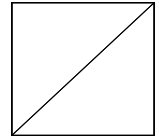


공개



의안번호	제 6 호	보 고 사 항
심 의 연 월 일	2018. 1. 19. (제 33 회)	

-혁신성장 촉진을 위한-
연구데이터 공유·활용 전략(안)

국가과학기술심의회
운 영 위 원 회

제 출 자	과학기술정보통신부장관 유영민
제출 연월일	2018. 1. 19.

1. 보고주문

- 「연구데이터 공유·활용 전략(안)」을 별지와 같이 보고함

2. 제안이유

- 데이터 중심 R&D 활성화로 연구데이터 활용 수요가 점증함에 따라, 국가 R&D를 통해 축적되는 연구데이터를 다양한 연구자가 공유·활용할 수 있는 체계를 구축하고자 함

3. 주요내용

□ 추진배경 및 국내외 현황

- 현대과학의 거대·복잡화와 디지털 기술의 발전으로 데이터 분석이 연구개발의 주요 추진 동력으로 부각*되고 이를 활용한 다분야융합·공동연구가 활성화

* 1914년 이후 노벨상 수상 연구의 87%가 대형장비의 대용량 데이터 분석 연구(KISTI, '16)

- 새로운 지식·가치 창출의 원천으로서 연구데이터의 가치가 높아지고 다양한 연구자간의 융합·공동연구 촉진을 위한 연구데이터 공유체계의 필요성이 점증

- 주요 선진국은 새로운 지식·가치의 창출과 공유·확산을 위해 연구성과와 과정을 개방화하는 오픈 사이언스 정책* 적극 추진

* (美) '연구데이터 관리·공유에 대한 지침'('13), (EU) 'OpenAIRE2020 Project'('15) 등

- 연구데이터의 관리 제도화, 연구데이터 공유·활용 인프라 지원, 연구자 가이드라인·교육 제공 등을 주요 정책으로 추진

- 국내는 바이오·소재 등 일부 분야의 연구데이터 공유가 추진 중이나 R&D 전반으로의 확산을 위한 제도·인력·인프라 미비

※ 연구개발 과정에서 산출된 '연구데이터'는 연구결과로서 가치가 높음에도 연구성과물(논문, 특허, 연구보고서, 장비·시설 등)에 미포함

□ 비전 및 목표

- (비전) 연구데이터 지식자산화·빅데이터화
- (목표) 법·제도 개선, 계층적 연구데이터 관리체계 및 인프라 구축, 인력양성을 통한 국가 연구데이터 공유·활용 활성화

□ 중점 추진과제

① 연구데이터 관리체계 구축 및 공유커뮤니티 형성 촉진

- (관리체계) 연구데이터의 체계적인 관리를 위해 국가연구데이터 센터 및 과학기술 분야별(대분야·소분야) 전문센터* 체계 구축

* 바이오, 미래소재, 대형연구장비, 인공지능 등 데이터 집약형 연구분야부터 우선 전문센터 지정을 추진하고 단계적으로 확산

- 국가연구데이터센터는 관리체계 총괄로서 국가연구데이터플랫폼 운영, 분야별 전문센터는 데이터 취합·관리 및 활용촉진 등 담당

< 연구데이터 관리 체계 >



- (커뮤니티 형성) 대규모 연구데이터를 구축·활용하는 집단연구 과제 신설*하고 既 구축된 데이터를 활용한 소규모 R&D 과제 지원

* 신설되는 집단연구과제 수행기관은 연구데이터 전문센터로 지정 추진

② 국가연구데이터플랫폼 구축 및 서비스 제공

- (통합 플랫폼) 연구데이터 공유·활용체계 전반의 연구데이터를 원스탑으로 검색하고 활용할 수 있는 통합 플랫폼 구축·운영

※ 데이터 유통이력의 투명한 관리와 위변조 방지 등을 위해 블록체인 기술을 접목하는 방안 검토

- 사용자에게 연구 데이터 융합 및 분석을 실시할 수 있는 가상
작업환경과 다양한 데이터 분석도구* 제공

* 데이터 분석도구 마켓 운영을 통해 다양한 데이터 분석도구의 자유로운 공유·유통환경 제공

- (인프라 서비스) 전문센터에 클라우드 형태로 분야별 컴퓨팅·데이터
인프라를 서비스*하고, 개별 연구자에도 데이터관리 클라우드 제공

* 국가연구데이터센터가 인프라를 보유하고 시스템 관리 책임을 맡되, 전문센터가
국가연구데이터센터의 인프라를 로컬 인프라처럼 활용 가능토 접근권한을 제공

