

# 보건의료정보 활성화를 위한 공공데이터 개방과 빅데이터 활용



이원석 교수  
연세대학교 공과대학

## 1. 공공데이터 개방과 활용

### 가. 정부3.0과 공공데이터 개방

2014년 한국에서 사는 우리는 언제 어디서나 손안의 스마트 폰으로 원하는 모든 정보를 활용할 수 있게 되었다. 이는 곧 지난 20년 동안 우리가 꿈꿔왔던 정보화 사회의 완성 단계임을 의미한다. 특히 UN 전자정부 대상을 3회 연속해서 받은 우리나라의 공공영역은 거의 모든 분야에서 빠른 인터넷으로 연결된 운영시스템을 이용하여 업무의 효율성을 극대화하고 있다는 것이다. 특히 이러한 운영시스템에는 우리 사회의 모든 변화가 방대한 양의 디지털 데이터로 축적되고 있다는 점이다.

최근 정부가 적극적으로 추진하고 있는 정부3.0은 개방, 공유, 소통 및 협력이라는 핵심 수단으로 투명한 정부, 유능한 정부, 그리고 서비스 정부를 실현하여 수요자 맞춤형 서비스를 제공하고 일자리와 新성장 동력을 창출함으로써 국민 모두가 행복한 대한민국을 만든다는 개념이다.

IT의 발전을 돌이켜 보면 하드웨어에서 소프트웨어로 옮겨 갔던 무게 중심이 이제 소프트웨어에서 데이터로 옮겨가고 있다. 따라서 현재 우리 공공 영역에서 축적된 다양하고 방대한 양의 데이터는 우리나라의 미래 新성장 동력을 위한 중요한 자원임을 모두가 공감하고 있다. 다만 데이터라는 새로운 IT 발전의 축을 어떻게 전개시켜야 하는지에 대한 해답을 다

방면으로 찾고 있는 단계이다. 이러한 상황은 IT를 선도하는 우리나라가 데이터 시대로 진입하는 퍼스트 무버(First Mover)로서 당연히 거쳐야할 단계이다. 지난 30여 년간 국가 차원의 정보화 추진 정책을 통해 성숙한 정보화 사회를 이룩한 우리나라에서 정부 3.0의 성공은 한마디로 데이터의 개방, 데이터에 대한 공유, 데이터를 통한 소통, 그리고 데이터에 의한 협력을 통해서 얼마나 효과적이고 창의적인 방식으로 경제 발전의 원동력을 찾을 수 있는가에 달려 있다. 따라서 데이터에 기반을 둔 투명한 정부, 유능한 정부 그리고 서비스 정부의 탈바꿈은 단기간에 가능하지 않으며 장기간 동안 사회 전반적인 구조와 역할을 변화시키면서 단계적으로 진행될 과업이다.

공공데이터란 개별 공공 기관이 자신의 고유한 공공 업무를 정보시스템을 활용하여 수행하면서 축적된 데이터를 의미하며, 해당 업무에 관련된 상세한 정보를 담고 있다. 지난 1년간 정부3.0의 초기 핵심 전략으로 적극적으로 추진하고 있는 공공데이터의 개방은 지금까지 공공 기관에서 폐쇄적으로 사용하면서 개별적인 정보공개 청구제도를 통해서만 접근이 가능했던 공공데이터를 특별한 사유나 근거가 없는 한 모두 개방한다는 정책이다. 즉 개별 공공 기관이 갖고 있는 공공데이터의 개방은 그 기관이 해당 업무를 수행한 내역을 모든 국민에게 상세히 알림으로써 기관의 업무 수행 투명성을 향상시키므로, 모든 공공 기관의 공공데이터 개방은 정부3.0의 투명한 정부를 실현하여 국가 정책의 투명성을 데이터를 통해 보장하겠다는 뜻이다.

