통계연구(2016), 특별호, 35-51

# 통계적 목적의 개인정보보호와 비식별화

김혜련[1)]

**요약**

빅데이터 시대에 개인정보의 활용가치는 증대하고 있지만, 개인정보보호법 등 법률에서 개인정보에 대한 활용을 엄격히 규제하고 있어, 개인정보가 포함된 자료를 민간에서의 활용은 어려운 현실이다. 하지만 개인정보보호법에서는 통계작성의 목적으로 비식별화된 개인정보를 제3자 제공 및 활용이 가능하고, 통계법에 따라 수집된 개인정보는 적용예외 규정을 두고 있어 통계청은 개인정보의 활용이 가능하여 빅데이터간 연계를 통한 다양한 통계적 정보 제공이 가능하다. 따라서, 통계청은 통계청 고유의 비식별화 방법과 개인정보보호 방법의 적용을 통해 현행법 체계하에서 개인정보를 보호하고 빅데이터를 연계·융합하여 다양한 수요자 맞춤의 통계적 정보를 제공하고 데이터의 잠재 가치를 최대한 활용할 수 있도록 공공데이터 및 민간데이터 거버넌스 구축을 위해 노력하여야 한다.

주요용어 : 빅데이터, 자료연계, 개인정보보호, 비식별화

## 1. 서론

최근 정보통신의 발전으로 데이터가 축적되고 대용량 데이터의 신속한 검색 및 분석기술이 개발되면서, 빅데이터는 새로운 부가가치를 창출하는 미래 성장의 원동력 및 새로운 통찰력을 제공하는 잠재력이 높은 자료로 주목 받고 있다.

빅데이터란 규모가 방대하고, 생성주기가 짧으며, 정형화된 수치 형태뿐만 아니라 문자, 영상 등 비정형화된 대용량의 데이터라고 일반적으로 말하지만, 단순히 용량이 큰 데이터라기 보다는 정보통신의 발전으로 과거에는 활용이 어려웠지만 이제는 분석과 활용이 가능한 자료 모두를 의미한다고 본다. 즉, 금융, 통신, 기업경영, 의료 등 민간이 보유한 데이터는 물론 통계자료, 국세자료 등 공공데이터도 모두 포함한다.

국내 빅데이터 시장규모는 2014년 약 2천억원으로 연 20% 이상의 성장세를 보일 것으로 전망되어[2)], 데이터사이언티스트 등 전문인력 양성, 데이터분석 기술 개발, 시스템 구축 등의 빅데이터를 관리·분석·활용 분야에서 성장이 기대된다.

한편 빅데이터는 대용량의 자료 자체만을 분석하는 것 뿐만 아니라 다양한 자료간 연계·융합을 통한 새로운 정보 제공시 보다 높은 가치를 가진다. 예를 들어, 개인신용평가기관의 부채정보와 통계청의 인구·가구·주택 모집단 정보와의 연계를 통해 가구특성별 부채정보를 제공하거나, 기업신용평가기관의 개별 기업의 대출정보와 국세청, 통계청 등의 사업체 매출정보간 연계를 통해 소상공인의 자금사정 및 보증위험도 등의 정보 제공이 가능하여 빅데이터 시대에 개인정보의 활용가치는 증대하고 있다.

---

1) 대전광역시 서구 둔산동 949, 통계청 빅데이터통계과. E-mail: khr@korea.kr

2) 한국정보화진흥원, “2014 빅데이터 시장 현황조사”

하지만, 빅데이터 활용시 대량의 개인정보 유출로 인해 헌법의 개인정보자기결정권[3]을 침해할 수 있다는 우려가 높으며, 개인정보보호법 등 법률에서 개인정보에 대한 활용을 엄격히 규제하고 있어, 개인정보가 포함된 자료를 민간에서 영리적 목적에서의 활용은 어렵다.

한편 미국, 일본 등 주요 선진국에서는 개인정보의 침해 가능성을 최소화하면서 빅데이터 활성화를 위한 정책을 추진하고 있어, 우리나라도 빅데이터의 이용 및 활용을 촉진하면서 개인정보를 침해하지 않는 방안을 마련하는 것이 필요하다. 이를 위해, 행정자치부, 금융위원회 등 정부기관에서 빅데이터 활성화를 위한 개인정보보호법 개정 등을 검토하고 있으나, 아직 현실화되기에는 시간이 걸릴 것으로 판단된다.

빅데이터는 산업 활성화 측면 뿐만 아니라 공공 및 민간 빅데이터를 활용한 통계적 정보를 제공하여 정책적으로 활용되고, 다양한 국민들의 요구에 맞는 시의성 높은 자료로 활용되는 것도 중요하다. 개인정보보호법에서는 통계작성의 목적으로 비식별화된 개인정보를 제3자 제공 및 활용이 가능하고, 통계법에 따라 수집된 개인정보는 적용예외 규정을 두고 있다. 우리나라와 유사하게 EU, 영국 등에서도 비식별화된 개인정보는 통계작성의 목적으로 동의없이 활용할 수 있어 통계적 목적의 개인정보가 포함된 빅데이터의 활용은 가능하다. 이런 측면에서, 통계청은 개인정보의 활용이 가능하여 빅데이터간 연계를 통한 다양한 통계적 정보 제공이 가능하다.

따라서, 본 논문에서는 개인정보보호 관련 법률을 검토하여 현재의 법체계하에서 빅데이터의 통계적 활용 가능성을 검토하고, 통계적 목적의 비식별화 방법을 제시하여 데이터간 연계·융합을 통한 빅데이터 활성화를 도모하고자 한다.

본 논문의 구성은 다음과 같다. 제2장에서는 개인정보보호 관련 국내외 법률현황, 비식별화의 개념 및 방법, 빅데이터 활용시 법적 한계를 살펴본다. 제3장에서는 통계적 목적의 빅데이터 활용과 비식별화를 위해 통계청의 빅데이터 활용의 법적 근거와 비식별화 방법을 제시한다. 마지막으로 제4장에서는 향후 통계적 목적의 빅데이터 활용 확대를 위한 제언으로 마무리한다.

