

# 의료분야 빅데이터 활용을 위한 개인정보 비식별화 규정 현황과 과제

*De-identification Policy of Personal Information  
and Tasks on Healthcare Big Data*



정영철 한국보건사회연구원 연구위원

정부3.0 국정패러다임에 맞추어 빅데이터 발굴 및 활용을 위한 정부정책이 활발히 추진되고 있는 가운데, '빅데이터 활용'과 '개인정보 보호'는 양면성을 지니고 있다. 이에 개인정보 침해위험을 최소화할 수 있는 방안으로 '개인정보 비식별화' 조치는 매우 중요한 수단으로 자리잡아 가고 있으며 이에 대한 가이드라인, 지침, 사례집 등이 발표되었다. 그러나 빅데이터 활용영역 중에서도 기대전망이 나날이 증가하고 있는 의료분야에서 이를 범용적으로 적용하기 위해서는 의료정보, 의료정보 종류, 개인의료정보 등과 같은 용어의 정의 및 범위, 구분 등의 명확화, 개인의료정보 중에서도 해당정보 자체로 식별성을 지닌 정보와 타 정보와 쉽게 결합하여 식별가능한 정보에 대한 실 사용항목을 예로 한 구분이 필요하며 이에 대해 다양한 주제간 합의 또한 선행되어야 한다.

## 1. 들어가며

최근 빅데이터에 대한 전세계적인 관심과 기대감이 수위를 차지하고 있는 가운데, 우리나라는 제 18대 정부의 정부3.0 국정패러다임과 맞물려 다양한 분야에서 정보자원, 빅데이터를 발굴하고 이를 활용하기 위한 정부정책 등을 활발하게 진행하고 있다. 이는 우리나라가 구글 트렌드서비스([www.google.com/trends](http://www.google.com/trends))에서 관심도가 높은 나라<sup>1)</sup>에 속하는 것을 보아서도 알 수 있다. 특히 빅데이터

활용에 있어 초석이라 할 수 있는 공공데이터에 대한 민간활용을 촉진하기 위해 정부에서는 2013년 『공공데이터의 제공 및 이용 활성화에 관한 법률』을 제정·시행하여 2015년 7월 현재 공공데이터 포털([www.data.go.kr](http://www.data.go.kr))에는 14,278건의 공공데이터가 등록되어 있고(표 1), 2015년 OECD가 조사한 공공데이터 개방지수<sup>2)</sup>에서는 0.98점으로 1위를 차지하는 등<sup>3)</sup> 긍정적 성과를 나타내고 있다.

그러나 이러한 빅데이터를 논할 때 자주 등장하는 이슈 중 하나는 '개인정보 보호' 문제로, 윤상

1) 구글트렌드([www.google.co.kr/trends/](http://www.google.co.kr/trends/))에서 'big data' 라는 검색어로 2004년 이후 현재까지 트렌드 검색을 시도해보면 결과 인도, 싱가포르, 우리나라 순으로 지역적 관심도가 높은 것으로 나타남(2015.7.10. 인출).

2) 진단항목은 가용성, 접근성, 정부지원등 3개분야 19개 임.

3) 행정자치부(2015.7.7.). 대한민국 OECD가 발표한 공공데이터 개방지수에서 1위 달성, 행정자치부 보도자료.

표 1. 우리나라 공공데이터 개방현황

(단위: 건)

분류	계	파일데이터	오픈API	데이터시각화
계	14,405	12,415	1,739	251
교육	829	684	127	18
국토관리	879	786	84	9
공공행정	2,664	2,432	184	48
재정금융	433	382	47	4
산업고용	1,417	1,148	222	47
사회복지	1,068	992	58	18
식품건강	481	386	87	8
문화관광	1,585	1,309	260	16
보건의료	999	898	78	23
재난안전	560	509	38	13
교통물류	1,020	835	172	13
환경기상	933	850	71	12
과학기술	619	472	135	12
농축수산	663	546	113	4
통일외교안보	211	154	51	6
법률	44	32	12	0

자료: 공공데이터포털(www.data.go.kr), [2015.7.10. 접속].

오<sup>4)</sup>는 정부정책 추진시 가장 우선적으로 대응방안을 마련해야 하는 위험 중 하나로 개인정보 침해를 제시하고 있으며 우리나라 빅데이터 정책 추진현황<sup>5)</sup>에서도 ‘개인정보 보호’ 문제를 빅데이터 활용에 있어 제약요인으로 제시하는 등 ‘빅데이터 활용’과 ‘개인정보 보호’는 양면성을 지니고 있다.

개인정보를 보호하는 목적은 개인정보 유출·오용·남용시 해당 개인에게 예상되는 신체적, 정신적, 사회적, 재정적 피해, 그리고 기업 혹은 국가차원의 심각한 피해를 예방하고 경감시

키기 위한 것으로 이인호<sup>6)</sup>는 개인정보(personal information)를 ‘보호’한다는 것은 개인의 사적정보(private information)에 대한 비밀 유지를 위한 프라이버시 보호와는 달리 개인정보를 무조건적으로 제한하기보다는 ‘안전하고 투명한 처리’를 통한 즉 ‘보호’와 ‘이용’의 두가지 가치를 적절히 고려하는 것임을 강조하고 있다. 그러므로 빅데이터에 있어 이와 같은 개인정보의 ‘보호’와 ‘이용’(혹은 활용)의 적절한 균형점을 찾는 것이 중요하다고 하겠다.