18세기 영국의 산업혁명을 계기로 분업화된 프로세스를 통해 효율적인 대량 생산 체계를 구축하면서 산업화 사회가 시작되었고, 오늘날 정보 사회 모든 영역에서도 이러한 분업화 프로세스를 인터넷과 컴퓨터를 활용하여 보다 효율적으로 업무를 수행하고 있다. 공공 기관들 또한 국가차원의 공적인 일을 수직 분업화하여 수행하고 있다. 즉 개별 공공 기관의 공공데이터에는 해당 기관이 분업적으로 담당하는 공공 업무나 서비스 수행 내용을 상세하게 담고 있다. 따라서 여러 공공 기관들의 공공데이터들을 의미적으로 결합하면 개별 기관의 업무 담당자가 파악하지 못했던 새롭고 유용한 업무 관련 지식을 발견할 수 있다. 이렇게 공공데이터를 연계 활용하여 더욱 새롭고 유용한 공공 업무관련 지식을 발견하고 실제 업무에 적극적으로 반영하면서 새로운 차원의 공공 서비스를 개발함으로써 정부3.0의 유능한 정부를 실현할 수 있다.

마찬가지로 공공데이터와 다양한 민간 데이터와의 연계 결합을 통해 공공과 민간 영역간의 새로운 지식을 발견하게 되면서, 공공 기관이 제공하는 공공 서비스보다 한층 더 업그레이드된 새로운 차원의 대국민 맞춤형 서비스를 제공하는 민간 기업이 나타나게 된다. 더 나아가 이러한 변화는 민간 영역 간의 데이터 결합 활성화를 통한 다양한 데이터의 활용을 촉

진시커 데이터를 유통하는 산업을 새롭게 만들고, 데이터 관련 민간 기업의 활성화 및 관련 분야의 많은 새로운 일자리 창출을 가져오게 된다. 즉 공공 기관이 민간 기업에서 발견한 새로운 공공 업무관련 지식이나 새롭게 개발한 대국민 맞춤형 공공 서비스를 적극적으로 수용함으로써 정부3.0에서 제시하는 새로운 차원의 서비스 정부를 실현하게 된다.

## 나. 공공데이터 개방 현황

공공데이터 전략위원회에 따르면 2014년 8월 기준으로 개방된 공공 데이터 셋의 수는 총 9,354개로 전년도 대비 96% 증가하였고, 공공데이터와 연결된 오픈API의 수는 총 649개로 조사되어 전년대비 28%가 증가하였다. 또한 공공데이터를 활용한 앱/웹 개발 사례는 2013년 42개에서 2014년 318개로 6.5배 증가하였다. 실제 데이터 활용 정도를 나타내는 개방된 데이터 셋의 다운로드 건수는 2013년 4,108건에서 2014년 8월 35,862건으로 7.7배 증가하였고, 오픈API로 제공한 월평균 트래픽 건수는 2013년 133백만 건에서 2014년 220백만 건으로 65.4% 증가한 것으로 조사되었다.

지난 1년간의 공공데이터 개방 정책은 양적으로 많은 성과를 이루었지만 무엇보다 중요한 성과는 공공데이터가 해당 공공 기관의 소유물이라는 과거의 인식에서, 정부3.0의 투명한 정부를 만들기 위해 국민에게 개방해야 할 공적인 대상이라는 인식으로의 변화이다. 앞으로 양적인 측면뿐만 아니라 질적인 측면으로 공공데이터 개방의 형태를 개선하면서 정부 3.0의 투명한 정부, 유능한 정부, 그리고 서비스 정부를 실현하기 위한 노력을 지속적으로 진행해야 할 것이다.

## 2. 보건 의료 빅데이터 개방과 활용

### 가. 정부 3.0과 빅데이터

빅데이터는 단순히 데이터의 양이 많은 것을 의미하기 보다는 기존 데이터에 비해 너무나 방대해서 현재까지의 방법이나 도구로 수집, 저장, 검색, 분석, 시각화하기 어려운 정형 또는 비정형 데이터 집합을 의미한다. 빅데이터는 방대한 규모(Volume), 빠른 처리 속도(Velocity), 다양한 형태(Variety), 새로운 가치(Value)의 '4V'로 그 특징을 정리할 수 있으며, 보다 구체적으로 표 1에서와 같이 대규모, 현실, 시계열성 그리고 결합성을 특성으로 갖는다. 빅데이터에 대한 다양한 정의가 논의되고 있고 개념 자체도 진화하고 있는 상황에서

단편적으로 빅데이터를 정의하기 보다는 여러 관점에서 빅데이터를 이해하는 것이 중요하며, 빅데이터 자체보다는 이를 활용하여 새로운 인사이트(Insight)를 창출하는 고급분석의 중요성이 더욱 부각되고 있다.

