술 등에 의해 특정한 개인을 식별 할 수 있는 것(다른 정보와 쉽게 조합할 수 있고, 그에 따라 특정한 개인을 식별하는 것이 가능하게 되는 것을 포함한다)을 말한다. |
| 호주<br>(Privacy Act) | 개인정보란, 진실여부 또는 기록된 형태에(데이터베이스를 구성하고 있는 정보 또는 의견을 포함한다) 상관없이, 개인에 관한 정보 또는 의견을 통하여 신원을 알 수 있거나 신원을 합리적으로 확인 가능한 경우, 그 정보 또는 의견을 말한다. |
| 캐나다<br>(Personal Information Protection and Electronic Documents Act) | 개인정보란 식별가능한 개인에 관한 정보를 말한다. 단 기관 직원의 이름, 직함, 직장주소 및 전화번호는 포함하지 않는다. |

□ 이상과 같은 세계적인 개인정보 보호법제를 둘러싼 논란은, 빅데이터(Big Data), 사물인터넷(Internet of Things) 등 새로운 정보통신 환경의 변화로부터 기인하는 것이라고 할 수 있을 것임

○ 필자는 개인적으로 이러한 개인정보 보호법제상의 논란이 발생하는 근원적인 이유 중 하나는 개인정보 또는 프라이버시 보호법제의 '패러다임 교착현상' 때문인 것으로 파악하고 있음

○ 현재는 개인정보자기결정권이라는 법적 권리를 근간으로 개인정보를 보호하는 법령체계를 취하고 있지만, 현실적으로 이러한 개인정보 보호법제의 궁극적인 목적은 개인의 사생활 또는 프라이버시의 부당한 침해를 방지하고자 하는 것이라고 할 수 있음

○ 현재의 기술적 발전상황에 비추어볼 때, '개인 식별 가능성'을 중심으로 한 개인정보의 법적 보호체계는 한계에 봉착할 수밖에 없음

- (개인정보 개념) 데이터의 급증은 물론이고, 이의 분석 및 예측 기술의 고도화로 인하여 식별 가능성 경계가 불명확해지고 있음
- (동의요건) 그 결과 식별 가능성을 전제로 정보주체의 개인정보자기결정권의 실현방식의 일환으로 동의권을 중심으로 규정하고 있는 체계 또한 한계에 봉착하게 됨

□ 국내에서도 빅데이터 등과 관련한 개인정보 보호입법에 대한 논의가 이루어지고 있는 것은 사실이지만, 해외 주요 사례들에서와 같이 진지한 입법(법률)적 차원의 성찰이 전제되어 있지 않음

○ 빅데이터 활용과 관련하여 현행 규제(수집제한 및 수집시 동의요건 등)가 장애 요인으로 등장하고 있는데, 대부분의 부처들은 비식별화(de-identification) 조치를 전제로 법률상 규정된 동의요건을 면제하는 데에 초점을 두고 있는 상황임

- 방송통신위원회는 2014년 12월 「빅데이터 개인정보보호 가이드라인」을 확정·공표한 바 있으며, 그 주요 내용은 개인정보의 비식별화를 거친 정보 활용에 관해 법률상 정

보주체의 동의요건을 면제해 주는 것이었음

- 이후 미래창조과학부,[2] 행정자치부[3] 등은 방송통신위원회의 가이드라인 내용과 거의 동일한 안내서(가이드라인)를 출간해 오고 있음

○ 또한 2015년 6월 3일 금융위원회는 금융분야(핀테크) 빅데이터를 활성화 정책을 공표하였으며, 여기에서도 다른 부처들과 유사하게 비식별화 정책기조를 제시함[4]

## 2. 동의요건 면제 중심의 가이드라인

□ 방송통신위원회가 2014년 12월 최종적으로 발표한 「빅데이터 개인정보보호 가이드라인」은 형식적으로만 보자면 상당한 정도의 의견수렴 절차를 거친 것으로 판단됨

○ 최초 방송통신위원회는 2013년 12월 18일, 「빅데이터 개인정보보호 가이드라인(안)」을 제시한바 있었으나, 이에 대해서는 시민단체의 강력한 비판이 있었음

○ 시민단체 등의 의견을 감안하여 방송통신위원회는 2014년 3월 19일, 「빅데이터 개인정보보호 가이드라인(안)」 수정안을 발표함

- 그러나 시민단체는 수정된 가이드라인은 여전히 △공개된 개인정보 및 이용정보를 정보주체의 동의 없이 수집·이용 할 수 있고 △이를 활용하여 새로운 정보의 생성할 수 있고 △개인정보·이용정보·생성정보를 자유롭게 제3자에게 제공할 수 있도록 해 사생활이 침해될 수 있다는 비판적 의견을 제시하였음

- 이와 관련하여, 방송통신위원회는 재차 수정안을 제시하겠다는 계획을 시사하였음

○ 이후 한 차례의 공청회를 더 거친 결과, 최종적으로 2014년 12월 23일 「빅데이터 개인정보보호 가이드라인」을 확정·발표하였음

- 확정된 가이드라인의 중심 내용은 개인정보 동의요건을 과거 초안에서와 같이 무조건 완화하는 것이 아니라, 비식별화 조치를 취한 경우 이를 완화시킨다는 내용임

- 향후 이와 관련하여, 비식별화 문제는 입법화하자는 논의가 쟁점으로 부각될 것으로 판단됨

□ 방송통신위원회가 제시한 「빅데이터 개인정보보호 가이드라인」의 신구조문 대비표는 다음과 같음[5]

### [표 2] 「빅데이터 개인정보보호 가이드라인」

| 초안 | 제1차 수정안 | 최종 |
|---|---|---|
| 제1조(목적) 이 가이드라인은 공개된 개인정보 또는 이용내역정보 등을 전자적으로 설정된 체계에 의해 조합, 분석 또는 처리하여 새로운 정보를 생성함에 있어서 이용자의 프 | 제1조(목적) (초안과 같음) | **제1조(목적)** 이 가이드라인은 공개된 개인정보 또는 이용내역정보 등을 전자적으로 설정된 체계에 의해 수집·저장·조합·분석 등 처리하여 새로운 정보를 생성함에 있어서 이용자 |

2) 미래창조과학부·한국정보화진흥원·K-ICT 빅데이터센터, 『빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서 (Ver 1.0)』, 2015.5.
3) 행정자치부·한국정보화진흥원, 『개인정보 비식별화에 대한 적정성 자율평가 안내서』, 2014.12.
4) 금융위원회, 「금융권 빅데이터 활성화 방안」, 2015.6.3.
5) 방송통신위원회, 「온라인 개인정보보호 세미나 결과 보고」, 2014. 3 및 방송통신위원회, 「빅데이터 개인정보보호 가이드라인」, 2014. 12. 23.

| | | |
|---|---|---|
| 라이버시 등을 보호하고 안전한 이용환경을 조성하는 것을 목적으로 한다. | | 의 프라이버시 등을 보호하고 안전한 이용환경을 조성하는 것을 목적으로 한다. |
| 제2조(정의) 이 가이드라인에서 사용하는 용어의 정의는 다음과 같으며, 본 조에서 정의되지 않은 용어는 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「개인정보 보호법」 등 관련 법률에서 정의한 바에 따른다.<br>1. “공개된 개인정보”란 정보주체 및 정당한 권한이 있는 자에 의해 제한 없이 일반 공중에게 공개된 부호·문자·음성·음향 및 영상 등의 정보로서 생존하는 개인을 식별할 수 있는 정보 및 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는 정보를 말한다.<br>2. “이용내역정보”란 정보통신서비스와 관련하여 이용자가 해당 서비스를 이용하는 과정에서 자동으로 발생되는 인터넷 접속정보파일, 거래기록 등의 정보로서 생존하는 개인을 식별할 수 있는 정보 및 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는 정보를 말한다.<br>3. “정보 조합·분석·처리시스템”이란 공개된 개인정보 또는 이용내역정보 등을 전자적으로 설정된 체계에 의해 조합, 분석 또는 처리하여 새로운 정보를 생성하는 시스템을 말한다.<br>4. “생성된 개인정보”란 정보 조합·분석·처리시스템 운용을 통해 생성된 정보로서 생존하는 개인을 식별할 수 있는 정보 및 다른 정보와 쉽게 결합하여 개인을 식별할 수 있는 정보를 말한다.<br>5. “비식별화”란 데이터 값 삭제, 가명처리, 총계처리, 범 | 제2조(정의) (초안과 같음)<br>1. ~ 6. (초안과 같음) | 제2조(정의) 이 가이드라인에서 사용하는 용어의 정의는 다음과 같으며, 본 조에서 정의되지 않은 용어는 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「개인정보 보호법」 등 관련 법률에서 정의한 바에 따른다.<br>1. “공개된 정보”란 이용자 및 정당한 권한이 있는 자에 의해 공개 대상이나 목적의 제한 없이 합법적으로 일반 공중에게 공개된 부호·문자·음성·음향 및 영상 등의 정보를 말한다.<br>2. “이용내역정보”란 이용자가 정보통신서비스를 이용하는 과정에서 자동으로 발생하는 서비스 이용기록, 인터넷 접속정보, 거래기록 등의 정보를 말한다.<br>3. “정보 처리시스템”이란 공개된 개인정보 또는 이용내역정보 등을 전자적으로 설정된 체계에 의해 조합·분석 등 처리하여 새로운 정보를 생성하는 시스템을 말한다.<br>4. “비식별화”란 데이터 값 삭제, 가명처리, 총계처리, 범주화, 데이터 마스킹 등을 통해 **개인정보의 일부 또는 전부를 삭제하거나 대체함**으로써 다른 정보와 쉽게 결합하여도 특정 개인을 식별할 수 없도록 하는 조치를 말한다. |

| | | |
|---|---|---|
| 주화, 데이터 마스킹 등을 통해 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 쉽게 결합하여도 특정 개인을 식별할 수 없도록 하는 조치를 말한다.<br>6. “재식별화”란 비식별화된 정보를 조합, 분석 또는 처리하는 과정에서 개인정보가 재생성되는 것을 말한다. | | |
| | | 제3조(개인정보의 보호) ① 정보통신서비스 제공자가 정보처리시스템을 통해 공개된 정보, 이용내역정보를 수집·저장·조합·분석 등 처리하고자 하는 경우, 개인정보의 보호를 위해 다음 각 호의 조치를 취하여야 한다.<br>1. 개인정보가 포함된 공개된 정보 및 이용내역정보는 비식별화 조치를 취한 후 수집·저장·조합·분석 등 처리하여야 한다.<br>2. 비식별화 조치된 공개된 정보 및 이용내역정보를 조합·분석 등 처리하는 과정에서 개인정보가 생성되지 않도록 하여야 한다. 다만, 개인정보가 생성되는 경우에는 지체없이 파기하거나 비식별화 조치를 취하여야 한다.<br>② 비식별화 조치된 공개된 정보 및 이용내역정보를 정보처리시스템에 저장·관리하는 경우 다음 각 호의 보호조치를 취하여야 한다.<br>1. 불법적인 접근을 차단하기 위한 침입차단시스템 등 접근 통제장치의 설치·운영<br>2. 접속기록의 위조·변조 방지를 위한 조치<br>3. 백신 소프트웨어의 설치·운영 등 악성 프로그램에 의한 침해 방지조치<br>4. 기타 안전성 확보를 위 |

| | | |
|---|---|---|
| | | 해 필요한 보호조치 |
| 제3조(공개된 개인정보의 수집) ① 정보통신서비스 제공자가 공개된 개인정보를 수집하고자 하는 경우에는 별도로 정보주체의 동의를 얻지 아니하여도 된다.<br><br>② 정보통신서비스 제공자는 제1항에 따라 공개된 개인정보를 수집하는 경우 공개된 개인정보의 수집 출처, 조합, 분석 또는 처리하는 사실 및 그 목적을 정보주체가 언제든지 쉽게 확인할 수 있도록 전자적 표시 방법 등을 통해 공개하여야 한다.<br><br>③ 정보통신서비스 제공자는 정보주체의 요구가 있으면 즉시 다음 각 호의 모든 사항을 정보주체에게 알려야 한다.<br>1. 개인정보의 수집 출처<br>2. 개인정보의 조합, 분석 또는 처리의 목적<br>3. 해당 개인정보의 처리 정지를 요구할 권리가 있다는 사실 | 제3조(공개된 개인정보의 수집) ① ----------. 다만 정보주체 및 정당한 권한이 있는 자가 공개 대상을 제한하거나 공개 목적을 설정한 경우에는 그러하지 아니하다.<br>② (초안과 같음)<br>③ (초안과 같음)<br>1. ~ 3. (초안과 같음) | 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.<br>② 정보통신서비스 제공자는 제1항에 따라 개인정보가 포함된 공개된 정보를 수집·이용하는 경우 공개된 정보의 수집 출처, 수집·저장·조합·분석 등 처리하는 사실 및 그 목적을 이용자 등이 언제든지 쉽게 확인할 수 있도록 개인정보 취급방침을 통해 공개하여야 한다.<br>③ 정보통신서비스 제공자는 이용자 등의 요구가 있으면 즉시 다음 각 호의 모든 사항을 이용자 등에게 알려야 한다.<br>1. 개인정보의 수집 출처<br>2. 개인정보의 수집·저장·조합·분석 등 처리의 목적<br>3. 해당 개인정보의 처리 정지를 요구할 권리가 있다는 사실 |
| 제4조(이용내역정보의 수집) ① 정보통신서비스 제공자가 정보통신서비스 제공과 관련하여 조합, 분석 또는 처리를 목적으로 이용내역정보를 수집하는 경우 별도로 정보주체의 동의를 얻지 아니하여도 된다. 다만, 이용내역정보를 수집하지 않으면 해당 정보통신서비스의 제공이 곤란한 경우를 제외하고 정보주체가 이용내역정보의 수집에 대한 거부 의사를 표시한 때에는 이용내역정보를 수집할 수 없다. | 제4조(이용내역정보의 수집) ① ----------. | 제5조(이용내역정보의 수집·이용) ① 정보통신서비스 제공자는 다음 각 호의 경우를 제외하고는 이용자의 동의를 받거나 비식별화 조치를 취한 후 이용내역정보를 수집·이용할 수 있다.<br>1. 정보통신서비스의 제공에 관한 계약을 이행하기 위하여 필요한 이용내역정보로서 경제적·기술적인 사유로 통상적인 동의를 받는 것이 뚜렷하게 곤란한 경우<br>2. 정보통신서비스의 제공에 |