③ 데이터 및 컴퓨팅 활용 R&D 인재 성장 지원

- (연구자 교육) 대학·연구소 등 연구자의 데이터 분석·활용능력
제고를 위해 빅데이터 교육과정 운영*

* 한국데이터진흥원 재직자 대상 빅데이터 전문인력 교육 프로그램 연계('18년~)

- 온라인 빅데이터 교육 프로그램 운영을 위해 데이터·컴퓨팅
활용 교육·실습 환경*을 제공하고 교육용 데이터뱅크** 구축

* 첨단사이언스교육허브(EDISON)이 보유한 소재·화학·열유체 등 6개분야 400종 SW 연계

** 국내외 공개 데이터, 연구보고서·논문 등 활용

- (인프라 운영인력) 연구데이터센터 등 ICT 인프라 운영인력 양성*

* 인프라 운영, 데이터 관리 및 병렬분석 환경, 시스템 모니터링, 장애분석 등 역량 강화

④ 연구데이터 관리·공유·활용에 관한 법·제도 마련

- 연구데이터의 법적 정의* 마련, 연구데이터 관리제도 도입,
데이터에 대한 연구자의 권리보장 등을 위한 법령정비** 추진

* 국가연구개발사업의 각종 실험, 관찰, 조사, 분석 등을 통하여 산출된 자료로 연구
성과의 재현에 필수적이고 객관적인 사실데이터

** 국가연구개발사업 공동관리규정(정의, 데이터관리계획), 연구관리 표준매뉴얼(연구자
권리) 등 개정 추진('18년~)

⑤ 연구데이터의 산업적 활용 촉진 및 일자리 창출

- (연구데이터 활용기반 조성) 기업 등이 연구데이터를 수집·활용할 수 있도록 API 제공하고 데이터활용 지원센터* 운영
 - * 대·소분야 전문센터별로 전담인력 배정하여 사용자 문의에 대응토록하고 정기적 활용교육 프로그램 운영, 데이터 활용 컨설팅 등 제공
- (연구데이터 사업화) 사업화 아이디어 발굴 및 육성을 위한 경진대회 개최, 연구데이터 사업화를 위한 후속 R&D 지원 등
- (연구데이터 서비스업 육성) 연구데이터 분석 SW 개발 지원 및 우수 분석 SW 조달 추진, 연구데이터 컨설팅 기업 육성 지원 등

4. 참고사항

□ 추진 경과

- 전략 수립을 위해 TF('모아서 새롭게 TF', 총괄 TF 및 5개 분과별 TF) 구성('17.7월), 총 26회 회의 개최('17.7월~12월)
- 공청회 개최('17.12.18.)
- 관계부처 협의 완료('17.12.27.)

□ 향후 계획 및 예산

- 우선 추진분야(바이오·미래소재·대형연구장비·인공지능) 대상 시범사업* 추진('18.1월~)
 - * 분야별 R&D 사업 내 관련 예산 활용 신약개발 효율화 플랫폼 구축, 열전소재분야 연구데이터 플랫폼 구축 등 추진('18년 총 41억원)
- 연구데이터 관리·공유·활용 제도화를 위한 법제 개선안 마련 추진('18.1월~)
- 국가 연구데이터 공유·활용 체계 구축·운영 사업* 기획**('18.1월~)
 - * 국가 연구데이터 관리 체계 구축·활성화, 인력양성, 지원 플랫폼·인프라 구축 등
 - ** (예산) 예산 당국과 협의 통해 신규 예산 프로그램 신설 추진(필요 시 예타 추진)

- 혁신성장 촉진을 위한 -

연구데이터 공유 · 활용 전략(안)

2018. 1. 19.



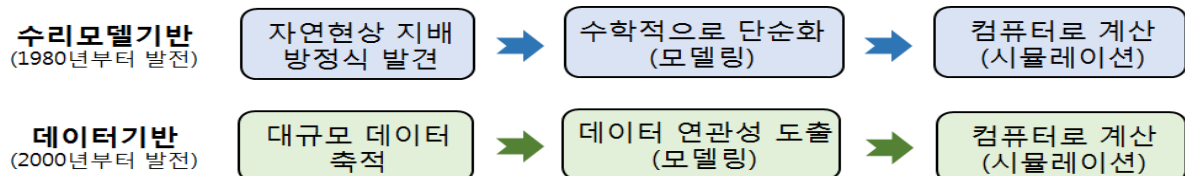
과학기술정보통신부

순서

I. 추진 배경	1
II. 국내외 현황	2
III. 비전 및 핵심 추진과제	5
1. 연구데이터 관리체계 구축 및 공유 커뮤니티 형성 촉진	6
2. 국가연구데이터플랫폼 구축·운영	7
3. 데이터 및 컴퓨팅 활용 R&D 인재 성장 지원	8
4. 연구데이터 관리·공유·활용에 관한 법·제도 마련	9
5. 연구데이터의 산업적 활용 촉진 및 일자리 창출	10
IV. 추진 로드맵	11
[붙임 자료] 우선 추진분야별 추진계획	
1. 바이오	12
2. 미래소재	14
3. 대형연구장비	16
4. 인공지능	18
5. 대형연구단	20