## 2. 개인정보보호 관련 법률과 한계

### 2.1 개인정보보호 관련법 현황

우리나라의 개인정보보호와 관련된 법은 민간 및 공공 부문에 적용되는 일반법으로서 개인정보보호법이 있다. 개인정보보호법은 개인정보에 대한 정의, 개인정보처리자가 준수해야 할 개인정보보호처리 원칙, 정보주체의 권리, 개인정보의 수집·이용, 목적외 이용 및 제공시 정보주체의 사전적 동의 등에 대해 규정하고 있어 동법 제6조

3) 헌법 제10조의 인간의 존엄과 가치, 행복추구권과 제17조의 사행활의 비밀과 자유에서 도출되는 것으로, 자신에 관한 정보가 언제 누구에게 어느 범위까지 알려지고 또 이용되도록 할 것인지를 그 정보주체가 스스로 결정할 수 있는 권리이다.

에 따라 다른 법률에 특별한 규정이 있는 경우를 제외하고는 개인정보보호에 관하여서는 개인정보보호법을 따라야 한다.

분야별 개별법으로 정보통신 관련 정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 정보통신망법), 금융거래 관련 신용정보의 이용 및 보호에 관한 법률(이하 신용정보법), 위치정보 관련 위치정보의 보호 및 이용 등에 관한 법률(이하 위치정보법), 의료 관련 의료법 등이 있다. 개별법에서도 개인정보보호법과 유사하게 개인정보의 정의, 개인정보의 수집·이용 등을 규정하고 있다.

### 2.1.1 개인정보보호법

개인정보보호법은 개인정보 보호의 중요성이 증가함에 따라 기존의 공공부문의 「공공기관의 개인정보보호에 관한 법률」을 기본으로 민간까지 포괄하는 통합법으로서 2011년 3월 제정되었으며, 개인정보에 대한 정의, 개인정보의 처리시 제한 등을 규정하고 있다.

개인정보보호법에서의 개인정보는 살아있는 개인에 관한 정보로서 성명, 주민등록번호, 영상 등을 통해 특정 개인을 알아볼 수 있는 정보이며, 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보까지 포함한다(제2조 제1호). 또한, 개인정보보호법은 개인정보의 수집, 이용, 제공 등 개인정보 처리시는 사전동의방식(Opt-in)[4]을 채택하고 있다. 즉, 개인정보처리자의 경우 정보주체의 사전동의를 받아야 수집·이용(제15조 제1항) 및 제3자 제공(제17조 제1항)이 가능하다.

개인정보의 목적외 이용·제공은 개인정보보호법에서는 제한하고 있지만, 예외적으로 통계작성 및 학술연구의 목적으로 필요한 경우 특정개인을 알아 볼 수 없는 형태로 개인정보를 제공할 수 있도록 규정하고 있다(제18조 제2항 제4호). 특히, 「통계법」에 따라 수집되는 개인정보는 제3장에서 7장까지의 적용의 예외를 규정하고 있어 통계청의 경우 통계작성을 목적으로 공공 및 민간의 개인정보를 수집·처리 할 수 있다.

### 2.1.2 개별법

정보통신망을 이용하는 자의 개인정보를 보호하는 정보통신망법은 1986년 '전산망 보급확장과 이용촉진에 관한 법률'로 시작되어 1999년 '정보통신망 이용촉진 등에 관한 법률'로 전부 개정되었다. 2001년 2월 '정보통신망 이용촉진 및 정보보호 등에 관한 법률'로 다시 전부 개정되었으며, 이후 2011년과 2012년에 두 차례 개정되어 오늘에 이르고 있다. 정보통신망법에서 개인정보는 생존하는 개인에 관한 정보로서 성명·주민등록번호 등 특정한 개인을 알아볼 수 있는 부호·음성·영상 등의 정보이며, 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보도 포함한다(제2조 제1항 제6호). 또한, 정보통신서비스 제공자는 개인정보의 수집·이용 및 제3자 제공시 사전동의를 받아야 하지만(제22조 제1항, 제24조의 2의 제1항), 다른 법률에 규정이 있는 경우는

4) 개인정보를 제3자에게 제공하기 위해 사전에 개인에게 동의를 구하는 방식을 말하며, 반대로 사전에 개인에게 동의를 구하지 않고 사후적으로 동의를 받는 사후동의방식(Opt-out)이 있다.

예외로 하고 있다(제24조의 2의 제2항).

신용정보법에서의 신용정보는 금융거래 등 상거래에 있어서 거래 상대방의 신용을 판단할 때 필요한 정보로서, 특정 신용정보주체를 식별할 수 있는 정보, 신용정보주체의 거래내용, 신용도 등을 판단할 수 있는 정보로 정의된다(제2조 1). 개인신용정보의 제3자 제공·활용을 위해서는 사전적으로 동의를 받아야 하지만(제32조 제1항), 과세목적, 법원의 제출명령 등과 다른 법률에 따라 제공하는 경우는 적용 예외조항으로 규정하고 있다(제32조 제6항). 또한 신용정보회사 등은 공공기관의 경우 관계법령이 정하는 공무상 목적으로 신용정보의 제공을 문서로 요청한 경우에는 신용정보를 제공할 수 있으며(제23조 제7항), 개인정보의 수집·이용 및 제3자 제공을 위해서는 사전적 동의를 받아야함을 규정하고 있다(제22조 제1항).

위치정보법에 의한 개인위치정보는 특정개인의 위치정보로서 다른 정보와 용이하게 결합하여 특정개인의 위치를 알 수 있는 것을 포함한다(제2조 제2호). 개인위치정보의 수집·이용 및 제3자 제공을 위해서는 개인위치정보주체의 동의를 얻어야 하지만(제18조 및 제19조), 예외조항으로서 통계작성, 학술연구 또는 시장조사를 위하여 특정개인을 알아볼 수 없는 형태로 가공하여 제공하는 경우를 두고 있다(제21조 제2호).

### 2.2 개인정보보호법 관련 해외사례

#### 2.2.1 OECD

OECD의 1980년 '프라이버시 보호와 개인데이터의 국제유통에 대한 가이드라인'(Guidelines on the Protection of Privacy and Transborder Flows of Personal Data)은 수집제한의 원칙, 정보품질의 원칙 등 8대 기본원칙[5]을 제시하여 개인정보보호와 관련하여 해외 각 국의 법률에 큰 영향을 주었다. 이후 2013년 7월 개정되어, 기본 원칙을 유지하되 프라이버시 집행기관 신설, 데이터 보안 관련 침해시 집행기관 및 정보주체에게 통지 의무화 등 높아진 개인정보 침해 가능성에 대비하였다.

