

보건복지분야 공공 빅데이터의 활용과 과제

Big Data and Its Applications in the Health and Welfare Sectors



이연희 한국보건사회연구원 책임전문원

보건복지분야에는 건강보험 표본코호트DB, 유전체자원, 사회보장정보시스템에 집적되는 사회보장정보, 사회서비스 전자바우처사업 트랜잭션DB 등 보건복지정책과 관련된 상당한 가치의 공공 빅데이터들이 있다. 공공 데이터의 공개·공유를 추구하는 정부3.0 추진에도 불구하고 보건복지부문의 공공 빅데이터는 개인정보나 민감정보가 포함되어 있어 제한적으로 공개되거나 아예 미공개된 데이터가 다수를 차지한다. 공공 빅데이터의 공개 환경이 이렇다보니 공공 빅데이터를 활용한 서비스 모델의 개발은 주로 데이터를 보유하고 있는 기관에 치우친 경향이 뚜렷하게 나타나고 있다. 보건복지부문 공공 빅데이터의 활성화를 위해서는 개인정보 보호기술의 고도화, 데이터 표준화, 빅데이터의 공개종류 및 범위 확대, 보건복지부문에 특화된 빅데이터 운영센터의 설치 등이 요구된다.

1. 들어가며

국민의 건강한 삶을 영위하고, 또한 삶의 질을 향상시키는데 보건의료 및 사회복지 공공영역에서의 서비스는 지속적으로 확대되고 있는 가운데 정형화된 형태의 관련 데이터 규모도 점차 방대해지고 있을 뿐 아니라 데이터의 가치도 여러 사례를 통해 증명되고 있어, 보건복지분야에서 수집되는 정형의 공공 빅데이터야말로 미래 성장동력으로 작용할 주요 영역중 하나로 꼽고 있다¹⁾. 이에 본고는 공공 서비스의 향상과 국가의 미래전략을 지원하는데 있어 기초자료가 될 보건의료 및 사회복지

부문의 빅데이터 활용을 활성화시키는데 요구되는 해결과제에 대해 짚어보고자 한다.

2. 보건복지분야 공공 빅데이터의 공개 현황

공공영역의 데이터 수집은 공공서비스의 향상과 평가, 위험을 예측함으로써 국민의 삶의 질을 높이는데 목적이 있다. 다양한 목적에 의해 수집되는 정보는 당초 정해놓은 일련의 데이터 생명주기를 거치는데, 빅데이터 개념의 출현은 바로 데이터

1) 박종수(2014), 빅데이터법제에 관한 비교법적 연구-독일-, 한국법제연구원.

의 생명주기를 장기화시키는 계기가 되고 있다. 이 유인 즉, 활용처를 찾지 못해 마냥 집적만 하고 있던 데이터가 새로운 가치를 찾는 수단이 되고 있기 때문이다. 이에 정부는 국민의 알권리 충족과 공공 정보의 이용 활성화를 위하여 「공공기관의 정보공개에 관한 법률」을 제정하는 등 공공이 보유하고 있는 정보를 적극적으로 공유·개방하도록 권고하고 있다.

보건복지분야에서 가장 대표적인 공공 빅데이터는 건강보험 빅데이터인 '건강보험 표본 코호트 DB'와 '환자데이터셋'으로서 전 국민의 진료내역을 담고 있어 세계에서도 찾아보기 힘든 상당한 가치를 지닌 데이터들이다. 그러나 두 데이터 모두는 개인의 진료정보를 가진 민감 정보라서 완전히 개방하지 않고 있어, 만일 데이터를 이용하려면 보유기관의 심사절차를 거쳐 공개가 결정되는 제한적 공개방식을 취하고 있다. 한편, 건강보험 빅데이터만큼 가치를 가진 보건의료부분의 또 다른 데이터로는 질병관리본부가 구축하고 있는 '한국인체자원' 정보로서 한국인의 유전체 정보가 있으며, 이 데이터 역시 제한적인 공개방식을 따르고 있다. 그 외에도 <표 1>에 나열한 바와 같이 주기적인 수집을 통해 점차 데이터 규모를 키워가고 있는 몇몇 데이터들이 있다. <표 1>에서 '건강보험 표본 코호트DB'나 '환자데이터셋'을 제외하면 규모면에서는 빅데이터로 보기 어렵지만 데이터 집적에 상당한 비용과 관리가 투입되어야 하고 데이터 자체의 가치뿐 아니라 타 데이터와의 연계를 통해 새로운 가치를 창출해 낼 수 있다는 측면에서 본다면 빅데이터로 간주되기도 한다. <표 1>에 기술한 바와

같이, 개인정보가 포함되지 않았거나 비식별화 조치된 기타 조사데이터들은 일반인이 쉽게 접근하도록 공개함으로써 다양한 영역에서의 활용을 유도하고 있다.