- 4차산업혁명에 인공지능, 빅데이터 등 디지털 기술로 촉발되는 초연결 기반의 지능화 혁명으로 데이터가 경쟁의 원천
- R&D에서도 연구성과와 과정을 개방화하는 오픈 사이언스가 확산
 - 연구데이터의 폭발적인 증가와 현대과학의 거대·복잡화로 다양한 분야간 융합연구 및 공동연구가 지속적으로 확대
 - ※ 대형 협업논문(공저자 1,000명 이상) 수 : 30('10) → 109('12) → 80('14) (Thomson Reuters, '15)
 - 주요선진국은 새로운 지식과 가치의 창출 및 공유·확산을 위해 적극적으로 오픈 사이언스 정책 추진
 - ※ OECD는 오픈엑세스(공공연구출판물), 오픈데이터를 중심으로 추진
- 연구개발 패러다임의 변화와 연구데이터의 가치 상승
 - 첨단연구장비 활용으로 인한 대용량 연구데이터의 생산, 데이터 처리 기술 등 디지털 기술의 발달로 연구활동이 본질적으로 변화
 - ※ 1914년 이후 노벨상 수상 연구의 87%가 대형장비의 대용량 데이터 분석 연구(KISTI, 2016)
 - 시뮬레이션과 데이터 마이닝 기법을 통해 데이터 융합·분석만으로 연구수행이 가능해지고, 이와 함께 연구데이터의 가치가 제고

< 디지털 기술로 인한 연구개발 활동의 변화 >



☞ 연구데이터에 대해서도 국내 실정을 고려한 정책 수립 및 이행 필요

- ※ ‘모아서 새롭게’ TF 운영과 연구현장 의견수렴을 통해 동 방안 마련('17.7~12)
 - 총괄 및 분야별(제도·바이오·미래소재·대형연구장비·대형연구단) TF 총 26회 운영

II

국내외 현황

1

해외 동향

- OECD는 오픈사이언스 실행수단으로 ①의무이행 조항 도입, ②연구성과 인용 표시 등 인센티브 도입, ③제도적·물리적 환경 등 역량강화 수단 구축 제시

법제도 · 정책

- 연구데이터를 연구결과물로 규정하여 공유·활용을 위한 제도적 기반을 마련하고, 분야별 특성을 반영한 공유·활용 계획 수립
 - (미국) 정보자유법('66), 미국경쟁력 강화법('10), 백악관 '연구데이터 관리 및 공유에 대한 지침'('13)에서 원칙을 제시
 - 공적자금이 투입된 연구데이터는 연구기관별(1억달러 이상) 관리·공개*, 연구 기획단계에서 데이터관리계획(DMP) 제출 의무화
 - * 연구 종료 후 12개월의 엠바고 기간 이후 공공에게 공개
 - (영국) 정보자유법에서 공개 정보의 범위에 데이터셋을 포함시켜 연구데이터를 포함한 광의의 데이터를 공개 대상으로 규정
 - 영국연구위원회의 공동원칙을 기반으로 각 연구위원회*가 독립적으로 분야별 특성을 반영한 데이터 관리 및 공유 정책 제시
 - * 7개 연구위원회 : 예술인문, 생명과학, 공학물리, 경제사회, 의학, 자연환경, 과학기술설비
 - (호주) 정보자유법의 공개대상에 모든 정부 정책 관련 데이터를 포함, '책임 있는 연구수행을 위한 국가지침'에 연구산출물* 관리 규정 마련
 - * 연구데이터를 포함한 연구소유권, 저작권·지적재산권 등
 - 호주연구위원회(ARC)의 재정지원을 받는 연구기관에 대하여 의무적으로 데이터관리계획 제출 규정('14)

인프라 · 지원

□ 연구데이터 수집·저장·연계·유통을 위한 국가데이터센터 구축 및 컨설팅·교육을 제공하는 지원조직 운영

○ (미국) 과학분야별 4개의 지역 허브를 구축하고, 커뮤니티(spoke) 중심의 연구데이터 활용 생태계 형성 추진

- 대학 차원에서 연구데이터 생산·관리 활동 지원 서비스* 제공

* 데이터관리계획(DMP) 작성, 메타데이터 표준, 데이터 저장·백업·보존 등 지원

명 칭	분 야
South Hub	보건의료, 해안재난, 산업 데이터, 재료 및 제조, 생물서식지
Northeast Hub	에너지, 금융, 교육 관련 데이터과학, 기후 및 환경
Midwest Hub	농업, 수자원, 스마트시티
West Hub	대용량 데이터 기술, 천연자원, 자연재해, 개인맞춤형 의료

