산업통상자원부 | 보도자료 | 다시 대한민국! 새로운 국민의 나라

보도시점 2023. 6. 19.(월) 12:00 < 6.20.(화) 조간 > 배포 2023. 6. 19.(월)

연구자 곁으로 성큼 다가온 바이오빅데이터

(국가바이오빅데이터 시범 사업 데이터 전면 개방)

- 국가 바이오 빅데이터 구축 시범사업 2.5만 명분 연구자원 전면 개방
- 국립중앙인체자원은행(보건의료연구자원정보센터)을 통해 만나볼 수 있어
- 데이터 기반 디지털 헬스케어 연구개발 촉진을 위한 연구자원 활용 기대

보건복지부(장관 조규홍), 과학기술정보통신부(장관 이종호), 산업통상자원부(장관 이창양), 질병관리청(청장 지영미)은 **100만 바이오 데이터 구축에 앞서 시행된 시범사업*이 종료**됨에 따라 **2.5만 명분 연구자원**을 6월 말부터 전면 개방하여 제공할 계획**이라고 발표하였다.

* (범부처 사업) 보건복지부(주관), 과학기술정보통신부, 산업통상자원부, 질병관리청

** 임상정보, 전장유전체분석데이터(Whole genome sequencing), 인체유래물(DNA, 혈청, 혈장, RNA 혈액, 소변)이 시범사업 연구자원으로 수집

국가 바이오 빅데이터 구축 시범사업('20.5.~'22.12.)은 **예방·예측·맞춤·참여의료 실현을 위해 임상·유전체 데이터를 통합하여 생산**하고, 디지털 헬스케어 기술개발(R&D)에서의 **활용체계를 구축**하고자 추진되었으며, 참여부처 및 수행기관*에서는 2022년 12월까지 2.5만 명 규모의 연구자원을 수집하였다.

* (수행기관) 한국생명공학연구원(KRIBB), 한국과학기술정보연구원(KISTI), 한국산업기술평가관리원(KEIT), 한국보건산업진흥원(KHIDI)

※ (연구자원 활용사례) 대규모 전장유전체 데이터 활용 희귀질환 진단 고도화, 한국인 전장유전체 기반 당뇨, 고혈압 발생과 합병증 연관 유전인자 탐색을 통한 병인 이해 및 질환 발생 위험도 예측 모델 개발 등 17개 연구과제에 데이터 제공

이번 바이오빅데이터 자원의 개방은 6.1.(목) 관계부처 합동으로 마련한 「첨단산업 글로벌 클러스터 육성방안」의 일환으로 보고된 '데이터 기반의 디지털 헬스케어 산업 고도화'의 후속 조치로써,

시범사업에서는 지난 2022년 7월 1차 개방*에 이어, **연구개발 목적으로 누적 2.5만 명분의 연구자원을 전면 개방**한다.

* (1차 개방) 국가 바이오 빅데이터 구축 시범사업 7,084명분 연구자원 개방(희귀질환자, 대장암, 자폐 스펙트럼 장애, 한국인유전체역학조사사업)

개방되는 연구자원은 **시범사업을 통해 신규로 수집한 희귀질환자 DB(누적 14,905명분 임상·유전 정보 및 인체유래물)**와 선행되었던 **일반연구과제*에서 기탁받은 임상·유전체 자료 DB(누적 9,797명분 임상·유전 정보)**이다.

* 기존 정부연구개발지원과제(울산 만 명 게놈프로젝트, 자폐 스펙트럼 장애, 대장암, 폐암, 치매) 및 선행사업(한국인유전체역학조사사업)

연구자원을 제공받고자 하는 연구자는 **국립보건연구원 국립중앙인체자원은행*을 통해 연구자원 제공 신청****을 해야 하며, 이후 국립중앙인체자원은행 분양위원회 심의를 거쳐 연구자원을 제공받을 수 있다.