한편, 사회복지분야의 대표적인 공공 빅데이터로는 사회보장정보시스템에 집적되는 사회보장 데이터가 있다²⁾. 사회보장 데이터는 각종 사회복지 급여 및 서비스 지원 대상자의 자격과 이력에 관한 정보를 통합 관리하고, 지자체의 복지업무 처리를 지원하는 목적으로 구축된 데이터이다. 또 다른 사회복지영역 공공 빅데이터로는 정부의 '사회서비스 전자바우처사업'에서 수집되는 서비스이용 트랜잭션 데이터가 있다. 사회서비스 전자바우처사업은 노인돌봄서비스, 장애인활동지원서비스, 지역사회서비스투자사업, 산모신생아도우미서비스, 가사간병지원서비스, 장애아동재활치료, 언어발달지원서비스 등 7개 공공 서비스별 이용 가능한 서비스 금액이나 수량, 시간범위 내에서 개인별, 제공기관별 실시간 이용 트랜잭션이 수집·집적되고 있다. 더욱이 서비스 제공규모 및 사업이 점차 증가하고 있어 전자바우처사업으로 수집되는 데이터는 규모면에서도 그야말로 빅 데이터이다. 이 외에 사회복지분야의 공공 빅데이터라고 할 수 있는 것으로는 <표 2>에 나열한 데이터들을 들 수 있다. 사회복지영역에 있는 공공 빅데이터의 공개율은 <표 2>만 보더라도 알 수 있듯 많이 낮은 편이다. 즉, 한국복지패널 데이터를 제외한 나머지 공공 빅데이터는 아예 공개가 되지 않고 있다.

선진국들은 정부 정책 및 서비스의 투명성을 제

2) <http://www.khwis.or.kr> 2015.06.29. 인출

표 1. 보건의료분야 공공 빅데이터 현황

구분	보유기관	내용	공개여부
건강보험 표본 코호트 DB ¹⁾	국민건강 보험공단	<ul style="list-style-type: none"> - 자격DB: 건강보험가입자 및 의료급여수급권자의 성, 연령대, 지역, 사회경제적 변수, 장애, 사망관련 등 - 진료DB: 요양급여 청구자료로서 진료, 상병, 처방 관련 변수 - 건강검진DB: 검진 주요결과 및 문진에 의한 생활습관 및 행태관련 자료 - 요양기관DB: 요양기관 종별, 설립구분, 지역, 시설, 장비, 인력관련 자료 	제한적 공 개
환자 데이터셋 ²⁾	건강보험 심사평가원	- 건강보험 청구자료를 기초로 진료개시일 기준 1년 간 진료 받은 환자대상의 표본 데이터	제한적 공 개
한국인체 자원 ³⁾	질병관리 본부	- 공여자로부터 기증받은 인체유래물(DNA, 조직, 혈액, 뇨 등)과 임상(진단명, 수술명, 병리조직검사결과, 혈액검사 등), 역학(성별, 생년월일, 음주력, 흡연력 등) 및 유전(SNP, CNV, Exome 등)정보	제한적 공 개
지역보건 의료정보	사회보장 정보원	- 전국 보건기관(보건의료원, 보건소/지소, 보건진료소)의 보건사업 및 행정업무, 전자의무기록 및 진료관련(진료내역 및 검진결과 등) 정보	미공개
지역사회 건강조사	질병관리 본부	- 지역 보건의료계획수립 및 보건사업 평가 활용 지표로서, 건강행태, 건강검진 및 예방접종, 질병이환, 의료이용, 사고 및 중독, 활동제한 및 삶의 질, 보건기관 이용, 사회 물리적 환경, 심정지, 교육 및 경제활동 등	공 개
국민건강 영양조사	질병관리 본부	- 국민의 건강 및 영양 상태에 관한 현황 및 추이 파악 - 신체계측, 비만, 고혈압 등 검진조사, 흡연, 음주, 비만 및 체중조절, 신체활동 등 건강설문조사, 식품 및 영양소 섭취현황, 식생활행태, 식이보충제 등 영양조사	공 개
한국의료 패널	한국보건 사회연구원	- 개인의 건강수준, 의료이용 및 의료비 지출 요인, 건강행태, 의료욕구, 보건의료서비스 수요행태 변화분석 - 사회경제적 특성, 의약품 구매, 경제활동, 건강수준, 의약품 복용행태, 민간의료보험, 건강기능식품, 건강행태 등	공 개

자료: 1) 표본코호트DB 사용자 매뉴얼, 국민건강보험공단
 2) <http://www.hira.or.kr/dummy.do?pgmid=HIRAA070001000410> 2015-07-08 15:15 인출
 3) <http://www.cdc.go.kr/CDC/contents/CdcKrContentView.jsp?cid=61074&menuIds=HOME001-MNU1136-MNU1826-MNU1830> 2015-07-08 10:12 인출

고하고, 나아가 공공영역의 정보를 재사용하여 신규 사업기회를 창출함으로써 시장을 활성화시키는데 행정정보 등 공공이 보유하고 있는 다양한 정보를 공개·개방하고 있다. 한편, 공공 데이터의 접

근성을 높이고 활용성을 확대하기 위해 정부주도의 포탈도 운영하고 있다. 우리나라 역시 공공데이터 포탈(data.go.kr)을 통해 2014년 6월말기준 9,378건³⁾의 데이터 셋을 공개하고 있고, 미국의

3) 송태민 외(2014), 보건복지 빅데이터 효율적 관리방안 연구, 한국보건사회연구원.