표 1. 빅데이터의 특성과 효과

특 성	효 과
대규모 (Huge Scale)	<ul style="list-style-type: none"> <li>기술 발전으로 데이터를 수집, 저장, 처리 능력 향상</li> <li>현실세계 데이터를 기반으로 한 정교한 패턴분석 가능</li> <li>데이터가 많을수록 유용한 데이터, 전혀 새로운 패턴의 정보를 찾아낼 수 있는 확률도 증가</li> </ul>
현실성 (Reality)	<ul style="list-style-type: none"> <li>우리사회 일상에서의 데이터 기록물의 증가 등 현실정보, 실시간 정보의 축적이 급증될 전망</li> <li>개인의 경험, 인식, 선호 등 인지적인 정보 유통 증가</li> </ul>
시계열성 (Trend)	<ul style="list-style-type: none"> <li>현시점 뿐만 아니라 과거 데이터의 유지로 시계열적인 연속성을 갖는 데이터의 구성</li> <li>과거, 현재, 미래 등 시간 흐름상의 추세 분석 가능</li> </ul>
결합성 (Combination)	<ul style="list-style-type: none"> <li>의료, 범죄, 환경 및 안보 등 타 분야, 이종 데이터간의 결합으로 새로운 의미의 정보 발견</li> <li>실제 물리적인 결합 이전에 데이터의 결합을 통한 사전 시뮬레이션, 안전성 검증 분야 발전 가능</li> </ul>

※출처: Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", WIRED MAGAZINE, 2008

지난 수년간 스마트폰과 같은 스마트기기의 빠른 확산과 함께 SNS 등 소셜 미디어(social media)가 급성장함에 따라 개인 정보와 소비 패턴, 위치 정보 등이 포함된 가치 있는 데이터가 매 순간 엄청난 양으로 생성되고 있으며, 사물인터넷(Internet of Things, IoT) 등이 활성화되면서 IT 기기들도 다량의 데이터를 직접 생성하기 시작했다. 이렇게 새로이 생성·유통되는 방대한 양의 데이터, 즉 빅데이터의 상당수는 비정형 데이터로서 전체 데이터량의 약 80%에 달하고 있다. 기존 방식으로는 이들 빅데이터 분석에 상당한 비용과 시간이 소요되고 분석 방식이 제한됨에 따라 빅데이터에 맞는 새로운 분석 방식과 기술이 등장하게 되었다.

점점 복잡해지고 많은 위험이 발생하는 최근의 사회·경제적 환경에서 특히 실시간 빅데이터 처리를 가능하게 하는 대규모 데이터 수집, 관리, 처리, 분석 기술의 발전은 정부와 기업이 시장에 보다 효과적으로 대응할 수 있는 통찰력과 가치를 제공해주기 시작했다. 실제로 구글, 페이스북, 아마존, 야후 등의 소셜 미디어나 인터넷 업체들이 빅데이터 기반의 소셜 분석 효과를 입증하면서 해외 주요국가의 정부 차원에서 빅데이터를 활용한 공공서비스 운영을 도모하고 있다. 민간기업도 빅데이터를 활용하여 수익성 향상과 프로세스 효율화는 물론 빅데이터 기반의 새로운 비즈니스 모델 발굴을 추진하면서 빅데이터 시장이 본격적으로 성장하고 있다.

현재까지의 정보시스템은 우리 사회의 정보를 미리 정해진 정형화된 틀로 저장하고 처리하면서 그 양이 너무 방대할 경우에는 표본 조사에 기반을 둔 통계적인 분석만이 가능하였다. 반면에 빅데이터 기술은 아무리 방대한 양의 데이터라도 그 크기에 상관없이 모든 종류의 데이터를 대상으로 빠른 분석이 가능하다는 점에서 이제 사회의 모든 변화를 아주 세밀한 단위로 상세하게 분석할 수 있게 되었다는 것을 의미한다. 따라서 산업화 사회에서 대중의 체형 통계 값으로 기성복을 만든다면 빅데이터 기술은 모든 개인 각자의 체형에 꼭 맞는 맞춤복을 만들 수 있는 기술이다.

