

# 헬스케어 빅데이터 현재 동향과 이슈

신수용 (서울아산병원)

목차	1. 서론
	2. 외국 정부 동향
	3. 국내 정부 동향
	4. 헬스케어 빅데이터 최신 사례
	5. 개인정보 보호 및 보안
	6. 결론

## 1. 서론

많은 사람들이 바이오 혹은 의료 빅데이터라고 할 경우에 유전체 데이터만을 가장 먼저 떠올린다. 실제로 human genome project의 성공적인 수행이후로 방대한 양의 인간 유전체 데이터가 생성되기 시작하고 있으며, 2014년 기준으로 미국 브로드 연구소 한 곳에서만 1년에 10 peta bytes의 유전체 데이터를 생성하고 있을 정도이다. 그러나 유전체 데이터 외에도 다양한 데이터들이 사람의 건강상태를 파악하는 데 필요한데, 전통적인 병원의 진료 기록 외에도 SNS, 환경, 운동량, 영양정보 등 다양한 정보들이 새롭게 부각되고 있다<sup>[1]</sup>. 과거에는 이런 데이터들이 전산화되어 있지 않아서 활용이 힘들었으나, 진료 기록만 하더라도 국내의 경우 이미 대형종합병원들이 전자

의무기록시스템을 도입하여 모든 진료 기록을 전산화하여 저장한지 10년이 넘었고, 2014년 기준으로 전체 의료기관의 전자의무기록시스템 도입률이 92%에 달하고 있다. 이러한 진료 기록 전산화를 통해 국민건강보험공단이나 건강보험심사평가원에서도 다양한 의료 빅데이터 사업을 진행하고 있는데, 대표적인 예로 국민건강보험공단은 100만명 전국민 코호트를 공개하였고<sup>[2]</sup>, 건강보험심사평가원은 의료보험청구 데이터 접근에 대한 API를 개발하여 공개하였다<sup>[3]</sup>. 진료 기록 외에 일상생활에서 생성되는 운동량, 식이 정보 등에 대한 데이터들은 최근 들어 급격히 발전하고 있는 웨어러블 및 IoT 장비로 인해서 필요한 데이터들을 손쉽게 측정할 수 있게 되었고, 애플의 Apple HealthKit 등과 같은 플랫폼이 개발됨에 따라 저장할 수 있는 기술도 개발되었다. 또한 날씨

\* 본 원고는 산업통상자원부 국가기술표준원 국가표준코디네이터사업의 스마트헬스 분야에서 수행된 연구를 기반으로 작성되었습니다.

와 같은 외부 환경 데이터도 정부 3.0 정책에 맞춰서 외부 공개가 진행됨에 따라 수집할 수 있게 되었다. IBM에서 발표한 인포그래픽에 의하면 한 개인이 평생 만들어내는 건강 관련정보는 병원 진료 기록이 0.4 tera bytes이고, 빅데이터로 알려진 유전 정보가 6 tera bytes인데 반해, 일상생활 및 외부 환경 데이터가 1,100 tera bytes에 달한다고 보고하고 있다.

본 원고에서는 이러한 헬스케어 빅데이터의 국내외 최신 동향에 대해서 소개하고, 해결해야 할 이슈들에 대해서도 간단히 지적하고자 한다.

## 2. 외국 정부 동향

### 2.1 미국

미국은 현재 전 세계에서 헬스케어 빅데이터 연구를 가장 활발히 추진하고 있는 나라이다. 미국은 고질적인 고비용 의료체제로 인한 재정 낭비에 직면해 있는데, 2013년에 발표된 맥킨지 보고서에 의하면 빅데이터 기술을 미국의 보건의료 분야에 적용할 경우 300조에서 450조의 비용을 절감할 수 있다고 예상하였다. 이런 예측을 바탕으로 헬스케어 빅데이터를 정책적으로 지원하면서 여러 정부 기관들이 R&D 예산을 투자하고 있다.

우선 체계적인 헬스케어 빅데이터 지원을 위해 미국 국립 보건원(National Institute of Health, NIH)에 데이터 과학 책임자를 임명하였고, 빅데이터 연구 분야를 세분화하여 NIH BD2K (NIH Big Data to Knowledge, <http://bd2k.nih.gov>) 센터를 12개 설립하여 올해 3천2백만 달러를 투자하였고, 향후 10년간 지속적인 투자를 약속하였다.