#### 2.2.2 EU

유럽연합(EU)위원회는 2012년 1월에 회원국간의 단일법으로서 일반정보보호규정(GDPR: General Data Protection Regulation)을 제안하였고 2015년 12월에 EU 위원회, 의회 및 이사회는 승인하였다. 이는 2005년 10월에 제정된 데이터보호지침[6]을 개정한 것으로, 일관되고 강화된 개인정보보호 체계를 구축하고 디지털 경제 발전을 도

5) 수집제한의 원칙(collection limitation principle), 정보품질의 원칙(data quality principle), 목적 명확화의 원칙(purpose specification principle), 이용제한의 원칙(use limitation principle), 안전확보의 원칙(security safeguards principle), 공개의 원칙(openness principle), 개인참여의 원칙(individual participation principle), 책임의 원칙(accountability principle)

6) Directive of the European Parliament of individuals with regard to the processing of personal data and on the free movement of such data, 95/46/EC

모하기 위한 목적으로 2018년 초부터 각 회원국에 직접 적용될 예정이다.

GDPR에서는 빅데이터 활용을 위해 특정인을 식별할 수 없는 개인정보인 가명화된(pseudonymized) 정보를 활용 가능하도록 하여 빅데이터 활용의 근거를 마련하였는데, 특히, 공익, 과학적 연구, 역사연구 및 통계작성의 목적으로는 정보주체의 동의없이 가명화된 정보를 사용가능함을 명시하였다.

이와 함께 정보주체의 권리도 강화하여 프로파일링(profiling)[7] 거부권과 온라인상의 개인정보에 대한 삭제할 수 있는 잊혀질 권리(right to be forgotten)[8]를 규정하였으며, EU내 개인정보의 제3국으로의 역외이전은 EU에 상응하는 수준의 정보보호체계를 가진 경우 또는 기업규칙(BCR: Binding Corporate Rule) 등의 경우 이전될 수 있도록 하였다.

### 2.2.3 일본

일본은 2003년 5월 제정된 개인정보보호법을 2015년 9월 개정하여 개인정보를 보호하면서 빅데이터 활용을 촉진하는 제도적 기반을 마련하였다. 이번 법률 개정에서는 특히 개인이 특정되지 않도록 하는 '익명가공정보화'를 도입하여 익명가공정보화시 본인의 동의없이 제3자에게 제공할 수 있도록 하여 빅데이터 활용을 도모하였다. 주요 개정 내용을 살펴보면 다음과 같다.

첫째, 개인정보에 대한 정의를 명확히 하였다. 기존의 개인정보는 성명, 생일 등 개인을 식별할 수 있는 살아있는 개인의 정보라고 정의 되었는데, 이번 개정에서는 지문, 안면인식 자료 등 특정 개인의 신체 정보와 여권번호, 운전면허번호 등 개인이 상품과 서비스를 받거나 구매시 부여받는 번호를 포함하는 개인식별부호가 담긴 것을 개인정보라고 명확히 정의하였다.

둘째, 적정한 규율하에서 개인정보 등의 유용성을 확보하였다. 빅데이터 관련 산업 활성화를 위해, 특정개인을 식별할 수 없도록 개인정보를 가공하고 복원할 수 없도록 만든 정보인 익명가공정보(anonymized information) 조항을 신설하여 사업자가 본인의 동의 없이 제공이 가능하도록 하였다.

셋째, 내각부(內閣府)의 외국(外局)으로 개인정보보호위원회를 신설하여 개인정보보호 취급에 대한 감시·감독 권한을 강화하였으며, 일본내의 개인정보를 취득한 외국 기관에 대해서도 개인정보보호법을 적용하고, 외국의 제3자 제공시 제한하는 조항을 신설하였다.

### 2.2.4 미국

미국의 경우 일반화된 개인정보보호법은 없으며, 공공·통신·온라인·의료 등 각

---

7) 개인을 평가하는 것 또는 직무상의 성과, 경제상황, 위치, 개인적 기호, 행동 등을 분석 또는 예측하기 위해 개인데이터를 처리하는 것을 말한다.

8) 개인정보를 보유할 합법적인 근거가 없는 한 온라인 등에서의 자신에 대한 개인정보를 삭제할 수 있는 권리이다.

영역별 법 및 제도가 있다. 공공부문에서는 예산관리실(OMB: the Office of Management and budget)에서 프라이버시법(Privacy Act)을 관장하고 있으며, 민간부문은 연방거래위원회(FTC: the Federal Trade Commission)에서 소비자보호관점에서 개별 프라이버시권 관련 관할을 하고 있지만 기본적으로 개인정보처리와 관련하여 민간의 자율규제에 맡기고 있다.

의료관련법 중 건강보험 이전과 책임에 관한 법(HIPAA: Health Insurance Portability and Accountability Act)에 따른 HIPAA 프라이버시 규칙(HIPAA Privacy Rule)에서 비식별화된 정보는 제한없이 사용가능하도록 하였으며, 전문가 결정 방식(expert determination)과 식별자를 제거하는 세이프하버(safe harbor) 두 가지 방식의 비식별화 방법을 제시하고 있다.[9)]

가족의 교육적 권리 및 프라이버시 법(FERPA: Family Educational Rights and Privacy Act)에서는 비식별 조치된 학생기록에 대해 별도의 동의없이 배포가 가능하고, 경제적·임상적 보건에 대한 건강 정보기술법(HITECHA: Health Information Technology for Economic and Clinical Health Act)에서는 비식별 조치된 건강정보에 대해 프라이버시 관련 규제를 적용하지 않고 있다.