4) 윤상오(2013.6). 빅데이터의 위험유형 분류에 관한 연구, 한국지역정보학회지 16(2), pp93~122, 한국지역정보학회.

5) 관계부처 합동(2014). 유능한 정부구현을 위한 빅데이터 활용 확대방안.

6) 이인호(2015). 개인정보 보호법 상의 ‘개인정보’ 개념에 대한 해석론, 정보법학 19(1), pp59~87, 한국정보법학회.

한편, 빅데이터는 분야별 활용성이 다양하여 재무, 교통, 에너지, 의료, 보안, 영업 등 많은 분야에서의 성공적인 사례들이 소개되고 있는 가운데<sup>7)</sup> 미국에서는 의료(Health care)분야에서 연간 3,000억달러(스페인 연간 의료비의 2배) 이상의 새로운 가치창출을 전망하고 있는 등<sup>8)</sup> 빅데이터 도입, 활용이 매우 기대되는 영역의 하나로 의료분야를 꼽고 있다<sup>9)</sup>. 우리나라에서도 분당서울대병원, 건강보험심사평가원, 건강보험공단, 국립중앙인체자원은행, 의약정보원 등 주로 공공기관에서 환자의 진료 및 처방정보, 인체자원에 의한 유전정보, 의약품정보와 같은 의료정보 빅데이터 활용사례가 소개되고 있는 가운데 질병치료 및 질병예방, 의약품오남용 방지, 의료서비스의 질 향상, 의료비용 절감 등과 같은 향후 가치에 대한 기대가 더욱 더 커지고 있다<sup>10)</sup>.

그러나 이와 같은 의료분야는 빅데이터에 대한 전망과 기대가 높은 반면, 그 어느분야보다도 개인의 민감한 정보를 많이 담고 있어 개인정보 침해, 개인정보 보호에 대한 논의와 대책 또한 더욱더 필요한 분야라 할 수 있다. 더욱이 빅데이터는 여러 영역에 걸친 데이터 융합으로 새로운 가치창출의 의미를 지니고 있어, 월스트리트 저널에서는<sup>11)</sup> 개인의료정보가 의료서비스 영역 뿐 아니라 신용카드정보, 쇼핑정보 등 개인의 일상적인 생활정보에서도 정교한 데이터 알고리즘에 의해 원치 않게 수

집·이용되고 있음에 우려를 나타내기도 한다.

한편, 많은 나라의 개인정보 보호에 관한 법 체계에서 보호대상이 되는 ‘개인정보’는 ‘개인이 식별되는(identified) 또는 식별가능성이 있는(identifiable)’ 즉 개인식별성(identification) 정보로써<sup>12)</sup>, 이러한 특정 개인을 식별할 수 없도록 하는 ‘비식별화’ 조치는 개인정보 침해위험을 최소화할 있는 방안의 하나로 알려져 있다.

그러므로 본 고에서는 빅데이터에 대한 필요와 수요가 증가하고 있는 환경에서 그 중에서도 가장 주목받고 있는 영역 중 하나인 의료영역에서의 빅데이터에 대한 안전하고도 활성화된 이용을 위해 개인정보 보호방안의 하나인 비식별화 조치에 대한 규정 현황과악과 더불어 이에 기반하여 의료영역에서의 시사점을 도출해내고자 한다.

## 2. 개념 고찰

본 고에서의 논의 대상인 개인의료정보는 엄격히 구분해보면 ‘개인정보’와 ‘의료정보’로 나누어 생각해볼 수 있다. 먼저 우리나라 『개인정보 보호법』에서는 보호대상이 되는 ‘개인정보’에 대해 다른 많은 나라와 유사하게, 살아있는 개인에 대해 해당 개인을 식별할 수 있는 정보(개인정보 보호법 제2조), 즉 ‘개인식별성(identification)’에 대한 의미

7) 한국정보화진흥원(2012), 빅데이터로 진화하는 세상 -Big Data 글로벌 선진사례 -.

8) McKinsey Global Institute(2011.06.), Big data: The next frontier for innovation, competition, and productivity.

9) 안창원, 황승구(2012.6), 빅 데이터 기술과 주요이슈, *정보과학회지* 30권6호, pp10~17, 한국정보과학회.

10) 김관형 외(2013), 의료 IT와 빅데이터의 융합, *한국컴퓨터정보학회지* 21(3), pp17~26, 한국컴퓨터정보학회.; 이지혜, 제미경, 조명지, 손현석(2014), 보건의료분야의 빅데이터 활용 동향, *한국통신학회지* 32(11), pp63~75, 한국통신학회.

11) WAJ Korea(2013.12.19.), 정교해진 의료데이터 마이닝, *사생활 침해논란*.

12) 이인호(2015), 개인정보 보호법 상의 ‘개인정보’ 개념에 대한 해석론, *정보법학* 19(1), pp59~87, 한국정보법학회.