NIH 외에도 미국 보건부 산하 ONC (The Office of the National Coordinator for Health Information Technology)에서도 기관의 미션 중

하나로 유전체, 소셜 네트워크 데이터, 환자 또는 소비자가 생성하는 데이터를 포함하는 헬스케어 빅데이터 분야를 선정하여 발표하였다. 미국 식약처에서도 빅데이터 기반 의약품 부작용 감시 프로젝트인 ‘Mini-Sentinel’을 2008년부터 추진하여 환자의 의료비 지불기록을 분석하여 효율적이고 빠른 의약품 감시를 추진하고 있다. 2014년 Patient-Centered Outcomes Research Institute는 미국에 거주하는 3천만명을 대상으로 보험 청구 데이터, 혈액검사 데이터, 과거진료기록 등을 수집하여 아스피린의 심장질환 예방 여부를 알아보기 위해서 천만달러를 투자하는 프로젝트를 시작하기도 하였다.

2015년 1월에는 백악관에서 Precision Medicine Initiative를 발표하여 개인의 건강을 위한 맞춤형 정보(유전자, 환경, 생활양식의 개인별 차이를 반영한 치료법 제안 등)를 제공하는 연구를 시작하였다. 단기적으로는 개개인의 유전자, 환경, 생활 방식을 고려한 맞춤형 치료법을 개발하고 약의 부작용을 줄이는 연구를 진행하고, 장기적으로는 이 프로젝트에 자발적으로 지원한 100만 명 참여자의 데이터를 공유하고 연구에 활용할 수 있는 시스템을 마련하는 연구이다. 이를 위해 빅데이터 기술을 활용하는 것을 기본으로 하여 데이터 수집/저장/분석/활용의 모든 단계에서 IT기술을 활용하도록 하였으며 개인정보 보호를 위한 노력도 강조하였다. 이 프로젝트의 일환으로 2015년 8월 미 식약처는 precisionFDA 프로젝트를 발표하였는데, 신약이나 치료법 개발에 활용되는 유전체, 환자 데이터를 연구자에게 공개하는 클라우드 플랫폼을 개발하는 과제이다.

연구비 투자 외에 정책 지원을 위한 활동도 활발히 진행되고 있는데, 대표적인 예로 빅데이터 분석에 항상 지적되는 개인정보 보호 및 보안 문제를 위해 백악관에서 2014년 5월 “Big Data:

Seizing Opportunities, Preserving Values”라는 보고서를 발표한 것을 들 수 있다<sup>4)</sup>. 이 보고서는 빅데이터 활용을 위한 개인정보 보호 및 보안 이슈에 대해서 전반적으로 점검하고 세부 가이드라인을 제시하고 있다. 특히 의료 분야에 대해서, 빅데이터 기술을 이용하여 전염병으로부터 매우 취약한 조산아들에 대해 발견하기 힘든 체온, 심장박동 등의 미세한 변화 등을 발견하여 전염병 감염을 방지할 수 있다는 예와 의료기관의 의약품 오남용 상황을 실시간으로 관리하여 세금을 줄이는 것이 가능하다는 예를 들었다.

## 2.2 영국

영국 CEBR (Centre for Economic and Business Research)의 2012년 자료에 의하면 영국의 경우 보건의료 분야에서 빅데이터의 가치는 2017년까지 5년간 총 204억 파운드에 달할 것으로 추정하였으며, 4,000개의 신규 일자리를 창출할 것으로 전망하였다. 따라서 영국 보건성도 국가에서 운영하는 영국 보건의료 체계의 특징을 살려 영국 전체의 약국, 병원의 처방 데이터를 DB화하였고, 이 빅데이터를 이용하여 국민 건강 예측 사업을 수행하고 있다. 또한 2012년 말 Genomic England 사업의 일환으로 100,000 genomes project를 발표하여 영국인 10만 명의 유전체를 분석하여 진료 기록과 결합하겠다는 야심찬 계획을 진행하고 있는 중이다.

## 2.3 EU

EU는 “Data Open Strategy”를 2011년에 발표하면서 빅데이터 처리기술 R&D에 2013년까지 2년간 1억 유로를 지원하였다. 이어서 2023년까지 10년간 보건의료향상을 위한 뇌 연구 프로젝트인

Human Brain Project에 10억 유로를 투자하였다.