## 나. 공공 빅데이터 활용 현황

2011년 11월 대통령소속 국가정보화전략위원회에서 빅데이터 활용의 필요성이 우리나라 최초로 제기된 이후 2012년에는 범부처 공공 빅데이터 활용 마스터플랜을 기획하고 2013년부터 정부3.0 차원에서 거의 모든 부처와 지자체에서 빅데이터의 활용이 적극적으로 시도되고 있다. 안전행정부와 미래창조과학부는 표 2와 표 3과 같이 다양한 공공 영역에서 빅데이터 기술을 시범적으로 적용하고 있다.

표 2. 안전행정부 빅데이터 시범사업

2013년	주요내용
통계청 물가지수 분석을 위한 시범 사업	20개의 사이트에서 295개의 품목에 대한 가격 정보를 수집하여 물가 변동요인을 파악하고, 오프라인 가격정보를 통해 월 단위로 물가 정보를 분석 했던 것을 일단위로 단축
2014년	주요내용
빅데이터 기반 산불 위험예보 및 확산서비스	산불발생시간 및 발생지 위치의 산불 진행시간에 따른 산불확산 방향, 산불강도 및 산불확산 속도 예측을 통한 산림피해 최소화
빅데이터 공통기반을 활용한 스마트 재난상황실	중앙·지자체 재난안전상황실에서 이용되고 있는 관련 정보와 수집한 온라인 소셜 데이터를 연계 분석하여 재난예측 및 의사결정을 할 수 있는 통합 재난 모니터링 체계 구축
의약품·의료정보 연계통합시스템 구축을 위한 시범사업	의약품 허가자료, 요양급여청구자료(심평원), 건강검진자료(건보공단), 사망통계자료(통계청), 환자 진료기록(의료기관)을 온라인상에서 수집한 약물부작용 데이터와 연계하여 의약품으로 인한 부작용을 분석 및 부작용 원인규명
빅데이터 기반의 개인정보 보호체계 및 침해예보제 운영기반 구축	약 350만개에 이르는 개인정보보호법 적용 대상에 대한 현황 파악 및 법 준수 모니터링 한계 및 개인정보 침해 발생 이후의 사후 처리에 국한된 기존 침해대응 방안을 개인정보 침해요소, 위험기관에 대한 사전조치로 침해·유출사고 사전예방 실현
GIS기반 맞춤형 교통안전 예보 시스템 구축	교통사고정보, 기상정보, 소통정보 등을 융합한 교통안전 예보 시스템을 개발하여 일반국민들에게 제공함으로써 기후, 시간대, 운전경로에 따른 위험지역의 운전을 회피토록 유도 및 교통사고 감소에 기여