표 2. 사회복지분야 공공 빅데이터 현황

구분	보유기관	내용	공개여부
사회보장정보	사회보장정보원	- 각 부처 및 정보보유기관에서 제공하고 있는 복지사업정보 및 지원대상자의 자격정보, 수급이력정보를 통합관리 - 복지대상자 선정·사후관리를 위해 45개기관 552종의 소득·재산자료 및 서비스 이력정보 연계	미공개
복지콜센터 상담 데이터	복지콜129센터 등	- 상담이력	미공개
사회서비스 전자바우처	사회보장정보원	- 정부와 지자체가 사회로부터 도움을 필요로 하는 사람에게 돌봄, 일상생활 지원, 사회적응지원, 문화체험 등의 서비스를 제공하는 것을 전산처리	미공개
보육통합정보	사회보장정보원	- 보육바우처 운영 및 행정지원 정보	미공개
한국복지패널	한국보건사회연구원	- 빈곤층, 근로빈곤층, 차상위층 등의 규모 및 생활실태 변화를 동태적으로 파악 - 인구집단별 생활실태 및 경제활동, 건강·의료, 주거, 소득, 복지서비스 이용, 복지욕구	공 개

경우 20만개 정도의 데이터 셋을 이용할 수 있으며, 영국은 1만개정도, 호주는 3,164개 데이터 셋을 이용할 수 있다⁴⁾. 특히 우리나라의 경우 공개된 정보 대부분은 통계나 요약정보 형태로 제공되고 있어 <표 1>이나 <표 2>에 나열된 정보는 공공 데이터 포털에 제공되지 않고 있는데 만일 공개된다고 하더라도 데이터를 보유하고 있는 각 기관이 직접 서비스하고 있다. 다시 말해서 통계정보나 요약정보는 공공데이터 포털을 통하여 공개되고 있어 접근이 용이하지만 심층 분석이나 수요예측 같은 고도의 시뮬레이션 작업에 사용될 <표 1>, <표 2>와 같은 원시자료(개인정보의 익명화 처리된 자료) 및 빅데이터는 공공데이터 포털에서 공개되는 경우가 극히 드문 것을 확인할 수 있다.

3. 보건의료 및 사회복지 빅데이터의 활용

보건복지분야 공공 빅데이터에 대한 공개 및 활용수준은 전체적으로 봤을 때 그다지 높다고 할 수 없는 상황인데, 그럼에도 보건의료분야는 사회복지분야와 비교하면 사정이 낫다. 데이터의 활용 정도는 이용 가능한 데이터의 수와 밀접한 관련이 있으며, 뿐만 아니라 데이터의 접근 용이성, 데이터의 가치 및 품질 우수성 등에 따라서도 활용도가 좌우된다고 할 수 있다. 보건복지부문의 공공 빅데이터는 <표 1>과 <표 2>에 나열되었듯이 그 수가 적고, 더군다나 거기에 미공개 데이터마저 포함하고 있다. 이것은 데이터에 접근할 수 있는 기회 의 희소함, 즉 활용 가능성이 낮음을 의미한다. 보

4) <http://www.ncoa.gov.au/report/phase-one/part-b/10-5-data.html> 2015.07.03. 인출

건복지 공공 빅데이터는 학문적인 연구나 의학적 전문연구분야에서 주로 활용되고 있을 뿐 상업적 비즈니스에 활용되는 사례는 전무하다. 국내의 공공 데이터를 활용한 비즈니스 모델 사례들을 정리해 보면, <표 3>에서 보여주는 바와 같이 총주에 소재한 장애인 복지시설 및 장애인정보 제공 서비스인 ‘총주 나눔의 집 서비스’를 제외하면 나머지는 보건의료와 관련된 서비스이다. 이것이 의미하는 바는 사회복지영역보다 보건의료영역에서 공개된 공공데이터가 많기 때문에 그만큼 활용사례도 많다는 것이며, 결국 공개 데이터가 많을수록 활용 수준도 높아진다는 것을 보여주는 것이다.

공공 빅데이터를 활용한 대표적인 서비스로는 건강보험 빅데이터가 사용된 ‘국민건강 알람서비스’, ‘건강검진 및 진료정보’, ‘대시증후군맞춤정보’, ‘뇌졸중위험예측프로그램’ 등과 ‘운전면허 발급 간소화 서비스’가 있다. 국민건강 알람서비스는 국민건강보험공단이 공공 빅데이터인 진료내역을 수집한 국민건강정보DB와 소셜미디어 정보를 융합하여 감기, 눈병, 식중독, 피부염 등 주요 질병의 위험도 동향 및 알람을 제공하는 대국민 서비스이다. 운전면허 발급 간소화 서비스는 국민건강보험공단의 건강검진 결과정보를 도로교통공단과 공동으로 이용하여 운전면허 취득·갱신 시 신체검사를 대신하는 서비스이다.

공공 빅데이터의 활용은 데이터를 보유하고 있는 기관 중심으로만 활발히 이루어지는 경향이 두드러진다(표 3). <표 3>에서 나열한 사례는 서비스 중심의 사례를 제시한 것에 불과하지만, 결

국, 보건복지분야 공공 빅데이터는 산업계나 연구기관, 혹은 학계 등 다양한 영역에서 폭넓게 활용되지 못하고 있다는 것을 의미하기도 한다. 물론 최근 논문이나 선행연구들을 보면 학술적·의학적·정책적 목적으로 공공 빅데이터를 활용하는 사례가 과거보다 증가하고 있는 경향을 보이고는 있다. 중요한 것은 공공 빅데이터가 공개적이지 않은 이유로 인해 데이터 접근성이 상당히 떨어지고, 이로 인해 적극적으로 활용할 기회가 줄어들어가는 점이다.