| | | |
|---|---|---|
| | | 따른 요금정산을 위하여 필요한 경우<br>3. 다른 법률에 특별한 규정이 있는 경우 |
| ② 정보통신서비스 제공자는 제1항에 따라 이용내역정보를 수집하는 경우 해당 정보가 수집, 조합, 분석 또는 처리되는 사실 및 목적을 정보주체가 언제든지 쉽게 확인할 수 있도록 전자적 표시 방법 등을 통해 공개하여야 한다. | ② (초안과 같음) | ② 정보통신서비스 제공자는 제1항에 따라 이용내역정보를 수집하는 경우 해당 정보가 수집·저장·조합·분석 등 처리되는 사실 및 목적을 이용자가 언제든지 쉽게 확인할 수 있도록 개인정보 취급방침을 통해 공개하여야 한다. |
| ③ 정보통신서비스 제공자는 이용내역정보의 수집, 조합, 분석 또는 처리를 거부할 수 있는 방법 및 절차를 마련하여야 한다. | ③ (초안과 같음) | ③ 정보통신서비스 제공자는 이용내역정보의 수집·저장·조합·분석 등 처리를 거부할 수 있는 방법 및 절차를 마련하여야 한다. |
| ④ 정보통신서비스 제공자는 이용자의 검색프로그램 등에서 이용자 또는 검색프로그램등 공급자가 설정해 놓은 이용내역정보의 수집 거부 선택을 이용자의 동의 없이 변경해서는 아니 된다. | ④ (초안과 같음) | ④ 정보통신서비스 제공자는 이용자의 검색프로그램 등에서 이용자 또는 검색프로그램 등 공급자가 설정해 놓은 이용내역정보의 수집 거부 선택을 이용자의 동의 없이 변경해서는 아니 된다. |
| 제5조(새로운 개인정보의 생성) ① 정보통신서비스 제공자는 정보 조합·분석·처리시스템을 통해 공개된 개인정보 및 이용내역정보 등을 활용하여 새로운 개인정보를 생성하는 경우 별도로 정보주체의 동의를 얻지 아니하여도 된다. 다만, 새로운 개인정보를 생성하지 않으면 해당 정보통신서비스의 제공이 곤란한 경우를 제외하고 정보주체가 새로운 개인정보의 생성에 대한 거부의사를 표시한 때에는 새로운 개인정보를 생성할 수 없다.<br>② 정보통신서비스 제공자는 제1항에 따라 개인정보를 생성하는 경우 개인정보가 생성된다는 사실 및 그 목적을 정보주체가 언제든지 쉽게 확 | 제5조(새로운 개인정보의 생성) ① ----------의 정당한 이익과 상당한 관련이 있고 합리적인 범위를 초과하지 아니한 경우, 정보통신서비스 제공자는 별도로 정보주체의 동의를 얻지 아니하고 정보 조합·분석·처리시스템을 통해 공개된 개인정보 및 이용내역정보 등을 활용하여 새로운 개인정보를 생성할 수 있다. 다만, 정보주체가 새로운 개인정보의 생성에 대한 거부 의사를 표시한 때에는 그러하지 아니하다.<br>② (초안과 같음) | 제6조(새로운 정보의 생성) ① 정보통신서비스 제공자는 비식별화 조치하여 수집한 공개된 정보 및 이용내역정보를 정보 처리시스템을 통해 조합·분석하여 새로운 정보를 생성할 수 있다. 다만, 새롭게 생성된 정보에 개인정보가 포함되어 있을 경우, 즉시 파기하거나 비식별화 조치를 취하여야 한다.<br>② 정보통신서비스 제공자는 제1항에 따라 개인정보가 포함된 정보가 생성될 수 있다는 사실 및 그 처리 방법을 이용자가 언제든지 쉽게 확인할 수 있도록 개인정보 취급방침을 통해 공개하여야 한다. |

| | | |
|---|---|---|
| 인할 수 있도록 전자적 표시 방법 등을 통해 공개하여야 한다. | | |
| 제6조(공개된 개인정보등의 조합·분석·처리) ① 공개된 개인정보, 이용내역정보, 생성된 개인정보(이하 “공개된 개인정보등”이라 한다)는 목적 달성을 위해 불가피한 경우를 제외하고 비식별화 조치를 취한 후 조합, 분석 또는 처리하여야 한다.<br>② 제1항에 따라 비식별화된 정보는 조합, 분석 또는 처리 과정에서 재식별화 되지 않도록 하여야 한다. 다만, 재식별화 되는 경우에는 해당 개인정보에 대하여 비식별화 조치를 취하여야 한다.<br>③ 공개된 개인정보등의 조합, 분석 또는 처리 과정에서 임시적으로 생성된 개인정보는 조합, 분석 또는 처리 목적을 달성한 경우 지체 없이 파기하거나 비식별화 조치를 취하여야 한다. | 제6조(공개된 개인정보등의 조합·분석·처리) ①---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 비식별화 조치를 취한 후 조합, 분석 또는 처리하여야 한다. 다만, 이용자의 동의를 받거나 법령상 허용하는 경우는 그러하지 아니하다.<br>② (초안과 같음)<br>③ (초안과 같음) | |
| 제7조(공개된 개인정보등의 저장·관리) ① 비식별화 조치가 취해진 공개된 개인정보등을 저장·관리하는 경우 다음 각 호의 보호조치를 취하여야 한다.<br>1. 불법적인 접근을 차단하기 위한 침입차단시스템 등 접근 통제장치의 설치·운영<br>2. 접속기록의 위조·변조 방지를 위한 조치<br>3. 백신 소프트웨어의 설치·운영 등 컴퓨터바이러스에 의한 침해 방지조치<br>② 이용내역정보의 조합, 분석 또는 처리를 위탁하는 경우, 해당 정보가 저장·관리되 | 제7조(공개된 개인정보등의 저장·관리) ① (초안과 같음)<br>1. ~ 3. (초안과 같음)<br>② (초안과 같음) | |

| | | |
|---|---|---|
| 고 있는 장소가 아닌 다른 곳으로 이전하여 조합, 분석 또는 처리할 수 없다. | | |
| 제8조(민감정보 생성의 금지) 특정한 개인의 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 정보주체의 사생활을 현저히 침해할 우려가 있는 정보의 생성을 목적으로 공개된 개인정보등을 조합, 분석 또는 처리하여서는 아니 된다. 다만, 정보주체의 사전 동의를 받거나 법률에 따라 허용된 경우에는 그러하지 아니하다. | 제8조(민감정보 생성의 금지) (초안과 같음) | 제7조(민감정보 생성의 금지) 특정한 개인의 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 그 밖에 이용자의 사생활을 현저히 침해할 우려가 있는 정보의 생성을 목적으로 공개된 개인정보 등을 수집·저장·조합·분석 등 처리하여서는 아니 된다. 다만, 이용자의 사전 동의를 받거나 법률에 따라 허용된 경우에는 그러하지 아니하다. |
| 제9조(통신 내용의 조합, 분석 또는 처리 금지) 정보통신서비스 제공자는 전송중인 이메일, 문자메시지 등의 통신 내용에 대하여 양 당사자의 동의를 얻은 경우를 제외하고, 통신 내용의 전부 또는 일부를 조합, 분석 또는 처리하여서는 아니 된다. | 제9조(통신 내용의 조합, 분석 또는 처리 금지) (초안과 같음) | 제8조(통신 내용의 조합, 분석 또는 처리 금지) 정보통신서비스 제공자는 전송중인 이메일, 문자메시지 등의 통신 내용에 대하여 양 당사자의 동의를 얻은 경우를 제외하고, 통신 내용의 전부 또는 일부를 조합, 분석 또는 처리하여서는 아니 된다. |
| 제10조(공개된 개인정보등의 이용) ① 정보통신서비스 제공자는 공개된 개인정보등을 정보주체의 별도 동의 없이 자신의 서비스 제공업무 수행을 위해 내부에서 이용할 수 있다. 다만, 정보주체가 거부 의사를 표시한 때에는 그러하지 아니하다.<br>② 정보통신서비스 제공자는 제1항에 따라 공개된 개인정보등을 이용하는 경우 해당 정보가 이용된다는 사실 및 그 목적을 정보주체가 언제든지 쉽게 확인할 수 있도록 전자적 표시 방법 등을 통해 공개하여야 한다. | 제10조(공개된 개인정보등의 이용) ① (초안과 같음)<br>② (초안과 같음) | 제9조(공개된 정보 및 이용내역정보의 이용) ① 정보통신서비스 제공자는 비식별화 처리된 공개된 정보 및 이용내역정보를 자신의 서비스 제공업무 수행을 위해 내부에서 이용할 수 있다. 다만, 이용자가 거부 의사를 표시한 때에는 그러하지 아니하다.<br>② 정보통신서비스 제공자는 제1항에 따라 공개된 정보 및 이용내역정보를 이용하는 경우 해당 정보가 이용된다는 사실 및 그 목적을 이용자가 언제든지 쉽게 확인할 수 있도록 개인정보 취급방침을 통해 공개하여야 한다. |
| 제11조(공개된 개인정보등의 제3자 제공) ① 정보통신서비스 제공자는 공개된 개인정보 | 제11조(공개된 개인정보등의 제3자 제공) ① (초안과 같음) | 제10조(제3자 제공) 정보통신서비스 제공자는 개인정보가 포함된 공개된 정보, 이용내역 |

| | | |
|---|---|---|
| 등을 제3자에게 제공하려면 다음 각 호의 모든 사항을 정보주체에게 알리고 사전 동의를 받아야 한다. 다만, 공개된 개인정보는 정보주체의 동의 없이 제3자에게 제공할 수 있다.<br>1. 공개된 개인정보등을 제공받는 자<br>2. 공개된 개인정보등을 제공받는 자의 이용 목적<br>3. 제공하는 공개된 개인정보등의 항목<br>4. 공개된 개인정보등을 제공받는 자의 보유 및 이용 기간<br>② 비식별화 조치를 취한 공개된 개인정보등은 제1항 각 호의 사항을 정보주체가 언제든지 쉽게 확인할 수 있도록 전자적 표시 방법 등을 통해 공개하고 제3자에게 제공할 수 있다. | 1. ~ 4. (초안과 같음)<br>② (초안과 같음) | 정보, 생성 정보의 경우, 이용자의 동의를 얻어 제3자에게 제공할 수 있다. 다만, 비식별화 처리된 공개된 정보, 이용내역정보, 생성 정보는 이용자 동의 없이 제3자 제공이 가능하다. |
| 제12조(적용범위) 이 가이드라인에서 규정하지 않은 사항은 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「개인정보 보호법」 등 관련 법률에 따른다. | 제12조(적용범위) (초안과 같음) | 제11조(적용범위) 이 가이드라인에서 규정하지 않은 사항은 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」, 「개인정보 보호법」 등 관련 법률에 따른다. |

□ “비식별화”라는 용어상의 문제를 별론으로 하더라도, 「개인정보 보호법」 및 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」 등 개인정보 보호법제의 (동의)요건을 행정지도로서의 성격을 가지는 가이드라인을 통해 면제할 수 있는지에 대해 진지한 검토가 필요함

○ 이러한 가이드라인을 통한 규제완화는 행정적인 국가권력 작용을 통해 의회가 제정한 법률의 해석적 한계를 넘어서는 것이라는 지적이 가능함

- cf. EU Directive와 EU Article 29 Data Protection Working Party

○ 동의 요건은 정보주체의 개인정보자기결정권이라는 기본권 행사방식의 일환이라는 측면에서, 이를 제한하는 결과에 이르기 위해서는 법률에 명시적으로 정해진 경우에만 가능하다고 할 수 있는데, 이것을 단지 해석의 여지가 있다는 이유만으로 그러한 해석의 여지를 구체화하는 가이드라인을 통해 해결하겠다는 것은 분명 문제가 있음