표 3. 미래창조과학부 빅데이터 시범사업

2013년	주요내용
빅데이터 분석을 통한 심야버스 노선 정책 지원	서울시의 교통데이터와 KT의 유동인구 데이터를 융합, 분석하여 최적의 심야버스노선 정책 수립
국민건강 주의 예보 시범서비스 구축	국민건강보험공단의 건강보험 DB와 SNS 분석 정보를 융합하여 독감, 눈병, 식중독 등 감염 병 발생 예측 모델을 개발하고, 상시 모니터링 하여 위험 징후 시 주의예보 서비스 제공
빅데이터 기본 의약품 안전성 조기경보 서비스	유해사례DB, 진료기록, SNS 등을 연계 분석하여, 유의 의약품을 추출하고, 이들의 위험도를 예측하여 병의원, 제약회사 및 유관기관 등과 정보 공유
보건의료 빅데이터 활용 시범사업	포털(다음), 질병관리본부 등과 협의된 데이터와 병원 자체 데이터를 활용하여 독감 유행 예측, 심실부정맥 예측, 입원병상 배정 최적화 등 서비스 제공
소상공인 창업성공을 제고를 위한 상권분석 및 점포평가 서비스	카드거래, 부동산, 상가이력 정보 등의 연계 분석을 통해 창업 관련 과거/현황 분석 및 미래예측 정보를 제공하여 창업 의사결정 지원
빅데이터 활용 스마트 뉴스 제공 모바일 앱 개발	대량의 기사DB에 대해 중요도, 관계도 등 다각도의 고급분석을 적용하여 지능형 뉴스 검색 서비스 제공
2014년	주요내용
유통 빅데이터를 통한 중소기업 지원	대형유통상의 판매정보를 분석하여 지역 슈퍼 마켓 등 중소기업인을 위한 시즌 별 인기상품 분석 정보, 날씨/지역 맞춤형 상품추천 등 데이터 기반 마케팅 정보 제공
빅데이터 분석 기반 외국인 관광산업 지원	내외국인 관광·소비 패턴, 중국인 관광객 트렌드를 복합 분석하여 추가 관광지 개발, 관광지 추천 등 지도 기반의 개인 맞춤형 관광 정보 제공을 통해 관광객 만족도를 향상시키고, 분석 정보를 관광 정책에 활용
의료정보 빅데이터 기반 맞춤형 유의질환 및 병원정보	진료정보 빅데이터 분석을 통해 발생 질환 별 예상 유의 질환정보 및 맞춤형 병원 정보 제공
자동차 부품기업의 생산성 향상을 위한 빅데이터 서비스	자동차 부품 제조사가 공동 활용할 수 있는 빅데이터 플랫폼을 제공하여 생산 공정 데이터의 정밀 분석을 통한 제품 품질향상에 활용

국내 민간 부분도 빅데이터를 활용한 서비스 모델을 활발하게 시도하고 있다. 신용카드사는 고객거래, 소셜 미디어, 가맹점 등에서 발생한 빅데이터를 이용하여 가맹점과 고객 소비 패턴 등을 분석해 실시간 마케팅에 활용할 방안을 모색하고 있다. 롯데마트, 롯데백화점, 롯데홈쇼핑 등을 통해 고객의 구매 정보를 손쉽게 분석할 수 있는 롯데카드의 경우, ‘빅데이터 마케팅’을 통해 쿠폰을 받은 고객이 실제 상품을 구입하는 비율인 ‘마케팅 반응률’이 기존 대비 크게 증가한 것으로 확인됐다. 평균 43.7%였던 롯데마트의 마케팅 반응률은 65.9%로 상승했고 롯데백화점은 30.8%에서 36.1%로, 롯데홈쇼핑은 11%에서 15.2%로 크게 상승했다. 이 중 장난감 매출은 전년 대비 7배나 증가했다. 삼성카드는 ‘m포켓’ 서비스를 통해 250만 가맹점에 대한 결제 데이터를 분석해서 매출액과 사용자 결제 건수를 기준으로 가맹점 이용패턴을 분석하고 있다. 분석 결과를 기반으로 고객에게 지역별, 연령별, 성별, 시간대별

및 주말/주중 등 조건에 따라 상황에 맞는 검색을 할 수 있도록 제공하고 있다. 하나SK카드는 빅데이터를 활용해 고객 맞춤형 이벤트와 쿠폰을 제공하는 ‘겟모어(get more)’서비스를 제공하고 있다. 가맹점과 관련된 이벤트 정보와 고객의 결제 데이터 분석을 통한 개별 고객의 카드사용 패턴에 따라 고객 맞춤형 쇼핑 정보를 제공하는 것이 특징이다.

2013년에 빅데이터 시범사업으로 진행된 서울시의 심야버스노선 최적화 모델은 민간 빅데이터인 KT의 이동통신 데이터와의 결합을 통해 공공 업무의 효과성을 향상시킨 모델로 2014년 여러 지자체에서 실시되고 있으며 카드사의 결제 내역 빅데이터를 결합하여 지자체의 축제 마케팅의 정확도를 향상시키고 있다. SK텔레콤은 티맵(T-Map) 내비게이션을 통해 콜택시, 유류 운반차량, 고속버스 등에 GPS를 장착하여 수집되는 방대한 실시간 교통정보를 실시간으로 분석하여 목적지까지의 최단거리와 정확한 도착시간을 제공하고 있다. 빅데이터 기반으로 지도와 유동인구, 업종별/월별 매출 등 10여 종의 정보를 분석하여 상권분석 서비스를 제공하고 있다.