컴퓨팅 파워의 향상, 고급분석, 그리고 빅데이터의 출현은 시민들의 욕구와 행태를 이해하는데 좋은 기회를 제공하고 있다. 전통적인 과학적 조사는 데이터 수집을 통해 검증되는 가설을 가지고 시작했으나, 빅데이터 분석은 기존의 대규모 데이터로부터 분석을 통해 도출해낼 수 있는 결론에서 통찰력을 가지는 기존의 방법과 반대방향에서 작업할 수 있다. 이제는 과거 3V(Volume, Variety, Velocity)의 개념만 가지고 빅데이터라고 온전히 말할 수는 없다. 3V개념 외에도 데이터의 복잡성, 적시성까지도 고려하여 빅데이터를 얘기한다⁵⁾. 따라서 빅데이터는 통찰력을 이끌어 내고, 즉각적인 조정이 가능한 변칙적인 정부 프로그램과 서비스들을 식별하는 가능성을 제공한다. 한편, 제약업계 전문가들, 의료 제공자 및 지불자들(건강보험공단, 보험사 등)은 빅데이터 분석을 통해 인사이트를 얻으려 하고 있다. 빅데이터는 산업계가 보건의료 및 사회복지 품질의 변동성, 혹은 보건의료 소비를 증가시키는 것과 관련된 문제들을 검토하는

5) Peter, G.,etc.,(2013) *The 'big data' revolution in healthcare*, Center for US Health System Reform Business Technology Office, McKinesy&Company.

표 3. 보건복지관련 공공데이터 활용 서비스 현황

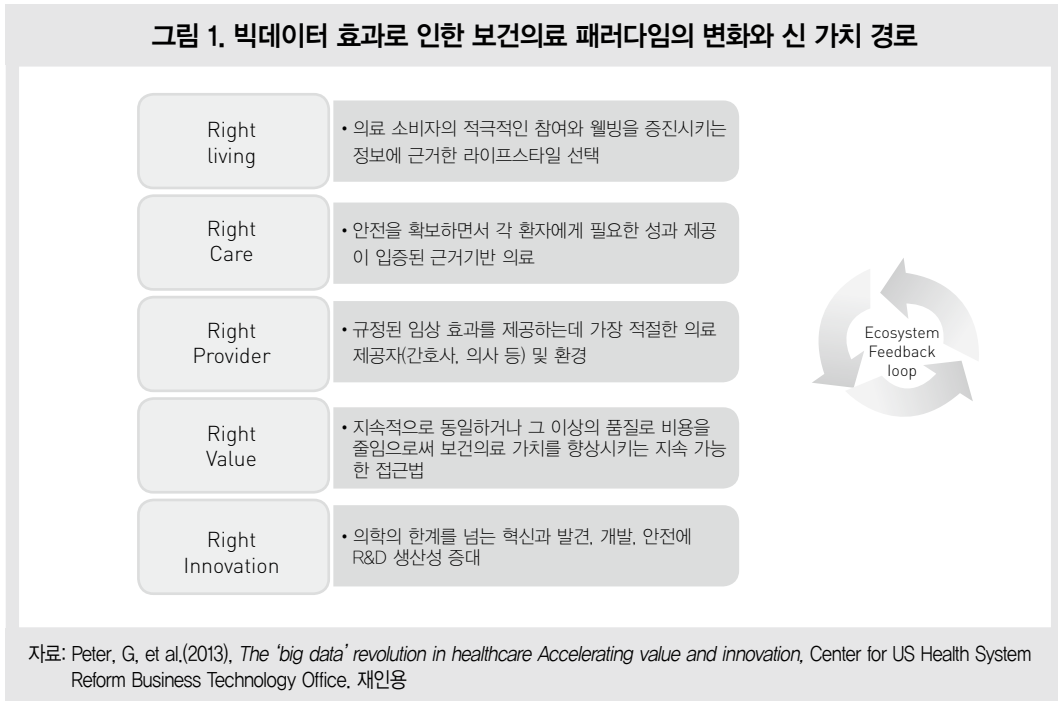
서비스	주요내용	데이터 종류
국민건강 알람서비스	4단계 건강위험 예보 발령, 개인건강기록 시스템을 통한 맞춤형 건강정보 제공	진료내역, 의약품 처방, 건강검진 정보 등 (국민건강보험공단)
건강검진 진료정보	검진결과 맞춤형건강서비스	건강검진 정보, 의료기관 이용 내역 등 (국민건강보험공단)
뇌졸중위험예측 프로그램	뇌졸중과 관련된 고혈압, 콜레스테롤, 생활습관, 가족 력, 환경요인 등을 기초로 10년 이내에 뇌졸중에 걸릴 위험도 평가	건강검진정보, 문진정보 등 (국민건강보험공단)
대사증후군 맞춤정보	대사증후군 요소와 관련한 건강상태 및 위험요인별 맞춤형 처방정보 제공	건강검진정보, 문진정보 등 (국민건강보험공단)
운전면허 발급 간소화 서비스	국가건강검진정보 중 운전면허 적성검사에 필요한 시력·청력 정보 공동이용	국가건강검진정보 (국민건강보험공단)
갑상선암 의사에 게 꼭 물어보세요	갑상선암 정보, 종합병원 전문의와 1:1 맞춤형 상담	병원정보 (건강보험심사평가원)
병원 약국찾기	위치기반 서비스와 연동하여 내 주변의 병원, 약국, 응급의료기관 위치정보 제공	전국 응급의료정보 (국립중앙의료원)
MediMap	서울 소재 병원 규모별, 카테고리별 병원 위치, 연락처 정보 제공	전국병원정보 (공공데이터활용지원센터)
처방약 토탈 검색	약국 위치, 개폐정보	약국정보 (국립중앙의료원)
4th-Life	요양병원 정보 제공	-
옴은 서비스	식단 재료의 합리적 구매기준을 위한 식단정보, 가격 정보, 성분정보	식단정보, 가격정보, 성분정보 (식품의약품안전처)
충주 나눔의 집	충주시 소재 장애인 복지시설 관련 정보 및 장애인 등록현황 정보 제공	장애인 등록정보 (충청북도 충주시)
대전당직병원	대전지역 당직병원 응급실정보, 전화번호, 주소, 병원 등급 등 정보 제공	당직의료기관정보 (대전광역시)
헬스온스토리	상황인식기술 통한 사용자 맞춤형 건강정보 자동제공	보건기상지수정보 (기상청)