○ (영국) 연구데이터 공유센터 및 연구데이터의 공유·활용을 지원하는 디지털 큐레이션 센터*(DCC) 구축·운영

* 연구데이터 관리 자료·교육 제공, 연구데이터 큐레이션·분석 전문가 양성 프로그램 운영

○ (호주) 호주국가데이터서비스(ANDS)에서 연구데이터 관리·공유 인프라 구축, 연구자 가이드·교육 제공 및 연구자간 협력 지원 담당

인식 · 연구공동체

□ 그동안 데이터·컴퓨팅 기반 연구공동체를 육성해 왔으며, 다수의 연구자가 연구데이터 공유활동 수행 경험*

* 미국 46%, 영국 43%, 호주 41% 연구데이터 공유 경험(Wiley 2014)

○ (미국) '95년부터 컴퓨팅 기반 교육·융합 연구 환경을 동적으로 지원하는 웹기반 플랫폼 허브제로(HubZero)* 구축·운영

* 60개 이상의 Hub('15년 기준)로 구성, 연간 60만 여명이 사용

○ (유럽) Horizon2020 사업에서 '14년부터 e-인프라에 기반한 분야별 전문센터(CoE, Center of Excellence) 운영('16년 9개)

○ (호주) '09년부터 분야별 가상실험실 플랫폼(Nectar) 운영('17년 12개)

- 실험·관측·분석 등 연구개발 과정에서 산출된 ‘연구데이터’는 연구 결과로서 가치가 높음에도 연구성과물에 미포함
- 연구자 개인적으로 데이터를 관리중으로, 관리·공유에 대한 연구 현장 인식이 부족(79%가 개인자산으로 인식)

< 정부R&D 연구성과물 관리전담기관 현황 >

연구성과물	관리전담기관	연구성과물	관리전담기관
논문	한국과학기술정보연구원	연구시설·장비	한국기초과학지원연구원
특허	한국특허전략개발원	화합물	한국화학연구원
보고서·원문	한국과학기술정보연구원	생명 자원	한국생명공학연구원
기술요약정보	한국산업기술진흥원		한국생명공학연구원
소프트웨어	한국저작권위원회		농업유전자원센터
	정보통신산업진흥원		농림수산식품교육문화정보원

- 데이터 집약형 연구 확산에 따라 특정 분야* 차원에서 일부 데이터를 수집·관리 중이나, 활용은 저조

* 유전체(국가생명자원정보센터), 소재 물성(소재정보은행) 등

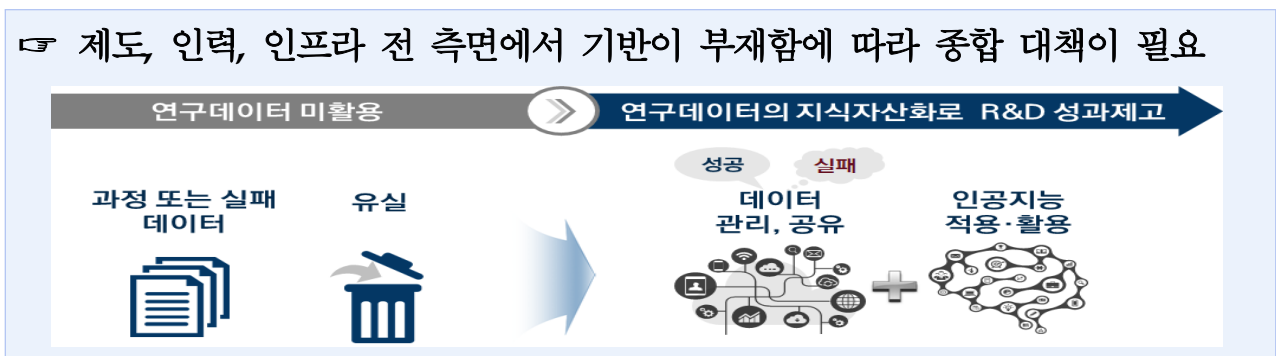
- 최근 대형연구장비의 완공·가동계획에 따라 데이터 저장·분석 수요 급증이 예상* 되나 데이터 저장·분석 인프라는 부족

