

## 데이터 3법과 보건의료 빅데이터



오미애

한국보건사회연구원 정보통계연구실  
빅데이터·정보연구센터 센터장

### I 들어가며

빅데이터는 의료, 금융, 복지, 경제 등 모든 분야에서 핵심 전략으로 활용되고 있다. 보건의료 분야의 경우, 보건의료 빅데이터 플랫폼 시범사업을 추진하고 있으며 지난 5월에는 ‘바이오·헬스산업 혁신전략’을 발표, 바이오·헬스 분야를 국민 생명과 건강을 보장하기 위한 핵심 산업으로 보고 차세대 선도 산업으로 육성할 계획이다. 또한 ‘국가 바이오 빅데이터’ 등 5대 빅데이터 플랫폼 구축도 구상하고 있다.

지난 1월에 데이터 3법이 통과되면서 보건의료 분야의 빅데이터 활용 가능성은 더 커졌지만 개정된 개인정보 개념 정의(입수가능성 등 개인을 알아보는데 소요되는 시간, 비용, 기술 등을 합리적으로 고려)부터 다양한 해석의 여지가 있다. 건강정보는 민감정보로 규정되며, “개인정보 보호법제 체계의 맥락”에서의 주요 개념은 다음과 같다.

표 1

건강정보 관련 주요 개념

주요 개념	관련 법	정의
의료정보	「보건의료기본법」 제3조 제6호	보건의료정보
유전정보	생명유리 및 안전에 관한 법률 제2조 제11호, 제14호	인체 구성물 또는 이로부터 분리된 물질인 인체유래물을 분석하여 얻은 개인의 유전적 특징에 관한 정보
생체정보	정보통신망 이용촉진 및 정보보호 등에 관한 법률 시행령 제9조의2 제1항제1호	지문, 홍채, 음성, 필적과 같이 개인을 식별할 수 있는 신체적 또는 행동적 특징에 관한 정보

출처: 고학수. 2019. 보건의료 빅데이터의 보호 및 활용을 위한 법적 기반 강화 연구. 요약 2페이지 재구성(<http://ai.re.kr/research-materials-and-publications/?pageid=1&lang=en&mod=document&uid=23>).

행정안전부, 방송통신위원회, 금융위원회, 개인정보보호위원회 등 데이터 3법 관계부처는 2월 중 시행령 개정안을 만들고, 3월에 행정규칙 개정안을 마련하는 등 법 시행 시점에 맞춰 가이드라인과 해설서를 배포해 가명정보 활용 범위, 데이터 결합 방법 및 절차 등을 명확히 한다는 계획을 발표하였다. 그리고 대통령직속 4차산업혁명위원회에서는 ‘4차위 데이터옴부즈만’ 서비스를 실시해 데이터 관련 질의 및 제언을 청취하고, 관계부처의 협조를 받아 검토의견을 전달하며, 부처에게 데이터 3법 후속조치를 권고할 예정이다.

이렇듯, 데이터 3법 개정안이 통과되었지만 하위법령 및 개인정보보호 정책을 구체화하는 작업이 진행 중에 있다. 본 고에서는 보건의료 빅데이터의 활용 현황과 데이터 3법에 따른 변화들을 살펴보고자 한다.

## II 보건의료 빅데이터 활용 현황

### 1. 보건의료 빅데이터 플랫폼<sup>1</sup>

보건의료 정보를 보유하고 있는 공공기관은 **건강보험공단, 건강보험심사평가원, 질병관리본부, 국립암센터** 등이 있다. 건강보험공단은 국민건강검진, 보험료 청구·지급내역 및 예방접종 등 건강에 관한 주요 빅데이터를 다수 보유하고 있으며, 건강보험심사평가원은 상세 보험료 청구·지급내역, 약제 처방내역, 의료기관 인력·장비 정보 등 진료현장의 상세한 데이터를 보유하고 있다. 질병관리본부는 인체자원은행, 만성질환·감염병질환·영양 등 조사 정보, 검역·감염병 등 공중보건에 관한 보건행정 정보를 다수 보유하고 있고, 국립암센터는 암 등록 정보, 암종별 레지스트리 정보, 암 유전체 정보, 국립암통계 정보 등 암에 관련된 종합적 정보를 보유하고 있다.