## 다. 보건 의료 데이터 개방 및 활용 사례

보건의료 분야의 데이터는 대부분 개인정보나 개인별 진료 정보를 포함하고 있어서 빅데이터 분석의 활용성은 매우 높으나 데이터의 개방이나 활용이 제한적이다. 건강보험심사평가원은 환자 개인정보를 제외하고 질병별, 특정 의료행위별, 치료재료별 등으로 집계된 진료정보와 병원정보 등 보유한 각종 보건의료정보를 API를 통해 민간에 개방하고 있고, 건강보험심사평가원이 보유한 연간 200억 개(49.5TB)에 달하는 보건의료 빅데이터를 활용한 다양한 서비스 개발과 산학연 연구가 가능한 보건의료 빅데이터센터를 운영하고 있다.

식품의약품안전처는 건강기능 식품, 검사 부적합 식품, 국내소송 의약품 등 85개의 데이터를 공개하고 있다. 국민건강보험공단에서는 5,000만 명의 건강보험 가입자 중 100만 명에 대한 2002년부터 2010년까지의 진료내역, 검진결과, 거주지 및 보험료, 요양기관 정보 등을 개인 식별이 불가능한 형태로 추출한 전 국민 건강정보표본 데이터베이스를 연구용으로 제공하고 있으며, 국립보건연구원에서는 중앙인체자원은행을 운영하여 전국 17개 병원에서 유전체 역학조사 등으로 확보한 일반인 32만 명, 병원에서 확보한 질환군 20만 명의 역학 정보, 혈액, 조직, DNA 등 정보를 수집하여 제공하고 있다.

국민건강보험공단의 국가건강검진 자료에 있는 개인의 시력검사 결과를 연계 활용하여 운전면허 발급 과정을 단순화하였고 메디벤처스 ‘메디라떼’는 건강보험심사평가원의 빅데이터를 이용하여 사용자에 따라 지역별, 혜택별, 거리순 등 실시간 업데이트 맞춤형 정보로 사용자가 원하는 최적의 병원을 찾아준다. 또한 실제 치료를 받은 고객만 리뷰를 작성할 수 있



는 시스템으로 허위 작성의 가능성을 원천 차단하였고, 진료를 받으면 치료비 중 비급여 항목에 대해 최고 10%까지 리워드도 되돌려줘 소비자들에게 의료비를 절약할 수 있도록 하며 적립된 포인트는 다양한 상품으로 교환이 가능하다.

2013년 미래창조과학부의 빅데이터 시범사업으로 추진된 국민건강 주의 예보 시범서비스 사업은 국민건강보험공단의 진료내역데이터와 트위터 데이터를 연계하여 질병위험을 예측하는 과제를 수행하였다. 의약품 안전성 조기경보 서비스 사업에서는 다음이나 네이버의 블로그 및 카페에 올라오는 의약품 부작용 관련 내용을 모아서 분석하여 의약품 부작용 여부를 조기에 알리는 서비스를 시범적으로 구현하였다. 또한 보건의로 빅데이터 활용 서비스 시범사업에서는 다음의 소셜 데이터와 서울 아산병원의 데이터를 활용하여 인플루엔자 유행 예측, 심실부정맥 예측, 마약류 감시, 그리고 병원 병상 배정 최적화 서비스를 시범적으로 구현하였다. 2014년에도 건강보험심사평가원의 진료정보를 기반으로 질병예측과 맞춤형 병원 정보를 제공하는 사업과 의약품과 의료정보를 통합하여 의약품 부작용의 원인 분석을 규명하는 사업을 진행하고 있다.