### 2.3 비식별화의 개념 및 방법

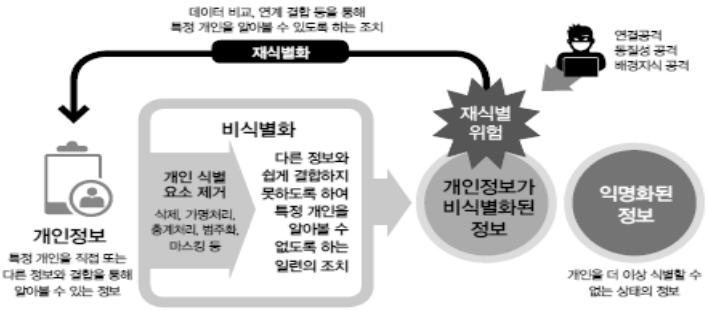
앞서 살펴본 바와 같이 일본, 미국, EU 등에서는 개인정보를 비식별화하여 더 이상 개인을 식별할 수 없으면 활용가능토록 하고 있다. 우리나라도 '14년 12월 방송통신위원회에서 빅데이터 산업의 활성화를 위해 빅데이터 처리 · 활용시 개인정보 보호와 활용을 위한 기준을 정한 「빅데이터 개인정보보호 가이드라인」(이하 가이드라인)을 발표하여 비식별화된 개인정보를 활용할 수 있도록 하였다. 가이드라인에는 개인정보가 포함된 경우 사전동의 없이 비식별화 조치시 수집·활용이 가능하도록 하였고, 이후 조합·분석 단계에서 다른 정보와 결합하여 재식별화 될 수 있는 가능성이 있는 경우 즉시 파기하거나 추가적인 비식별화 조치를 취하도록 명시하였다. 하지만, 가이드라인은 강력한 법적 효력을 가지고 있지 않아 빅데이터 활용에는 한계가 있다.

비식별화는 개인정보의 일부 또는 전부를 삭제하거나, 다른 정보로 대체함으로써 특정 개인을 식별하기 어렵도록 하는 일련의 조치를 말한다. 적용대상은 그 자체로 개인을 식별할 수 있는 정보와 다른 정보와 결합하여 개인을 알아 볼 수 있는 정보이다.[10)]〈그림 2.1〉

---

9) 전문가 결정 방식은 통계적, 과학적 원칙과 방법에 대한 지식과 경험을 보유한 전문가가 개인식별의 위험을 최소화하는 방법을 적용하는 것으로 반복가능성(replicability), 데이터소스이용가능성(data source availability), 구별가능성(distinguishability) 및 위험평가(assess risk) 원칙을 활용하여 판단한다. 세이프 하버 방식은 이름, 주소정보, 개인과 직접 관련된 날짜정보(생일, 합격일 등), 전화번호, 팩스번호 등 18개 주요 식별자를 제거하여 개인식별이 가능하지 못하도록 하는 방식이다.

10) 유사용어로 익명화(anyonymisation)가 있는데, 익명화는 개인에 대한 재식별이 더 이상 불가능한 상태를 말한다.




〈그림 2.1〉 개인정보 비식별 및 재식별 개념
* 출처 : 개인정보 비식별화 기술 활용 안내서

비식별화 기법으로는 가명처리, 총계처리, 값 삭제, 범주화, 마스킹의 5개 기법이 있다. 가명처리(Pseudonymisation)는 개인 식별이 가능한 데이터를 직접적으로 식별할 수 없는 다른 값으로 대체하는 기법으로 휴리스틱 가명화, 암호화, 교환 방법이 있다. 총계처리(Aggregation)는 통계값(전체 혹은 부분)을 적용, 특정 개인을 식별할 수 없도록 하는 것으로 총계처리, 부분총계, 라운딩 등의 기법이 있다. 데이터 삭제(Data Reduction)는 개인 식별이 가능한 데이터를 삭제 처리하는 것으로 식별자 삭제, 식별자 부분삭제, 레코드 삭제 등의 기법이 있다. 데이터 범주화(Data Suppression)는 특정 정보를 해당 그룹의 대표값으로 변환(범주화)하거나 구간값으로 변환(범주화)하여 개인 식별을 방지하는 것으로 감추기, 랜덤 라운딩, 범위 방법 등이 있다. 데이터 마스킹(Data Masking)은 데이터의 전부 또는 일부분을 대체값(공백, 노이즈 등)으로 변환하는 것으로 임의 잡음 추가, 공백과 대체 기법이 있다.

〈표 2.1〉 비식별화 기법

| 처리기법 | 예 시 | 세부 기법 |
|---|---|---|
| 가명처리 | ㅇ 홍길동, 35세, 서울 거주, 한국대 재학<br>→ 임꺽정, 30대, 서울 거주, 국제대 재학 | 휴리스틱 가명화<br>암호화, 교환 방법 |
| 총계처리 | ㅇ 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm<br>→ 물리학과 학생 키 합 : 660cm, 평균키 165cm | 총계처리, 부분총계<br>라운딩, 재배열 |
| 데이터 삭제 | ㅇ 주민등록번호 901206-1234567<br>→ 90년대 생, 남자<br>ㅇ 개인과 관련된 날짜정보(합격일 등)는 연단위로 처리 | 식별자 삭제<br>식별자 부분삭제<br>레코드 삭제<br>식별요소 전부삭제 |
| 데이터 범주화 | ㅇ 홍길동, 35세 → 홍씨, 30 ~ 40세 | 감추기, 랜덤 라운딩<br>범위 방법, 제어 라운딩 |
| 데이터 마스킹 | ㅇ 홍길동, 35세, 서울 거주, 한국대 재학<br>→ 홍○○, 35세, 서울 거주, ○○대학 재학 | 임의 잡음 추가<br>공백과 대체 |

* 출처 : 개인정보 비식별화 기술 활용 안내서(한국정보화진흥원)

### 2.4 빅데이터 활용시 현행 법의 한계

디지털 시대에 개인정보는 항상 SNS, 블로그 등 온라인상에 존재하고 노출되어 있어 프라이버시가 침해될 가능성이 높으며, 데이터의 해킹 등으로 인해 개인정보가 유출될 수 있다. 또한 정보기술의 발달으로 인해 비식별 정보도 데이터간 연계를 통해 개인 식별이 가능할 수 있다. 따라서, 개인정보보호 강화에 대한 사회적 요구가 증가하고 있다.

반면, 민간에서는 빅데이터를 이용한 새로운 서비스 창출과 신산업 활성화를 위해 개인정보의 활용가치가 증대하고 있으며, 정부에서는 공공정보의 개방·공유를 통해 투명하고 효율적 정책 집행 및 맞춤형 서비스 제공을 위해 빅데이터 활용이 필요하다.