를 부여하고 있으며<sup>13)</sup> 이와 같은 ‘개인정보’는 시대와 기술, 인식 등이 변화하면서 대상과 범위 또한 변화하기도 한다<sup>14)</sup>. ‘개인정보’는 일반 인적정보, 교육 및 훈련정보, 병역정보, 소득정보, 신용정보, 고용정보, 의료정보 등으로 구분되기도 하며<sup>15)</sup>, 개인정보 중에서도 이름, 전화번호, 고유식별정보(주민등록번호, 운전면허번호, 외국인등록번호, 여권번호), 생체정보(지문, 홍채, DNA정보 등), 기관 및 단체등의 이용자계정(등록번호, 계좌번호 등) 등은 그 자체로 개인을 식별할 수 있는 정보로 구분하기도 한다<sup>16)</sup>.

다음으로 ‘의료정보’는 많은 연구에서 보건의료정보, 의료정보, 의학정보, 진료정보, 건강정보, 헬스케어정보 등 다양한 용어로 명확한 구분없이 사용되고 있으며 심지어 동일연구에서도 다양한 용어가 혼재되어 있는 것을 쉽게 발견할 수 있다. 그 중에서 해당분야의 개인정보보호와 관련된 문헌들을 일부 살펴보면 백윤철<sup>17)</sup>은 ‘의료정보’를 정보의 용도에 따라 환자의 기본인적정보, 건강보험정보, 진료정보, 진료관리정보, 요약정보, 사망기록정보 등으로, 생성시점에 따라 환자가 직접작성 혹은 객관적 사실에 의한 1차의료정보와 이러한 1차의료정보를 기초로 생성된 2차의료정보(가공의료정보)로 구분하기도 하며 유전자정보의 중요성

을 제시하기도 하였다. 전영주<sup>18)</sup>는 ‘의료정보’를 환자에 의해 작성되는 주관정보, 검사 및 진료 등에 의한 객관정보, 그리고 의료인의 전문성에 기초하여 작성된 가치판단정보로 분류하고 있다. 그 외 정규원(2002), 박윤희(2005), 장석천(2007), 박지용(2012) 등<sup>19)</sup>도 공통적으로 의료현장에서의 관련정보들을 정의하고 있다. 한편 보건의료기본법 제3조에서는 ‘보건의료정보’를 국민의 건강을 보호·증진하기 위해 국가·지방자치단체·보건의료기관 또는 보건의료인 등이 행하는 모든 활동과 관련된 지식 및 자료라고 정의하여 앞에서 언급한 의료정보와는 차이를 보이고 있으며 의료정보, 보건의료정보에 속하는 정보의 종류가 매우 다양하다는 것을 알 수 있다.

이와 같은 용어에 대한 정의, 차이, 범위 등을 명확히 규정하는 것은 매우 중요하고도 필요한 작업이나 이는 추후 별도 연구에서 보다 심도깊은 논의가 이루어질 수 있기를 기대해 보며, 본 고에서는 임의적으로 ‘의료정보’라는 용어를 채택하여 사용하고, 기존 문헌들을 참고하여 ‘개인의 질병 치료·관리 및 예방을 위한 정보, 그리고 이에 기반하여 연구·분석된 정보’로서 포괄적인 의미로 이해하기로 한다.

이처럼 ‘개인정보’, ‘의료정보’에 대해 살펴본 결

13) 이인호(2015)는 이와 같은 ‘개인정보’ 해당여부 및 식별가능성을 판단함에 있어 지나친 의미확장과 적용을 우려하기도 함.

14) 개인정보보호위원회(2012). 2012 개인정보보호 연차보고서.

15) 개인정보보호 종합지원포털, www.privacy.go.kr [2015.06.25. 인출].

16) 안전행정부 개인정보보호과(2013.9.). 공공정보 개방·공유에 따른 개인정보보호지침.

17) 백윤철(2005). 우리나라에서 의료정보와 개인정보보호, *헌법학연구* 11(1), pp395~442, 한국헌법학회.

18) 전영주(2006). 의료정보와 개인정보보호, *법학연구* 23, pp521~540, 한국법학회.

19) 정규원(2002). 의료정보의 활용 및 보호, *정보법학* 6(1), pp1~25, 한국정보법학회.; 박윤희(2005). 진료정보와 국민 사생활 보호정책의 현황과 전망, *한국의료법학회지* 13(1), pp87~98, 한국의료법학회.; 장석천(2007). 의료정보보호와 민사법적 문제, *법학연구* 28, pp159~180, 한국법학회.; 박지용(2012). 환자의 프라이버시 및 정보보호의 법적 근거 고찰, *한국의료법학회지* 20(2), pp163~190, 한국의료법학회.

과, 본 고에서 살펴보고자 하는 '개인의료정보'는 '의료정보' 중에서도 개인식별성을 지닌 정보에 초점이 맞추어질 수 있겠다.

### 3. 개인정보에 대한 비식별화 규정

앞에서 언급하였듯이 '개인정보'란 특정 개인을 식별할 수 있는, 더 나아가 정보들을 결합하고 조합하여 특정 개인을 구별할 수 있는 특성을 지니고 있으므로 이러한 '개인정보'에서 식별성을 제거한 즉 '비식별화'하는 것이 개인정보 침해위험을 줄일 수 있는 방안으로 부각되고 있다. 이에 비식별화 규정 중 가장 대표적인 『HIPAA 프라이버시 규칙(HIPAA Privacy Rule)』에서의 개인의료정보에 대한 비식별화 규정과, 최근 주요 외국 입법례를 기반으로 하여 우리 정부에서 제공하고 있는 가이드라인, 지침, 사례집 등을 살펴보고자 한다.

