산업통상자원부 보도참고자료 다시 대한민국! 새로운 국민의 나라

보도시점 2023. 6. 29.(목) 18:00 배포 2023. 6. 29.(목)

“100만 명 바이오 빅데이터 구축사업” 예비타당성 조사 통과

- 1단계 사업으로 2024년부터 5년간 총 6,065.8억 원 투자 -
- 100만 명 규모 한국형 바이오 빅데이터를 기반으로 바이오 강국 도약 기대 -

보건복지부(장관 조규홍), 과학기술정보통신부(장관 이종호), 산업통상자원부(장관 이창양), 질병관리청(청장 지영미)은 6월 29일(목)에 개최된 국가연구개발사업평가 총괄위원회*(위원장 : 과학기술혁신본부장)에서 범부처 사업으로 기획된 ‘국가 통합 바이오 빅데이터 구축사업’ 이 예비타당성 조사*를 통과** 하였다고 밝혔다.

* 과학기술정보통신부 국가연구개발사업평가 총괄위원회 보도자료 (’23.6.29.(목)) 참고

** ´22.9.2. 예비타당성 조사(이하 예타) 신청 → ´22.11.1. 예타 대상 선정 → ´23.6.29. 예타 통과

“국가 통합 바이오 빅데이터 구축사업” 은 정밀의료 기술개발 등 의료혁신과 바이오헬스 혁신성장을 위해 100만 명 규모의 임상정보, 유전체 등 오믹스 데이터, 공공데이터, 개인보유건강정보를 통합하여 구축·개방하는 연구개발(R&D) 사업이다. 100만 명 규모의 대규모 구축에 앞서 2.5만 명 규모로 지난 2년여간 국가 바이오 빅데이터 구축 시범사업(’20.5.~’22.12.)이 추진되었다.

지난 6월 1일(목) 「첨단산업 글로벌 클러스터 육성방안」의 일환으로 발표된 ‘데이터 기반의 디지털 헬스케어 산업 고도화’의 후속조치로 2.5만 명 규모의 시범사업의 구축 데이터가 전면 개방(’23.6.20.)된데 이어 100만 명 규모의 “국가 통합 바이오 빅데이터 구축사업” 의 추진이 확정되면서 첨단 바이오 분야를 육성하려는 정부의 강력한 의지가 다시금 표명되었다.

100만 명분의 통합 바이오 빅데이터를 구축·개방하는 사업으로 기획된 “국가 통합 바이오 빅데이터 구축사업”은 국가연구개발사업평가 총괄위원회의 예비타당성 조사 결과에 따라 사업기간 9년을 2단계(5년+4년)로 분할하여 우선 77만여 명의 바이오 데이터를 구축·개방하는 1단계 사업을 5년간('24년~'28년, 사업비 6,065.8억 원) 추진한다.

이미 영국, 미국 등 주요국*은 바이오 빅데이터를 신약·의료기기 개발, 예방·맞춤의료 등에 활용되는 국가전략자산으로 인식하고, 디지털 헬스케어 분야의 글로벌 시장 선점을 위해 바이오 빅데이터 구축에 힘쓰고 있다.

* (영국) 일반인 50만 명(UK Biobank, '06-'10년), 암·희귀질환 10만 명(Genomes Project, '12-'17년) 구축 완료, 500만 명 게놈사업으로 확대 계획
(미국) 일반인 100만 명(All of Us, '18~'26년) 중 65만 명 모집

'24년부터 추진되는 “국가 통합 바이오 빅데이터 구축” 1단계 사업의 모집 규모는 77.2만 명(질환자 18.7만 명, 일반인 58.5만 명) 수준으로 일반인 50만 명(UK Biobank, '06~'10년), 암·희귀질환 10만 명(Genomes Project, '12~'17년)을 나눠 실시한 영국의 사례를 본따라서 글로벌 격차 해소를 위해 영국과 유사한 규모를 한번에 추진하는 것이다.

한국형 바이오 빅데이터 구축의 시작을 알리는 이번 사업*은 무엇보다 통합 데이터를 특징으로 한다. 참여자의 자발적인 참여와 동의에 기반하여 혈액, 소변 등 검체를 채취하고 임상정보와 유전체 등 오믹스 데이터를 생산할 계획이다. 여기에 공공데이터와 개인보유건강정보가 연계되면서 개인 중심의 통합 데이터가 구성되고 지속적으로 관리된다.

* 1단계 사업('24~'28년)에서 임상정보·공공데이터·개인보유건강정보는 77.2만 명분 축적, 전장유전체분석데이터는 혈액에서 34만 명분, 암 조직에서 4.1만 명분 생산

통합 데이터 구성·관리를 통해 바이오 데이터가 개별 연구자, 연구기관, 공공기관 등에 분산 및 파편화되어 있어 쓸만한 통합 데이터를 확보하기가 쉽지 않았던 문제점이 개선될 것으로 기대된다.