그러므로, 개인정보를 보호하면서 빅데이터 활용 방안 모색이 필요한데, 개인정보보호법 등의 현행 법에는 빅데이터 활용시 한계가 존재한다. 첫째, 개인정보에 대한 모호한 정의로 인해 결합가능성만으로도 개인정보로 간주하고 있다. 개인정보보호법과 위치정보법에서 개인정보는 개인을 알아볼 수 있는 정보뿐만 아니라 다른정보와 쉽게 결합하여 알아 볼 수 있는 정보까지 포함한다. 즉, 정보처리자가 결합가능한 정보를 처리하는지의 여부를 불문하고 잠재적 결합 가능성만을 가지고 개인정보로 판단하고 있다.

둘째, 개인정보처리시 엄격한 사전동의방식(Opt-in)을 채택하고 있다. 개인정보보호법에서 개인정보를 처리하기 위해서는 정보주체에게 고지 후 동의를 받도록 하고 있으며, 위반시 5천만원 이하의 과태료를 부가하는 등 강력한 규제를 가하고 있어 빅데이터 활용에 장애요인이 되고 있다.

세째, 비식별화된 개인정보 활용에 대한 정확한 정의가 없어 비식별화 방법의 적용시 논란이 된다. 개인정보보호법에서는 비식별화를 특정 개인을 알아볼 수 없는 형태로만 정의하고 있어, 어느 정도까지가 비식별화가 완료된 것인지에 대한 판단기준이 불분명하다. 비록 방통위의 가이드라인에서 비식별화에 개념이 제시되었으나, 명확한 기준은 제시되지 않아 실질적인 활용에는 한계가 있다.

## 3. 통계적 목적의 빅데이터 활용과 비식별화

### 3.1 통계청의 공공 및 민간 빅데이터 활용 근거

빅데이터 활용과 개인정보보호는 상충되고 있어 개인정보가 포함된 빅데이터를 정보통신업자, 신용정보업체 등 민간에서 활용하기는 현재는 어려운 현실이다. 하지만, 통계청의 경우는 통계법에 따라 개인정보가 포함된 자료를 수집하여 비식별화하고 철저한 보안조치하에서 관리하여 통계 작성에 활용하고 있다. 특히, 국세청, 행정자치부 등 정부기관이 보유하고 있는 개인정보가 포함된 행정자료[11]를 통계법 제24조와 제24

11) 공공기관이 직무상 작성·취득하여 관리하고 있는 문서·대장 및 도면과 데이터베이스 등 전산자료를 말하며 통계자료는 제외(통계법 제3조)

조 2에 따라 수집·활용하고 있다. 2014년 현재 약 141종의 행정자료를 입수하여 약 58종의 통계 중 36종의 통계에 조사항목 대체·검증 및 신규통계 작성에 활용하고 있다.

현재 통계법에 민간빅데이터 활용에 대한 직접적인 규정은 없지만, 통계청은 기존의 법체계하에서도 충분히 비식별화된 민간빅데이터 활용이 가능하다. 즉, 개인정보보호법 제18조 및 위치정보법 제21조에 따라 통계작성 및 학술연구 등의 목적을 위해서는 비식별화한 경우 목적외 이용 및 제3자 제공이 가능하다. 또한, 통계청에서 수집하는 개인정보는 개인정보보호법 제58조에 따라 「개인정보보호법」 제3장부터 제7장까지의 적용을 받지 않아 개인정보 수집·이용·제공 등이 가능하다.

통계법에서도 제25조 '지정통계 작성을 위한 개인, 단체 등에 자료 제출' 및 제26조 '통계작성을 위한 실질조사를 위한 자료 제출'에 따라 개인정보가 포함된 자료를 입수할 권한을 가진다. 또한, 신용정보법 제23조 제7항에서는 공공기관의 장이 공무상 목적으로 신용정보의 제공을 요청한 경우 신용정보를 제공토록 하고 있어 통계청은 기관고유의 업무인 통계작성을 위해서는 신용정보의 입수가 가능하다. 자료입수와 함께 통계청은 통계법 제30조 및 제31조에 따라 통계작성 및 학술연구 등의 목적으로 통계청의 비식별화된 자료를 민간에게 제공도 가능하다.

그러므로, 통계청은 민간빅데이터를 제공받고 통계청 보유의 자료를 민간에게 제공하여 공공빅데이터와 민간빅데이터를 연계하여 공익적 목적의 다양한 통계적 정보 제공이 가능하다.

〈표 3.1〉 통계법의 자료 입수 및 자료제공 규정

| 조항 | 내 용 |
|---|---|
| 제24조<br>행정자료의 제공 | ①중앙행정기관의 장 또는 지방자치단체의 장은 통계의 작성을 위하여 필요한 경우에는 공공기관의 장에게 행정자료의 제공을 요청할 수 있다.<br>②공공기관의 장은 제1항에 따라 행정자료의 제공을 요청받은 때에는 국가기밀, 개인과 기업의 중대한 비밀의 침해 등 대통령령으로 정하는 정당한 사유가 없는 한 이에 응하여야 한다. |
| 제24조의 2<br>사법기관 등의<br>자료 제공 | ①통계청장은 통계의 작성을 위하여 필요한 경우에는 가족관계등록전산자료의 제공을 법원행정처장에게 요청할 수 있다.<br>②통계청장은 통계의 작성을 위하여 필요한 경우에는 사망원인통계에 관련된 형사사법정보의 제공을 국민안전처장관 및 경찰청장 등에게 요청할 수 있다. |
| 제25조<br>자료제출명령 | ①중앙행정기관의 장 또는 지방자치단체의 장은 지정통계의 작성을 위하여 필요하다고 인정되는 경우에는 개인이나 법인 또는 단체 등에 관계 자료의 제출을 명할 수 있다. |
| 제26조<br>실질조사 | ①통계의 작성에 관한 사무에 종사하는 자는 통계의 작성을 위한 조사 또는 확인을 위하여 제18조에 따라 통계청장의 승인을 받은 사항에 관하여 관계인에게 관계 자료의 제출을 요구하거나 질문을 할 수 있다. |
| 제30조<br>통계자료의 제공 | ①통계작성기관의 장은 통계의 작성을 위하여 필요한 경우에는 다른 통계작성기관에 통계자료의 제공을 요청할 수 있다. 이 경우 요청을 받은 통계작성기관의 장은 특별한 사유가 없는 한 이에 응하여야 한다. |
| 제31조<br>통계자료의 이용 | ①특정의 대상에 관한 수량적 정보를 작성하거나 학술연구를 위한 목적으로 통계자료를 이용하고자 하는 자는 대통령령으로 정하는 바에 따라 통계작성기관의 장에게 통계자료의 제공을 신청할 수 있다. |

출처: 국가법령정보센터

해외의 경우도 통계작성을 위해서는 개인정보보호하에 활용할 수 있음을 예외적으로 규정하고 있다. EU의 GDPR에서는 역사, 통계, 과학연구 목적으로 개인정보 처리 가능함을 명시하였으며, 영국의 자료보호법(Data Protection Act), 네덜란드의 개인정보보호법(Personal Data Protection Act), 캐나다의 개인정보보호법(Personal Information Protection and Electronic Documents Act) 등 많은 나라의 개인정보보호법에서 통계 등 연구목적의 경우 개인정보를 비식별화하여 처리 및 활용할 수 있도록 규정하고 있다(부록 참조).