* 포항방사광가속기 연간 10PB 생산, 중이온가속기 연간 1.5PB 생산 예상

- 소재, 바이오 분야에서는 데이터 기반 프로그램이 일부 진행중*이나 아직 초기단계이며, 연구공동체 형성에 대한 지원이 매우 부족

* 촉매 분야 데이터 기반 환경구축(*17, KIST), 열전소재 빅데이터 플랫폼 시범구축(*17) 등

☞ 제도, 인력, 인프라 전 측면에서 기반이 부재함에 따라 종합 대책이 필요



Ⅲ

비전 및 핵심 추진과제

비전	연구데이터 지식자산化·빅데이터化
목표	국가 연구데이터 공유·활용 활성화
추진 전략	① 연구데이터 공유·활용 기본원칙 제시 ② 융합·공동연구 역량 강화를 위해 분야별 연구공동체 형성 ③ 인프라 구축, 지원으로 연구자 부담 경감 ④ 데이터 집약형 연구부터 우선 추진, 단계적 확장 ⑤ 데이터 공유·활용의 가치를 단기간에 보여줄 시범사업 추진

핵심 추진과제

1. 연구데이터 관리 체계 구축 및 공유 커뮤니티 형성 촉진	○ 계층적 연구데이터 연계체계 마련 ○ 데이터 기반 연구커뮤니티 형성 촉진 및 확산
2. 국가연구데이터 플랫폼 구축·운영	○ 국가 연구데이터 플랫폼·인프라 구축·운영 ○ 대·소분야 전문센터 플랫폼·인프라 지원
3. 데이터 및 컴퓨팅 활용 R&D 인재 성장 지원	○ 연구자에 대한 데이터 및 컴퓨팅 활용 교육 제공 ○ 데이터 기반 연구참여를 통한 고급인력 양성 추진 ○ 전문센터 기반 강화를 위한 전문인력 양성
4. 연구데이터 관리 공유·활용 관련 법제도 마련	○ 연구데이터의 관리·공유·활용 근거규정 마련 ○ 데이터 관리계획(Data Management Plan) 도입
5. 연구데이터의 산업적 활용 촉진 및 일자리 창출	○ 연구데이터 활용기반 조성 ○ 연구데이터 사업화 촉진 ○ 연구데이터 서비스업 육성

1

연구데이터 관리체계 구축 및 공유 커뮤니티 형성 촉진

□ 계층적 연구데이터 연계체계 마련

- ◇ 다양한 분야에서 개별적으로 생성되는 연구데이터를 체계적으로 취합·관리·공유할 수 있도록 계층적 공유·활용 체계를 구축·운영

< 국가 연구데이터 공유·활용체계 개념도 >



- (국가연구데이터센터) 연구데이터 공유·활용 체계의 총괄센터로서 국가연구데이터플랫폼 운영, 대·소분야 전문센터 지원* 등 기능

* 클라우드 컴퓨팅 인프라 제공, 플랫폼SW·분석SW 개발, 데이터 활용 교육 제공 등

- (대분야 전문센터) 소분야 전문센터의 연구데이터를 취합·관리, 특화된 데이터 관리·분석SW 개발, 교류·협력 프로그램 운영 등 기능

* 소분야 전문센터로 지정되지 않은 R&D 과제 또는 대형연구개발사업에서 수집된 데이터의 1차 아카이빙(Archiving)도 대분야 전문 센터가 담당

- (소분야 전문센터) 데이터 기반 공동연구의 추진 주체*로서 연구 데이터 활용 방향을 설정하고, 데이터 표준화, 활용지원 등 제공

* 신약개발센터, 국가생명자원정보센터(유전체), 천문연구원(위성정보) 등

□ 데이터 기반 연구 공동체 형성 촉진 및 확산

- (형성 촉진) 데이터 기반의 시뮬레이션·AI 등 통한 난제 해결(신소재·신약 개발, 과학적 발견 등)을 목표로 하는 집단연구 지원 프로그램 신설

※ 미국의 허브제로 및 EU Horizon 2020 사업의 CoEs 프로그램 벤치마킹

- (성과 확산) 데이터 기반 연구 수요 수용 및 확산을 위해 현재 활용 가능한 데이터를 활용하는 소규모 연구 과제 신설(2~3억원/년, 3~5년)