자료: 1) <http://forecast.nhis.or.kr/index.do> 2015.07.14. 인출
2) 한국정보화진흥원(2014), 공공데이터 활용 사례집

데 도움을 줄 수 있을 것으로 본다. 특히, 해외의 보
건의료분야 연구자들은 특이한 조건에 있는 환자
에게 투입된 치료가 효과적인지를 보는데 데이터
를 마이닝 하는가 하면, 약물 부작용이나 혹은 병
원 재입원과 관련된 패턴을 찾는데, 그리고 비용

절감이나 환자를 지원함에 있어 다른 중요정보를
얻는데 빅데이터를 적극 활용하고 있다. 빅데이터
는 [그림 1]과 같이 정보에 근거한 생활스타일을
선택하게 되고, 근거 중심의 의료, 최적의 임상 효
과를 볼 수 있는 의료제공자 및 환경, 보건의료 가

그림 1. 빅데이터 효과로 인한 보건의료 패러다임의 변화와 신 가치 경로



치를 향상시키는 지속 가능한 접근법, 의학의 한계를 넘어서는 혁신 등 패러다임 변화를 이끌고 있으며, 이것은 새로운 가치의 경로가 되고 있다.

사실 선진국들의 공공 빅데이터 활용수준에 비하면 우리나라는 한참 뒤쳐져 있다고 봐야 한다. 예컨대 영국의 경우, 민간영역일지라도 공공의 이익과 지역사회의 안녕을 목적으로 한다면 공공 빅데이터의 접근을 허용하고 있을 만큼 국가차원의 적극적인 빅데이터 활용을 권장하고 있음과 동시에 공공 빅데이터를 국가의 미래 성장동력으로 인식하고 있다⁶⁾. 우리나라도 조금 늦은 감이 있지만 미래창조과학부를 중심으로 빅데이터 기술 및 서비스의 개발, 법·제도 개선 그리고 빅데이터를 이

용한 시범사업 등을 추진하고 있다. 보건의료 및 사회복지영역에서도 공공 빅데이터를 지금보다 적극적으로 활용할 수 있는 환경과 기반이 조성된다면 정부가 추진하고 있는 맞춤형 복지를 제공하는데 있어 정보에 기반한 생애주기별, 취약조건별 개인화된 보건·복지 서비스의 과학적 접근이 가능해진다. 즉, 건강이력 및 사회경제적 능력 등 개인의 정보를 종합적으로 프로파일링함으로써 개인의 상황과 환경에 최적화된 서비스를 제안할 수 있다. 뿐만 아니라 공공 빅데이터는 민간영역의 SNS 데이터와 연계함으로써 정부의 지원 경계에서 진입하지 못하고 있는 복지사각지대 조기발굴 등 정부정책의 사각지대를 모니터링하고 정책

6) Health and Social Care Information Centre(2014), *A strategy for the Health and Social Care Information Centre 2013-2015*, Health and Social Care Information Centre.

표 4. 보건의료 및 사회보장 공공 빅데이터 활용 가능한 R&D 사업 예시

구분	모델
사회보장 ¹⁾	<ul style="list-style-type: none"> - 사회보장정책 부적정 이용 대응 모델 - 빅데이터를 활용한 복지 사각지대 조기발견 모델 - 빅데이터를 활용한 복지자원 관리 모델 - 상담데이터를 활용한 지자체 통합사례관리 담당자 의사결정 지원 모델
보건의료 ²⁾	<ul style="list-style-type: none"> - 유전체 의료 빅데이터 마이닝을 통한 개인별 질환 발생예측 기술 개발 - 지역별 일차의료지표 산출을 통한 지역주민의 건강 및 질병 감시 및 예보시스템 개발 - 건강보험·진료비의 허위·부당청구 및 의료서비스 오남용 감시 및 관리 시스템 개발 - 의료보장성 강화를 위한 건강보험 재정규모 감시 및 예측 시스템 개발 - 건강보험 수가 적합성 및 부과 형평성 분석시스템 개발 - 보건의료 빅데이터 시계열 분석 기반 복합만성질환 예측 기술 - 진단/처방의 신뢰도 향상을 위한 보건의료 빅데이터 기반 진단결정 지원시스템 개발 - 빅데이터 기반 적정 의료기관 이용모형 개발