* 보건의료연구자원정보센터 : https://coda.nih.go.kr

** 연구자원이용계획서, IRB 심의용 연구계획서 등 분양신청서류 제출

공개 대상 연구자원에 대한 상세한 정보는 **보건의료연구자원정보센터 누리집*에서** 안내받을 수 있다

* 누리집 주소 : https://coda.nih.go.kr

지영미 질병관리청장은 “이번 시범사업은 연구자 서랍 속에만 있던 데이터를 인체유래물은행을 통해 연구 목적으로 공유한 뜻깊은 사례로서, 바이오 빅데이터 공유 문화 확산에 기여해주신 희귀질환 협력 병원 및 데이터 기탁 연구자분들께 감사드린다.”고 전했으며,

“시범사업 성과를 토대로 **2024년부터는 100만 규모의 한국형 바이오 빅데이터 구축에 박차**를 가하고, **디지털 헬스 시장에서 기술개발(R&D)을 적극 지원하여 글로벌 경쟁력을 확보**해 나갈 것이다.”고 밝혔다.

<붙임> 1. 국가 바이오 빅데이터 구축 시범사업 개방 DB 주요 정보
2. 국가 바이오 빅데이터 구축 시범사업 연구자원 제공 절차
3. 국가 바이오 빅데이터 구축 시범사업 연구자원 활용 현황

담당 부서 <총괄>	질병관리청 바이오빅데이터과	책임자	과 장	채희열 (043-249-3003)
		담당자	연구관	이정현 (043-249-3065)
담당 부서	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)

대한민국 정책브리핑
OPEN 출처표시
공공누리 공공저작물 자유이용허락

붙임 1 **국가 바이오 빅데이터 구축 시범사업 개방 DB 주요 정보**

개방 DB	대상자 수	세부내용
국가 바이오 빅데이터 구축 시범사업 (희귀질환)	▸ 시범사업에 참여한 희귀질환자 및 가족 14,905명	▸ (임상·역학정보) 성별, 연령, 희귀질환 분류 등 24개 변수 ▸ (유전정보) 14,905명의 WGS(GVCF) ▸ (인체유래물) 혈청, 혈장, RNA보존혈액, gDNA, 뇨
국가 바이오 빅데이터 구축 시범사업 (KoGES)	▸ 일반인 5,000명	▸ (임상·역학정보) 기본정보, 일반사항, 질병과거력 등 202개 변수 ▸ (유전정보) 5,000명의 WGS(GVCF)
국가 바이오 빅데이터 구축 시범사업 (선도사업_울산게놈)	▸ 일반인 2,504명	▸ (임상·역학정보) 116개 변수 ▸ (유전정보) 2,504명의 WGS(GVCF)
국가 바이오 빅데이터 구축 시범사업 (선도사업_자폐증)	▸ 자폐증환자 및 가족 892명	▸ (임상·역학정보) 환자정보, 한국판 사회적 의사소통 등 71개 변수 ▸ (유전정보) 892명의 WGS(GVCF)
국가 바이오 빅데이터 구축 시범사업 (선도사업_대장암)	▸ 대장암 322명 * 암조직/정상조직 pairs	▸ (임상·역학정보) 환자정보, 진단시정보, 생존정보 등 92개 변수 ▸ (유전정보) 322명의 WGS(VCF)
국가 바이오 빅데이터 구축 시범사업 (선도사업_폐암)	▸ 폐암 84명 * 암조직/정상조직 pairs	▸ (임상·역학정보) 환자정보, 진단시정보, 생존정보, 치료정보 등 19개 변수 ▸ (유전정보) 84명의 WGS(VCF)
국가 바이오 빅데이터 구축 시범사업 (선도사업_치매)	▸ 치매 995명	▸ (임상·역학정보) 환자정보, 치매선별검사, 신경심리검사 등 139개 변수 ▸ (유전정보) 995명의 WGS(VCF)

1 희귀질환자 DB (신규 수집)

○ (대상자) 희귀질환 연구 참여자 14,905명

○ (역학·임상정보) 일반정보 및 임상정보 등 24개 변수

연번	구분(변수 수)	변수 설명
1	일반정보(4)	성별, 연령, 가족 관계, 가족질환분류
2	임상정보(20)	질환분류(대, 중, 소), 산정특례코드(의심질환)(1-3), ICD10코드(의심질환)(1-14)