#### 1) HIPAA 프라이버시 규칙(HIPAA Privacy Rule)

먼저 미국의 『HIPAA(Health Insurance Portability and Accountability Act, 건강보험 이전과 책임에 관한 법)』는 1996년 제정된 연방법으로 의료관련 행정 및 금융자료의 전자교환을 표준화하는 법률이다. 이러한 법률에 따라 개인의료정보 보호를 위해 마련된 규칙이 『HIPAA Privacy Rule』로 의료분야 개인정보보호에 대한 상세한 규정을 담고 있다.

『HIPAA Privacy Rule』에서는 해당 법에 의해 보호를 받는 의료정보(PHI, Protected Health

Information)를 규정하고 있을 뿐 아니라, 보호 대상에서 제외하여 의료정보의 자유로운 이용 및 제공을 허용하는 전면적 규율면제와 부분적인 규율면제에 대해 설정하고 있다. 먼저 보호를 받는 의료정보(PHI)는 '개인을 식별할 수 있는(individually identifiable) 정보' 혹은 '개인을 식별할 수 있는 합리적인 근거가 있는(reasonable basis) 정보'이며 이와 반하여 개인을 식별할 수 없는 즉, 비식별화된 의료정보(de-identified health information)는 전면적 규율면제로 누구나 자유롭게 이용 및 제공할 수 있음을 명시하여<sup>20)</sup> 의료정보 이용에 있어 개인정보 비식별화가 중요 수단임을 명확히 하고 있다.

한편 전면적 규율면제를 위한 비식별화 방식에는 '전문가 결정방식(expert determination method)'과 18가지 식별자를 제거한 '피난처방식(safe harbor method)'이 있다. 미국 HHS의 인권보호국(OCR, Office for Civil Rights)에서 2012년 11월 발표한 가이드에서는 '전문가 결정방식'의 전문가는 통계적, 과학적 원칙과 방법에 대한 적절한 지식과 경험을 보유한 자로서, 식별을 통한 위험을 최소화하는 원칙과 방법을 적용하며 그러한 결정에 이르게 한 분석방법과 결과를 문서화할 것을 제시하고 있다. 또한 '피난처방식'의 ① 이름(Names), ②주소정보(우편번호 등), ③개인 과 직접 관련된 날짜정보(생일, 자격취득일 등), ④전화번호(Telephone numbers), ⑤팩스번호(Fax numbers), ⑥이메일주소(Email addresses), ⑦사회보장번호(Social security numbers), ⑧의료기록번호(Medical record numbers), ⑨건강보험번호

20) 이인호(2015). 개인정보 보호법 상의 '개인정보' 개념에 대한 해석론, 정보법학 19(1), pp59~87, 한국정보법학회.

호(Health plan beneficiary numbers), ⑩계좌번호(Account numbers), ⑪자격취득번호(Certificate/license numbers), ⑫자동차번호(Vehicle identifiers and serial numbers), ⑬각종 장비 식별번호(Device identifiers and serial numbers), ⑭URL 정보(Web Universal Resource Locators), ⑮IP 주소(Internet Protocol addresses), ⑯생체정보(Biometric identifiers), ⑰얼굴사진 등(Full-face photographs etc.), ⑱기타 고유특징 등(Any other

unique identifying number) 18가지 항목을 제거하는데 있어 부가적인 설명을 제공하고 있다<sup>21)</sup>.

다음으로 부분적인 규율면제에 있어서는 의료정보의 '제한적 데이터집합(limited data set)'을 사용하는 경우로서, 앞의 18가지 식별자 중 16개의 직접적 식별자(direct identifiers)는 제거한 상태에서 허용되는 데이터 이용 및 제공목적, 데이터 이용 및 수령자의 범위 제한, 데이터 식별 복원불가 등의 내용을 담아 데이터 제공자와 제공받는자 간

표 2. 개인식별요소 제거 처리기법 및 상세규칙

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

자료: 미래창조과학부, 한국정보화진흥원(2014.05.), 빅데이터 활용을 위한 개인정보 비식별화 사례집;  
미래창조과학부, 한국정보화진흥원(2015.05.), 빅데이터 활용을 위한 개인정보 비식별화 기술활용 안내서;  
방송통신위원회(2014.12.), 빅데이터 개인정보보호 가이드라인;  
안전행정부(2013.09.), 공공정보 개방·공유에 따른 개인정보보호지침;  
행정자치부, 한국정보화진흥원(2014.12.), 개인정보 비식별화에 대한 적정성 자율평가 안내서.

21) HHS OCR(2012.11.26.), Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act(HIPAA) Privacy Rule.

데이터 이용계약(data use agreement)을 체결하여 이용할 수 있다<sup>22)</sup>.

## 2) 우리나라의 관련 규정

우리나라에서는 개인정보 비식별화에 대해 주요 외국 입법례를 기반으로 하여 작성된 가이드라인, 지침, 사례집 등을 제공하고 있다(표 2).