또한 “국가 통합 바이오 빅데이터 구축사업”은 뱅킹시스템(Banking

system)을 특징으로 한다. 참여자는 병원, 건강검진센터 등 참여자 모집 기관을 통해 자신의 혈액, 소변 등 검체와 임상정보 등 바이오 데이터를 본 사업의 바이오뱅크와 데이터뱅크*에 기탁하게 된다. 의료계·산업계·학계 등에서는 정밀의료 기술, 혁신 신약, 디지털 헬스 신제품 및 서비스 개발 등 연구 목적에 맞는 한국형 바이오 빅데이터를 필요한 양과 종류만큼 데이터뱅크로부터 제공받을 수 있다.

* ①질병관리청의 국립중앙인체자원은행을 바이오뱅크(검체 관리 총괄)로 지정, ②한국보건의료정보원의 인체유래물은행(설립 예정)을 데이터뱅크(데이터 관리 총괄)로 지정하고 한국과학기술정보연구원(KISTI)에 바이오빅데이터플랫폼 인프라를 구축하여 분석환경 지원

지금까지는 연구자들이 데이터 기반의 연구설계를 하면서 데이터 보유 기관을 직접 탐색하고, 보유 기관별로 데이터 제공 신청 후 기관생명윤리위원회의 심의를 거치는 등 불편함이 있었으나 데이터뱅크를 통해 필요한 데이터를 원스톱으로 제공받을 수 있게 되면서 시간비용과 탐색비용 등의 거래비용이 획기적으로 절감될 것으로 기대된다.

이번 한국형 바이오 빅데이터 구축사업은 주관부처인 보건복지부가 과학기술정보통신부, 산업통상자원부, 질병관리청과 합동으로 기획하였으며, 한국보건산업진흥원, 한국보건의료정보원, 한국과학기술정보연구원, 한국생명공학연구원, 한국산업기술평가관리원과 함께 추진할 계획이다.

<붙임> 1. 「국가 통합 바이오 빅데이터 구축사업」 개요
2. 용어설명

담당 부서 <총괄>	보건복지부 보건의료데이터진흥과	책임자	과 장 심은혜 (044-202-2940)
		담당자	서기관 김수영 (044-202-2945)
담당 부서	과학기술정보통신부 첨단바이오기술과	책임자	과 장 김영혜 (044-202-6180)
		담당자	사무관 임영주 (044-202-6182)
담당 부서	산업통상자원부 바이오융합산업과	책임자	과 장 송영진 (044-203-4290)
		담당자	주무관 이경봉 (044-203-4293)
담당 부서	질병관리청 바이오빅데이터과	책임자	과 장 채희열 (043-249-3003)
		담당자	연구관 이정현 (043-249-3065)

대한민국 정책브리핑
OPEN 출처표시
공공누리 공공저작물 자유이용허락

붙임 1 **「국가 통합 바이오 빅데이터 구축사업」 개요**

□ 추진 배경

○ 바이오 빅데이터는 신약·의료기기 개발, 맞춤의료 등에 활용되는 국가전략자산 → 글로벌 격차* 해소를 위해 구축 시급

* (미국) **All of US**('18~'26년, **100만 명 목표**, 65만 명 모집)
(영국) **UK Biobank**('06~'10년, 50만 명 완료) → **500만 명** 게놈사업으로 확대 시행

□ 사업 내용

○ **임상정보·유전체 데이터, 공공데이터** 및 **라이프로그** 등을 **통합**한 **100만 바이오 빅데이터 구축·개방***('24~'32. 복지부[주관]·과기부·산업부·질병청)

* 1단계 사업 : 77.2만 명(질환자 18.7만, 일반인 58.5만)의 바이오 데이터 구축·개방
1단계 사업기간/사업비 : '24년~'28년 / 6,065.8억 원 (국비 6,039.5억 원)