자료: 1) 사회보장부 빅데이터(Big Data) R&D 사업기획 연구
 2) 보건의료 Big Data R&D 사업기획 연구

을 보완하는데 중요한 근거가 된다. 따라서 보건복지지역의 공공 빅데이터 활용영역은 다양할 것이며, 이미 보건의료 및 사회보장영역의 빅데이터를 활용한 미래 R&D 사업기획에 대한 연구가 진행된 바 있다. 선행 연구들에 의하면 공공 빅데이터는 정부의 보건복지 지원·서비스 전달체계의 적정성을 평가하는 수단으로서 그 가치가 있는 것으로 파악되었다. 요컨대 의료나 혹은 복지서비스의 중복 또는 부적정 수급을 모니터링하는데 공공 빅데이터의 활용가치가 강조되고 있다. 향후 보건복지 공공 빅데이터를 활용할 때 가능한 R&D 사업 모델을 정리해보면 <표 4>와 같다.

4. 보건복지분야 공공 빅데이터 활성화를 위한 과제

고령화가 급격히 진행되고 있는 우리나라에서

빅데이터의 영향은 인구고령화(건강 서비스)의 장기적 도전과 재정적 공공부채 위기(재정, 부동산, 행정), 지식경제 구현(지식 집약형 섹터) 등에 도움을 줄 수 있다. 이미 의료영역의 많은 이해관계자들은 최상의 과학적 근거에 바탕을 두고 환자 개개인에게 적합한 치료방법을 결정하는데 근거기반의 의료 개념을 받아들이기 시작하였다. 선진 각국은 빅데이터가 국가의 미래 성장동력중 하나로 인식하고 있으며, 특히 빅데이터 수집 및 분석과 관련된 기술이나 기법 등의 개발에 선제적 위치를 점유하려는 경쟁이 치열해지고 있다. 이러한 흐름 속에서 세계적으로도 찾아보기 힘든 우리나라만이 보유하고 있는 엄청난 가치의 건강보험정보를 비롯하여 보건의료 및 사회보장영역 공공 빅데이터는 지금보다 훨씬 많이 공개되고 활성화될 필요성이 있다. 따라서 공공 빅데이터의 활성화를 위하여 해결되어야 할 과제로는 개인정보 보호기술의 고도화, 데이터 표준화, 빅데이터의 공개종류 및 범

위 확대, 그리고 빅데이터 운영센터의 설치가 절실하다.

우선 공공 빅데이터의 활성화에서 가장 큰 걸림돌은 개인정보보호법의 준수에 있다. 공공 빅데이터는 고유식별정보뿐 아니라 건강이나 복지급여와 관련된 개인정보 및 민감 정보를 상당히 포함하고 있다. 이러한 개인정보는 「개인정보보호법」 제17조제1항에 의해 민감 정보와 같은 개인정보를 제3자에게 제공하려면 정보주체의 동의를 받아야 하는데 엄청난 수의 정보주체에게 동의를 받는 것은 상당한 시간과 노력 외에도 정보 제공 시마다 반복적으로 동의를 받는 것이 사실상 불가능하다. 한편, 빅데이터의 처리과정에서 수집된 데이터의 개인정보는 비식별화 조치하여 재식별의 가능성이 없을 때 제3자에게 제공 가능하다(「빅데이터 개인정보보호 가이드라인」 제5조, 제10조). 이 때 비식별화 조치라 함은 고유식별정보와 같은 개인정보 및 민감 정보에 대하여 암호화기술이나 익명화기술로 개인을 식별할 수 없도록 처리한 후 원본에서 대체하는 것을 말한다. 비식별화 메카니즘은 빅데이터 활용단계별로 다른 익명화기술을 적용할 수도 있고 데이터가 가진 특성과 분석수준 등에 의해 상황별로 적용기술을 달리한다. 이미 개인정보 비식별화 처리에 데이터 마스킹이나 총계처리, 가명처리 같은 익명화 기술이 적용되고 있기는 하지만 처리방식이나 기술에 따라 재식별의 가능성을 완전히 배제할 수는 없다. 최근 금융사의 개인정보 대량유출 등 개인정보 불법유통 및 이용에 따른 피해가 갈수록 늘어나면서 개인정보보호법 역시 더욱 강화되고 있는 마당에 개인정보에 대한 재식별의 위험성을 최소화하는 암호화내지 익명화기술의 개발은 한층 더 정교하고 고도화 되어야만 한

다. 이러한 암호화·익명화 기술이 뒷받침된다면 정보주체는 개인정보의 안전한 사용이라는 신뢰를 바탕으로 개인정보의 공익적 활용을 묵인할 수 있을 것이다. 따라서 암호화 기술이나 혹은 익명화 기술의 수준을 높이는 신기술의 개발 등과 관련된 정부의 재정적 지원과 인력의 양성이 요구된다. 한편, 영국의 Motivated Intruder Test 같은 비식별화 조치에 대한 재식별 위험 모니터링 기술도 개발, 발전시켜 개인정보의 올바른 활용과 제공을 유도해야 한다.