먼저 2013년 제정·시행된 『공공데이터의 제공 및 이용 활성화에 관한 법률』 후속조치로 작성, 발표된 『공공정보 개방·공유에 따른 개인정보보호 지침』<sup>23)</sup>에서는 공공정보 처리, 분석 시 개인정보를 비식별화하여 보호조치토록 하고 있으며 비식별화 처리원칙, 비식별화 단계별 조치사항, 비식별화 처리기법 등을 제시하고 있다. ‘비식별화란 “개인정보의 일부 또는 전부를 삭제하거나 다른 정보로 대체함으로써 다른 정보와 쉽게 결합하여서도 특정 개인을 식별하기 어렵도록 하는 일련의 조치”라 정의하고 비식별화를 위한 기법으로는 가명처리, 총계처리(또는 평균값 대체), 데이터값 삭제, 범주화, 데이터마스킹 등을 소개하고 있다.

『빅데이터 활용을 위한 개인정보 비식별화 사례집』<sup>24)</sup>에서는 빅데이터 활용단계별 비식별화 조치사항, 비식별화 처리기법, 그리고 각 기법에 대한 상세규칙 등을 제시하고 있으며(표 3) 빅데이터 활용사례별 비식별화 처리 실예를 제공하고 있다. 의료분야 실무 적용사례에서는 성명, 주민등록번호, 연령, 주소, 전화번호, 이메일주소, 외국인등록번호,

여권번호, 등록번호, 건강보험증번호, 은행계좌번호, 자격/면허번호, 차량번호, 바이오정보(지문, 얼굴, 홍채, 정맥, 음성, 필적 등), 유전자정보, 홈페이지 회원 ID, 사번, 비밀번호, 요양기관기호, 소득, 민감상병, 아이디, 진단명, 약처방날짜, 진단검사날짜, 검사수행날짜 등과 같은 항목에 대한 비식별조치 적용 내용을 제시하고 있다.

한편 『빅데이터 개인정보보호 가이드라인』<sup>25)</sup>에서는 정보통신서비스 제공자가 공개된 정보 및 이용내역정보를 처리시에는 개인정보를 보호하기 위해 비식별화 조치를 취할 것을, 또한 비식별화 조치를 한 공개정보는 이용자의 동의없이 수집·이용할 수 있음을 없음을 명시하고 있다.

『개인정보 비식별화에 대한 적정성 자율평가안 내서』<sup>26)</sup>는 개인정보 비식별화에 대한 적정성 평가를 위한 절차와 세부평가방법, 재식별 위험관리 방안 등을 주 내용으로 구성되어 있다. 이 중 ①성명(한자, 영문성명 포함), ②주소, ③고유식별정보(주민등록번호, 여권번호, 외국인등록번호, 운전면허번호), ④연월일(생일, 기념일, 사망일, 자격증 취득일등), ⑤전화번호(휴대폰번호, 집전화, 회사전화), ⑥팩스번호, ⑦전자메일, ⑧의료기록번호, ⑨건강보험번호, 복지수급자번호, ⑩계좌번호, 카드번호, ⑪각종 자격증 및 면허번호, ⑫자동차번호, ⑬각종 기기의 등록번호 및 일련번호, ⑭IP주소, Mac주소, ⑮홈페이지 URL, ⑯사진(정지, 동영상, CCTV 영상 등), ⑰신체식별정보(지문, 음성, 홍채

22) 이인호(2015), 위의 책

23) 안전행정부(2013.09.), 공공정보 개방·공유에 따른 개인정보보호지침.

24) 한국정보화진흥원(2014.05.). 빅데이터 활용을 위한 개인정보 비식별화 사례집. 미래창조과학부.

25) 방송통신위원회(2014.12.). 빅데이터 개인정보보호 가이드라인.

26) 행정자치부, 한국정보화진흥원(2014.12.). 개인정보 비식별화에 대한 적정성 자율평가 안내서.



