질병관리청 **보도참고자료** 다시 도약하는 대한민국 함께 잘사는 국민의 나라

보도 일시	2022.12.22.(목) 배포 즉시	배포 일시	2022.12.22.(목) 08:50
담당 부서	미래의료연구부 바이오빅데이터과	책임자	과 장 채희열 (043-249-3003)
		담당자	연구관 김정은 (043-249-3072)

국립보건연구원장, 한국과학기술정보연구원과 보건의료연구데이터 활용 협력체계 강화를 위한 현장방문

주요 내용

□ 권준욱 국립보건연구원장, 국립보건연구원(보건의료 연구데이터)과 한국과학기술정보연구원(슈퍼컴퓨터 기반 정보분석 인프라)의 협력을 위해 국가슈퍼컴퓨팅 본부 방문(12.21.)

□ 국립보건연구원 권준욱 원장은 12월 21일(수), 슈퍼컴퓨터 기반의 '보건의료 연구데이터 활용 협력체계' 강화를 위해 한국과학기술정보연구원(KISTI, Korea Institute of Science and Technology Information)을 방문하였다.

□ 이날 권준욱 국립보건연구원장은 국가바이오빅데이터 구축 시범사업을 비롯한 대규모 데이터를 활용할 수 있도록 연구자에게 용이한 분석환경을 제공 중인 국가슈퍼컴퓨팅본부를 방문하여,

○ 국립보건연구원 보건의료 연구데이터의 공유, 분양 및 한국과학기술정보연구원과의 공동 활용연구 관련 협력 계획 등에 대해 논의하고, 국가슈퍼컴퓨터 5호기인 '누리온' 등을 참관했다.

OPEN 출처표시 공공누리 공공저작물 자유이용허락

1339 질병관리청 콜센터

□ 국립보건연구원은 보건의료연구자원정보센터(CODA)*를 통해 임상정보, 유전정보, 멀티오믹스** 정보 등을 포함한 다양한 의료영상정보를 수집하고,

* 보건의료연구자원정보센터(CODA, Clinical Omics Data Archive) : 2017년 생명연구자원기탁등록보존기관(제21호) 및 생명연구자원정보센터(제1호)로 지정되어 있으며, 국가연구개발사업에 축적되는 연구자원정보와 연구자가 자발적으로 제공한 정보자원의 수집·관리·활용을 지원하는 보건의료 분야의 데이터 센터

** 멀티오믹스(Multi-Omics) : 유전적, 기능적, 환경적, 면역반응 등의 이해를 위해 최신의 기술을 활용하여 인체자원에서 생산되는 종합 데이터

○ 이를 국내 연구자에게 공개*하여, 활용할 수 있게 하였으며,

* 공개 현황 : 총 용량 1,218.87TB, 연구대상자 116,151명, 임상·역학정보 864개 변수

○ 한국과학기술정보연구원 국가슈퍼컴퓨팅본부와 국립보건연구원이 보유한 유전체, 임상정보 등 보건의료분야의 대규모·대용량 바이오빅데이터의 임상적·과학적 활용을 촉진하기 위해 협력의향서(LOI, Letter of Intent)를 체결('22.2.17.)하는 등 지속적인 노력을 기울이고 있다.

- 국립보건연구원은 앞으로도 한국과학기술정보연구원과의 긴밀한 협력을 통해 보건의료연구 데이터를 더욱 활발하게 공유함으로써, 미래의료 연구를 위한 인프라 지원을 끊임없이 이어나갈 예정이다.

<붙임> 1. 한국과학기술정보연구원장, 국가슈퍼컴퓨팅본부 현장방문 사진
2. 한국과학기술정보연구원 국가슈퍼컴퓨팅본부 개요
3. 보건의료연구자원정보센터 공개 연구데이터 현황

OPEN 출처표시 공공누리 공공저작물 자유이용허락

1339 질병관리청 콜센터

붙임 1 국립보건연구원장, 한국과학기술정보연구원 현장방문 사진



국립보건연구원장, 한국과학기술정보연구원장과 인사



슈퍼컴퓨팅본부 실무진과의 미팅





슈퍼컴퓨터 누리온 5호기 참관

OPEN 출처표시
공공누리 공공저작물 자유이용허락

1339 질병관리청 콜센터

붙임 2 한국과학기술정보연구원(KISTI) 국가슈퍼컴퓨팅본부 개요

□ 개요

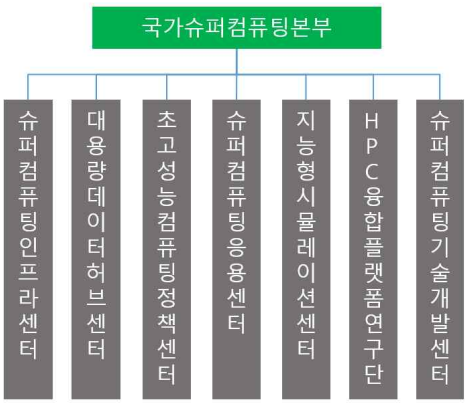
○ (설립) 한국과학기술정보연구원(이하 KISTI)은 국가 과학기술 정보 분야의 전문연구기관으로서 국가 과학기술 진흥과 산업의 발전에 기여하기 위해 설립