2 한국인유전체역학조사사업(KoGES) 코호트 DB (질병관리청)

○ (대상자) KoGES 코호트(지역사회, 농촌) 연구 참여자 5,000명

○ (역학·임상정보) 생활습관, 일반정보, 임상정보 등 202개 변수

연번	구분(변수 수)	변수 설명
1	기본정보(6)	단위데이터, 조사년도, 성별, 만 나이, 추적조사 만나이, 추적조사 조사년도
2	일반사항(4)	직업 종류, 교육수준
3	질병과거력(13)	고혈압/당뇨병/갑상선질환/고지혈증/각종 종양 진단여부 및 처음 진단받은 나이, 고혈압/당뇨병/고지혈증 발생여부
4	가족력(18)	부모, 형제의 고혈압/당뇨병/뇌졸중 진단여부 및 처음 진단받은 나이
5	생활습관(24)	음주 여부, 과거음주 금주기간, 지금까지 총 음주기간, 막걸리/소주/맥주/양주의 지난 1년간 평균 섭취횟수 및 1회 음주량, 흡연 여부, 금연 나이, 총 흡연기간, 하루 흡연량, 규칙적 운동 여부 등
6	식품섭취빈도조사(영양소)(23)	에너지, 단백질, 지방, 탄수화물, 칼슘, 인, 철, 칼륨, 비타민A, 나트륨, 비타민 B1, 비타민 B2, 니아신, 비타민 C, 아연, 비타민 B6, 엽산, 레티놀, 카로틴, 회분, 섬유, 비타민 E, 콜레스테롤
7	신체계측(8)	혈압, 맥박수, 허리둘레, 엉덩이둘레, 신장, 체중, BMI
7	임상검사(66)	Glucose, Albumin, BUN, Creatinine, AST, ALT, r-GTP, Total Cholesterol, Triglyceride, Total protein, Total Bilirubin, Calcium, Sodium, Potassium, Chloride, CRP, WBC, RBC, Hemoglobin, Insulin 등
8	수면력(40)	불면증 여부, 잠들기 어려운지 여부 및 회수, 자다깨면 다시 잠들기 어려운지 여부 및 횟수, 다시 잠드는데 걸리는 시간, 평소 수면시간, 수면시기, 보통 취침시간, 낮잠을 자는지 여부, 가족의 수면습관, 밤에 소변을 보는 횟수 등

3 울산 만명 게놈프로젝트 DB

○ (기탁 과제) “울산 만명 게놈프로젝트”과제 (울산과학기술원 이세민 교수)