표 3. 개인식별요소 제거 처리기법 및 상세규칙

처리기법	내용 및 상세규칙
가명처리 (Pseudonymisation)	<ul style="list-style-type: none"> <li>○ 개인식별 값을 다른 값으로 대체(대체하는 일정규칙 노출위험 방지 필요) (예) 홍길동, 35세 → 임꺽정, 30대</li> <li>① 시계열 데이터마이닝(k-익명화): 데이터 공개시 동일한 값의 데이터를 k개 이상으로 유지하는 방법</li> <li>② 부분그래프 익명화: 그래프수정을 통해 전체 그래프 중에서 특정부분의 그래프가 k개 이상 존재하게 만드는 방법</li> <li>③ 차수익명화</li> <li>④ 부분그래프+차수익명화</li> <li>⑤ 매크로기법: 민감한 셀의 행과 열 주변값을 동시에 감추는 방법</li> <li>⑥ 휴리스틱익명화: 준식별항목 값을 정해진 규칙 혹은 사람 판단에 따라 가공</li> <li>⑦ 교환: 미리 정해진 항목들을 데이터베이스 레코드와 연계하여 교환</li> </ul>
총계처리 (Aggregation) 또는 평균값대체 (Replacement)	<ul style="list-style-type: none"> <li>○ 개별 데이터값이 아닌 데이터 총 합계값 또는 평균값으로 표현 (예) 임꺽정 180cm, 홍길동 170cm, 김팔쥐 150cm → 물리학과 학생들 키의 합:500cm, 키의 평균:167cm</li> <li>① 프라이버시 모델: 일정 규칙을 통해 데이터를 가공</li> <li>② 마이크로기법: 식별항목의 기존 값을 타 값으로 대체하여 식별정보를 희석</li> </ul>
데이터값 삭제 (Data reduction)	<ul style="list-style-type: none"> <li>○ 개인식별에 중요한 값 삭제 (예) 주민등록번호 901206-1234567 → 90년대생, 남자 (예) 홍길동, 35세, 서울거주, 한국대졸업 → 35세, 서울거주</li> <li>① 식별자 제거: 개인식별항목 제거</li> <li>② 준식별자제거: 준식별항목 제거</li> </ul>
범주화 (Data suppression)	<ul style="list-style-type: none"> <li>○ 데이터값을 범주값으로 변환 (예) 홍길동, 35세 → 홍씨, 30대</li> <li>① 데이터레인지: 개인식별항목의 값을 범위로 설정</li> <li>② 랜덤라운드: 개인식별항목의 값을 임의의 수 기준으로 올림 또는 절사</li> <li>③ 제어라운드: 행과 열을 맞게 제어하여 올림 또는 절사</li> <li>④ 정점/간선 클러스터링</li> <li>⑤ 세분정도(subdivide level) 제한</li> </ul>
데이터마스킹 (data masking)	<ul style="list-style-type: none"> <li>○ 개인식별자를 보이지 않도록 처리 (예) 홍길동, 35세 → 홍**, 35세</li> <li>① 임의 잠음추가: 임의의 숫자를 더하거나 곱함</li> <li>② 공백(blank)과 대체(impute)</li> </ul>

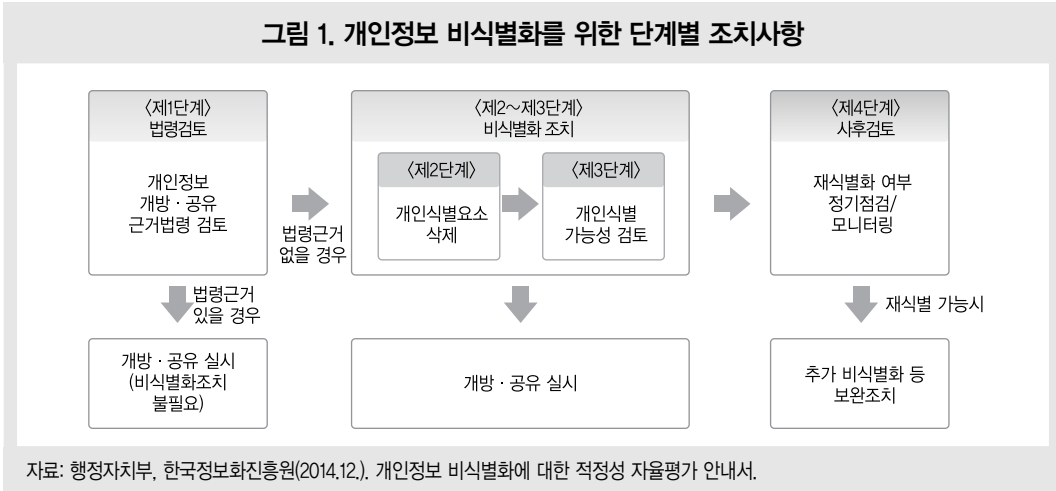
자료: 안전행정부 개인정보보호과(2013.9.), 공공정보 개방공유에 따른 개인정보보호지침;  
미래창조과학부, 한국정보화진흥원(2014.05.), 빅데이터 활용을 위한 개인정보 비식별화 사례집.

등), @기타 유일 식별번호(군번, 사업자등록번호, 식별코드 등) 등 18개 식별자를 예시로 들고 있는 바, 이는 HIPAA 프라이버시 규칙에서 제시한 18개 식별자를 고려하여 우리나라 상황에 맞게 수정하여 제시하고 있다. 또한 비식별화를 위한 단계

별 조치사항으로는 관계법령 등 검토(제1단계)하여 근거가 없을 경우에는 표 3의 기법을 활용하여 개인식별요소를 삭제(제2단계)하며, 이에 대해 통계 및 수학 등 관련분야 전문가들로 하여금 개인식별 가능성을 검토(제3단계) 한 후 데이터를 활용토



그림 1. 개인정보 비식별화를 위한 단계별 조치사항



특 하고 있다. 이후 재식별화 여부에 대한 지속적 인 정기점검/모니터링(제4단계) 단계를 두고 있다 (그림 1).