### 3. 입법시 고려요소

□ 최근 국내에서는 “익명화(Anonymisation)”와 “비식별화(de-identification)”라는 개념을 두고 논쟁이 발생한 바 있었는데, 필자는 편의적 관점에서 비식별화라는 용어를 사용하는 데에는 반대하지만, 아직까지 두 용어의 구별의 실익에 대해서는 의문이 있음

□ 비식별화(de-identification)이라는 단어의 번역에 있어 이를 단순히 “식별” 對 “비식별”의 의미로 파악하게 되면, 이 단어가 가지는 본래의 취지를 오인하게 만드는 것임

○ 영국의 「익명화지침」[6](해설)이나 미국 백악관의 「빅데이터 보고서」[7]에서도 “de-identification”이라는 단어를 사용하고는 있을 뿐만 아니라, 최근 미국 NIST(National Institute of Standards and Technology)의 경우도 마찬가지임[8]

○ 관련 보고서들의 내용을 일관해 보면, “de-identification”이라는 단어는 ‘비식별성’이라는 의미를 지칭하기보다는 비식별“화(-cation)”라는 지점에 방점을 두고 있는 기술적 수단을 표현하는 것이라고 보아야 함

□ 그런데 문제는 비식별화라는 단어가 가이드라인이나 각종 안내서 등에서 활용되면서, 이를 비식별성 또는 익명성과 등치(동의요건 면제)시키고 있는 잘못된 법해석 관행을 유발하고 있다고 할 수 있음

○ 예를 들어, 최근 금융위원회는 지침의 미비로 인하여 비식별화할 때 이에 대한 명확한 지침이 없이 금융회사가 비식별화 정보의 활용에 주저하고 있다고 하면서, 이를 개선하겠다고 공표하고 있는데, 이와 관련한 금융위원회의 설명을 그대로 인용하면 다음과 같음[9]

**[표 3] 금융위원회 「금융권 빅데이터 활성화 방안」 설명내용(인용)**

< 비식별정보 활용 관련 국내외 제도 >

□ (외국) 영미법계, 대륙법계 국가 모두 **비식별정보**를 개인정보로 보지 않고 있어 빅데이터 활용 가능

□ (한국) 일반법인 개인정보보호법에 **비식별화할 경우** 동의 목적 외(빅데이터 활용)로 활용이 가능하도록 예외 조항이 있으나

- 특별법인 신용정보법에 예외 조항이 없어, 개인정보보호법에 따라 비식별정보를 동의 목적 외로 활용 가능한지 불명확
- 신용정보법령은 비식별정보라도 개인신용정보로 정의하고 있어 빅데이터 활용시 개인의 동의가 필요

○ 개인정보의 동의 목적외 활용이 가능하도록 하고 있는 예외 규정이 무엇인지를 금융위원회에 문의한 결과, 「개인정보 보호법」 제18조 제2항 제4호에서 “통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우” 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다

6) UK ICO(Information Commissioner's Office), *Anonymisation: managing data protection risk code of practice*, 2012.11

7) US Executive Office of the President, *BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES*. 2014.5.

8) Simson L. Garfinkel(Information Access Division, Information Technology Laboratory), *De-Identification of Personal Information*, NIST, 2015.10

9) 금융위원회, 「금융권 빅데이터 활성화 방안」, 2015.6.3.

고 규정하고 있는 사실을 제시하였음

○ 그러나 빅데이터 활용 기업에서 학술연구를 위한 목적으로 관련 정보를 활용하는 것은 당연히 있을 수 없으며, 통계작성인 경우가 있을 수 있는데 이 또한 기업이 영리를 목적으로 하는 경우(cf. 서울중앙지방법원 2014.11.4, 2013나49885)가 해당할 수 있는지는 재차 논의가 필요하며, 더 나아가서는 "비식별화 조치"가 "특정 개인을 알아볼 수 없는 형태"로의 변환에 해당하는지에 대한 고민도 필요함

○ 결국 국내의 '비식별화 가이드라인'을 비롯한 각종 안내서들은 입법적으로 해소되어야 할 사항을 가이드라인 등에 정함으로써, 의회가 제정한 법률에 정하지 않은 사항의 의미를 왜곡하고 있는 측면이 강함

□ 동의요건은 개인정보자지결정권 보장을 위한 최소 요건이라고 할 수 있으며, 궁극적으로 빅데이터 환경에서 문제시 되는 것은 막대한 양의 정보를 처리하는 과정에서 실질적으로 정보주체의 동의를 받을 수 없는 경우가 증가하고 있는 상황에서 정보주체의 개인정보자기결정권을 어떠한 방식으로 실현할 것인지에 입법 작업의 초점을 맞추어야 함

○ 동의는 말 그대로 '최소 요건'이기 때문에, 동의가 있었다고 하여 정보주체의 (개인)정보를 사업자 등이 자의적으로 처리할 수 있다는 것을 의미하는 것은 아니라는 점을 명확하게 해야 할 필요가 있음

○ 이와 관련해서는, 최근 EU의 GDPR 논의에 있어 우리나라에서는 익명화 또는 비식별화 정보의 일환으로 가이드라인에 규정하고 있는 가명처리정보(pseudonymous data)도 개인정보의 일종으로 포함시켜 규제하기 위한 시도가 이루어지고 있다는 점도 유의할 필요가 있음

- 즉 이는 비식별화 또는 익명화된 정보, 특히 가명처리정보의 경우에는 재식별화의 위험성이 있기 때문에, 이를 개인정보에 준해서 보호하겠다는 의도로 판단됨

□ 최근 해외 주요 국가들의 개인정보 보호에 관한 미래지향적 논의에서 "동의권(notice & consent) 패러다임"을 넘어서야 한다는 지적은 이러한 요건을 면제하자는 의미가 아니라, 동의권만으로는 정보주체의 권리(개인정보자기결정권)을 실질적으로 보장할 수 없게 되었다는 지적이라고 할 수 없음

○ 실제 업계 관행에 있어서는 빅데이터 분석을 위한 정보수집이 필요한 경우에는 형식적으로라도 동의(향후 이용 가능성을 확보하기 위하여 다소 추상적인 약관 및 개인정보처리방침상 문구를 전제로 함)를 받고 있다는 점을 감안한다면, 가이드라인 등에서 제시하고 있는 바와 같이 광범위한 동의 요건 면제가 실질적으로 필요한 이유가 무엇인지를 명확히 할 필요가 있음

○ 물론 모두에 언급했던 바와 같이 사업자들의 입장에서만 보자면, 다소 광범위하게 해석 가능한 현재의 개인정보 개념정의(개인 식별 가능성 + 용이한 결합을 통한 개인 식별 가능성)를 수정하자는 논의까지 제시되고 있지만, 해외 입법례도 '개념정의'에 한정해 보자면 우리 입법과 유사한 내용을 가지고 있다고 볼 수 있어 이러한 주장은 다소 문제가 있음

□ 결국 향후 개인정보 보호법제 개선방향의 핵심은 정보주체의 동의가 없더라도, 정보주체의 개인정보자기결정권을 어떻게 보호할 수 있는지 여부라고 할 수 있음

○ 최근 국제적인 입법동향에 있어 프로파일링(profiling) 금지 및 설계시 프라이버시 고

려(privacy by design) 규정 등의 도입이 논해지고 있는 이유가 바로 여기에 있음

- "보호영역을 넓게, 사안에 따른 행정적·사법적 제재조치는 구체적으로"
- 개인정보자기결정권의 의미와 결부시켜서 보자면, 이러한 기본권 보장의 실현방식이 반드시 수집 및 이용시 동의요건 설정에만 한정되는 것이 아니라는 점을 명확히 할 필요가 있음(ex. 처리정지(거부) 및 사후 동의철회 등)

○ 그러나 우리나라의 경우 개인 식별 정보(주민등록번호, 아이핀, 휴대전화)의 활용이 제도적으로 광범위하게(공공+민간) 허용되거나 일부에 있어서는 강제되고 있는 상황이어서, 이에 대한 개선 없이는 해외 주요 국가들의 경우와 같이 미래지향적 입법을 고민하기에는 한계가 있음(달리 말하여, 우리나라에서는 재식별화의 위험이 다른 나라에 비하여 매우 높다고 할 수 있음)

## 4. 개인정보 보호법제 패러다임 변화

□ "비식별화" 개념을 둘러싼 논쟁은 사실상 개인정보의 법적 보호체계 또는 패러다임의 개편에 관한 논쟁을 의미한다고 볼 수 있음

○ 물론 단순히 개인정보 개념의 구체화 및 정보주체 동의 면제 등을 요구하는 수정담론의 견해가 실질적인 패러다임 변화라고는 볼 수 없을 것임

○ 여기에서 논하고자 하는 패러다임은 권리중심 개인정보 보호체계에서 위험기반 개인정보 또는 프라이버시 보호체계로의 개편임(risk-based approach)[10)]

**[표 4] 개인정보 보호 패러다임**

| 구분 | 프라이버시 | 개인정보자기결정권 | 新프라이버시 |
|---|---|---|---|
| **보호대상** | 식별성비전제<br>(사생활 영역) | 식별성 전제<br>(개인정보) | 식별성+ 비식별성<br>(포괄성) |
| **권리주장** | 소극적 | 적극적 | 소극적 + 적극적 |
| **규제 및 집행** | 맥락적 형량 | 개인정보 (확정)개념 기반<br>(해석적 형량 불가피) | 실질적 위험 기반<br>(Risk Management) |

□ 결론적으로 산업 활성화의 견지에서 일방적인 개인정보 보호요건(식별성, 동의요건 등) 완화에만 치중할 것이 아니라, 실질적인 이용자 프라이버시 보호방안이 무엇인지에 대한 진지한 고민이 필요할 것으로 보임

10) EU GDPR 합의안의 경우, 전형적으로 이러한 패러다임 전환의 과도기에 위치하고 있는 것으로 판단되는데, 예를 들어 가명처리정보를 개인정보의 범주에 포함시키면서, 프로파일링을 금지시킨다는 형량의 결과는 새로운 프라이버시 관념에 입각한 것으로 이해됨. 또한 대부분 국가들의 개인정보 보호법제 등이 참조하고 있는 기준이라고 할 수 있는 1980년 「OECD 프라이버시보호가이드라인」은 2013년에 개정이 이루어졌다. 크게 원칙적 사항들의 수정이 이루어지지는 않았는데, 다만 기존 가이드라인에서 두 가지의 측면이 고려되어 증보되었다. 첫째는 위험관리(risk management)에 기반한 프라이버시 보호의 실천, 둘째는 국제적 차원에서의 규범간 상호운용성(interoperability) 증진의 필요성이 그것이다.

# 참고자료

빅데이터 관련 이른바 '비식별화' 입법에 대한 의견서

# 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>

박경신 토론문 - <오픈넷포럼> 비식별화된 개인정보 문제 2016.3.21.

개인정보는 식별가능성이 요건이다. 비식별화된 정보는 당연히 개인정보가 아니며 개인정보의 적용을 받지 아니 한다.
하지만 식별가능성은 누구의 기준으로 보는가에 따라 달라진다.
예를 들어, 이통사가 가진 각 고객의 통화기록은 어느 통화기록이 누구의 것인지 식별이 가능하므로 당연히 이통사에 대해서는 개인정보이다.
하지만 고객신원정보를 가지고 있는 제3자들은 이용자들의 통화기록이라고 할지라도 전화번호, 통화시간, 상대방 번호 만으로는 어느 이용자의 것인지 알 수 없기 때문에 개인정보가 아니다.
그렇다면, 이통사는 고객의 통화기록에서 신원정보를 제거하여 즉 비식별화하여 제3자에게 제공한다면, 이통사는 개인정보를 준 것이지만 제3자는 개인정보가 아닌 것을 받은 셈이 된다.
이와 같은 정보제공은 개인정보보호법이 적용되어야 하는가?

**문제는 이통시가 비식별화한 카피를 만드는 것이지 원본데이터베이스는 식별화된 상태로 남아 있게 된다는 점이다.** 즉 비식별화를 한 카피를 만들었다고 할지라도 원본 통화기록 데이터베이스에 대한 검색기능을 이용해 '언제 누구와 몇시에 통화한 사람'을 검색해보면 통화자의 신원을 확인할 수 있다. 그렇다면 이를 넘겨준 후에 제3자가 그것이 개인정보가 아니므로 자유롭게 가공하여 이통사가 원래 생각하지 못했던 목적의 다양한 연구나 마케팅 용도로 사용하였다고 하자. 그렇다면 나중에 제3자의 연구결과나 마케팅 자료를 보면 이통사는 원본데이터베이스를 가지고 있는 한 고객을 식별해낼 수 있게 된다.

즉 이통사가 원래 고객들과 계약할 때 상정했던 목적 이외의 용도로 정보가 이용된 결과물을 접할 수 있게 되는 것이고, 이는 개인정보보호법의 취지에 반하는 것이 된다.