○ (대상자) 울산 만명 게놈프로젝트 참여자 2,504명

○ (역학·임상정보) 신체계측, 심폐기능, 혈액검사 등 116개 변수

연번	구분(변수 수)	변수 설명
1	기본정보(4)	샘플 일련번호, 시점, 동의서 작성일, 성별, 만나이
2	신체계측(8)	신장, 체중, 표준체중, 비만도, 신체계측검사, 체지방율, BMI, 허리둘레
3	혈압검사(3)	수축기혈압, 이완기혈압, 맥박수
4	심폐기능(7)	노력성폐활량, 노력성폐활량백분율, 1초간노력성폐활량, 1초율노력성폐활량백분율, 1초율, 폐기능 진단, 심전도 결과
5	청력검사(8)	왼쪽 인지 데시벨 수치(1000Hz, 2000Hz, 3000Hz, 4000Hz), 오른쪽 인지 데시벨 수치(1000Hz, 2000Hz, 3000Hz, 4000Hz),
6	생화학검사(25)	혈중 총콜레스테롤, 혈중 중성지방, 혈중 HDL-Cholesterol, 혈중 LDL-Cholesterol, 혈중 Lipoprotein(a), 공복 혈중 혈당, 당화혈색소, 혈중 총단백, 혈중 총빌리루빈, 혈중 직접빌리루빈, 혈중 aspartate aminotransferase, 혈중 alanine aminotransferase, 혈중 alkaline phosphatase, 혈중 Gamma Glutamyl Transpeptidase, 혈중 α-fetoprotein, 혈액요소질소, 혈중 크레아티닌, 혈중 알부민, 알부민/글로불린 비, 혈중 Lactate dehydrogenase, 혈중 homocysteine, 혈중 나트륨, 혈중 칼륨, 혈중 칼슘, 혈중 무기인
7	혈액검사(20)	혈중 백혈구 수, 혈중 적혈구 수, 혈색소, 혈중 헤마토크리트, 적혈구용적분포폭, 평균적혈구용적, 평균 적혈구혈색소량, 평균 적혈구혈색소, 혈중 혈소판 수, 혈중 백혈구 중 호중구 (Polyseg neut) 비율, 혈중 백혈구 중 림프구 (Lymphocyte) 비율, 혈중 백혈구 중 단구 (Monocyte) 비율, 혈중 백혈구 중 호산구 (Eosinophil) 비율, 혈중 백혈구 중 호염기구 (Basophil) 비율, Rh식 혈액형, ABO식 혈액형, 혈중 글로불린, 적혈구침강속도, 혈중 아밀라제, 혈중 creatine phosphokinase
8	면역검사(5)	B형간염바이러스 표면항원 , B형간염바이러스 표면항체, C형간염 항체, 혈중 C 반응성 단백, 혈중 류마티스인자
9	갑상선검사(3)	혈중 T3, 혈중 Free T4, 혈중 TSH
10	소변검사(16)	요산, 소변 백혈구반응 , 요질산염, 요산도, 요단백, 요비중, 요당, 요케톤체, 유로빌리노겐, 요빌리루빈, 요잠혈반응, 요백혈구, 요적혈구, 요상피세포(Squamous), 요상피세포(Transitional), 요상피세포(Renal)
11	안과검사(9)	시력:나안(좌, 우), 시력:교정(좌, 우), 시력:안압(좌, 우), 안저(좌, 우), 안구질병 진단
12	암표지자(4)	혈중 carcinoembryonic antigen, 혈중 CA 19-9, 혈중 CA 125, 혈중 Prostate-specific antigen
13	자궁경부암(1)	자궁 경부와 질 분비물 액상세포 도말검사
14	영상검사(2)	요추x선, 복부초음파
15	소화관련(1)	위 헬리코박터균

4 자폐스펙트럼장애 DB

○ (기탁 과제) “자폐증 분자진단을 위한 바이오마커의 임상적 검증” 과제 (분당서울대병원 유희정 교수)

○ (대상자) 자폐 스펙트럼 장애환자 및 환자 가족 총 892명

○ (역학·임상정보) 환자 정보 및 지능검사 등 71개 변수

연번	구분(변수 수)	변수 설명
1	환자정보(8)	샘플일련번호(가족), 검사시행 기관번호, 가족 관계, 성별, 생년월일, 평가 시행일, 평가 당시 개월 수, 나이
2	한국판 사회적 의사소통(2)	(SCQ(한국판 사회적 의사소통 설문지)) 일생, 현재 총점
3	사회성 반응성(1)	SRS(사회성 반응성 설문지) T점수
4	아동기 행동평가척도(21)	(K-CBCL(아동기 행동평가척도)) 사회성, 학업 수행, 총 적응도, 불안/우울, 위축/우울, 신체 증상, 사회적 미성숙, 사고 문제, 주의 집중 문제, 규칙 위반, 공격 행동, 기타문제, 내재화, 내면화, 외현화, 총 문제행동, 정서 문제, 불안 문제, 신체화 문제, ADHD, 반항 행동, 품행문제
5	지능검사(14)	지능검사 종류, 전체 지능지수 (지능검사 종류가 1, 2일 경우) 언어성, 동작성 지능지수 (지능검사 종류가 3일 경우) 언어이해, 시공간, 유동추론, 작업기억, 처리속도지수 (지능검사 종류가 4, 5일 경우) 언어이해, 지각추론, 작업기억, 처리속도지수 Leiter-R full-scale IQ(전체 동작성지능지수)
6	바인랜드 적응 행동척도(5)	(VABS(바인랜드 적응행동척도)) 의사소통, 일상생활기술, 사회화, 운동기술, 전체적응행동 t점수
7	정밀주의력검사 (8)	(ATA_시각) 누락, 오경보, 반응 시간 평균, 표준편차 t점수 (ATA_청각) 누락, 오경보, 반응 시간 평균, 표준편차 t점수
8	자폐증 진단 관찰 스케쥴(5)	(ADOS) 모듈, 사회적 정동 알고리즘 점수, 반복적이고 상동적인 행동 및 관심 점수, 총점, 비교점수
9	자폐증 진단 면담지(6)	(ADI-R) 사회적 상호작용 영역의 알고리즘 점수, 언어가 있는 아동의 의사소통 영역 알고리즘 점수, 언어가 제한된 아동의 의사소통 영역 알고리즘 점수, 비언어성 부분 알고리즘 점수, Restricted, repetitive, and stereotyped patterns of behavior, Abnormality of development evident at or before 36 months
10	아동기 자폐증 평정척도(1)	CARS(아동기 자폐증 평정척도) 총점