마지막으로 『빅데이터 활용을 위한 개인정보 비 식별화 기술활용안내서』<sup>26)</sup>에서는 빅데이터 활용 을 위해 각 부문별 개인식별 정보항목과 18가지

비식별화 기술활용방법, 그리고 부문별 활용사례 를 안내하고 있다. 이에 비식별화 적용대상 정보 를 해당정보 자체로 개인식별 가능한 정보와 다른 정보와 쉽게 결합하여 개인식별 가능한 정보로 나 누어 예시를 들고 있으며(표 4), 의료분야 실무 적 용사례는 앞의 『빅데이터 활용을 위한 개인정보

표 4. 비식별화 적용대상 정보

구분	주요 내용
해당정보 자체로 개인식별 가능한 정보	<ul style="list-style-type: none"> <li>○ 개인식별이 용이한 정보: 이름, 전화번호, 주소, 생년월일, 사진등</li> <li>○ 고유식별정보: 주민등록번호, 운전면허번호, 외국인등록번호, 여권번호</li> <li>○ 생체정보: 지문, 홍채, DNA 정보등</li> <li>○ 기관,단체등 이용자계정: 등록번호, 계좌번호, 이메일주소등</li> <li>○ 기타 유일한 식별번호: 군번, 사업자등록번호, 식별코드(아이디, 아이핀값) 등</li> </ul>
다른 정보와 쉽게 결합하여 개인식별 가능한 정보	<ul style="list-style-type: none"> <li>○ 개인특성: 성별, 생년, 생일, 연령, 국적, 고향, 거주지, 시군구명, 우편번호, 병역여부, 결혼여부, 종교, 취미 등</li> <li>○ 신체특성: 혈액형, 신장, 몸무게, 허리둘레, 혈압, 장애유형, 병명 진료내역등</li> <li>○ 신용특성: 신용등급, 건강보험료 납부액, 신용등급 등</li> <li>○ 경력특성: 학교명, 학과명 학년, 성적, 학력, 직급 등</li> <li>○ 전자적 특성: PC 사양, 접속일시, 방문일시, 서비스이용기록, 단말기 ID 등</li> <li>○ 가족특성: 배우자, 자녀, 부모, 가족정보, 법정 대리인 정보 등</li> <li>○ 위치특성: GPS데이터, RFID리더 접속기록, 인터넷접속기록 등</li> </ul>

자료: 미래창조과학부, 한국정보화진흥원(2015.05.), 빅데이터 활용을 위한 개인정보 비식별화 기술활용안내서.

26) 미래창조과학부, 한국정보화진흥원(2015.05.), 빅데이터 활용을 위한 개인정보 비식별화 기술활용안내서.

표 5. 우리나라 의료관련 법령기반 개인정보 취급서식 현황

관련법	개인정보 취급서식 및 주요 정보
의료법	<ul style="list-style-type: none"> <li>- 진단서(등록번호, 연번호, 환자성명 주민등록번호, 환자주소, 전화번호, 주상병, 부상병, 질병분류번호, 치료내용 등)</li> <li>- 처방전(요양기관기호, 환자성명 및 주민등록번호, 의료기관(명칭, 전화번호, 팩스번호), 질병분류기호, 의료인(성명, 면허종류, 면허번호, 처방의약품명칭 등)</li> <li>- 진료기록부(환자주소, 성명, 연락처, 주민등록번호), 주된증상, 진단결과 또는 진단명, 진료경과, 치료내용, 진료일시) 등</li> <li>- 기타</li> </ul>
의료급여법	<ul style="list-style-type: none"> <li>- 의료급여의뢰서(의료급여 의뢰종류, 보장기관기호, 보장기관명, 세대주성명, 수급권자성명 및 주민등록번호, 주소, 상병명, 상병분류기호, 진료기관 등)</li> <li>- 수급권자 자격취득통보서(보장기관기호 및 기관명, 읍면동(시설기관) 기호 및 명칭, 세대주(성명, 주민등록번호), 취득수급권자(관계, 주민등록번호, 성명) 등)</li> <li>- 의료급여증(세대주(성명, 주민등록번호, 구분, 자격취득일, 보장기관기호 및 명칭), 부양가족(성명, 주민등록번호, 관계, 자격취득일), 급여기록(성명, 진료일자, 의료급여기관명) 등)</li> <li>- 기타</li> </ul>
국민건강보험법	<ul style="list-style-type: none"> <li>- 건강보험증(성명, 진료일, 요양기관기호)</li> <li>- 요양비 지급청구서(접수번호, 접수일, 처리기간, 수진자(성명, 주민등록번호), 요양기관(명칭, 등록번호), 진료구분, 처방전발행일, 진료기간 또는 출산일, 상병명 등)</li> <li>- 건강보험료 분기납부 신청서(접수번호, 접수일, 처리기간, 납부의무자(성명, 주민등록번호, 고지번호, 주소, 전화번호), 신청내용(분기납 희망기간)</li> <li>- 기타</li> </ul>
감염병 예방 및 관리에 관한 법률	<ul style="list-style-type: none"> <li>- 감염병 발생신고서(감염병환자등 또는 사망자(성명, 주민등록번호, 전화번호, 직업, 성별, 주소), 감염병명, 발병일, 확진검사결과 등)</li> <li>- 예방접종증명서(성명, 생년월일, 성별, 주소, 접종명, 접종차수 등)</li> <li>- 진료확인서(진료자(성명, 생년월일, 주소), 질병명 또는 주요조건, 진료일수 등)</li> <li>- 기타</li> </ul>
응급의료에 관한 법률	<ul style="list-style-type: none"> <li>- 응급환자진료의뢰서(환자성명, 생년월일, 주소), 보호자(성명, 주소, 전화번호, 도착시간, 응급처치 전 환자상태 등)</li> <li>- 응급환자진료비 미수급 대지급 청구서(청구기관(요양기관번호, 기관명, 대표자, 전화번호, 주소), 응급환자(성명, 주민등록번호, 전화번호, 주소, 우편번호), 대지급금 수령계좌(예금주, 은행명, 계좌번호) 등)</li> <li>- 기타</li> </ul>
후천성 면역결핍증 예방법	<ul style="list-style-type: none"> <li>- 검진통지서(일련번호, 검진대상자(성명, 주민등록번호, 주소), 검진사유, 검진일시, 검진기관명 등)</li> <li>- 인체면역결핍바이러스 감염인/후천성면역결핍증 환자 발견(사망) 신고·보고(문서번호, 시행일자, 주신, 제목, 신고구분, 성별, 생년월일, 최초진단일, 확인진단일, 검사조건, 사망여부, 주요사망원인 등)</li> <li>- 후천성면역결핍증 검사대장(일련번호, 검진일자, 인적사항(검진목적, 성명, 가검물번호, 주소, 전화번호), 검사완료(일자, 결과, 회보일), 확인검사(의뢰일, 회보일, 결과) 등)</li> <li>- 기타</li> </ul>
장기등 이식에 관한 법률	<ul style="list-style-type: none"> <li>- 장기등 이식대기자 등록신청서(이식대기자(성명, 주민등록번호, 주소), 신청내용(이식이 필요한 장기등)</li> <li>- 장기등 및 조직 기증희망자 등록신청서(접수번호, 접수일자, 처리기간, 신청인(성명, 주민등록번호, 전화번호, 전자우편주소, 주소, 정보수신여부), 신청내용(기증형태, 기증희망자 표시여부, 법정대리인의 동의(성명, 주민등록번호, 관계, 서명)</li> <li>- 뇌사조사서(뇌사판정대상자(성명, 주민등록번호, 주소), 뇌사판정기관(명칭, 전화번호, 소재지), 뇌사조사전문(성명, 주민등록번호) 등)</li> <li>- 장기등 이식통보서(장기등을 이식받은 사람(성명, 생년월일, 주소, 전화번호), 이식한 장기등, 이식한 장기등의 상태 등)</li> <li>- 기타</li> </ul>
생명윤리 및 안전에 관한 법률	<ul style="list-style-type: none"> <li>- 배아생성 등에 관한 동의서(동의서 관리번호 배아생성의료기관명, 체외수정시술대상자(성명, 생년월일), 해당배우자(성명, 생년월일), 상담자성명, 담당의사성명등)</li> <li>- 생식세포 기증 동의서(배아생성의료기관명, 생식세포기증자(성명, 생년월일), 해당배우자(성명, 생년월일), 수증자와의 관계 등)</li> <li>- 연구이용 동의서(동의서 관리번호, 배아생성 의료기관명, 잔여배아 등의 기증자(성명, 생년월일, 해당배우자(성명, 생년월일), 잔여배아정보(배아생성 동이일, 잔여배아수, 잔여배아상태) 등)</li> <li>- 기타</li> </ul>