2

국가연구데이터플랫폼 구축·운영

□ 국가 연구데이터 플랫폼·인프라 구축·운영

- (플랫폼) 연구데이터 공유·활용체계 전반의 연구데이터를 원스탑으로 검색하고 활용할 수 있는 국가연구데이터 플랫폼 구축·운영

* 해외의 데이터 공유·활용 프로젝트와 연계하여 해외 데이터 리포지터리와 상호운용성 확보

- 각종 연구데이터에 대한 색인·검색 기능과 연구자가 데이터 융합 및 분석에 활용할 수 있는 분석도구* 및 가상작업환경 구축

* 다양한 데이터 분석도구의 자유로운 공유·유통을 위해 데이터 분석도구 마켓 운영

- 데이터 유통이력의 투명한 관리와 위변조 방지 등을 위해 데이터 저장·유통 시스템에 블록체인 기술을 접목하는 방안 검토

- (인프라) 국가연구데이터 플랫폼의 운영과 대·소분야 전문센터 임무수행 지원을 위한 컴퓨팅·스토리지·네트워크 인프라 구축*

* 연구현장 수요 변화에 따른 탄력적 운용성 확보 등을 위해 민간 클라우드 적극 활용

< 국가연구데이터플랫폼 개념도 >



□ 대·소분야 전문센터 플랫폼·인프라 지원

- (플랫폼 지원) 전문센터의 연구데이터 공유·활용 플랫폼 및 데이터 분석 SW 개발·운영을 지원하고 컨설팅* 및 교류·협력 프로그램 제공

* 연구자 가이드·교육 제공 및 연구자간 협력 등 컨설팅·교육 프로그램

- (인프라 지원) 대분야 전문센터의 영구보존이 필요한 데이터를 백업하고 데이터 관리 시스템 운영(인프라 미보유 센터의 경우)을 지원

- 소분야 전문센터(개별 연구자 등)가 보유한 데이터를 저장·관리·검색* 할 수 있도록 저장공간(예: 개인당 100TB)·데이터분석 도구 등 제공

* 종료과제 및 수행중인 과제에서 생산된 데이터의 공개 여부와 수준 등은 연구자가 결정하되, 성실실패 과제 과정에서 생산된 데이터는 공개토록 우선 독려

□ 연구자에 대한 데이터 및 컴퓨팅 활용 교육 제공

- 대학생·대학원생 및 현직 연구자들을 대상으로 전문분야에서의 데이터·컴퓨팅 활용을 위한 초급 및 중급 교육 프로그램 제공*

* 한국데이터진흥원 재직자 대상 빅데이터 전문인력 교육 프로그램 연계('18년~)

- 전 분야에서 공통적으로 활용되는 데이터 분석기법과 분야별 문제해결에 특화된 데이터 분석 및 활용 사례 교육 등 제공

- 온라인 데이터·컴퓨팅 활용 교육·실습 환경을 제공하고 분야별 교육용 데이터 बैं크를 구축

- 수준별 교재 개발하고 교육과정에 기계학습 도구, 계산과학 SW*, 광범위한 교육용 데이터** 등 교보재 탑재

* 첨단사이언스교육허브(EDISON) 보유의 소재, 화학, 열유체 등 6개 분야 400여종 SW 활용

** 보고서·논문 등 연구개발 성과물, 국내외 공개데이터, 계산데이터 등

□ 데이터 기반 연구참여를 통한 고급인재 성장 지원

- 데이터 기반 집단연구를 통해 각 분야의 문제해결이 이루어지는 소분야 전문센터에서 R&D와 인력양성 동시 추진으로 고급인재 육성

- 대학원생 인턴쉽 및 산·학·연 방문연구자 프로그램 등을 통해 집단연구에 참여하게 함으로써 고품질 데이터의 활용 기회 제공

□ 전문센터 기반 강화를 위한 전문인력 확보

- 데이터 기반 연구 및 전문센터의 증가와 함께 수요가 크게 증가할 것으로 예상되는 ICT 기반 관리·운영·서비스 인력 양성

- 인프라 운영, 데이터 관리 및 병렬분석 환경, 효율적 전송, 시스템 모니터링, 장애분석, 보안 등에 대응 가능한 실무형 관리자 양성

□ 추진 전략

- (단기) 국가연구개발사업의 중간산출물인 연구데이터의 수집·관리 및 공유·활용을 위한 법·제도적 기반 마련
- (중·장기) 분야별 연구데이터 관련 정책의 추진경험을 바탕으로 국가 연구데이터 법제 정비 및 개선

□ 제도화 방안(안)

- 연구데이터의 정의(「공동관리규정」 개정)
 - 국가연구개발사업의 각종 실험, 관찰, 조사, 분석 등을 통하여 산출된 자료로 연구성과의 재현에 필수적이고 객관적인 사실 데이터
 - ※ 美 OMB(Office of Management and Budget)의 정의 참조
- 데이터관리계획(DMP : Data Management Plan)의 도입(「공동관리규정」 개정)
 - 주관연구기관이 비영리기관이고, 중앙행정기관의 장이 필요하다고 인정하는 과제에 적용
 - 협약체결 단계에서 연구제안자로부터 데이터관리계획(DMP) 접수
 - 연구데이터 시범사업(18)에 한하여 데이터관리계획을 우선 제출받고 향후 표준서식 등을 마련하여 제도화 추진
 - ※ 美 OSTP(Office of Science and Technology Policy)의 Memorandum : 연간 1억 달러 이상 R&D 지원기관은 Data 공유·활용 계획(DMP)을 수립·시행하여야 함
- 데이터에 대한 연구자 권리의 근거 확보(연구관리 표준매뉴얼에 반영)
 - 논문, 특허와 마찬가지로 타인의 연구데이터를 인용(citation)한 경우 출처를 명시