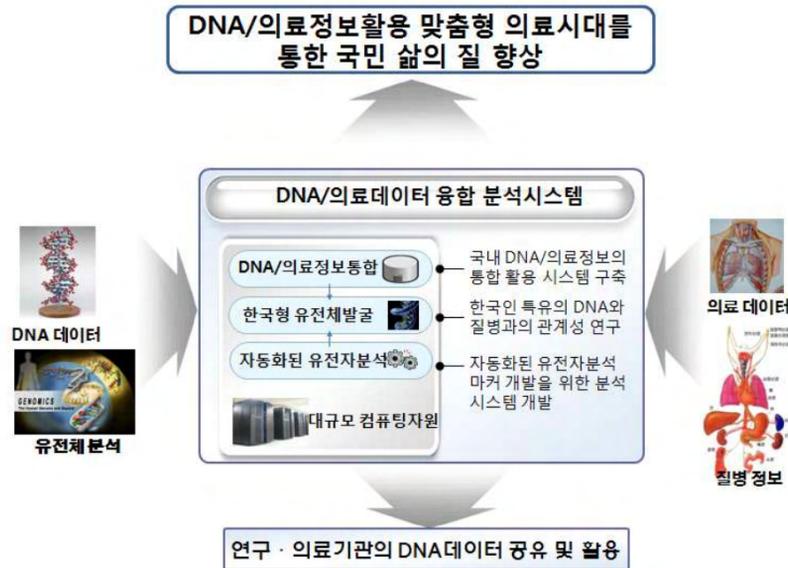
미국처럼 빅데이터의 정책적 지원을 위해 2014년 7월에는 유럽 위원회 명의로 빅데이터의 잠재력을 유럽의 경제 개발에 활용하기 위해 회원국들의 정책 개발을 위한 결의안을 채택하기도 하였다.

## 3. 국내 정부 동향

국내에서도 헬스케어 빅데이터에 대한 지원을 위해 다양한 정책을 발표하였다. 2011년에 발표한 국가정보화전략위원회의 ‘빅데이터를 활용한 스마트정부 구현(안)’에서 (그림 1)과 같은 빅데이터를 활용한 과학기술·의료 선진화계획을 발표하였다. 해당 안에서는 DNA·의료 데이터 공유 및 활용 촉진으로 개인 맞춤형 의료시대 실현을 목표로 하여, 폭증하는 바이오데이터의 수집·관리 및 분석을 바탕으로 의료데이터와의 연계 시스템 구축을 추진하기로 하였다.

2012년에는 관계부처합동으로 마련한 ‘빅데이터 마스터플랜’에서 의료데이터 분석을 통한 국민 건강증진을 핵심 과제 중 하나로 선정하였다. 해당 과제는 DNA, 진료기록, 질병정보의 융합 분석을 통한 건강정보 진단과 질병 발생 가능성의 예측 및 암·당뇨·비만 등 100여개 질환의 유전 위험도를 예측하며 맞춤형 사전 예방 서비스 제공을 추진하기로 하였다.

2013년부터 미래창조과학부는 빅데이터 시범 사업을 추진하고 있는데, 2013년 선정 6개 과제 중 3개 분야가 보건의료 분야(질병주의 예보 서비스, 의약품 안전성 조기 경보 서비스, 심실부정맥 예측 등 의료서비스) 이었으며, 2014년 선정 과제 4개 과제에도 보건의료 분야가 1개 (국민 질환 특성 분석을 통한 맞춤형 유의질환 및 병원



출처: 국가정보화전략위원회, 빅데이터를 활용한 스마트정부 구현(안), 2011

(그림 1) 의료분야 빅데이터 활용 방안

정보 제공 서비스)가 포함되어 있는 등, 활발히 지원을 하고 있다. 그 외에도 여러 정부부처에서 헬스케어 빅데이터 산업 발전을 위해 ICT 힐링 플랫폼 과제, 개인건강기록 기반 개인 맞춤형 건강관리 시스템 개발 과제 등을 진행하고 있는 등 다양한 헬스케어 빅데이터 R&D 과제를 지원하고 있다.