이와 같은 경우 어떻게 대비해야 할까?
2015년 12월에 유럽에서 합의된 GDPR은 이에 대해 다음과 같은 해결책을 제공한다.
위와 같이 재식별화가 가능한 정보를 개인정보로 정의하되 몇가지 개인정보보호의무에 대해 예외를 둔다.

제6조 제3a항 - "암호화 및 가명화(pseudonymization)" 등을 고려하여 수집목적과 다른 목적으로 이용가능.
단, 제60a조 - 가명화할 때는 실명화되지 않기 위한 기술적 조치 요구, 예) 암호화
Article 29 Working Party - "가명화는 익명화는 아니다. 그러나, 재실명화가 합리적으로 개연성이 있을 때("reasonably likely") 익명화가 아니라는 뜻이다."

제83조 - 과학, 역사, 통계 목표의 이용을 위해서는 가명화를 할 것.
제23조 - 디자인에 의한 개인정보(data protection by design)의 핵심이 가명화
제30조, 제38조 - 되도록 가명화 상태로 유지할 것 요구

자, 만약 원본까지 모두 비식별화한 경우는 어떠할까?

토론문 : 익명화, 비식별화, 개인정보에 관하여

전응준(유미 법무법인 변호사)

## 1. 용어의 정의

### 1.1 익명화

○ Anonymisation
(6) “Rendering anonymous” shall mean the alteration of personal data so that information concerning personal or material circumstances cannot be attributed to an identified or identifiable natural person or that such attribution would require a disproportionate amount of time, expense and effort(BDSG 3 (6))

○ Data are anonymised if all identifying elements have been eliminated from a set of personal data. No element may be left in the information which could, by exercising reasonable effort, serve to re-identify the person(s) concerned. Where data have been successfully anonymised, they are no longer personal data.(데이터보호법 핸드북 44면)

○ EU 95년 Directive는 본문 조항에서 익명화에 관한 정의를 두고 있지는 않으나 서문에서 익명화 정보가 데이터 규제의 대상이 되지 않음을 기술함

### 1.2 가명화

○ GDPR에서 가명화의 정의와 일정 영역의 법적 규율을 시도하고 있음

○ 'pseudonymisation' means the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person(GDPR 4 (3a))

○ “Aliasing(pseudonymization)” shall mean replacing a person's name and other identifying characteristics with a label, in order to preclude identification of the data subject or to render such identification substantially difficult. (BDSG 3 (6a) )

○ Personal information contains identifiers, such as a name, date of birth, sex and address. When personal information is pseudonymised, the identifiers are replaced by one pseudonym. Pseudonymisation is achieved, for instance, by encryption of the identifiers in personal data(데이터보호 핸드북 45면)

Example: The sentence “Charles Spencer, born 3 April 1967, is the father of a family of four children, two boys and two girls” can, for instance, be pseudonymised as follows:
“C.S. 1967 is the father of a family of four children, two boys and two girls”; or
“324 is the father of a family of four children, two boys and two girls”; or
“YESz320l is the father of a family of four children, two boys and two girls”.

○ 방통위 가이드라인이 EU에서 Pseudonymous data를 익명화로 보고 마치 개인정보보호 관계 규정을 전부 면제하는 것으로 이해하는 것은 오류에 가까움.

### 1.3 비식별화

○ 미국 Department of Education
- Anonymization [of data] refers to the process of data de-identification which produces de-identified data, where individual records cannot be linked back to an original student record system or to other individual records from the same source, because they do not include a record code needed to link the records.
- De-identification [of data] refers to the process of removing or obscuring any personally identifiable information from student records in a way that minimizes the risk of unintended disclosure of the identity of individuals and information about them.
- While it may not be possible to remove the disclosure risk completely, de-identification is considered successful when there is no reasonable basis to believe that the remaining information in the records can be used to identify an individual
출처 : http://ptac.ed.gov/sites/default/files/data_deidentification_terms.pdf

○ 주로 미국에서 비식별화(de-identification)이라는 용어를 사용하고 익명화(anonymization)와 일부 구분되는 것으로도 보이나, 의식적으로 명확한 용어정의를 하고 있다고 보이지는 아니함.

○ 개인정보보호법 제18조 제2항 제4호

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. 다만, 제5호부터 제9호까지의 경우는 공공기관의 경우로 한정한다.
4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우

- ‘특정 개인을 알아볼 수 없는 형태’이면 개인정보성을 상실한다고 볼 수 있으므로 위 제4호를 주의적 규정으로 이해할 수도 있음. 그러나 제18조의 전체적인 지위를 볼 때 제4호는 제3자 제공의 예외를 둔 열거적 규정으로 보는 것이 자연스럽기 때문에 ‘특정 개인을 알아볼 수 없는 형태’는 가명처리, 총계처리, 데이터 마스킹, 범주화, 데이터 일부 값 삭제등의 비식별화 처리라고 이해하는 것이 타당(사견).

## 2. 비식별화 정보의 개인정보 해당성

### 2.1 EU 데이터 보호법 핸드북 2014

○ Data are personal data if they relate to an identified or at least identifiable person, the data subject.

○ A person is identifiable if additional information can be obtained without unreasonable effort, allowing the identification of the data subject.

○ Data are anonymised if they no longer contain any identifiers; they are pseudonymised if the identifiers are encrypted.

○ In contrast to anonymised data, pseudonymised data are personal data.

### 2.2 GDPR

○ EU 95년 Directive는 pseudonymization에 대한 규율이 없었음.
- 독일 BDSG등 회원국 개별법에서는 일부 규정이 있음

○ GDPR(안)은 pseudonymized data에 대한 일부 규정을 마련하고 있음
- 규정안이 계속 수정되면서 일부 문언 변경이 있음

○ 2014년 GDPR(안)
- Article 10
1. If the data processed by a controller do not permit the controller or processor to directly or indirectly identify a natural person, or consist only of pseudonymous data, the controller shall not process or acquire additional information in order to identify the data subject for the sole purpose of complying with any provision of this Regulation.
2. Where the data controller is unable to comply with a provision of this Regulation because of paragraph 1, the controller shall not be obliged to comply with that particular provision of this Regulation. Where as a consequence the data controller is unable to comply with a request of the data subject, it shall inform the data subject accordingly.

○ 2015. 12. GDPR안
- Article 10 Processing not requiring identification
1. If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.
2. Where, in such cases the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, Articles 15 to 18 do not apply except where the data subject, for the purpose of exercising his or her rights under these articles, provides additional information enabling his or her identification.

- 제23조 : pseudonymisation을 appropriate technical and organisational measure로 인식함(data protection by design and by default)
- 제30조 : 데이터 처리의 보호조치로서 pseudonymasation 예시
- 제32조 : 암호화 등 데이터를 인식불가능하게 하였다면 데이터 유출(data breach)시 정보주체에게 통지 면제
- 제38조 : code of conduct로서 pseudonymisation 강조

### 2.3 HIPAA(Health Insurance Portability and Accountability Act)

○ Since 1996 the US Federal Health Insurance Portability and Accountability Act of 1996 (HIPAA) differentiate between:
- Protected Health Information (Patient name + medical records = sensitive personal data)
- Limited Data Set (a form of pseudonymized data still covered by HIPAA)
- De-identified Data (a form of anonymized data not any longer covered by HIPAA)
- 다만 비식별화 정보, LDS의 기준에 관하여 명확한 통계적, 실증적 근거가 있다고 보이지는 아니함

○ 의료정보관기관(covered entity)가 취할 수 있는 비식별화 방법
- 아래의 2가지 경우에 해당하면 HIPPA 프라이버시 규율에서 벗어남
- 관련 전문가의 개별적인 판단(very small risk) : 통계학적, 과학적 원칙을 적용
- 식별자들 18개를 데이터에서 제거하는 방법(safe harbor)

○ Limited Data Set(LDS)
- HIPAA 프라이버시 규칙에서 부분적인 규율면제
- 제한적인 요건 하에서 정보주체의 동의, 허가 없이 이용하고 제공할 수 있음
- 18개 식별자 중 직접적 식별자에 해당하는 16가지 식별자를 제거하거나(나머지 간접적 식

별자는 생년월일, 치료나 처방일자, 일부 지리적 정보라고 함)
 - 데이터를 제공받는 자와 데이터이용계약(data use agreement)를 체결하면 LDS 인정

**3. 우리 법체계상의 익명 정보, 비식별화 정보**

○ 개인정보의 범위를 해석할 때 결합가능한 다른 정보의 보유가능성을 합리적인 범위내에서 제한 해석할 필요가 있다고 생각됨
 - EU 데이터 보호규정도 같은 입장에 서 있다고 보임
 - 다른 정보의 획득가능성을 합리적인 범위내에서 제한하지 않고 데이터의 속성에 따른 결합가능성의 용이성만을 따지는 것은 타당하지 아니함.

○ 원본 데이터로 회복불가능한 익명화 정보는 개인정보법의 대상이 아님. 문제는 이론상으로 보면 원본 데이터로 연결이 불가능한 익명화 정보는 없다는 점에 있음.
 - 독일 BDSG, EU 핸드북과 같이, 합리적 수단, 비용, 시간 범위내에서 원본 데이터로 복원할 수 없는 익명화 정보는 더 이상 개인정보로 보지 않는 해석론이 우리 법 해석에서도 가능하다고 생각됨(사견)

○ 일본 법과 같은 '익명가공정보', '특정성 저감 데이터' 등의 개념 정의 도입이 바람직하겠지만, 단기간내에 불가능하다면 법 해석론으로 개인정보보호법의 규율을 받지 않는 익명화 정보를 정의할 수 있지 않을까 함.
 - 일본 법은 익명가공정보에 대하여 제3자 제공시 정보주체의 동의를 받지 않아도 되도록 설계하였는데, 제3자 제공은 전면적인 데이터 유통을 가능하게 하므로 초기 단계에서는 수집 및 이용과 제3자 제공의 경우를 분리하여 규율하는 것이 정책적으로 바람직할 수 있다는 생각을 해 봄.

○ 미국 FTC 동의의결 사례를 참조
 - 형사처벌을 규정하고 있는 우리 법률 체계상 사법조치보다는 행정처분이 유연한 대응을 가져올 수 있음.

○ 통계법에 의하여 수집된 정보에 대하여는 개인정보보호법이 적용되지 않으므로(개인정보보호법 제58조) 이를 기초로 공공영역에서 빅데이터 분석은 일정 부분은 가능하지 않을까 함.

**4. 위험성의 상존**

○ 우리나라의 약학정보원 사례
 - 비식별화 논리가 무분별하게 이루어지는 제3자 제공의 근거가 될 수 있음
 - 위 사건은 개인정보처리자가 취한 비식별화(암호화) 정도가 실제 현실에서 쉽게 복호화될 수 있었던 사안으로 추측됨

○ 이론적 논의 외에 비식별성 방법론(알고리듬)과 재식별성 정도에 관한 통계적, 실증적 연구가 필요함

# 개인정보보호법의 기술적 문제점에 대하여

이영환(건국대 정보통신기술대학원 교수)

개인정보보호법등 27개 법률이 촘촘히 금지하고 있는 개인정보의 유통금지의 가장 큰 희생자는 약탈적 금리에 시달리는 서민들과 그에 의해 거리로 내몰리고 있는 청년실업자들이다. 본 토론에서는 특히 개인정보가 차단될수록 안전하다는 식의 오도된 믿음에 대해서 문제를 제기한다.

문제는 현재 개인정보보호법 등 27개 법률에서 촘촘하고 꼼꼼하게 개인정보를 포함하는 빅데이터의 유통을 금지하고 있다. 개인정보보호의 중요성은 아무리 강조해도 지나치지 않다. 하지만 과도하게 금지되어 인터넷전문은행 등의 사업자를 빅데이터 유통이 불가능하다.

이런 점을 인식하고 있는 정부가 내놓는 대책은 비식별화를 전제로 한 빅데이터유통 허용이다. 비식별화를 할 경우 데이터를 유통할 수 있도록 하겠다는 것이다. 문제는 비식별화가 기술적으로 “어렵다”는데 있다. 더군다나 기하급수적으로 증가하는 컴퓨팅파워는 날로 재식별화 기술의 발전을 보이고 있다. 이런 상황에서 빅데이터 산업의 발목을 붙잡고 개인정보의 비식별화를 하자는 것은 기술적으로 “어렵다”는 점에서 모순성을 논의하고 이 법이 현실적인 문제점을 설명한다.

본 토론은 개인정보보호와 빅데이터 활성화라는 모순적인 목표를 달성하는 방법에 대해서 생각해 본다. 특히 개인정보보호를 위해 기 제안된 바 있는 k-익명성, l-다양성, t-근접성에 대해서 검토하고 한계를 지적한다.

이러한 한계를 극복하기 위한 대한 해답으로 개인정보보호법의 27개 법안의 취지를 살리고 빅데이터를 활성화 시키기 위한 대안으로 데이터의 유통에 관한 법률의 제정과 옵트아웃의 도입이 절실히 필요하다.