### 3. 보건의료정보 활성화 방안

현재까지 국민들의 진료내역 정보는 보건의로 행정만을 위해 관리되었으나 앞으로는 국민 개개인을 위한 맞춤형 보건의료 서비스를 실현하는 미래 핵심 자원으로 활용될 것으로 예상되므로 새로운 용도에 적합하도록 보건의료 데이터의 생산 및 관리 과정과 개방 범위를 시급히 정비해야 한다. 정부3.0의 투명한 정부를 위한 공공데이터 개방은 정부의 행정 프로세스를 명확히 보여주는 데이터를 대상으로 한다. 반면, 유능한 정부나 서비스 정부를 위한 공공 기관이나 민간 기업의 빅데이터 활용은 그 범위와 대상이 보다 넓고 다양할 수 있다. 원석에 비유되는 빅데이터는 산업적 민간 활용성이 매우 높으므로 현재의 공공 데이터 개방 방식으로 누구나 활용할 수 있게 개방한다는 것은 국내에서 생산되는 주요 자원을 외국 업체가 아무런 조건 없이 채광해 가도록 하는 것과 같은 경우가 된다. 또한 공공 빅데이터의 활용성을 보다 향상시키기 위해서는 모두에게 공개하기 어려운 민감한 개인정보나 개별 의사의 진료행위 정보를 공익 차원에서 제한된 범위와 방법으로 활용할 수 있어야 한다. 이는 마치 위험성이 매우 높은 원자력을 안전하게 사용하는 방식을 찾은 것과 같다. 이러한 관점에서 볼 때 보건의료 분야에서 공공데이터 개방과 빅데이터 활용이란 두 가지 명제를 앞으로는 보다 명확하게 구분해야 한다고 생각한다.

빅데이터의 가장 큰 장점은 분업화된 정보화 사회의 서로 다른 영역에서 생산되는 데이터



의 결합으로 그 분야의 전문가도 파악하지 못하던 새로운 사실을 발견할 수 있다는 것이다. 공공 영역의 보건의로 행정 데이터와 민간 병원의 의료 임상 데이터, 그리고 인간 유전체 및 단백질 연구 데이터 등이 하나로 연결되어 개인별 맞춤형 실시간 빅데이터 분석이 가능해진다면 인류 문명에 엄청난 파급효과를 발생시키는 계기가 될 것으로 생각한다.

이러한 특성을 최대한 활용하기 위해서는 보건의로 분야의 빅데이터 생산을 담당하는 컨트롤 타워(control tower)와 보건의로 분야에 적합한 새로운 활용 방안을 창의적으로 도출하는 싱크 탱크(think tank)가 필요하다. 각 기관에서 운영되고 있는 개별 시스템들로부터 보건의로 분야의 융합적인 관점에서 활용성이 높은 주요 빅데이터를 발굴하고, 빅데이터 활용에 적합한 빅데이터 품질 지표를 설정하여 새롭게 유입되는 빅데이터를 상시 모니터링하는 시스템을 갖춰야 한다. 제도적인 문제와 기관들 간의 이해관계로 인해 데이터의 수집과 융합이 방해되지 않도록 해야 하고, 빅데이터의 역기능을 현명하게 해결하면서 주요 빅데이터를 개인 정보 및 영업 정보 민감도에 따라 등급화하고, 빅데이터 관리 주체를 지정하여 컨트롤 타워를 중심으로 귀중한 데이터를 자원화 하는 보건의로 빅데이터 거버넌스 체계를 확립하여야 한다.

보건의로 빅데이터 싱크 탱크는 보건의로 빅데이터를 총체적으로 활용하는 새롭고 다양한 방법을 지속적으로 찾아 기존의 보건의로 행정 업무에 반영하여 대국민 행정 업무 프로세스를 지속적으로 최적 고도화할 때 정부 3.0의 유능한 정부가 실현된다. 이와 동시에 귀중한 자산인 보건의로 빅데이터를 활용하는 민간 기업이 새로운 비즈니스 모델로 지속적으로 성장할 때 정부3.0의 서비스 정부가 실현된다. 정부는 이와 같이 공공데이터를 제공하여 민간 기업을 통해 국민에게 새로운 일자리와 새로운 차원의 맞춤형 보건의로 서비스를 제공함으로써 정부3.0의 궁극적인 목표인 국민 행복 시대를 달성하는 것이다.