#### 4. 헬스케어 빅데이터 최신 사례

세계 유명 IT 기업체에서도 차세대 사업으로 다양한 헬스케어 빅데이터 서비스와 솔루션을 개발하고 있는데, 이 중 가장 기술적·사업적으로 두각을 나타내고 있는 것이 IBM의 Watson이다. IBM Watson은 Wellpoint와 같은 미국 대형보험 회사를 비롯하여, MD Anderson 암센터, Memorial Sloan Kettering 암센터, Cleveland Clinic 등의 세계적인 의료기관, 뉴욕 게놈 센터와 같은 유전체

연구소를 포함하여, 미국 보훈청과 같은 공공기관에게도 서비스를 제공하고 있다. 2014년 미국 종양학회 발표 자료에 의하면 IBM Watson이 다양한 암종 (대장암, 췌장암 등 9개 암)에 대해서 83% ~ 100%의 진단 정확도를 보일 정도로 우수한 성능을 보였다<sup>5)</sup>. 또한 2015년 8월에는 성능을 향상시키기 위해 의료영상 분석 전문업체인 Merge를 10억 달러에 인수하는 등 부족한 부분을 계속 채워 나가고 있다.

다양한 스타트업들도 헬스케어 빅데이터 분야에 뛰어들었는데, 이러한 트렌드로 인해 RockHealth (<https://rockhealth.com>)이라는 헬스케어 전문 스타트업 지원 펀드도 운영되고 있다. 여러 스타트업들 중에 헬스케어 빅데이터 사업과 관련해서 가장 성공적인 회사 중 하나는 Flatiron Health (<http://www.flatiron.com/>)를 들 수 있다. Flatiron Health는 OncologyCloud라는 제품을 통해 병원과 의사, 환자를 대상으로 암 진단 및 치료 관련 분석 결과를 제공하고 있으며, 전자의무기록에

기록되어 있는 방대한 비정형 데이터를 실시간으로 처리하는 기술에 장점을 보이고 있다. 현재 미국 내 천7백 개 이상의 의료기관에 서비스를 제공하고 있으며, 매달 4십만 명 이상의 암환자가 해당 서비스를 사용하고 있다.

국내의 경우에도 삼성SDS와 KT의 유전체서비스 사업을 포함하여 대기업들이 차세대 성장 동력사업으로 헬스케어를 선정하여 큰 투자를 하고 있으며, 외국과 마찬가지로 다양한 빅데이터 헬스케어 전문기업들 및 스타트업들이 등장하고 있다. 그 중에서도 뷰노(<http://www.vuno.co/>), 클디(<http://www.cldi.io/>)가 대표적인데, 두 회사 모두 deep learning 기술을 이용한 의료 영상 데이터 분석을 진행하고 있다.

이 외에도 국민건강보험공단은 자체적으로 보유한 헬스케어 빅데이터를 SNS를 포함한 여러 외부 데이터와 연계하여 다양한 대국민 서비스를 제공하고 있다. 그 중 대표적인 것이 Google의 Flu Trends와 유사한 국민건강 주의 알람서비스(<http://forecast.nhis.or.kr/index.do>)로, SNS, 블로그 등의 정보와 공단이 가지고 있는 진료 정보를 통합 분석하여 인플루엔자, 눈병, 식중독, 피부염 등에 대해서 지역별 위험도를 제공하고 있다.

## 5. 개인정보 보호 및 보안

헬스케어 빅데이터 사업이나 연구의 경우 반드시 개인정보 보호와 보안에 관한 이슈가 제기된다. 그래서 미국과 같은 선진국에서는 국가에서 정책적으로 가이드라인을 제시하고 있다. 국내의 경우에도 개인정보보호법, 생명윤리 및 안전에 관한 법률 등과 2014년 12월에 방송통신위원회에서 발표한 ‘빅데이터 개인정보보호 가이드라인’<sup>6)</sup>을 통해, 개인정보를 보호하기 위한 노력

을 하고 있다. 해당 법률 등에 의하면 개인정보 보호를 위해 개인들에게 명문화된 동의를 받거나, 아니면 데이터를 익명화하여 개인을 알아볼 수 없도록 하는 것을 요구하고 있다. 빅데이터 분석의 경우 개인들에게 동의를 받는다는 것이 불가능하기 때문에, 데이터를 익명화하여 사용할 수밖에 없고, 이를 위해서 2015년에는 미래창조과학부가 ‘빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서’를 발표하였다<sup>7)</sup>.