비식별화 사례집』 예시와 거의 동일한 내용을 담고 있다.

## 4. 개인의료정보 비식별화에 대한 시사점

앞에서 개인정보에 대한 비식별화 규정들을 살펴본 결과 비식별화 처리원칙, 비식별화 단계별 조치사항, 비식별화 처리기법 및 상세 활용방법, 부문별 빅데이터 처리시 비식별화 적용사례, 비식별화에 대한 적정성 평가절차와 세부평가방법, 재식별 위험관리 방안 등 개인정보 비식별화에 대한 전반적인 내용을 담고 있다. 더욱이 이는 HIPAA 프라이버시 규칙 내용을 기반으로 하고 있어 사실상 의료부문, 의료정보에 범용적으로 적용한다 하여도 큰 무리가 없을 것으로 여겨질 수도 있다.

그러나 우리나라 의료관련 법령인 의료법, 국민건강보험법, 의료급여법, 감염병 예방 및 관리에 관한 법률, 응급의료에 관한 법률, 후천성면역결핍증 예방법, 혈액관리법, 장기 등 이식에 관한 법

률, 생명윤리 및 안전에 관한 법률 등에서 규정하고 있는 의료정보는 진단서, 의료급여증, 건강보험증, 예방접종증명서, 응급환자 진료의뢰서, 후천성면역결핍증 예방관리자증, 장기 등 이식통보서, 배아생성 등에 관한 동의서 등 다양한 서식에 인적정보, 질병정보 뿐 아니라 기관정보, 행정정보 등 다양한 정보가 포함되어 있다(표 5).

이와 같이 개인의 질병치료, 관리 및 예방을 위한 정보들을 ‘의료정보’라는 하나의 용어로 표현하다보니 개인의료정보 비식별화 조치범위 등을 명확히 정하기가 어려워 현장에서는 세부영역에 따라, 적용하는 관계자에 따라 매우 혼란스러울 수밖에 없다. 더욱이 개인의료정보란 ‘의료정보’ 전체가 아닌 개인식별성을 갖춘 정보에 국한하여야 하므로 의료정보 및 의료정보 종류에 대해 보다 명확한 정의와 더불어 일반 의료정보와 개인의료정보, 개인의료정보 중에서도 해당정보 자체로 식별성을 지닌 정보와 타 정보와 쉽게 결합하여 식별가능한 정보에 대한 실 사용항목을 예로 한 구분이 필요하며 이에 대해 다양한 주체간 합의 또한 선행되어야 할 것이다. ■