□ 연구데이터 활용 지원기반 조성

- (데이터 API 개발) 기업이 사업화 R&D 및 시제품 개발에 연구데이터를 손쉽게 활용할 수 있도록 연구데이터 API* 개발·제공

* (사례 추가) 미국 국립보건원(NIH)에서는 연구데이터가 신약개발, 질병진단 등에 활용될 수 있도록 화합물데이터 125만건을 API로 개발하여 제공(PubChem 프로젝트)

- (데이터 활용 지원센터) 기업 등이 연구데이터를 활용하는 과정에서 발생하는 애로사항 해소를 위한 데이터 활용 지원센터*를 운영

* 대·소분야 전문센터별로 전담인력 배정하여 사용자 문의에 대응토록하고 정기적 활용교육 프로그램 운영, 데이터 활용 컨설팅 등 제공

□ 연구데이터 사업화 촉진

- (아이디어 사업화) 공개적 경진대회를 통해 축적된 연구데이터를 활용한 사업화 모델을 발굴·지원

- 발굴된 우수 아이디어에 대해 초기 사업화 자금 및 멘토링 제공하여 연구데이터 응용 제품·서비스* 개발 및 스타트업 활동 촉진

* (예시) 스탠다임, 파로스BT등 국내 스타트업은 화합물·약효·논문 등 연구데이터에 자체 개발 알고리즘을 적용하여 신약 후보물질의 효능·구조·작용기전 등을 예측하는 서비스 제공

- (제품·서비스 개발 촉진) 기업이 연구데이터 원저자와 공동으로 응용 제품·서비스 개발을 추진할 수 있도록 R&D 연계 지원*

* 공공연구성과사업화 지원사업 중 후속R&D지원('18년 52억원) 등 연계 검토

□ 연구데이터 서비스업 기업 육성

- (분석도구 개발 지원) 상용 연구데이터 분석 SW R&D를 지원하고 우수 SW는 국가연구데이터센터가 라이선스 구입하여 보급·활용

- (연구데이터 컨설팅) 수요자 맞춤형 연구데이터 정제·분석 등에 특화된 연구데이터 컨설팅 기업 육성*

* 연구개발서비스업 혁신역량 강화지원 사업('18년 40억원) 등 연계하여 R&D 서비스 도구 개발, 컨설팅·시험분석 기법 개발 등 역량강화 지원

IV

추진 로드맵

◇ 데이터 집약형 분야에서 시범사업을 추진하여 성공사례를 조기에 창출

- 시범사업을 통해 구체화된 실행계획 도출 후 본사업 추진

시범사업 추진('18년)

- 바이오, 미래소재, 대형연구장비, 인공지능 분야에서 시범사업 추진

< 추진 분야별 시범사업 추진(안) >

<바이오>	유전체 빅데이터 경진대회
	인공지능 활용 신약개발 플랫폼 구축
<소 재>	열전소재 분야 연구데이터 플랫폼 구축
	축매분야 데이터 기반 R&D 환경 구축(KIST)
<대형장비>	이기종 대형연구장비 데이터 융합 분석사업
<인공지능>	인공지능 오픈이노베이션 허브 구축

- 연구데이터 관리·공유를 위한 제도적 기반 마련(공동관리규정 개정* 등)

* 국가연구과제 제안 시 데이터관리계획(DMP) 제출 등 제도 도입

- 연구공동체 육성을 위해 데이터·컴퓨팅 기반 집단연구 신규지원

※ 선도연구센터사업에서 3개 과제 예정, 예산 및 인프라 클라우드 지원

- 국가연구데이터플랫폼 구축예산 등 예산 확보 추진

- 연구현장의 인식제고와 공감대 확산을 위해 성공사례 발굴과 정례 포럼 개최 등 홍보 강화

성과확산 ('19년~)

- 시범사업 성과분석을 통해 제도를 개선·보완하고 파급성 등 우선 순위에 따라 전 분야 확산

- 국가연구데이터플랫폼과 국가과학기술정보서비스(NTIS)간 메타 데이터 연계 등 연동방안 마련

- ◇ 과학기술 분야에서 중심이 되는 실험, 관측, 참조 데이터를 중심으로 컴퓨팅을 비롯한 IT 인프라 활용이 많은 분야, 해외 사례 등을 고려하여 우선 추진분야 선정