두 번째 데이터의 공개·공유는 데이터 표준화를 전제로 한다. <표 1>과 <표 2>를 보면 보건복지분야에도 다양한 빅데이터가 있음을 알 수 있는데, 빅데이터는 타 정보와의 연계·결합으로 새로운 빅데이터를 생산하고, 이를 활용하여 소스 데이터에서 발견할 수 없던 신 가치를 창출하는 것이 특징이다. 그러나 개인정보보호법의 강화로 인해 데이터간 연계·결합에 필요한 고유식별정보의 수집이 제한적일뿐 아니라 원천적으로 이용할 수 없는 상황이 발생하고 있다. 다행히 최근 학계에서는 연계의 정확성을 높일 수 있는 기법이나 방식에 관한 논의가 활발히 이루어지고 있는 것이 고무적이기는 하다. 타 정보와의 연계나 결합은 결합하려는 각 데이터에 개인을 식별할 수 있는 정보가 있다면 가장 정확한 결합을 이끌어내지만 개인정보보호법에 의해 고유식별정보의 사용이 불가능하다면 그것을 대신할 대체키가 있어야 하는데, 데이터가 표준화되었다면 대체키를 선정하는데 훨씬 용이할 수 있다. 한편 개별적으로 수집하고 있던 정보, 예컨대 최근 헬스케어 및 PHR 서비스에 대한 관심이 높아지면서 비즈니스 목적에 따라 수집되는 데이터의 속성과 타입 등이 상이할 수 있으므로 데

이터 표준화는 병원단위 또는 비즈니스 조직의 집적된 빅데이터를, 혹은 조직내에서의 흩어져 있던 데이터를 통합하는데 반드시 요구되는 필수 조치이다. 데이터 표준화가 적용된 데이터라면 누가 사용하더라도 동일한 데이터는 똑같은 결과를 산출하는 것이 원칙이므로 분석결과 및 해석에 대한 오해나 오류를 방지할 수 있고, 한편으로는 이용자의 혼란을 최소화할 수 있으며, 결과나 통계에 대한 상호 비교가 명확해 질 수 있다. 결국, 데이터 표준화는 데이터의 공개·공유나 데이터간 결합·연계를 용이하게 함으로써 궁극적으로는 데이터의 활용성까지도 높이는 결과를 낳는다.

세 번째 빅데이터의 공개 종류 및 범위 확대와 관련하여 건강보험공단이 집적하고 있는 진료정보는 방대한 원시 빅데이터를 표본코호트로 구성하였고, 표본코호트는 심사를 통해 제한적으로 공개가 허용되고 있다. 그 외에도 유전체 빅데이터를 비롯하여 보건의료분야 공공 빅데이터는 타 영역보다 공유수준이 높다. 해외 사례들을 보더라도 각국의 보건의료분야 공공 빅데이터는 국가차원의 공개를 유도하고 있을 뿐 아니라 ICT와의 융합을 통해 다양한 영역에서 활용되고 있다. 그러나 사회보장영역의 공공 빅데이터는 <표 2>만 보더라도 현재 공개된 데이터가 거의 없다고 해야 맞을 것 같다. 사회보장정보시스템을 통해 수집되는 사회보장 빅데이터는 사회보장서비스의 수요예측이나 복지사각지대 패턴분석, 복지서비스 및 자원의 수요·공급 적정성 평가, 개인화된 맞춤형 복지급여 설계 및 실행에 대한 평가 등 사회보장부문 가치창출의 기초자료로서 활용 가치를 가진다. 다양한 소스로부터 풍부한 데이터를 가지고 분석을 하게 되면 보다 나은 정부정책이 나올 수 있고, 그 결과 보

다 나은 정책성과가 생산될 수 있다. 결국, 다양하고 풍성한 데이터를 활용할 수 있을 때 급변하는 환경에서 정확한 진단과 대응을 가능케 한다. 따라서 사회보장영역도 건강보험 빅데이터와 마찬가지로 사회보장 공공 빅데이터의 공개·공유가 요구된다. 공공 빅데이터는 민간영역이 수집할 수 없는 정보들을 집적하고 있기 때문에 민간부문의 관심이 상당히 높다. 국민의 건강과 행복한 삶을 유지하는데 민간의 역할이 빠질 수는 없다. 이에 정부는 민간영역과의 협력 및 보조를 유지해야 하며, 거기에 빅데이터라는 도구를 사용하여 정책성과 및 서비스를 개선하는 기회로 잡아야 한다. 그러려면 민간영역까지도 공공 빅데이터의 이용을 허용할 수 있어야 한다.