하지만 해당 가이드라인들은 다양한 영역을 취급하고 있기 때문에, 가장 민감한 개인정보인 의료데이터에 대한 가이드라인으로는 미흡한 부분들이 많이 있다. 따라서 미국의 HIPAA (Health Insurance Portability and Accountability Act)의 익명화 가이드라인과 같은 의료 데이터에 특화된 가이드라인이 필요한 상황이다<sup>8)</sup>.

## 6. 결론

불과 몇 년 전까지만 하더라도 헬스케어 빅데이터라고 하면 유전체 데이터, 진료 데이터 등 밖에 없었으나, 이제는 IoT의 발전으로 인해 소위 말하는 디지털 헬스케어 시대가 열리고 있다. 기존의 유전체 데이터, 진료 데이터들에 대한 수집/저장/분석/활용에 대해서 많은 기술들이 개발되어 있지만, 여전히 데이터 표준화, API 개발 등 다양한 이슈들이 여전히 남아있다. 또한 의료기관의 정밀한 의료기기보다는 부정확하지만, 웨어러블을 비롯하여 다양한 소형화된 진단기기들이 개발되어 의료기관 외부에서도 여러 가지 건강정보 및 의료정보를 본인이 직접 측정할 수 있게 되었다. 새롭게 부각되는 IoT 분야는 이제 막 시작한 단계라 데이터 표준화도 전혀 되어 있지 않고, 해당 데이터를 획득하기 위한 open API도 없는

상황이다.

보다 나아가 진료 데이터, 유전체 데이터, 자가 측정 데이터 등과 같은 이종의 데이터들을 통합하여 분석하는 기술, 개인당 1,000 tera byte 사이즈에 달하는 방대한 양의 빅데이터를 저장하기 위한 기술 등 많은 도전적인 과제들이 남아 있다. 주목할 것은 해당 과제들이 헬스케어 분야의 문제가 아니라, 전부 IT 분야의 문제들이라는 것이다. 따라서 이러한 과제들을 해결하기 위해서 ICT 전문가의 연구가 강하게 요구되고 있기 때문에, 국내 연구자들의 적극적인 참여를 통해 초창기 분야인 헬스케어 빅데이터 분야의 세계선도적인 기술을 개발하는 것이 반드시 필요해 보인다.

cation of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, <http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html>

### 저 자 약 력



신 수 용

이메일: sooyong.shin@amc.seoul.kr

### 참 고 문 헌

- [ 1 ] G. M. Weber, K. D. Mandl, and I. S. Kohane, Finding the missing link for big biomedical data, *The Journal of the Americal Medical Association*, 311(24): 2479-2480, 2014.
- [ 2 ] 국민건강보험자료 공유서비스, <http://nhiss.nhis.or.kr/>
- [ 3 ] 보건의료빅데이터개방시스템, <http://opendata.hira.or.kr/>
- [ 4 ] The White House, Big Data: Seizing Opportunities, Preserving Values, [https://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf)
- [ 5 ] 최윤섭, 헬스케어 이노베이션, 클라우드나인, 2015
- [ 6 ] 방송통신위원회, 빅데이터 개인정보보호 가이드라인, <http://www.kcc.go.kr/download.do?fileSeq=41264>
- [ 7 ] 미래창조과학부, 빅데이터 활용을 위한 개인정보 비식별화 기술 활용 안내서, [https://kbig.kr/sites/default/files/pds/빅데이터\\_비식별화\\_기술\\_활용\\_안내서\\_ver\\_1.0.pdf](https://kbig.kr/sites/default/files/pds/빅데이터_비식별화_기술_활용_안내서_ver_1.0.pdf)
- [ 8 ] Guidance Regarding Methods for De-identifi-

- 1998년 2월 서울대학교 컴퓨터공학과 학사
- 2000년 2월 서울대학교 컴퓨터공학부 석사
- 2005년 8월 서울대학교 전기, 컴퓨터공학부 박사
- 2006년 4월~2008년 3월 National Institute of Standards and Technology, Guest Researcher
- 2008년 5월~2010년 2월 서울대학교병원 의료정보센터 연구교수
- 2010년 3월~2011년 8월 삼성SDS Bioinformatics Lab, 수석연구원
- 2011년 9월~2012년 8월 서울아산병원 의학통계학과 연구조교수
- 2012년 9월~현재 서울아산병원 의생명정보학과 연구조교수
- 관심분야: Machine Learning, Data Mining, Translational Bioinformatics, Clinical Informatics, Clinical Research Informatics