4개의 공공기관에 분산된 정보를 개인단위로 연계한다면 그동안 데이터 분절로 할 수 없었던 연구가 가능해진다. 보건의료 데이터의 연계에 따른 연구 범위 사례는 다음과 같다.

<sup>1</sup> 주요 내용은 “보건복지부 사전공표정보: 보건의료 빅데이터 시범사업 계획(최종), 등록일:2018-12-07”을 참고하여 기술하였음.

표 2

데이터 연계 여부에 따라  
수행 가능한 연구 범위  
확대 사례

연계되는 데이터	기관별 분절 상태	데이터 연계 후
암등록(국립암센터) ↔ 건강보험 가입자 정보(건강보험공단) ↔ 영양조사(질병관리본부)	암-소득-영양상태 간 관계 조사 불가	소득별 암증 분포, 항암치료 대 응 시간, 영양상태와의 관계 등 파악
가입자 정보(건강보험공단) ↔ 건강검진(건강보험공단) ↔ 세부진료내역(건강보험심사평가원) ↔ 암등록(국립암센터) ↔	건강검진 상 위험신호와 진료내 역, 암 유형 간 관계 조사불가	가입자 소득·재산의 여건에 따 라 건강검진 시 발생된 의심증 세와 진료내역 간 관계, 암종류 및 발생시점, 예후까지 분석
가입자 정보(건강보험공단) ↔ 약처방·진료내역(건강보험심사평가원) ↔ 지역별 의료기관 현황(건강보험심사평가원) ↔ 건강검진(건강보험공단)	약, 가입자 정보, 검진 등 기관 내 정보만 연계 가능	지역별 의료기관 현황, 가입자 소득·재산 등의 여건에 따라 건강검진 이상신호 대응방법에 의 사회적 격차 파악 가능

출처: 보건복지부 사전공표정보. 2018.12.07. 보건의료 빅데이터 시범사업 계획(최종).

보건복지부는 건강보험공단, 건강보험심사평가원, 질병관리본부, 국립암센터 4개 기관의 정보를 연계하여 연구를 지원하는 보건의료 빅데이터 시범사업을 추진 중에 있으며, 시민, 환자 단체를 포함한 보건의료 빅데이터 정책심의위원회를 구성하고 시범사업의 추진방안, 입법 필요사항, 데이터 활용방안 등의 논의가 진행되고 있다.

얼마 전, 정부는 데이터 3법 개정안 시행이 보건의료 빅데이터 플랫폼에 미치는 영향 분석을 통해 플랫폼 개선안을 마련하고자, ‘보건의료 빅데이터 플랫폼 관련 기준개발·평가 연구’ 사업을 발주한다고 발표하였다.<sup>2</sup> 이 연구는 도출된 시범사업 평가결과, 데이터 3법 시행 시 파급효과, 분석결과 등을 종합적으로 고려하여 플랫폼 발전방향을 제시하게 된다. 이 연구가 잘 수행된다면 데이터 3법 개정안에 따른 보건의료 빅데이터 플랫폼 지향점이 명확해질 수 있을 것이다.

## 2. 국가 바이오 빅데이터 구축 시범사업

지난 5월, 정부는 ‘바이오·헬스산업 혁신전략’을 발표할 때 최대 100만 명 규모의 ‘국가 바이오 빅데이터’를 구축하고, 정부 R&D투자에 2025년까지 연간 4조 원 이상을 투입할 예정이라고 밝혔다.

### [국가 바이오 빅데이터 구축 개요]

- (대상) 환자(암, 희귀난치질환 등) 40만 명, 건강인(환자 가족 포함) 60만 명 등
- (수집 및 축적) 병원을 통해 희망자에게 유전체 검사서비스 제공, 데이터(유전체+의료기록+건강정보) 수집, 국립중앙인체자원은행 등에 보관·관리
- (사업기간) 10년(3단계, 2+3+5년, '20~'29, 2단계부터는 예타 추진)