– KISTI 국가슈퍼컴퓨팅본부는 1967년 9월 KIST의 전자계산실에서 기원, 이후 1982년 KAIST 전산개발센터, 1989년 KIST 부설 시스템공학센터의 전산운영부, 1996년 ETRI의 슈퍼컴퓨터센터를 거쳐 현재에 이름, 2011년 초고성능컴퓨터[1]법 제정에 이어 국가초고성능컴퓨팅센터로 KISTI가 지정되었고 국가슈퍼컴퓨팅본부가 전담 중임

○ **(목적)** 국가 차원의 초고성능컴퓨팅 공동활용 체계 구축을 통해 과학기술 공공·산업 분야 초고성능컴퓨팅 활용 증진과 과학기술 R&D 혁신에 기여하기 위한 국가 슈퍼컴퓨팅 연구·서비스

□ 조직 및 주요 인프라 등

○ **(조직)** 6개 센터, 1개 연구단, 총 136명



1) 법률 용어임, 초고성능컴퓨터=슈퍼컴퓨터

OPEN 출처표시 공공누리 공공저작물 자유이용허락

1339 질병관리청 콜센터

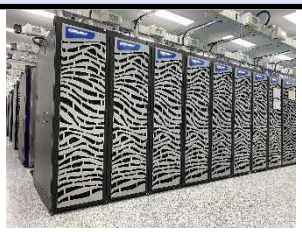
○ **(중점 분야)**

- **(인프라)** 대규모 계산자원이 소요되는 R&D, 사회현안 등 환경변화에 적기 대응하고 수요 확대를 감안하여 충분한 계산성능, 데이터컴퓨팅 등을 갖춘 국가 플래그십 초고성능컴퓨팅 인프라 선제적 확보(슈퍼컴퓨터 6호기 도입 및 GPU기반 보조 시스템 뉴론 확충) 및 국가 차원의 초고성능컴퓨팅 공동활용 서비스

- **(응용활용)** 초거대 계산기술·초고성능컴퓨팅 활용기술 확보 및 선순환형 과학·공학·산업 분야의 초거대문제해결 연구·지원을 통한 세계적인 연구성과 창출

- **(기술개발)** 자원특성·운영환경 등이 다른 공동자원을 긴밀히 연동하고 사용자 접근성을 높인 클라우드 기반 플랫폼인 KI 클라우드의 개발

○ **(주요 인프라) 슈퍼컴퓨터 5호기 누리온 및 뉴론 시스템**

메인시스템(누리온)	GPU기반 보조시스템(뉴론)	스토리지	기반시설
			
• 계산노드 8,305대 • CPU-only 노드 132대 • 575,300 CPU코어 • 100Gbps 고속인터커넥트	• 계산노드 83대 • GPU 203개 • 이론성능 2PF(FP64)	• 23PB 저장 용량 • 8TB SAS 3,200개 • 1.2TB SSD 768개 • 10PB Tape Library	• 냉동기 400RT 4EA • 프리쿨링 250RT 2EA • 냉각탑 450CRT 4EA • 항온항습기 30RT 11EA

OPEN 출처표시 공공누리 공공저작물 자유이용허락

1339 질병관리청 콜센터

붙임 3 보건의료연구자원정보센터 공개 연구정보(데이터) 현황

□ 공개대상 정보 현황(총 1,218.87TB)

('22년 12월 기준)

연번	대상자	임상·역학정보		유전정보		
		연구대상 수(명)	변수 수(개)	연구대상 수(명)	종류	용량
1	KoGES 전장유전체 코호트	2,500	199	2,500	Fastq	161TB
					BAM	116TB
					BAI	22.52GB
					gVCF	7TB
		총 용량				284TB
2	대장암 DB	197	88	197	BAM	30.2TB
					gVCF	2GB
		총 용량				30.17TB
3	자폐증 DB (환자 149명, 가족 351명)	500	69	500	BAM	21.5TB
					BAI	4.4GB
					gVCF	2.8TB
		총 용량				24.3TB
4	희귀질환 DB (환자 가족 포함)	3,887	19	3,887	Fastq	270TB
					BAM	182TB
					BAI	35GB
					gVCF	14.7TB
		총 용량				466.7TB
5	코로나19 확진자 DB (확진자 및 일반인)	420	245	420	WGS(420)	63TB
					Cytokine(840)	0.1GB
					COVID-seq(279)	19GB
					HLA typing(420)	412GB
					Bulk TCR-seq(418)	1TB
					Bulk BCR-seq(418)	1TB
					scRNA-seq(483)	18TB
					SNP array(124)	2.4GB
		총 용량				83.4TB
6	포스트게놈 다부처유전체 사업 (49개 과제)	약 10만	-	약 10만	DNA methylation	7.5GB
					Epigenome-seq	5.73TB
					Gene copy number variation	4.5GB
					Gene expression	161.8GB
					metagenome	435.9GB
					RNA-seq	4.34TB
					RNA-seq(miRNA)	72.4GB
					SNP detection	13.47TB
					Targeted-seq	5.25TB
					WES	61.9TB
					WGS	151TB
		총 용량				242.3TB
7	K-MASTER DB	7,305	128	7,305	Fastq	26TB
					BAM	22TB
					VCF	6.2GB
		총 용량				47TB
8	울산게놈프로젝트 DB	1,342	112	1,342	BAM	34TB
					BAI	11.7GB
					gVCF	6TB
		총 용량				41TB

OPEN 출처표시 공공누리 공공저작물 자유이용허락

1339 질병관리청 콜센터