5 대장암 DB

○ (기탁 과제) “대장암 다중유전체 통합 분석 기반의 정밀 의료 원천기술 개발” 과제 (연세대학교 김영준 교수)

○ (대상자) 대장암 환자 322명

○ (역학·임상정보) 환자정보, 진단시정보, 생존정보, 치료정보 등 92개 변수

연번	구분(변수 수)	변수 설명
1	환자정보 (5)	성별, 나이, 진단일, 중복암(여부, 진단연도)
2	진단시 정보 (36)	암위치 상세(ascending, transverse, descending, sigmoid, rectal, rectosigmoid, cecum, other), pathology, 암 분화도, MSI 상태, KRAS 상태, KRAS 상태(MT 상세내용), NRAS 상태, NRAS 상태(MT 상세내용), BRAF 상태, BRAF 상태(MT 상세내용), 진단시 병기, 세부 병기, 진단시 T stage, 진단시N stage, 림프관 침윤, 혈관 침윤, 신경주위 침윤, 진단시 M stage, 전이장기(간, 폐, LN, bone, peritoneal seeding, brain, 기타)
3	생존정보 (4)	마지막 추적관찰(일자, 상태), 생존여부, 사망일
4	치료정보 (47)	수술(일자, 수술명), 보조항암화학요법 여부, 방사선치료 여부(수술전, 후), 보조항암/방사선 complete, 보조항암 시행일 또는 기간, 재발관련(여부, 일자, 패턴, 장기), 1st line chemotherapy(방법, 시작일, 종료일, 반복수, best response, 종료 사유), 2nd line chemotherapy(방법, 시작일, 종료일, 반복수, best response, 종료 사유), 3rd line chemotherapy(방법, 시작일, 종료일, 반복수, best response, 종료 사유), 4th line chemotherapy(방법, 시작일, 종료일, 반복수, best response, 종료 사유)

6 폐암 DB

○ (기탁 과제) “차세대 염기서열분석법을 이용한 비소세포폐암의 치료 타겟 유전체 분석” 과제 (국립암센터 한지연 교수)

○ (대상자) 폐암 환자 84명

○ (역학·임상정보) 환자정보, 진단시정보, 생존정보, 치료정보 등 19개 변수

연번	구분(변수 수)	변수 설명
1	환자정보(5)	샘플일련번호(개인), 시점, 성별, 연령, 가족력
2	진단시정보(3)	MSS/MSI 상태, 진단시 병기, 1기 예후 (good or poor)
3	생존정보(2)	생존여부, 생존기간(추적 기간)
4	치료정보(9)	재발여부, 재발 후 추적기간, 재발장기(폐, 간, 뼈, 뇌, 흉막, 기타, 기타내용)

7 치매 DB

○ (기탁 과제) “MRI기반 알츠하이머성 치매 특이 뇌손상 규명 및 뇌지도 구축” 과제 (조선대학교 이건호 교수)