특히, 영국의 경우 데이터의 연계·공유를 위해 별도의 행정자료연구센터(ADRC: Administrative Data Research Center)를 설립하여 비식별화 후 연계된 행정자료에 대해 연구자들이 접근할 수 있도록 하고 있다. ADRC는 영국내에 총 4개의 센터[12]가 존재하며, 영국통계청(ONS: Office for National Statistics)의 감독을 받는다. ADRC의 데이터 연계 절차와 방법은 우선 개별자료 보유자는 고유참조번호와 개인식별 정보가 포함된 자료를 제3의 기관의 데이터 관리자에게 제공하고 데이터관리자는 제공받은 자료에서 개인식별 정보를 삭제하고 ADRC에 송부한다. 이후 ADRC는 고유참조번호를 이용하여 자료를 연계하고 고유참조번호를 삭제 후 연구자들이 보안구역에서 활용할 수 있도록 자료를 제공한다.


〈그림 3.1〉 영국의 ADRC의 데이터 연계 절차 및 방법

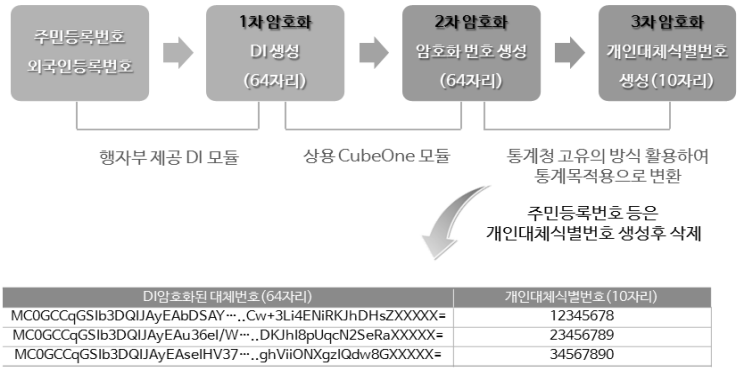
### 3.2. 통계청의 비식별화 방법 및 보안조치

통계청에서는 관리적, 물리적, 기술적 보안이 구비된 별도의 독립된 공간에서 주민등록번호, 외국인등록번호 등 개인정보를 3차에 걸친 암호화방법을 사용하여 비식별화하고, 내부망인 행정자료관리시스템의 DB에 수록하여 관리하고 있다.

비식별화 단계를 살펴보면 먼저, 행자부가 제공하는 DI 모듈을 사용하여 단방향 암호화(One-way Encryption)하여 일차암호화한다. 이 방법은 일정규칙의 알고리즘을 통해 암호화하여 64비트의 랜덤한 번호를 생성하고, 복원 key를 삭제하여 복호화를

12) 잉글랜드, 북아일랜드, 스코틀랜드, 웨일스

불가능하게 함으로써 개인정보의 식별성을 완전히 제거한다. 일차암호화된 번호를 일반 상용 암호모듈인 CubeOne을 이용하여 64바이트의 이차암호번호를 생성하고 최종적으로 통계청 고유의 암호화 방법을 이용하여 10바이트의 개인대체식별번호를 생성한다. 생성된 개인대체식별번호는 자료간 연계를 위한 연계 key로서 활용된다.〈그림 3.2〉




〈그림 3.2〉 통계청의 비식별화 단계

통계청은 통계법 및 내부규정을 규정하여 철저한 개인정보보호보조치를 실시하고, 체계적으로 개인정보를 관리하여 통계를 작성하고 있다. 통계법 제33조, 제34조, 제39조 등에 따라 통계법에서 수집된 자료에 대한 통계청 직원들의 비밀보호 의무와 위반 시 벌금, 과태료 등 처벌을 규정하고 있다. 또한, 행정자료 뿐만 아니라 빅데이터를 포함하는 자료의 입수·보관·활용 전 과정에 대한 개인정보보호를 위해 「행정자료의 정보보호를 위한 운영규정」을 제정하였다.[13)]

또한, 통계청은 시스템적인 보안을 위해 일반망(인터넷)과 분리된 업무전용망을 설치하고, 원격분석시스템〈그림 3.3〉을 도입하여 서버에서 분석 후 결과만 활용토록 하고 있다. 즉, 행정자료통합관리시스템에 통계생산시스템을 구축하여 통계 작성·분석을 시스템 내에서 수행하고, 원자료의 다운로드를 금지하고 있다. 또한, 입수된 자료는 서버DB에 저장 후 자료접근 제한 및 접근기록을 보존하고, 행자부 사이버안전센터 등을 통해 실시간 감시하고 있다.