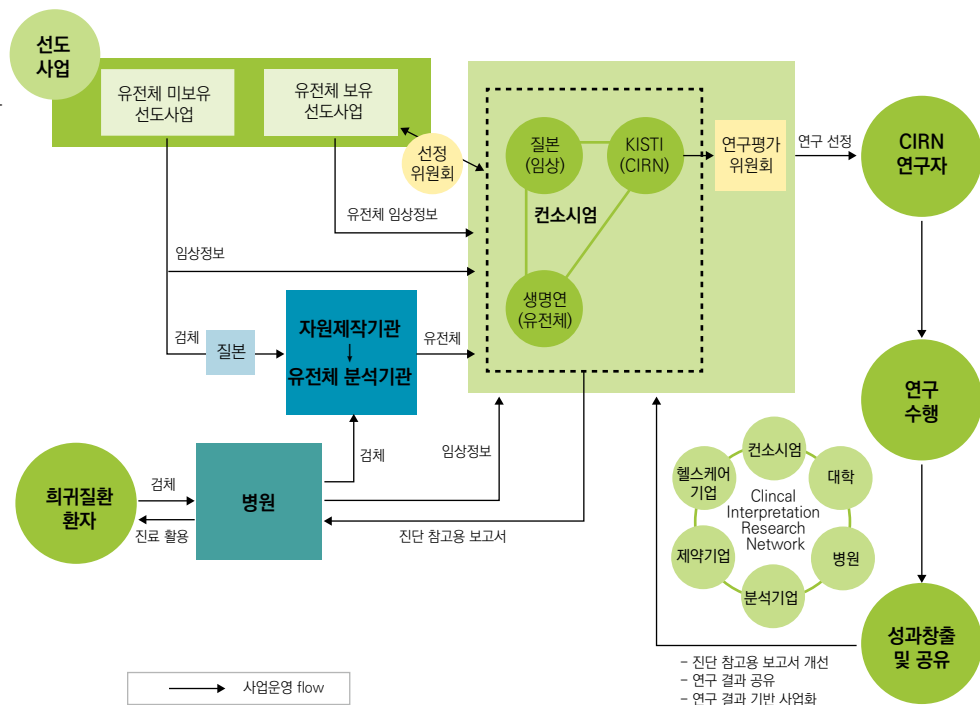
출처: 관계부처 합동. 2019.05.22. 바이오헬스 산업 혁신전략.

2 데일리메디. 백성주 기자. 2020.03.22. ‘빅데이터 플랫폼 기준개발·평가 연구 미국 등 선진국 사례 제시.’

앞으로 2년간의 시범사업(’20~’21) 동안 희귀질환자(’20년, ’21년 각 5,000명) 모집 및 선도사업<sup>3</sup>과의 연계(’20년, ’21년 각 5,000명 이상)를 통해 총 2만 명 이상의 임상 정보와 유전체 데이터를 구축할 예정이다. 바이오 빅데이터에는 민감정보인 임상정보·유전체 데이터가 포함되는데, 민감정보 수집에 따른 개인정보 유출 우려를 해소하기 위해 수집·생산 단계에서 각각의 식별자(임상정보 ID, 검체 ID, 유전체 ID) 부여 및 폐쇄망 구축을 통해 데이터를 관리한다.<sup>4</sup>

그림 1

국가 바이오 빅데이터  
구축 시범사업 운영 절차



출처: 관계부처 합동 보도자료. 2020.03.20. 국가 바이오 빅데이터 구축 시범사업을 위한 첫걸음.

보건복지부, 과학기술정보통신부, 산업통상자원부가 협력하여 시행되는 국가 바이오 빅데이터 구축 시범사업은 2020년 3월 12일에 추진위원회를 출범하였으며, 운영위원회, 사무국, 전문위원회, 5개 정부출연연구기관이 컨소시엄을 구성하여 운영 및 지원할 계획에 있다. 국가 바이오 빅데이터 구축 사업도 보건의료 빅데이터 플랫폼처럼 시범사업을 시행하면서 본사업으로 가기 위해 필요한 사항들을 점검하고, 데이터 3법 개정안을 발판으로 바이오헬스 산업의 기반을 마련하고자 한다.

3 한국인유전체역학조사사업, 울산 1만 명 게놈 프로젝트 등

4 관계부처 합동 보도자료. 2020.03.20. 국가 바이오 빅데이터 구축 시범사업을 위한 첫걸음.