다섯째 보건복지 빅데이터 운영센터의 설치와 관련해서 현재 정보 수집기관별로 운영되고 있는 빅데이터는 개별 데이터 나름대로 역할을 하고 있지만 기관별로 집적된 정보 내에서만 가지는 가치에 국한되어 있다. 그러나 복잡하게 변하고 있는 사회를 다각적으로 통찰하려면 투입되는 정보가 다양해야 하므로 한쪽에 치우친 정보로만 분석하는 것보다 보건의료와 사회보장영역을 아우르는 통합된 빅데이터를 이용하려는 욕구가 증가하고 있다. 만일 빅데이터 운영센터가 설치되면 빅데이터 수집, 공유, 분석, 활용에 필요한 인프라를 구축하여 보건의료와 사회보장 영역간 경계를 뛰어넘는 데이터의 통합을 가능케 하고, 이런 데이터는 미래 성장동력으로서 학문적, 정책적, 비즈니스적, 의학적인 엄청난 가치를 창출하는데 기여하게 된다. 그 밖에도 보건복지 빅데이터를 운영하는 전담센터가 설치되면 개인정보의 보호와 관련된 익명화 기술의 적용에 있어서도 명확한 기준과 일관된

원칙을 적용·유지하는 장점을 가진다. 한편, 데이터의 표준을 개발하고 체계적인 적용을 통해 각 영역별 빅데이터를 수집하는 단계부터 표준을 준수함으로써 공공 빅데이터의 품질을 유지·향상시키는 데 기여할 수 있다. 대표적인 해외사례로 영국의 보건복지정보센터(Health and Social Care Information Centre)를 들 수 있다. 영국 보건복지정보센터는 빅데이터를 수집·관리·운영함에 있어 데이터 보안 유지, 데이터의 품질 개선, 정보 표준화를 통한 상호운용성 개선, 정보수집 현장의 부담 최소화를 위한 역할을 수행한다. 뿐만 아니라 이종 형태의 다양한 데이터를 연계·결합하는 업무를 관여하고 있고, 보건복지정보센터가 집적하고 있는 보건의료 빅데이터를 민간영역에까지도 공개하는 역할도 맡고 있다. 따라서 보건복지부문의 빅데이터 운영센터의 설치에 공공 빅데이터의 품질을 보장하면서 데이터 활용의 극대화를 리드하고, 복잡한 사회현상과 급변하는 실세계를 반영하는 R&D분야, 다각적인 서비스 및 정책의 발굴 뿐만 아니라 평가를 하는 기초자료로써 제공이 용이해진다.

5. 나가며

보건복지 공공 빅데이터의 활용은 정책 및 제도를 설계하고 집행하는 공급자·생산자 측면과 이를 소비하고 활용하는 수요자·소비자 측면에서 비교했을 때, 수요자의 빅데이터 이용에 대한 욕구는 공급자보다 훨씬 강하다고 봐야 한다. 왜냐하면 공공 빅데이터 공급자인 정부나 기관 등은 해당 데이터의 수집부터 폐기에 이르는 데이터 라이프사

이클을 설계할 때 이미 활용분야까지 염두에 두고 설계되므로 활용수준은 어느 정도 정해지지만, 다양한 분야에 있는 수요자는 크게 고려되지 않기 때문이다. 즉, 빅데이터 수요자는 데이터를 이용할 수 있는 기회가 생산자보다는 낮다는 것이다. 더군다나 최근 공공 빅데이터의 이용 요구가 학계 및 연구기관, 나아가 민간부문에서도 크게 증가하고 있는 실정이다. 그럼에도 개인정보보호법이나 데이터 미공개 등과 같은 법·제도적 한계로 인해 수요자의 공공 빅데이터 활용에 대한 욕구는 충족되지 못하고 있다. 특히, 공공 빅데이터의 경우 감히 민간부문은 염두조차 낼 수 없는 엄청난 물적, 인적자원이 투입된 미래 경쟁력을 결정짓는 핵심요소로 인식되고 있으며, 복잡하고 불확실한 미래를 예측하여 사회적 합의를 이끌어내는 기초자료로서 활용가치도 상당하다. 그렇다고 공공 빅데이터를 활용하여 그 가치를 인증 받은 사례가 현재까지 그리 많은 것도 아니다. 다만 가치 창출의 가능성을 추측할 뿐이다. 우리나라는 어떤 나라도 가지고 있지 못한 전 국민의 의료정보와 검진자료, 사회보장정보 등 방대한 양의 공공 빅데이터를 집적하고 있다. 뿐만 아니라 최신 정보통신기술에 있어서도 선진국과 비교하여 경쟁우위를 선점하고 있어 보건의료 및 산업, 사회보장영역과 관련된 비즈니스 모델을 창출하는데 선도적 위치를 차지할 기회가 충분하다. 그럼에도 불구하고 선진국보다 활용 측면에서 뒤쳐져 있어, 가치 있는 공공 자원의 투입대비 효과를 극대화하기 위해서는 적극적인 사용을 권장해야 하는데, 이 때 데이터 사용으로 인한 리스크나 혹은 피해가 데이터 사용에 따른 공공의 안녕과 공익적 이익보다 작다고 한다면 공공 빅데이터의 공개 및 활용을 허용해야 마땅하다. 물

론 공공 빅데이터의 공개·제공으로 인해 정보주체의 손해나 혹은 리스크가 발생할 경우 불법적 사용에 대한 현재의 처벌수준보다 훨씬 강력하고 엄중한 중벌에 처할 수 있는 법적 근거를 마련해야만

공공 데이터의 올바른 사용과 개인정보의 공익적 사용에 대한 국민의 부정적 시각을 누그러뜨릴 수 있고, 다양한 분야에서 공공 빅데이터의 이용을 극대화시킬 수 있으리라 사료된다. ■