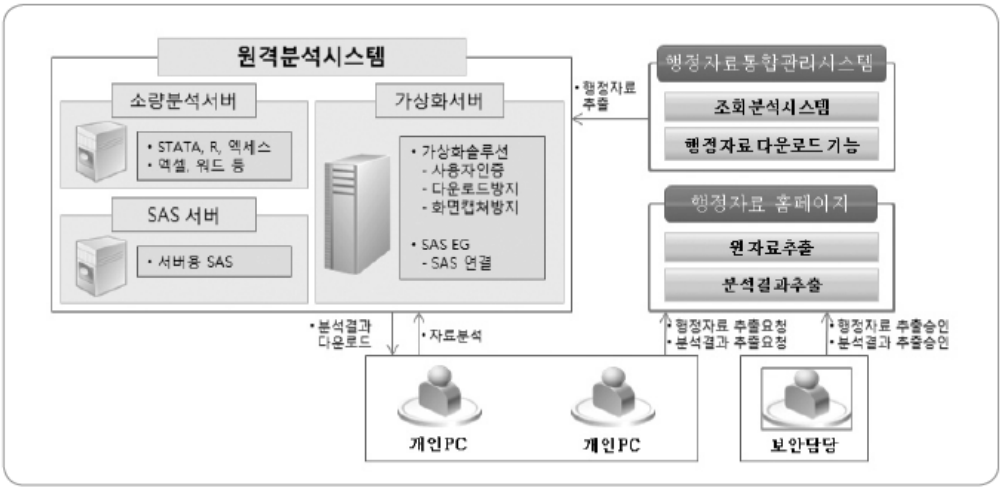
〈표 3.2〉 통계법의 개인정보보호 관련 법적 규정

| 조항 | 내 용 |
|---|---|
| 제33조<br>비밀의 보호 | ①통계의 작성과정에서 알려진 사항으로서 개인이나 법인 또는 단체 등의 비밀에 속하는 사항은 보호되어야 한다.<br>②통계의 작성을 위하여 수집된 개인이나 법인 또는 단체 등의 비밀에 속하는 자료는 통계작성 외의 목적으로 사용되어서는 아니 된다. |

13) 2009년 1월 제정하였으며 2015년 11월 6차 개정하여 행정자료와 민간자료간 연계 분석을 위한 DB구축 시 또는 민간자료 제공기관이 요청한 경우에는 별도의 독립된 서버(또는 PC)에 구축하도록 하였다.

| 조항 | 내 용 |
|---|---|
| 제34조<br>통계종사자 등의<br>의무 | 통계종사자, 통계종사자이었던 자 또는 통계작성기관으로부터 통계 작성업무의 전부 또는 일부를 위탁받아 그 업무에 종사하거나 종사하였던 자는 직무상 알게 된 사항을 업무 외의 목적으로 사용하거나 다른 자에게 제공하여서는 아니 된다. |
| 제39조<br>벌칙 | 다음 각 호의 어느 하나에 해당하는 자는 3년 이하의 징역 또는 1천만원 이하의 벌금에 처한다.<br>1. 통계의 작성을 목적으로 수집되거나 제공(제31조제2항에 따른 제공을 포함한다)을 받은 개인이나 법인 또는 단체 등의 비밀에 속하는 사항을 그 목적 외의 용도로 사용하거나 이를 다른 자에게 제공한 자<br>2. 통계의 작성을 목적으로 수집되거나 제공(제31조제2항에 따른 제공을 포함한다)을 받은 개인이나 법인 또는 단체 등의 비밀에 속하는 사항을 속임수나 그 밖의 부정한 방법으로 열람하거나 제공받은 자 |
| 제41조<br>과태료 | ③ 다음 각 호의 어느 하나에 해당하는 자에게는 100만원 이하의 과태료를 부과한다.<br>1. 제24조제4항을 위반하여 공공기관으로부터 제공받은 행정자료(비밀에 속하는 사항을 제외한다)를 제공받은 목적 외의 목적으로 사용하거나 다른 자에게 제공한 자 |

* 출처 : 국가법령정보센터




〈그림 3.3〉 통계청의 원격분석시스템 구성도

현재, 통계청은 통계청이 보유한 인구·가구 및 사업체 모집단 등 공공빅데이터와 신용정보, 신용카드 등 민간의 빅데이터간 연계 시범사례를 구축 중에 있다. 이에 따라 최근 통계청은 민간신용평가기관의 부채정보와 통계청의 인구센서스, 인구동향 자료 등과 연계한 신혼부부 부채 DB를 구축하였다. DB 구축은 통계법, 개인정보보호법 및 신용정보법을 근거로 통계청 고유의 방식에 따라 비식별화된 민간의 자료를 제공 받

고 철저한 정보보안조치 하에서 추진되었다. 신혼부부의 경우 국가의 주택 수요의 주요 수요층이자 출산과 관련한 주된 정책대상자로서 이들의 정확한 부채 현황 파악이 중요하다. 하지만, 신혼부부의 출산, 부채 등 관련 통계가 없어 주택구입 등 결혼준비 비용 부담으로 인한 만혼, 출산율 감소의 대책을 수립하기 위한 통계가 그동안 부족하였다. 따라서, 신혼부부 부채 DB는 공공 및 민간 빅데이터간 연계한 최초의 사례로서 지역별·결혼연차별 신혼부부 가구의 부채 및 주택구입 현황, 출산자녀수 변동, 부채상환 능력 평가 등 다양한 분석을 통해 저출산 관련 정책 수립·집행에 활용 가능하므로 향후 유용한 정책 기초자료가 될 것이다.

## 4. 결론

정보통신의 발달과 함께 정보의 자유로운 흐름속에 개인정보는 반드시 보호되어야 하지만, 빅데이터를 활용하여 새로운 부가가치를 창출하고 가치있는 정보를 제공하는 것은 반드시 필요하다. 하지만, 개인정보보호와 빅데이터 활용간의 상충으로 인해 현재 민간 등에서는 빅데이터를 활발하게 활용하지 못하고 있다.

하지만, 통계청은 국가통계작성기관으로서 통계법에 따라 비식별화된 개인정보를 수집하여 철저한 개인정보보호 조치하에서 통계를 작성하여 왔다. 즉, 개인정보보호법과 통계법에 따라 빅데이터 활성화의 가장 큰 제약인 개인정보 활용 논란에서 통계청은 예외로 적용되고 있다. 그러므로 통계청은 개인정보보호법 등 법률을 개정하지 않고도 현재의 법체계하에서 민간빅데이터를 활용하여 가치있는 정보를 제공하여 빅데이터 활성화를 추진 할 수 있는 유일한 기관으로서, 빅데이터를 활용하여 새롭고 가치있는 정보를 제공할 수 있다. 다만, 개인정보보호법은 일반법으로서 개인정보보호에 관하여는 다른 법률에 규정이 있을시 개별법이 우선 적용되고, 개별법인 통계법에서 민간자료 활용에 대한 명확한 규정이 없어 엄격한 유권해석시는 민간빅데이터 활용에 어려움이 있을 수 있다.