### 3. 암 빅데이터 플랫폼

과학기술정보통신부는 ‘빅데이터 플랫폼 및 센터 구축’ 사업으로 분야별 플랫폼 10개소, 이와 연계된 기관별 센터 100개소를 구축하는데 3년간 총 1,516억 원을 투입한다고 발표하였다.<sup>5</sup> ‘빅데이터 플랫폼 및 센터 구축’ 사업은 공공과 민간이 협업해 빅데이터 센터 등에서 수집된 데이터를 분석·유통할 수 있는 플랫폼을 구축하고, 혁신 서비스를 발굴 확산해 데이터 기반의 가치창출 생태계를 조성하기 위한 것이다. 보건의료 분야에서는 국립암센터 주도 하에 10대 암종별 임상데이터를 융합하여 암 진단·치료 의사결정 및 항암 치료제 연구개발 등에 활용하고자 국내 10개 병원이 함께 사업을 진행하였다. 지난달에는 각 병원에 흩어진 암 관련 임상데이터를 수집하는 다기관 임상 라이브러리 플랫폼 ‘커넥트(CONNECT)’를 공개하였고, 3년 안에 주요 10개 암종 임상데이터 30%를 빅데이터로 구축, 치료에 활용할 계획이다.

데이터 3법과 관련해서 국립암센터 암빅데이터 센터장은 “개인정보 보호법 개정으로 가명 처리된 데이터를 연구용으로 활용하는 법적 근거가 마련된 만큼 협의를 통해 구축된 데이터를 사전 신청한 외부 기업이나 연구자도 활용할 수 있도록 공개 범위를 확대하는 것을 검토하고 있다”고 설명하였다.<sup>6</sup>

## III 나아가며

이 글에서는 보건의료 분야 내 정부에서 추진하고 있는 빅데이터 관련 사업과 데이터 3법으로 인한 변화를 살펴보았다. 정부의 보건의료 분야 빅데이터 추진에 있어서 건강정보가 민감정보이기 때문에 데이터 3법과 관련하여 여전히 우려의 목소리가 나오고 있다. “개정된 개인정보 보호법이 가명조치를 거치면 개인의 건강정보를 산업적·상업적 목적으로 활용할 길을 열어주었다고 평가하기 어렵다”는 의견도 있다.<sup>7</sup>

보건의료 분야에서 데이터 3법 개정안이 정책지원을 위한 빅데이터의 활용 가치 증대로 이어지기 위해서는 빅데이터 활용과 개인정보보호에 대한 국민의 정부신뢰도 제고가 우선되어야 할 것이다. 4차산업혁명위원회의 데이터옴부즈만이 기업을 위한 사회적 논의의 시작이라고 볼 수 있다면, 국민을 위한 사회적 논의의 장도 필요하다. 보건의료

5 IT DAILY. 김성수 기자. 2019.05.13.(월) 17:30:20.

6 전자신문. 정현정 기자. 2020.2.16. “암 빅데이터 플랫폼’ 가동…진단·치료 새 길 개척” ([https://www.etnews.com/20200214000128?mc=em\\_008\\_0001](https://www.etnews.com/20200214000128?mc=em_008_0001))

7 김재용. 개인 건강정보의 일차적 활용과 이차적 활용: 사회적 맥락과 거버넌스 확립 문제. 이화여자대학교 생명의료법연구소. <생명윤리정책연구>. 2020년 13권 2호 게재예정.

빅데이터 사업의 경우 2017년 말부터 사업을 추진하였으나, 정부가 국민건강정보를 상업적 목적으로 기업에 제공할 수 있다는 우려 때문에 시민사회단체에서 반대하였다. 그 후, 정부는 설득 작업을 거쳐 시민단체, 전문가가 참여하는 보건의료정책심의위원회를 구성하고, 여러 차례 회의를 진행하면서 안전장치를 마련한 후 시범사업을 추진해 나가는 것으로 의견을 모았다. 개인건강정보는 민감정보 중 하나로 유출이나 오용될 경우 사회적으로 큰 파장을 일으킬 수 있기에, 데이터 3법 개정안으로 인한 빅데이터 활용의 우려사항을 면밀히 검토하고, 공청회 등을 통해 국민들의 의견수렴 자리가 마련되어야 할 것이다. 서로에 대한 신뢰를 바탕으로 한 충분한 의견수렴 과정이 이루어져야 데이터 3법 개정에 따른 성공적인 보건의료 빅데이터의 활용 기회가 될 것이다.

#### 참고문헌

- 오미애. 2019. 국내외 빅데이터 활용 사례 분석과 시사점. 제 2차 사회보장포럼.
- 대통령직속 4차산업혁명위원회 홈페이지. <https://www.4th-ir.go.kr/ombudsman>