## 4. 나가며

빅데이터는 인터넷 기술 이후에 다가오는 미래 사회로의 중요한 변혁을 시작하는 기술이다. 향후 20년 이내에 인류는 600억 개의 스마트 기기들을 사용하면서 매일 15페타바이트(PB) 분량의 개인별 생활 정보를 발생시킬 것으로 예상된다. 빅데이터 기술이 선도하는 미래는 사회가 작동하면서 발생하는 모든 데이터를 상세하게 모니터링하면서 개인의 가치와 대중의 가치를 최적으로 조화시키면서 각자에게 주어진 상황에 가장 적합하고 합리적인 방식으로 업무가 처리되고 서비스되는 최적화 사회가 될 것이다.

이미 인터넷 시대에서 경험했듯이 신기술로 인한 급격한 변화는 사회 각 분야에 새롭고 다양한 비즈니스 모델을 창출시키고 이를 통해 개인이나 기업 그리고 국가가 발전할 수 있는 소중한

기회를 제공한다. 더욱이 성숙한 정보 사회를 구현한 우리나라에서, 개별 공공 기관의 정보 시스템을 통해 지속적으로 수집되고 있는 데이터를 국가차원으로 부가적인 활용 방법을 찾는 것은 우리나라의 미래 경쟁력을 좌우하는 핵심 역량이 될 것이다. 정부 3.0의 공공데이터 개방 노력은 이러한 관점의 시작이며 데이터가 중심이 되는 미래 시대를 여는 중요한 시발점이라고 생각한다.

지난 30년 이전부터 우리나라의 상위 1%의 인재들이 의학 분야로 진출하여 IT 강국인 우리나라의 정보시스템 환경에서 보건의료 빅데이터를 생산하고 있다. 따라서 우리나라 의료 행정 빅데이터의 정확도는 매우 높을 것으로 예상된다. 18세기 수학자 콩도르세의 배심원 이론에 의하면, 배심원 개개인이 정답을 맞힐 확률이 1/2이 넘으면 배심원의 수가 많을수록 정답을 찾을 확률이 1로 수렴된다는 것이다. 상위 1%의 능력을 갖춘 우리나라 의사 개개인의 진료행위가 정확할 확률은 당연히 1/2을 넘을 것이고 이러한 의사들의 진료행위를 모두 모은 빅데이터를 분석하면 항상 정답을 찾는다라는 뜻이다.

국가의 우수한 인력이 지속적으로 유입되고 있는 보건의료 분야는 이제 공공데이터 개방을 넘어 우리나라의 미래 핵심 성장 동력을 창출하는 보건의료 빅데이터 활성화를 위한 진지한 발걸음을 시작해야 할 때이다. 모든 자료가 디지털 데이터화되는 정보사회이면서 국민 개개인들의 IT 기기 활용도가 세계 최고 수준인 우리나라는 보건의료 빅데이터로 시작되는 인류의 미래 사회 진화를 주도적으로 이끌어갈 수 있는 충분한 잠재력을 갖고 있다.

돌이켜보면 지금까지의 정보화는 산업화의 업무 프로세스를 그대로 컴퓨터에 이식하는 단순 작업으로 강의 상류에서 하류로 강을 따라 운행하는 것이라면, 빅데이터를 활용하는 미래 최적화 사회로의 전이는 마치 컬럼버스가 신대륙을 찾아 떠나는 바다 항해에 비유할 수 있다. 예상하지 못한 다양한 형태의 도전을 긍정적인 사고로 현명하게 극복하면서 새로운 문제 해결 방식을 창의적으로 찾는 노력을 경주해야 할 것이다. 우리가 IT강국이 되기 위해서 컴퓨터 HW와 SW에 대한 연구개발(R&D)에 많은 노력을 경주하였듯이 데이터가 중심이 되는 미래 시대에 적합한 새로운 형태의 데이터 R&D 체계를 만들어 실제 데이터와 HW 및 SW가 융합되는 새로운 데이터기반 첨단 도구를 선도적으로 만들어야 한다. 이와 동시에 민간과 공공의 다양한 빅데이터를 안전하고 활발하게 유통시킬 수 있는 새로운 관점의 데이터 스타일 보건의료 행정 체계와 데이터 스타일 법제도의 마련이 필수적이다. IT 강국이면서 매년 수능 성적의 상위 1%가 의료 인력이 되는 우리나라가 가까운 장래에 인류의 보건의료와 생명산업을 책임지는 국가로 우뚝 서는 날을 기대해 본다. 🌐