따라서, 통계청은 개인정보를 보호하고 빅데이터를 연계·융합하여 다양한 수요자 맞춤의 통계적 정보를 제공하고 데이터의 잠재 가치를 최대한 활용할 수 있도록 공공데이터 및 민간데이터 거버넌스 구축을 위한 노력을 지속하여야 한다. 예를 들어, 영국의 ADRC와 같은 데이터 연계센터를 구축하여 통계자료, 행정자료 및 민간자료간 연계를 통한 다양한 데이터를 공유하고 제공하는 것이 필요하다. 또한, 통계법을 개정하여 민간자료 활용할 수 있음을 명확히 규정하여 통계적 목적의 빅데이터 활성화를 추진하여야 할 것이다.

## 〈부록〉

□ 통계목적의 개인정보 활용 가능 관련 해외 법률

| 국가 | 관련 법(규정) | 주요 내용 |
|---|---|---|
| EU | General Data Protection Regulation | Chap Ⅸ(특정 자료처리 관련 규정), Art 83(역사, 통계, 과학연구 목적 처리)<br>1. 다음의 경우 역사, 통계, 과학연구 목적으로 개인정보 처리 가능<br>(a) 정보주체가 식별되지 않으면 자료처리가 불가한 경우<br>(b) 정보주체의 식별정보를 다른 정보와 분리하여 보관해도 처리를 할 수 있는 경우<br>2. 역사, 통계, 과학연구를 수행하는 기관은 다음의 경우 개인 정보를 공표 가능<br>(a) 7조에 따라 정보주체의 동의를 받은 경우<br>(b) 정보주체의 이익, 기본권, 자유보다 역사, 통계, 과학연구로 인한 이익이 더 큰 경우로써, 연구 결과를 제시하거나 연구를 용이하게 하기 위해 필요한 경우<br>(c) 정보주체가 이미 해당 정보를 공개한 경우 |
| 영국 | Data Protection Act | Chap 29, Part Ⅳ(면제), 33(연구, 역사 및 통계)<br>(1) “연구목적”은 통계 및 역사적 목적을 포함<br>“관련 조건”은 개인정보 처리에 관한 다음 조건을 의미함<br>(a) 특정인에 대한 조치나 결정을 지원하기 위해 처리되지 않음<br>(b) 정보주체에게 중대한 손해 및 고통을 야기하는 방법으로 처리되지 않음<br>(4) 연구목적에 한해 처리되는 개인정보는 다음의 경우 Section7로부터 면제됨<br>(a) 관련 조건을 준수하여 처리되는 경우<br>(b) 연구 결과가 정보주체를 식별할 수 있는 형태로 공개되지 않는 경우 |
| 핀란드 | Personal Data Act | Chap3. Sec12. Sub6.<br>역사적, 과학적 또는 통계 조사 목적으로는 민감정보 처리 금지 예외 |
| | | Chap3. Sec13. Sub1.<br>역사적, 과학적 또는 통계 조사 목적으로는 개인식별정보 처리 가능 |
| | | Chap4. Sec15.<br>개인정보는 다음의 경우 통계목적으로 처리 가능 |

| 국가 | 관련 법(규정) | 주요 내용 |
|---|---|---|
| | | 1. 개인정보 없이는 통계를 작성할 수 없을 때<br>2. 통계 작성이 담당자의 소관 업무일 때<br>3. 통계목적으로만 사용되고 공식 통계 외에는 개인이 식별되지 않는 형태로 공개될 때 |
| 덴마크 | Data Confidentiality Policy at Statistics Denmark | 1.1. 개인정보 처리법에 의하면 덴마크 통계청 직원만이 개인식별정보에 직접적 접근 권한을 가짐 |
| 네덜란드 | Personal Data Protection Act | Chap2. Sec2. Art23<br>Art16과 관련한 개인정보 처리 금지 규정은 다음의 과학적 연구 및 통계 목적인 경우 적용되지 않음<br>1. 연구가 공익 목적인 경우<br>2. 연구 및 통계작성을 위해 필요한 경우<br>3. 동의를 구하는게 불가능하거나 너무 많은 노력이 요구되는 경우<br>4. 개인 프라이버시를 충분히 보증할 수 있을 때 |
| 스웨덴 | Information on the Personal Data Act | 7. 개인정보 처리 기본 조건<br>역사, 통계 및 과학목적의 개인정보 처리는 특정 규정을 적용함. 만약 이러한 목적으로 처리되는 개인정보가 나중에 처리된다면, 자료를 원래 수집했을 때의 원래 목적과 양립하지 않는 것으로 간주함. 역사, 통계 및 과학 목적인 경우 자료는 더 오래 보유 가능하나 필요이상으로 길게는 안됨. |
| 캐나다 | Personal Information Protection and Electronic Documents Act | Part 1(민간영역의 개인정보 보호), Division 1(개인정보 보호), 7(인지 및 동의 없는 수집), (2) 인지 및 동의 없는 사용. 다음의 경우 개인의 인지 및 동의 없이 기관은 개인정보를 사용할 수 있음<br>(c) 통계, 학술연구, 연구 목적으로 개인정보 없이 목적 달성이 어렵고, 자료비밀보호가 보장되는 방법으로 사용되며, 동의를 구하는 것이 실현불가능하여 기관에서 위원장에게 개인정보 사용전에 사용을 알리는 경우 |

## 참고문헌

방송통신위원회 (2014). 빅데이터 개인정보보호 가이드라인

손영화 (2014). 빅데이터 시대의 개인정보보호방안 〈기업법연구〉, 28(3), 355-393.

한국정보화진흥원 (2015). 빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서

한은영 (2015). 일본 개인정보보호법의 개정 내용 및 평가, 〈정보통신방송정책〉, 27(17), 41-51.

Health Human Services (2012). Guidance Regarding Methods of De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule.

# De-identification and Privacy protection for Statistical Purpose

Hae Ryun Kim[1)]

## Abstract

Although the value of the use of personal information increases in big data era, it is difficult for private sector to use personal information under the Privacy Act and other Act. However, Statistics Korea can use personal information because de-identified personal information is allowed to be used for statistical research purpose. In addition, personal information collected by Statistics Law is an exception under the Privacy Law. Therefore Statistic Korea is able to provide statistical information through the linkage between public-big data and private-big data, and should build big data governance in order to promote the utilization of big data with privacy protection.

Key words : big data, data linkage, privacy protection, de-identification

1) Deputy Director, Big data & Statistics Division, Statistics Korea, 949 Dunsan-Dong, Seo-Gu, Daejeon 302-120. Korea. E-mail: khr@korea.kr