# 개인정보보호에서 비식별화의 기술적 문제점과 트렌드

이영환교수
건국대학교
경영대학 기술경영학과
정보통신기술대학원 금융IT 학과

# Agenda

I. 비식별화와 법규제

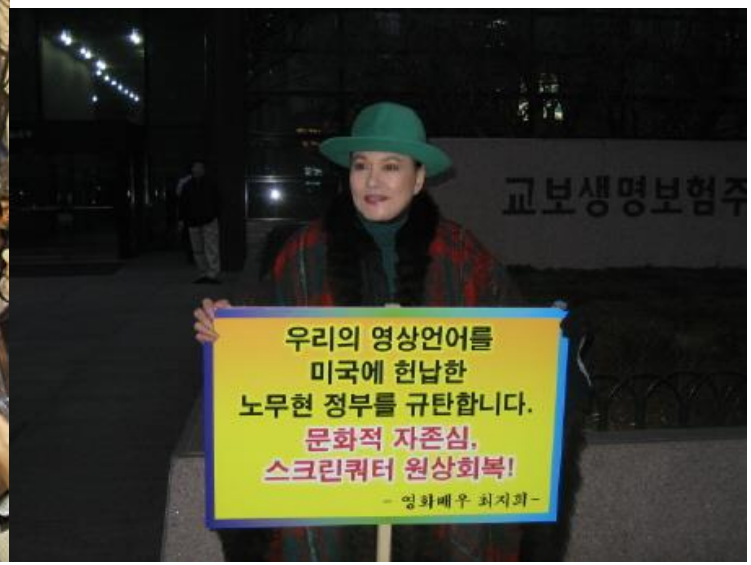
II. 진화의 관점에서 본 데이터 산업

III. 결론

부록. 식별정보와 인증정보

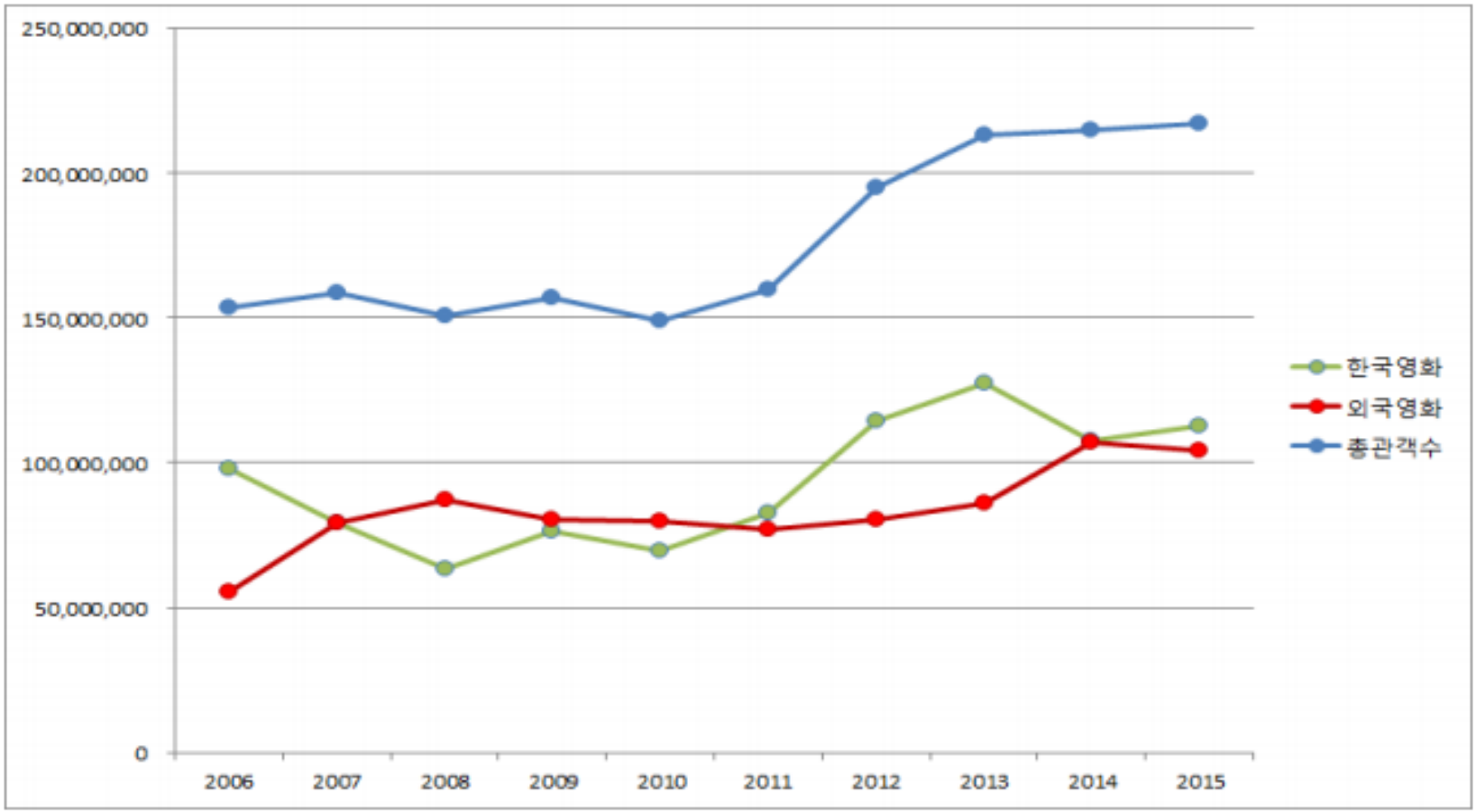
# 영화 쿼터제





# 영화 쿼터제

<그림 1> 2006년-2015년 한국영화vs외국영화 극장 관객 수 추이




출처: 영화진흥위원회 산업정책연구팀. 2015 한국 영화산업 결산.

# I. 비식별화와 법규제

## ● 개인정보보호: 식별정보와 인증정보 모두 보호

**개인정보보호법이 보호하고 있는 것?**

| 구분 | | 근거 | 개인 식별 정보 항목 |
|---|---|---|---|
| 일반 | | · 개인정보보호법 제18조, 제23조, 제24조제1항, 제24조제3항, 제24조제2 | · 주체자의 사생활을 침해할 수 있는 식별정보 (ex. 의료정보, 정신적 성향 등)<br>· 주체자의 신분 확인을 위한 일반 식별정보 (ex. 이름, 주민등록번호, 주소 등) |
| 공공부문 | | · 전자정부법 제42조 | · 정당한 사용자임을 인증하는 식별정보 (ex. 인증서 일련번호, 유효기간 등) |
| | | · 주민등록법 10조 | · 신분 확인정보와 가족구성원 정보를 통해 확인될 수 있는 식별정보 (ex. 성명, 성별, 세대주와의 관계 등) |
| | | · 공공기관의 정보공개에 관한 법률 제18조<br>· 공공기록물 관리에 관한 법률 제37조 | · 주체자의 신분 확인을 위한 일반 식별정보 (ex. 이름, 주민등록번호, 연락처 등) |
| | | · 민원사무처리에 관한 법률 제26조<br>· 국가정보화 기본법 제39조 | · 본인·대리인 확인을 위한 식별정보 (ex. 주민등록번호, 대리인 신분증 등) |
| 민간 | 정보통신 | · 정보통신망 이용촉진 및 정보보호 등에 관한 법률<br>· 전자서명법 제24조 | · 회원제 관리를 위한 사용자 식별 정보 (ex. 이름, ID, PWD 등)<br>· 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<br>· 통신비밀보호법 | · 업무 수행 및 처리를 위한 통신상의 식별정보 (ex. 접속 IP정보, GPS 정보 등) |
| | | · 청소년보호법 제29조, | · 제한된 연령 확인에 대한 식별정보 |

# III. 비식별화와 법규제

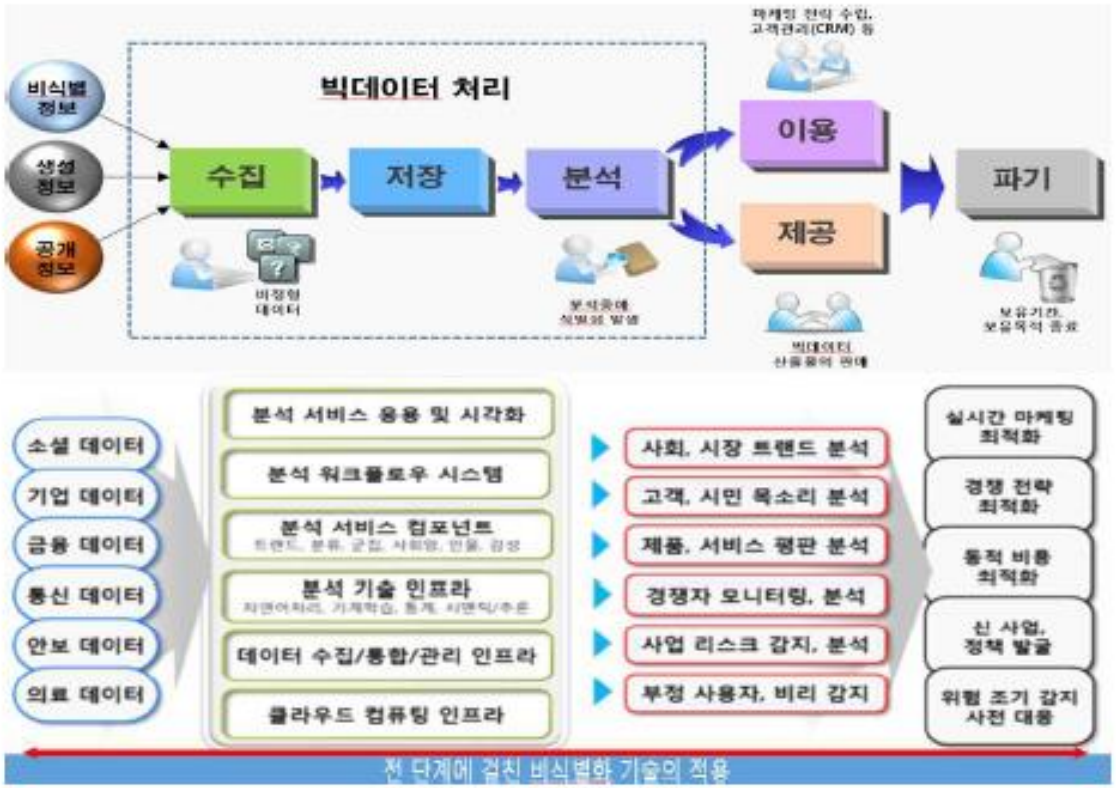
| 구분 | | 근거 | 개인 식별 정보 항목 |
|---|---|---|---|
| | | 제16조 | (ex. 법정 생년월일, 법정 대리인 정보 등) |
| | 상거래 | · 전자문서 및 전자거래기본법 제12조<br>· 정보통신망 이용촉진 및 정보보호 등에 관한 법률 제23조, 제24조<br>· 전자상거래 등에서의 소비자보호에 관한 법률 제12조 | · 전자문서 서비스를 위한 식별정보 (ex, 공인전자주소, 송신자, 수신자 등)<br>· 통신의 안전한 조치를 위해 확인할 수 있는 식별정보 (ex. 비밀번호, 계좌번호, 주민등록번호 등)<br>· 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) |
| | 금융·신용 | · 신용정보의 이용 및 보호에 관한 법률 제33조<br>· 금융실명거래 및 비밀 보장에 관한 법률 제4조 | · 신용정보 및 거래능력을 판단할 수 있는 식별정보 (ex. 재산, 소득, 대출 보증 등)<br>· 금융기관의 거래내역을 판단할 수 있는 정보 (ex. 주민등록번호, 계좌번호, 거래실적 자료 등)<br>· 이용자 및 거래내용의 정확성을 확인하기 위한 식별정보 (ex.전자금융업자에 등록된 이용자번호, 이용자의 생체정보, 등) |
| | | · 전자금융거래법 제26조<br>· 전자금융 감독규정 제5조의 3 | |
| | | · 특정 금융거래 정보의 보고 및 이용 등에 관한 법률 제5조의 3 | · 자금이체를 수행을 위한 식별정보 (ex. 송금인 성명, 계좌번호, 수취인의 정보) |
| | 보건·의료 | · 의료법 제21조<br>· 응급의료에 관한 법률 제22조의 2조<br>· 산업안전보건법 | · 정확한 환자의 진료를 위해 확인가능 한 식별정보 (ex. 주민등록번호, 의료기록, 가족력 등)<br>· 신체의 질병정보를 통해 인지될 수 있는 식별정보 (ex. 감염병명, 혈액정보, 조직정보 등) |
| | | · 후천성면역결핍증예방법<br>· 감염법의 예방 및 관리에 관한 법률 제74조 | |
| | | · 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) |

## ● 무차별적 개인 정보 보호

**비식별화 기술 활용 안내서 (미래부, 정보화진흥원 공동 발간)**

o 빅데이터는 '수집 → 저장 → 분석 → 이용·제공 → 파기' 단계를 거쳐 활용

< 빅데이터 활용 단계 정의 >



## ● 무차별적 개인정보 보호

**비식별화 기술 활용 안내서 (미래부, 정보화진흥원 공동 발간)**

1. 빅데이터 분석 및 활용은 법률상 허용되는 목적 및 범위 내에서만 가능
2. 개인정보가 포함된 정보는 비식별화 조치 필요
3. 생성된 개인정보는 목적 달성 후 파기 또는 비식별화 해야 함
4. 전송중인 이메일 문자메시지 등 통신 내용은 조합, 분석 또는 처리 불가
5. 공개 개인정보도 수집-이용 및 제공을 위하여 정보주체의 동의 필요

자급자족 형태 이외는 데이터 산업이 허용되지 않음!