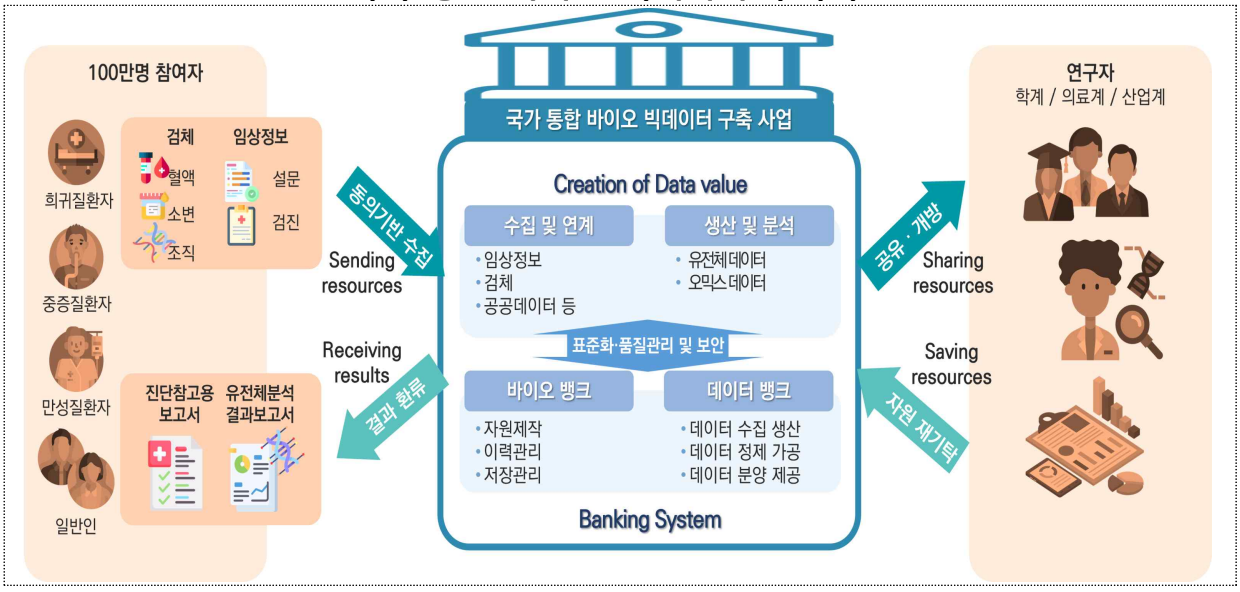
▸ **참여자 동의를 기반**으로 **검체**(혈액, 소변 등)**를 확보**하고, 100만 명 규모의 **임상·유전체 데이터를 생산**하고, **공공데이터와 라이프로그를 수집·연계**하여 **통합 데이터 구성**

① (**동의 기반 참여자 100만 명 모집**) 건강인, 질환 이환자, 질환자를 확보하기 위해 일반인(검진센터), 타겟중증질환자 및 희귀질환자(병원) 모집

② (**검체와 바이오뱅크**) 100만 명분 자동화 인체자원 제작 및 전수 검수체계 운영을 위한 기존 국립중앙인체자원은행의 고도화

③ (**데이터와 데이터뱅크**) 개인을 중심으로 통합된 바이오 데이터 축적 및 빅데이터가 원활하게 흐를 수 있도록(Data Flow)하는 데이터뱅크 구축

< 국가 통합 바이오 빅데이터 구축사업 >



붙임 2 용어설명

□ 빅데이터(Big Data)

○ 대용량, 짧은 생성 주기, 빠른 축적 속도 등을 특징으로 하는 다양한 정형 및 비정형 데이터로 구성된 대규모 데이터

□ 임상정보

○ 본 사업의 참여자 모집기관에서 수집·생산한 건강상태 등 설문 항목, 검진 항목, 내원정보, 진단 및 치료정보 등으로 구성된 정보

□ 공공데이터

○ 건강검진결과, 진료내역, 약물처방내역, 예방접종내역 등 건강보험공단, 건강보험심사평가원 등 공공기관이 보유하고 있는 데이터

□ 개인보유건강정보(Patients Generated Health Data, PGHD)

○ 참여자가 생성, 기록 또는 수집한 활동량, 혈압, 혈당, 수면정보, 신체정보, 그리고 일상정보(라이프로그) 등 건강 관련 데이터

□ 전장유전체분석데이터(Whole Genome Sequencing, WGS)

○ 인체유래물로부터 얻은 개인의 유전적 특징에 관한 데이터로서 인간의 모든 염기서열에 대해 온전히 해독한 데이터

□ 오믹스 데이터

○ 인체유래물로부터 얻은 유전체, 단백체, 대사체, 전사체 등의 데이터

○ ‘유전체’ 데이터는 인간의 유전정보에 관한 데이터

○ ‘단백체’ 데이터는 단백질의 서열·양에 관한 데이터

○ ‘대사체’ 데이터는 대사물질의 종류·양에 관한 데이터

○ ‘전사체’ 데이터는 발현되는 유전자의 종류·양에 관한 데이터

□ 통합 데이터

○ 참여자의 임상정보, 유전체정보, 공공데이터, 라이프로그 등을 개인 중심으로 통합한 데이터

□ 인체유래물은행(생명윤리법 제2조 제13호)

○ 인체유래물 또는 유전정보와 그에 관련된 역학정보, 임상정보 등을 수집·보존하여 이를 직접 이용하거나 타인에게 제공하는 기관

□ 데이터뱅크

○ 본 사업에서 데이터를 담당하는 인체유래물은행으로서 한국보건의료정보원의 인체유래물데이터은행을 데이터뱅크로 지정

□ 바이오뱅크

○ 본 사업에서 검체(인체유래물)를 담당하는 인체유래물은행으로서 질병관리청 국립중앙인체자원은행을 바이오뱅크로 지정