○ (대상자) 알츠하이머, 정상인 995명

○ (역학·임상정보) 환자정보, 치매선별검사, 신경심리검사 등 139개 변수

연번	구분(변수 수)	변수 설명
1	환자정보 (5)	성별, 진단나이, 진단정보, 아밀로이드 양전자방출단층촬영 검사 결과(해당정보 존재 시), Apolipoprotein E 유전자형 검사 결과
2	치매선별검사 (31)	시간지남력, 장소지남력, 기억등록/즉각기억, 주의집중 및 계산, 기억회상, 언어, 시공간구성, 총점
3	신경심리검사 (103)	검사나이, 학력, 교육년수, 주된 이전 직업, 주로 사용하는 손, 교육수준, 연령대, 기억력(언어 기억, 시각 기억), 언어기능(유창성, 언어 이해력, 이름대기, 따라 말하기), 시공간능력(RCFT의 베껴 그리기, 육면체 조각 그림 맞추기, 시계 그리기), 전두엽기능(Stroop Test, Digit Symbol Test, Trail Marking Test), 주의집중력(digit span test, letter cellation), 실행증, 정서(shot form geri-atric depression scale), CDR 척도(인지, 사회기능 정도 측정)

붙임 2 국가 바이오 빅데이터 구축 시범사업 연구자원 제공 절차

□ **(검색 및 신청)** 국립보건연구원 보건의료연구자원정보센터(coda.nih.go.kr) 또는 국립중앙인체자원은행 누리집 및 온라인 분양데스크에서 분양대상자원 검색 및 신청

1. 데이터 제공 신청시

○ **(검색)** 질병관리청 국립보건연구원 보건의료연구자원정보센터(coda.nih.go.kr) 회원가입 및 권한신청(문의: 043-249-3055) → **보건의료연구자원정보센터 접속 후 통합검색 메뉴 → 분양 대상 자원 검색**

○ **(신청)** 질병관리청 국립보건연구원 보건의료연구자원정보센터(coda.nih.go.kr) 회원가입 및 권한신청(문의: 043-249-3055) → **온라인 분양 신청**

2. 데이터 및 인체유래물 제공 신청시

○ **(검색)** 질병관리청 국립중앙인체자원은행 누리집(biobank.nih.go.kr) 접속 → **분양 대상 자원 검색**(분양상담: 1661-9071)

○ **(신청)** 질병관리청 질병보건통합관리시스템(is.kdca.go.kr) 회원가입 및 권한신청(문의: 043-719-6544) → **온라인 분양데스크 분양신청서 작성**

□ **(분양 신청서류)** 국립중앙인체자원은행 운영·관리 규정(질병관리청 예규 제42호)에 따른 분양과제 심의를 위한 서류 제출

1. 인체자원이용계획서(온라인 작성)
2. 서약서(온라인 작성)
3. 개인정보수집이용동의서(온라인 작성)
4. IRB 심의 결과서
5. IRB 심의용 연구계획서
6. 분양신청공문

□ **(제공심의 원칙) 국립중앙인체자원은행 운영·관리 규정(질병관리청 예규 제42호)에 따라 진행하며, 세부적인 지침을 아래와 같이 진행**

○ **(데이터)** 정보의 민감성, 데이터 분석환경 운영의 효율성 등을 고려하여 분양신청 과제 계획과의 연관성, 결과 도출에 필요한 분양신청 대상자수 등 고려하여 심의 진행

○ **(인체유래물)** 인체유래물 대상 분양과제 심의시, 인체유래물 유한한 특성을 고려한 분양원칙 수립

- **(높은 활용성)** 공개 대상 전체에서 추가 데이터를 생산하는 경우
- **(오믹스 데이터)** 단백체, 대사체 등 오믹스 수준 데이터를 생산하는 경우
- **(이용성과물 제출)** 데이터 생산시 이용성과물 제출 의무 부여