## ● 식별정보 vs. 인증정보

**비식별화 기술 활용 안내서 (미래부, 정보화진흥원 공동 발간)**

1. 빅데이터 분석 및 활용은 법률상 허용되는 목적 및 범위 내에서만 가능

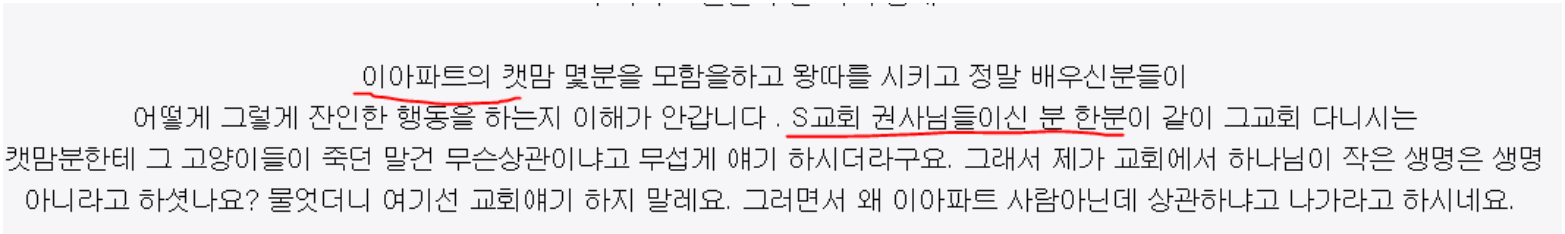
2. 개인정보가 포함된 정보는 비식별화 조치 필요

   - **비식별화 조치된 정보가 조합, 분석 처리 과정에서 재식별화되면 안됨.**

3. 생성된 개인정보는 목적 달성 후 파기 또는 비식별화 해야 함

4. 전송중인 이메일 문자메시지 등 통신 내용은 조합, 분석 또는 처리 불가

5. 공개 개인정보도 수집-이용 및 제공을 위하여 정보주체의 동의 필요


이아파트의 캣맘 몇분을 모함을하고 왕따를 시키고 정말 배우신분들이
어떻게 그렇게 잔인한 행동을 하는지 이해가 안갑니다 . S교회 권사님들이신 분 한분이 같이 그교회 다니시는
캣맘분한테 그 고양이들이 죽던 말건 무슨상관이냐고 무섭게 얘기 하시더라구요. 그래서 제가 교회에서 하나님이 작은 생명은 생명
아니라고 하셧나요? 물엇더니 여기선 교회얘기 하지 말래요. 그러면서 왜 이아파트 사람아닌데 상관하냐고 나가라고 하시네요.

어느 정도로 비식별화해야 재식별화가 안되나요?

# ● 제안된 비식별화 기술

## k-익명성

### 나. 정의

⊙ 주어진 데이터 집합에서 준식별자 속성값들이 동일한 레코드가 적어도 k개 존재 해야 함

### 다. 의미

⊙ 데이터 집합의 일부를 수정하여, 모든 레코드가 자기 자신과 동일한(구별되지 않는) k-1개 이상의 레코드를 가짐

⊙ 예를 들어, 〈표 1〉의 의료 데이터가 익명화 된 〈표 3〉에서 1~4, 5~8, 9~12 레코드는 서로 구별되지 않음[17]

⊙ 예를 들어, 〈표 1〉의 의료 데이터가 익명화 된 〈표 3〉에서 1~4, 5~8, 9~12 레코드는 서로 구별되지 않음[17]

| | 준식별자 | | | 민감한 정보 |
|---|---|---|---|---|
| | 지역 코드 | 연령 | 성별 | 질병 |
| 1 | 130** | < 30 | * | 전립선염 |
| 2 | 130** | < 30 | * | 전립선염 |
| 3 | 130** | < 30 | * | 고혈압 |
| 4 | 130** | < 30 | * | 고혈압 |
| 5 | 1485* | > 40 | * | 위암 |
| 6 | 1485* | > 40 | * | 전립선염 |
| 7 | 1485* | > 40 | * | 고혈압 |
| 8 | 1485* | > 40 | * | 고혈압 |
| 9 | 130** | 3* | * | 위암 |
| 10 | 130** | 3* | * | 위암 |
| 11 | 130** | 3* | * | 위암 |
| 12 | 130** | 3* | * | 위암 |

⊙ 따라서, 익명화된 데이터 집합에서는 공격자가 정확히 어떤 레코드가 공격 대상인지 알아낼 수 없도록 하여 "프라이버시 보호"

예) 김민준 → 레코드 1~4 → 전립선염 또는 고혈압

출처: NIA. "개인정보비식별화 안내서"

## 〈k-익명성의 취약점〉

⊙ 데이터가 k-익명화 되었더라도 민감한 정보가 충분히 다양하지 않으면 프라이버시 문제 발생 가능

⊙ 취약점 1. 동질성 공격 (Homogeneity attack)

- 데이터 집합에서 동일한 민감한 정보를 이용하여 공격 대상의 민감한 정보를 알아내는 공격
- 〈표 3〉에서, 레코드 9~12의 민감한 정보는 모두 '위암'이므로, k-익명성 모델이 적용되었음에도 불구하고 민감한 정보가 직접적으로 노출됨

⊙ 취약점 2. 배경지식에 의한 공격 (Background knowledge attack)

- 주어진 데이터 이외의 공격자의 배경 지식을 통해 공격 대상의 민감한 정보를 알아내는 공격
- 〈표 2〉와 〈표 3〉에서, 공격자가 '이지민'의 질병을 알아내려고 할 때, 준식별자 조합(13068, 29, 여)에 따라 '이지민'은 1~4 레코드 중 하나이며, 질병은 전립선염 또는 고혈압임을 알 수 있음
- 이 때, '여자는 전립선염에 걸릴 수 없다'라는 배경 지식에 의해 공격 대상 '이지민'의 질병은 고혈압으로 쉽게 추정 가능함

⊙ k-익명성의 취약점의 원인

- 다양성의 부족 (lack of diversity)
  - 익명화할 때 민감한 정보의 다양성을 고려하지 않음
  - 동일한 민감한 정보를 가진 (다양하지 않은) 레코드가 익명화되어 하나의 '동질 집합'으로 구성될 경우, 동질성 공격에 무방비
- 강한 배경지식 (strong background knowledge)
  - k-익명성은 '여자는 전립선염에 걸리지 않는다', 또는 '남자는 자궁암에 걸리지 않는다'와 같은 공격자의 배경지식을 고려하지 않아 이를 이용한 공격에 취약함

## ℓ-Diversity

정의

– 주어진 데이터 집합에서 함께 익명화되는 레코드들은 (동질 집합에서) 적어도 1개의 서로 다른 민감한 정보를 가져야 함

다. 의미

⦿ 익명화 과정에서, 충분히 다양한(1개 이상) 서로 다른 민감한 정보를 갖도록 동질 집합을 구성

⦿ 민감한 정보가 충분한 다양성을 가지므로, 다양성의 부족으로 인한 공격에 방어가 가능하고, 배경지식으로 인한 공격에도 일정 수준의 방어력을 가짐

⦿ 예를 들어, 〈표 4〉에서 모든 동질 집합은 3-다양성을 통해 익명화 되어 3개 이상의 서로 다른 민감한 정보를 가짐

– 〈표 3〉과 같이 동일한 질병으로만 구성된 동질 집합이 존재하지 않음

– 공격자가 질병에 대한 배경지식(예: 여자는 전립선염에 걸리지 않음)이 있더라도, 어느 정도의 방어력을 가지게 됨 (예: 여성 이지민이 속한 동질 집합 2, 3, 11, 12에서 전립선염을 제외하더라도 고혈압, 위암 중 어느 질병이 이지민의 것인지 여전히 알 수 없음)

## ℓ-Diversity

| | 준식별자 | | | 민감한 정보 |
|---|---|---|---|---|
| | 지역 코드 | 연령 | 성별 | 질병 |
| 1 | 1305* | ≤ 40 | * | 전립선염 |
| 4 | 1305* | ≤ 40 | * | 고혈압 |
| 9 | 1305* | ≤ 40 | * | 위암 |
| 10 | 1305* | ≤ 40 | * | 위암 |
| 5 | 1485* | > 40 | * | 위암 |
| 6 | 1485* | > 40 | * | 전립선염 |
| 7 | 1485* | > 40 | * | 고혈압 |
| 8 | 1485* | > 40 | * | 고혈압 |
| 2 | 1306* | ≤ 40 | * | 전립선염 |
| 3 | 1306* | ≤ 40 | * | 고혈압 |
| 11 | 1306* | ≤ 40 | * | 위암 |
| 12 | 1306* | ≤ 40 | * | 위암 |

표 4 l=3

## 마. 정보 유용성 (data utility)

◉ 정보 손실 (information loss)

– 데이터를 익명화하게 되면, 프라이버시를 보호하는 대가로 일정량의 정보가 필연적으로 손실됨

– 〈표 3〉, 〈표 4〉의 '*' 나 '〉' 등과 같이 일부 정보를 지우거나, 원본 값을 구간 값 또는 더 상위 개념의 값으로 일반화(generalization)하는 과정에서 원본 데이터의 정보가 일부 손실됨

– 예: 연령 '23'을 구간 값 (20 ~ 25)으로 익명화, 성별 '여성'을 '*' (남성/여성을 구분 없이 모두 지칭함)로 익명화

◉ 프라이버시 보호–정보 손실 간의 관계

– $k$-익명성과 $l$-다양성 모델에서 $k$, $l$값은 곧 프라이버시 보호 수준을 의미

– $k$, $l$값이 증가할수록 프라이버시 보호 수준은 증가하지만, 이에 따라 많은 정보가 손실되므로 정보 유용성은 감소하는 경향을 보임

## *t*-Closeness

### 나. 정의

◉ 동질 집합에서 민감한 정보의 분포와, 전체 데이터 집합에서 민감한 정보의 분포가 이하의 차이를 보여야 하는 것

### 다. 의미

◉ 민감한 정보의 분포를 고려

- 각 동질 집합에서 '민감한 정보의 분포'가 전체 데이터 집합의 그것과 비교하여 너무 특이하지 않도록 함
- 〈표 5〉에서, 전체적인 급여 값의 분포는 30 ~ 110
- 이 때, 레코드 1, 2, 30이 속한 동질 집합에서 급여의 분포는 30 ~ 50으로, 이는 전체 급여 값의 분포(30 ~ 110)와 비교할 때 극히 일부 → 공격자는 근사적인 급여 값을 알 수 있음
- *t*-근접성 모델은 이러한 동질 집합과 전체 데이터 집합 사이 분포의 과도한 차이를 *l*-다양성 모델의 취약점으로 규정함

## *t*-Closeness

◉ '민감한 정보의 분포'를 조정하여 프라이버시를 보호
- 민감한 정보가 특정 값으로 쏠리거나, 유사한 값들이 뭉치는 경우를 방지
- 〈표 6〉에서 *t*-근접성 모델에 따라 레코드 1,3,8이 하나의 동질 집합으로 익명화됨
- 이 때, 레코드 1, 3, 8의 급여의 분포는 (30 ~ 90)으로 전체적인 급여의 분포(30 ~ 110)와 큰 차이가 나지 않음
- 또한, 레코드 1, 3, 8의 질병 분포는 (위궤양, 만성위염, 폐렴)으로 병명이 서로 다르면서, 질병이 '위'와 관련된 것 이외에 '폐'와 관계된 것이라서, 특정 부위의 질병임을 유추하기 어려움
- 따라서 〈표 5〉의 경우와 비교하여, 공격자가 공격 대상의 민감한 정보를 추측하기가 더욱 어려워짐

◉ 민감한 정보의 의미(semantics)까지 파악하는 프라이버시 모델
- 민감한 정보의 의미를 고려하여 값의 분포를 계산함
- 연속 속성(continuous attribute)의 경우 숫자 값을 통해 의미가 유사한 정도를 파악 (예: 급여)
- 범주 속성(categorical attribute)의 경우 분류 트리(taxonomy tree, 〈그림 1〉 참고)를 이용해 의미가 유사한 정도를 파악 (예: 질병)

- **k-Anonymity, l-Diversity, t-Closedness의 문제점**

  - **General k-Anonymity is NP-Hard**

    - Practical algorithms exist using approximation!

## ● 비식별화된 정보가 재식별화가 가능한지를 둘러싼 논쟁

### 비식별화가 불가능하다고 보는 논문 및 보고서

1. **백악관 보고서, "Big Data: Seizing Opportunities, Preserving Values," Executive Office of the President, May 2014**

   "When data is initially linked to an individual or device, some privacy-protective technology seeks to remove this linkage, or 'de-identify' personally identifiable information—but equally effective techniques exist to pull the pieces back together through 're-identification.'"

2. Public Knowledge et al., "Petition for Declaratory Ruling Stating that the Sale of Non-Aggregate Call Records by Telecommunications Providers without Customers' Consent Violates Section 222 of the Communications Act," December 11, 2013

3. Timothy Lee, "There's No Such Thing as an Anonymized Dataset," TechDirt, November 30, 2007

## ● 비식별화된 정보가 재식별화가 가능한지를 둘러싼 논쟁

**비식별화가 불가능하다고 보는 논문 및 보고서**

4. **Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen & Vincent D. Blondel, "Unique in the Crowd: The Privacy Bounds of Human Mobility," Scientific Reports 3.**

   The study found that when an individual's location is specified hourly within the reception area of a mobile phone antenna, knowing as few as four random spatio-temporal points was enough to uniquely identify 95 per cent of the mobility traces in the dataset of one and a half million individuals.

   It is admittedly very difficult to de-identify mobility traces, while maintaining a sufficient level of data quality necessary for most secondary purposes, due to their high degree of uniqueness.

5. **EU Data Protection Working Party (Article 29), "Opinion 05/2014 on Anonymisation Techniques," April 10, 2014.**

   The removal of direct identifiers alone is generally insufficient to properly de-identify datasets.

## ● 비식별화된 정보가 재식별화가 가능한지를 둘러싼 논쟁

**비식별화가 불가능하다고 보는 논문 및 보고서**

6. **Latanya Sweeny, Uniqueness of Simple Demographics in the U.S. Population, Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA, 2000.**

   It used 1990 U.S. census data to show that 87 per cent of the U.S. population could be uniquely identified through the combination of gender, date of birth, and ZIP code.

7. Phillippe Golle, Revisiting the Uniqueness of Simple Demographics in the US Population, Palo Alto Research Center, 2006

   - Only 63 per cent of the U.S. population is uniquely identifiable given those data categories.
   - If an individual's date of birth is replaced with only the month and year of birth, the percentage of those uniquely identifiable drops to 4.2 per cent.
   - If one further replaces the ZIP code with an individual's county, then the percentage of the population capable of being uniquely identified drops dramatically to 0.2 per cent.1

## ● 비식별화된 정보가 재식별화가 가능한지를 둘러싼 논쟁

**비식별화가 불가능하다고 보는 논문 및 보고서**

8. Nate Anderson, "'Anonymized Data Really Isn't—and Here's Why Not," Ars Technica, September 8, 2009.
9. Caroline Perry, "You're Not So Anonymous," Harvard Gazette, October 18, 2011.
10. Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets," Proceedings of the 2008 IEEE Symposium on Security and Privacy, 2008.

## ● 비식별화된 정보가 재식별화가 가능한지를 둘러싼 논쟁

**비식별화가 가능하다고 보는 논문 및 보고서**

1. Jane Yakowitz, “Tragedy of the Data Commons,” Harvard Journal of Law and Technology 25 (2011): 1–40.

2. Felix T. Wu, “Defining Privacy and Utility in Data Sets,” University of Colorado Law Review 84 (2013): 1117–1175.

3. Khaled El Emam, Guide to the De-Identification of Personal Health Information (Boca Raton, FL: CRC Press, 2013).

4. Daniel C. Barth-Jones, “The ‘Re-identification’ of Governor William Weld’s Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now.” June 4, 2012.

5. Ann Cavoukian and Khaled El Emam, “Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy,” June 2011.

# III. 비식별화와 법규제

## ● 비식별화된 정보가 재식별화가 가능한지를 둘러싼 논쟁

**비식별화가 가능하다고 보는 논문 및 보고서**

6. Ann Cavoukian and Daniel Castro, “Big Data and Innovation, Setting the Record Straight: De-identification *Does* Work,” Jun 2014.

“While nothing is perfect, the risk of re-identification of individuals from properly de-identified data is significantly lower than indicated by commentators on the primary literature.”

## ● 비식별화 기술이 처리하는 문제

1. They protect an individual's records from being uniquely identified in the dataset.
2. They prevent an individual's records from being linked to other datasets.
   - If a set of attributes uniquely identifies an individual within a de-identified dataset and those same attributes are found in a personally identifying dataset, then that individual may be re-identified by linking the two datasets together.
3. They make it difficult to infer sensitive information about an individual from the de-identified dataset.
   - If groups of individuals are identified in a dataset and all the individuals in a certain group have a certain property, then if an individual is known to belong to that group, one could easily find out the value of his/her group property.

Source: EU Data Protection Working Party (Article 29), "Opinion 05/2014 on Anonymisation Techniques," p. 11–12

## ● 데이터 유통 금지에 따른 가장 큰 피해자는?

1. 개인 정보 부재에 따른 중금리 대상자 (서민)
   - 은행 안전 대출 선호로 인해 우리나라는 중금리시장이 형성되지 못한 채 제2금융권, 대부업체, 불법사금융으로 내몰리고 있음.
   - 중금리 대출시장의 부재로 신용도로 5~6등급 (중신용자)에 해당하는 1180만명이 연 15%~34.9 (50%이상이 30%이상임)의 약탈적 금리에 시달리는 중.
   - 금리 양극화를 해소하기 위해서는 데이터가 유통되어야 함.
2. 제2금융권 및 대부업체
   - 대출자에 대한 정보가 전무함으로 인해 높은 대출 부실률에 시달리는 중.

12.6%(2016.2)에 달하는 청년실업자들이 가장 큰 피해자!

## ● 데이터 생산자는 누구?

1. 기간 산업-인프라 제공 서비스사
- 전기통신, 전화, 수도, 철도, 고속도로, 인터넷

2. 편의제공 서비스 제공사
- 은행-신용카드-보험 등 금융사 및 금융관련 서비스 제공사
- 식품점, 마트

3. 온라인 플랫폼
- 페이스북, 트위터, 카카오톡, 유튜브, 위키피디아 등 소셜미디어
- 네이버, 다음 등 포털 서비스
- 아마존, 11번가, G Market 등 e-Commerce 서비스
- 구글 등 검색제공사

데이터 생산자로 보면 빅데이터(=벌크데이터)는 상품화할 이유 없음!

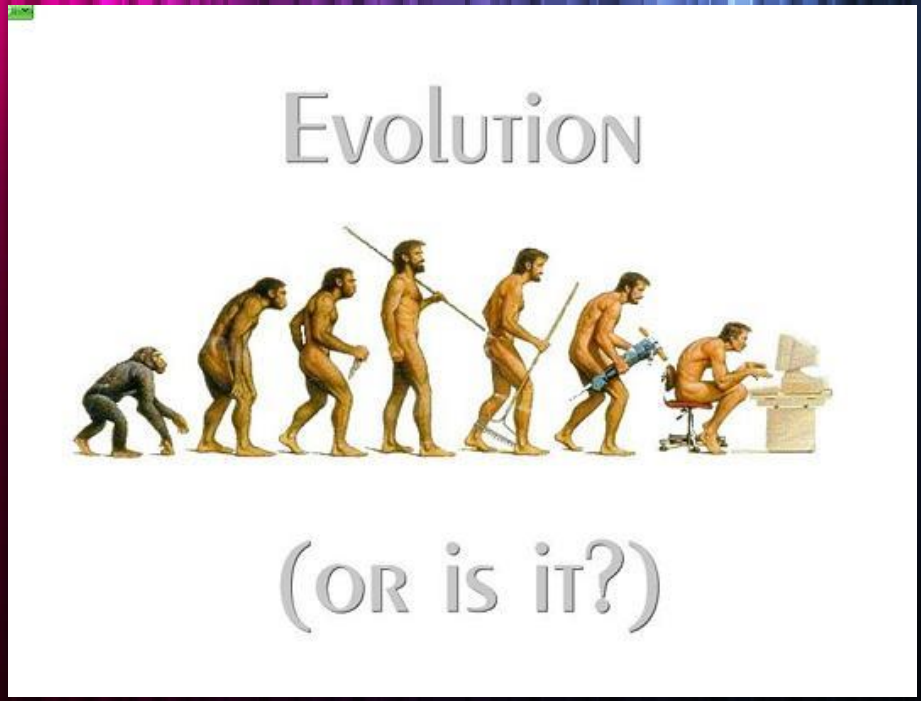
## ● 개인정보자기결정권: 판매결정도 자기가 할 수 있어야 함

개인정보의 공개와 이용에 관하여 스스로 결정할 수 있는 권리

- 헌법재판소는 개인정보자기결정권을 헌법상 기본권으로 인정.
- 개인정보자기결정권의 보호대상이 되는 개인정보 : 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보 등.

개인정보자기결정권은 개인정보 판매 유통권이 포함되어야 함!

# II. 진화의 관점에서 본 데이터 산업



- 위험하면 막아야 할까요?

# II. 진화의 관점에서 본 데이터 산업

- 산업의 진화

산업사회 시대

가내수공업 시대

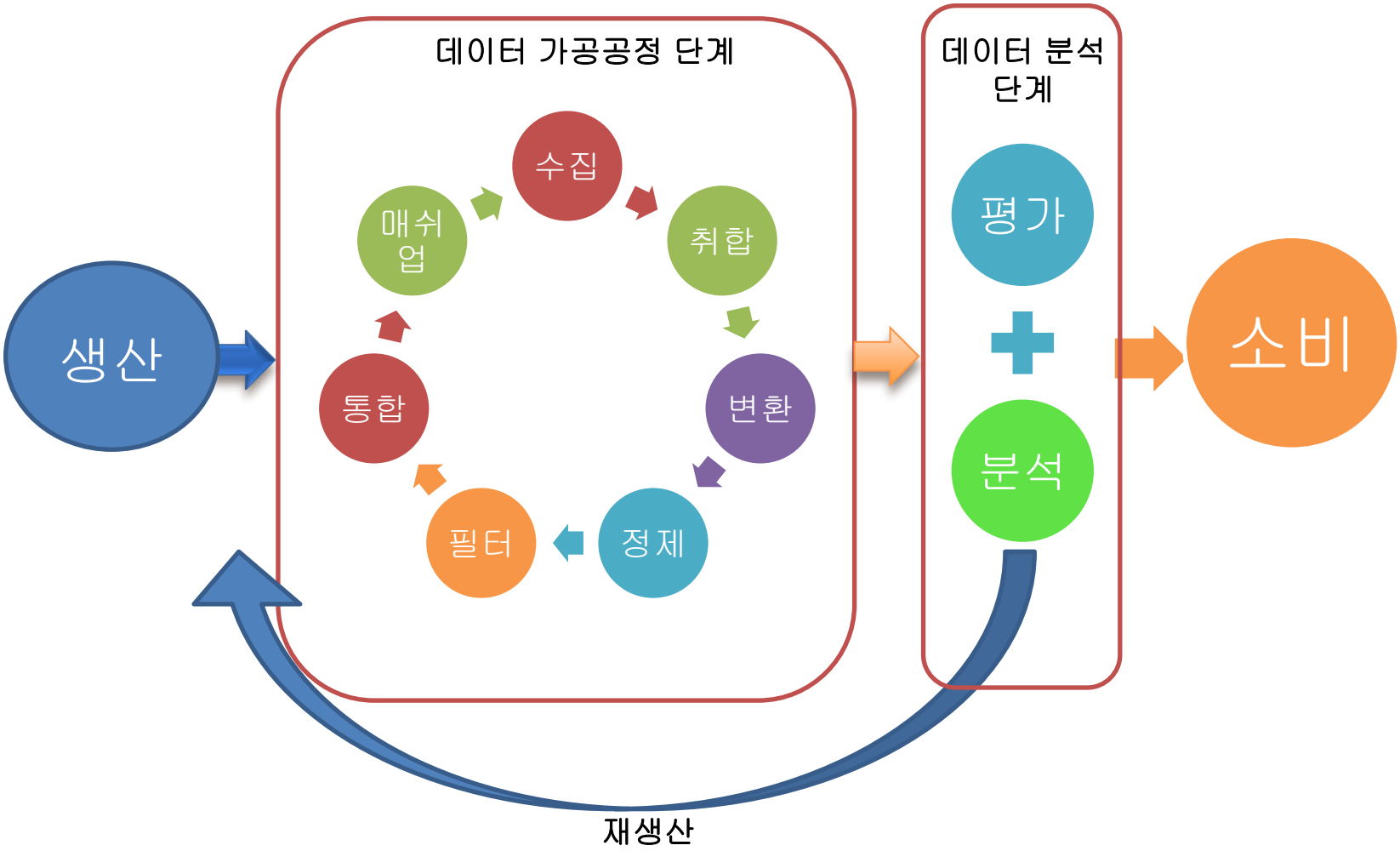
자급자족 시대

수렵채취시대

진화로 본 한국 데이터 산업은?