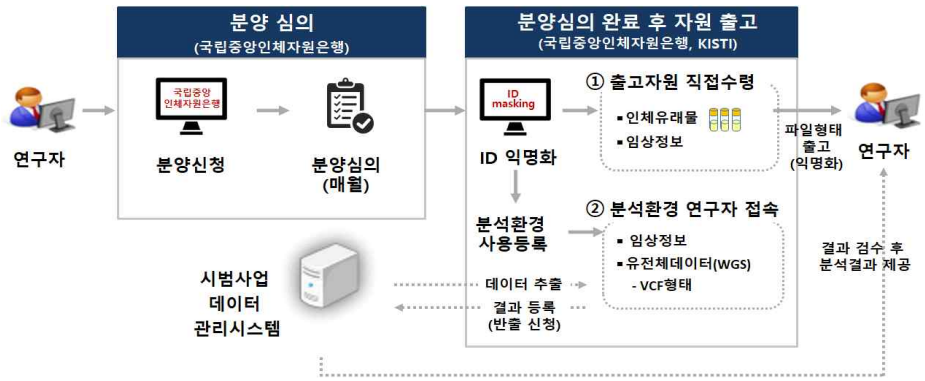
□ **(분석환경 신청·활용)** 국립중앙인체자원은행 분양위원회 심의 완료 후 한국과학기술정보연구원 **국가 바이오 빅데이터 연구포털** 내에서 연구계정 신청 및 데이터 분석 진행

○ **(연구과제 신청)** 연구포털(https://www.cirn.re.kr) 내에서 과제 정보 등록 및 **계정신청**

○ **(분석환경)** 연구자는 연구포털에서 제공하는 **원격접속** 기반 **보안 연구환경** 내에서 **접근 승인된 데이터** 범위 내에서 연구를 수행

○ **(데이터 반입·반출)** 연구환경 내부로 연구자 데이터의 **반입** 또는 분석결과의 **외부반출**은 **별도의 승인 절차**를 거쳐 진행

□ **국가 바이오 빅데이터 구축 시범사업 수집자원 제공절차 모식도**



붙임 3 국가 바이오 빅데이터 구축 시범사업 연구자원 활용 현황

연번	연구 과제명	활용 DB
1	한국인 전장유전체 기반 당뇨, 고혈압 발생과 합병증 연관 유전인자 탐색을 통한 병인 이해 및 질환 발생 위험도 예측 모델 개발	KoGES, 울산게놈
2	한국인 유전적 다형성 및 메타유전체 데이터베이스 구축	KoGES, 대장암, 울산게놈
3	인공지능을 활용한 질병유발 변이 진단율 향상과 환자맞춤형 치료 적용가능 희귀질환 환자 발굴	KoGES, 자폐증, 희귀질환, 울산게놈
4	한국인 대장암 전장유전체에서 치료표적이 될 수 있는 actionable marker 카탈로그 구축	대장암
5	한국인 대장암 유전체 및 종양 내 마이크로바이옴 특성 규명	대장암
6	대규모 전장유전체 데이터 활용 희귀질환 진단 고도화	희귀질환
7	전장유전체분석 데이터 세트를 이용한 희귀질환의 원인 기전 분석	희귀질환
8	유전성 뇌졸중을 유발하는 단일 유전자 돌연변이 유병률-국내연구	KoGES, 대장암, 희귀질환
9	다양한 혈액형군들의 유전형 분포 분석	KoGES
10	일반인의 임상유전체 데이터 표준화 및 건강상태를 나타내는 위험인자 예측을 위한 기계학습기반의 분석	KoGES, 울산게놈
11	유전성 난청 및 어지럼증 극복을 위한 환자 맞춤형 진단법 개발	KoGES, 대장암, 희귀질환, 자폐증, 울산게놈
12	유전체 비교 분석을 통한 희귀질환 진단 및 치료 타겟 발굴	KoGES, 자폐증, 희귀질환
13	차세대 시퀀싱 기반 한국 자폐스펙트럼 장애 유전자 바이오 마커 발굴	자폐증
14	연령관련 황반변성 유전학적 연구	KoGES
15	Noncoding mutation 연구를 위한 범주형 연관성 방법론 개발 및 신경발달장애 원인 pathogenic enhancer mutation 규명	KoGES, 울산게놈, 자폐증, 희귀질환
16	한국인 자폐스펙트럼장애 가족에서 유전자기반 진단 기술 개발	희귀질환
17	바이오빅데이터 활용을 위한 한국인 전장유전체 데이터 특성 분석 및 데이터 셋 구축	KoGES, 희귀질환