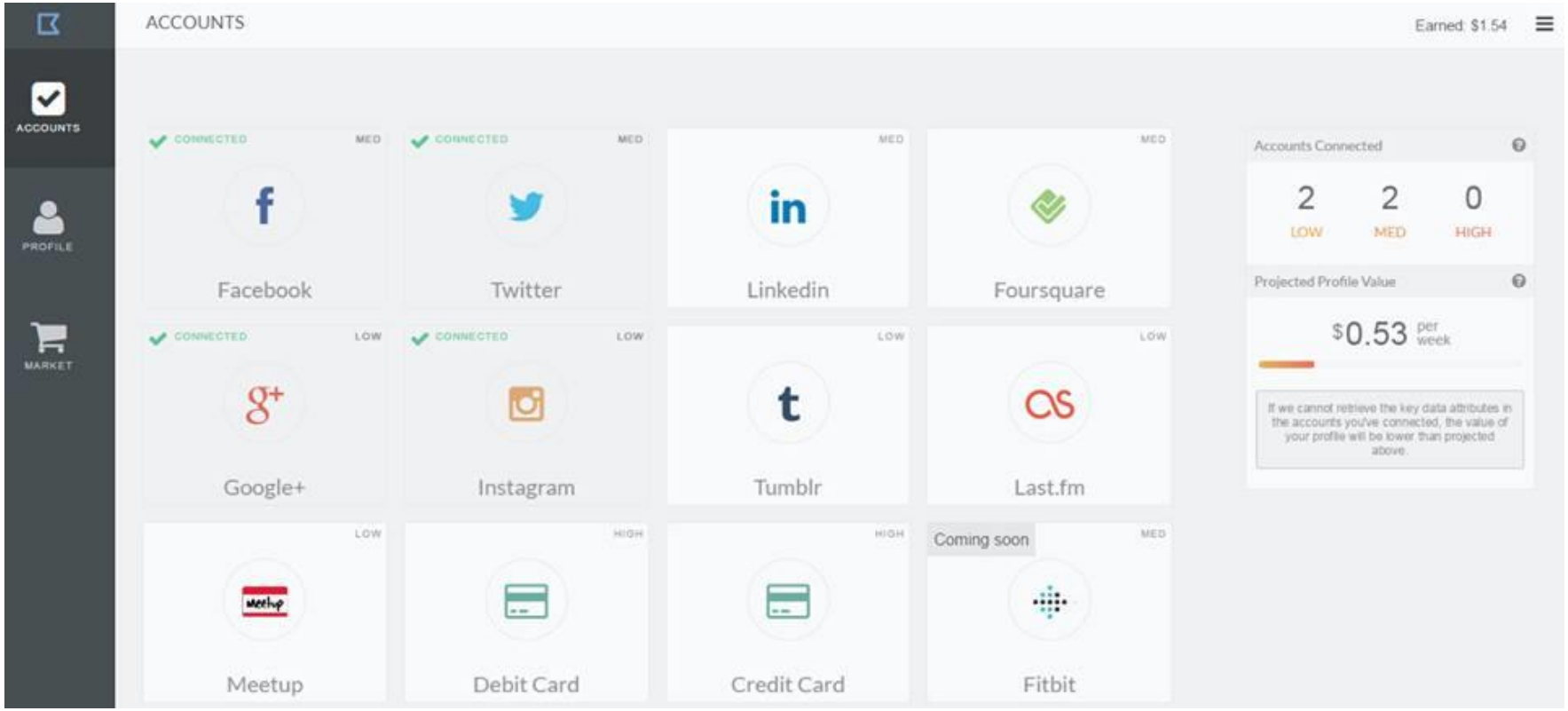
# I. 진화의 관점에서 본 데이터 산업

## ● 이상적인 데이터 생태계



# II. 진화의 관점에서 본 데이터 산업

- 외국의 경우: DataCoup




개인 정보 일주일에 0.53센트!

# II. 진화의 관점에서 본 데이터 산업

- 외국의 경우: Dutch 대학생 Shawn Buckles

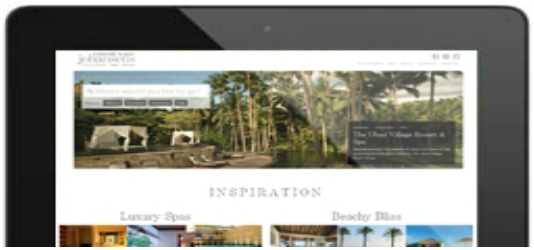
WIRED.CO.UK

## 'Data soul' of Shawn Buckles sells for £288

TECHNOLOGY / 15 APRIL 14 / by OLIVIA SOLON








Dutch student Shawn Buckles has auctioned all his personal data to the highest bidder and earned a grand total of €350 (£288).

In March, Buckles set up a website with an online bidding system in order to make a comment about privacy and the value of personal data. He put all of his personal data

Shutterstock

출처 : http://bit.ly/1m5djw3

2014년 한 대학생이 경매를 통해 개인정보 판매로 벌어들인 돈은 350유로!

# III. 결론

# ● 데이터산업 활성화

- **과도한 정보보호는 데이터산업 발전을 저해**
  - 정보보호에 대한 정확한 인식이 필요
  - 현행 개인정보보호법은 위헌 소지 있음
- **데이터 산업은 유통 생태계부터 시작**
  - 데이터를 유통시킬 수 있는 법제도 정비필요
  - 완벽한 비식별화를 요구하는 것은 무리라는 인식이 필요
- **데이터 유통을 위한 거래소에 투자할 필요있음**

- **참고논문**
  - 이영환, 전희주, 윤정연. 데이터 산업에서 창업 활성화를 위한 데이터 거래소 제안 : 금융거래소형 데이터거래소를 중심으로. 창업학회지.  2015년 6월.
- **신문칼럼**
  - 개인정보보호법은 무효다. TechM (2016/1/6).
  - "'비식별화'라는 코끼리" 전자신문. (2015/11/9)
  - "사하라 사막에 인터넷을 공급하려면" 서울경제. (2015/10)
  - ."인터넷 전문은행, 빅데이터 분석이 특화전략 핵심" 머니투데이. (2015/8)


# Q&A

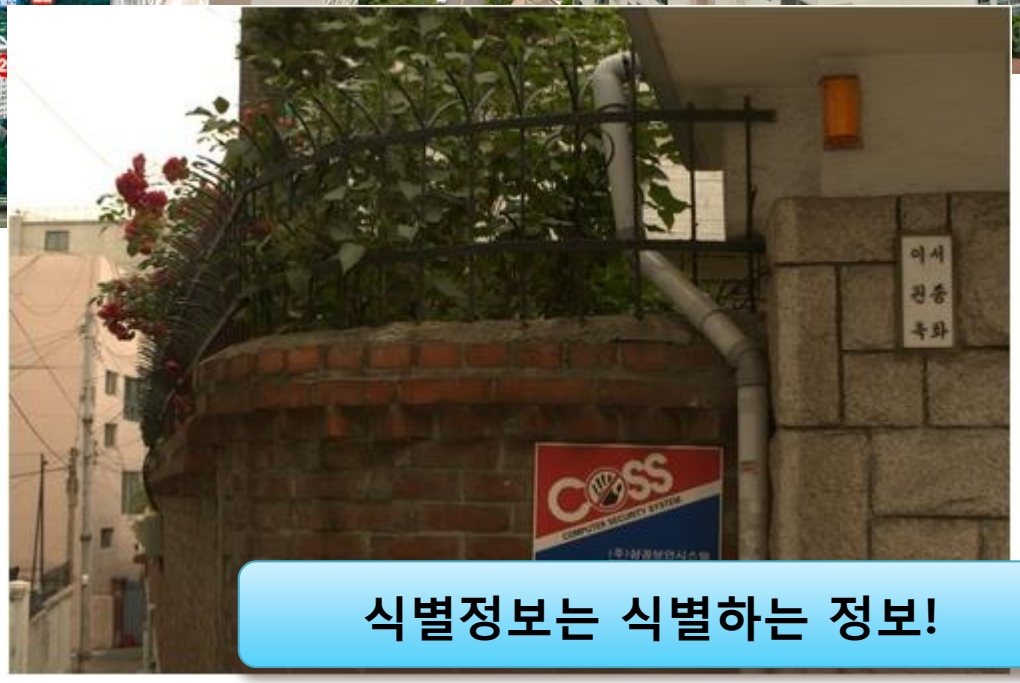
# 부록

얼굴이 식별 정보 인가요, 인증 정보인가요?

# 식별정보와 인증정보

## 식별정보의 예






**식별정보는 식별하는 정보!**

# 식별정보와 인증정보

## 식별정보의 예



# 식별정보와 인증정보

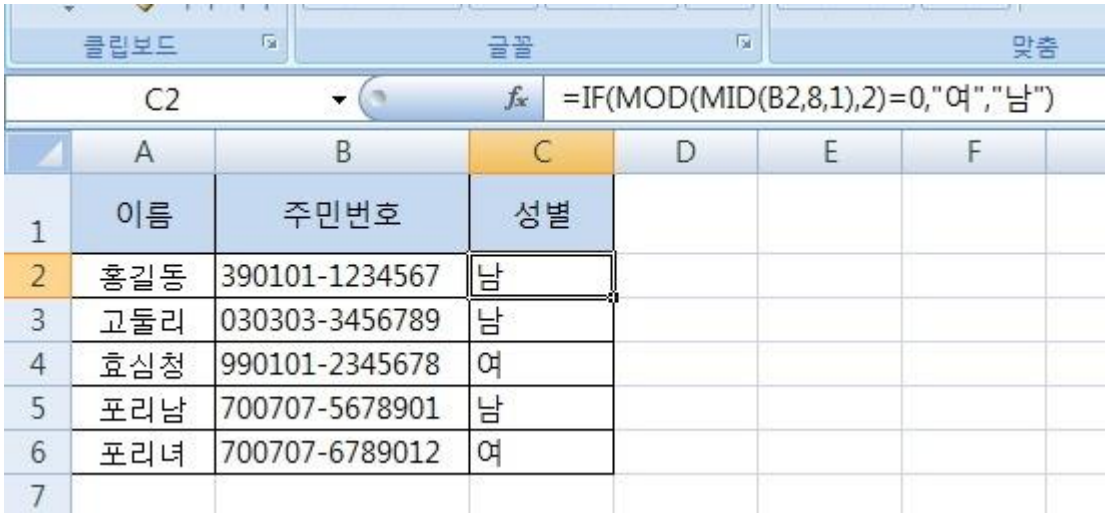
## ● 식별정보 vs. 인증정보

### 식별정보의 예


식별정보를 감추면?

# 식별정보와 인증정보

## 식별정보의 예




| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | 이름 | 주민번호 | 성별 | | | |
| 2 | 홍길동 | 390101-1234567 | 남 | | | |
| 3 | 고둘리 | 030303-3456789 | 남 | | | |
| 4 | 효심청 | 990101-2345678 | 여 | | | |
| 5 | 포리남 | 700707-5678901 | 남 | | | |
| 6 | 포리녀 | 700707-6789012 | 여 | | | |
| 7 | | | | | | |

주민번호는 식별정보!

# 식별정보와 인증정보

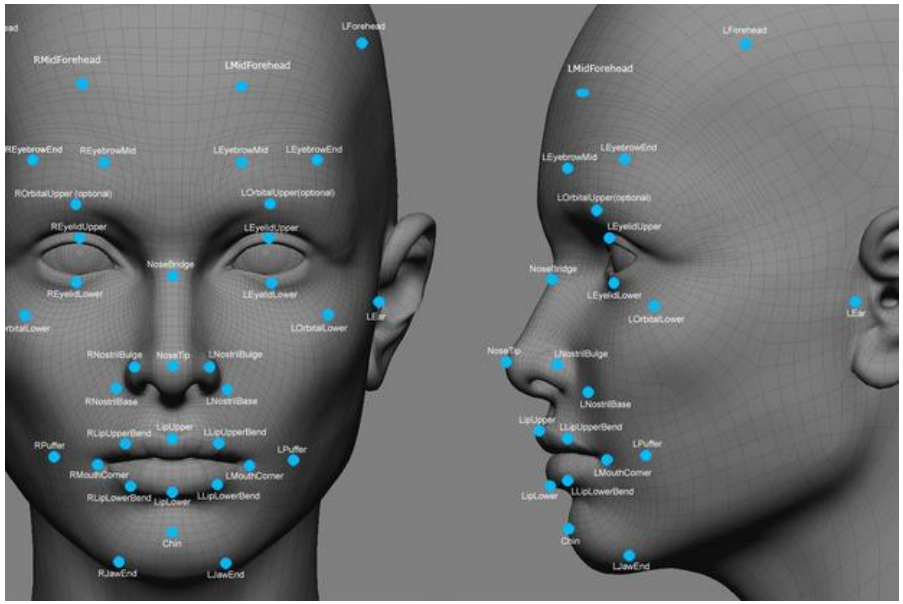
## 인증정보의 예




**Q 지금까지 알려진 인증 기술이 안전한가요?**

# 식별정보와 인증정보

## 식별정보와 인증정보의 혼동 존재




안면 인식이 인증이 되나?

opennet.or.kr
opennetkorea.org