1 바이오

□ 데이터관리 현황

- 국가연구개발사업에서 생산된 ‘생명정보(유전체·단백질 정보 등)’를 수집·공유·활용하는 통합DB 운영 중(생명연)

※ 현재 약 120만건, 91TB의 생명정보 및 분석 프로그램 제공

□ 추진 방향

- 분야별 전담기관을 중심으로 한 데이터 관리·활용 및 관련 사업과의 연계 강화를 통한 데이터 관리체계 고도화

<분야 및 활용방향 예시>

분야	신약개발	유전체	뇌연구
데이터 형태*	화합물 화학 구조	DNA 염기서열	MRI 등 영상데이터
활용방향	질병치료	질병예측	질병진단

* 데이터 수집 시 표준화된 데이터 원본 뿐 아니라, 이와 관련된 효능정보, 실험정보, 생물종 등의 연구정보(메타데이터) 필수 포함

□ 추진내용

- (수집·관리) 과제 평가 시 DMP 양식에 맞춘 연구데이터 등록을 의무화하며, 전담기관은 분야별 DB 구축 및 데이터 큐레이션 담당

- (공유·활용) 공유·활용의 성공 모델 창출을 위한 분야별 단기 시범사업('18) 및 중장기 집단연구('19~) 추진

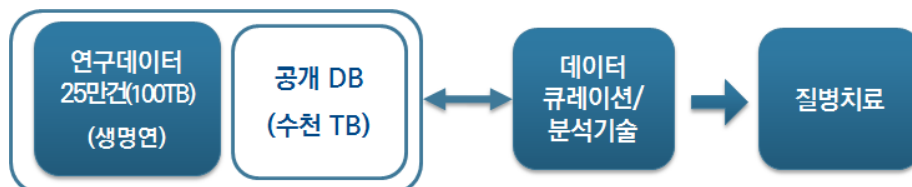
□ 단계별 추진계획

- ('18) 데이터 활용모델 발굴을 위한 유전체·신약개발분야 시범사업 추진
- (1단계) 중장기 집단연구 및 기존 연구커뮤니티와의 연계 착수
- (2·3단계) 연구데이터 고도화 및 성공적 활용모델 창출

< 2018년도 시범사업 추진계획 >

▶ 유전체 빅데이터 경진대회('18년, 2.5억)

- (목적) 국가연구개발사업을 통해 생산된 25만 건(100TB)의 유전체 연구 데이터를 활용한 커뮤니티 내 데이터 공유 문화 확산 및 활용모델 창출
 - (내용) 유전체 분석·활용기술 경진대회를 통해 우수 분석기술을 확보하여, 유전체 정보를 활용한 질병 치료시간·비용 감소 도모
- * 우수 아이디어에 대해 포상하며(2팀 내외), 추후 연구과제로 선정하여 연구그룹에 대해 지속 지원(과제당 1.25억원 내외)



▶ 인공지능 기반 신약개발 플랫폼 구축('18년, 10억)

- (목적) 국가연구개발사업을 통해 생산된 50만 건의 화합물 연구데이터에 인공지능을 적용하여 신약개발 시간·비용을 1/3로 단축
- ※ 하나의 신약개발에는 1조원, 15년 이상의 막대한 시간과 비용 소요
- (내용) 인공지능 연구자-신약개발 연구자가 한 팀이 되어 연구데이터를 활용한 인공지능 기반 신약 효능·독성 예측 모델 등 개발·검증



□ 연구 및 데이터관리 현황

- (과기정통부) 미래소재디스커버리('15~, 16개 연구단), 웹기반 계산과학 플랫폼 시범사업('16~), 소재 빅데이터 플랫폼 시범 구축('17~) 등 추진
- (산업부) 산업용 소재의 물성데이터를 수집·제공하는 소재정보은행('07~), 물성데이터를 측정·제공하는 국가참조표준센터('06~) 구축·운영

□ 추진 방향

- 다양한 구조의 소량 데이터로 구성된 소재 연구데이터 특징을 고려하여 응용분야별 연구데이터 수집 및 활용체계 구축
- 연구데이터 표준화 및 메타데이터 공개로 데이터 활용도 증대
 - 데이터 큐레이션 시스템 구축으로 연구자의 데이터 접근성 향상
 - 데이터 공개여부, 비공개 기간 선택 등 연구자의 저작권 보호방안 마련
- 소재와 데이터 전문가의 응용 융합연구(머신러닝 등)를 추진하여 새로운 소재연구 활용도 향상
 - 산업소재 물성 중심의 '소재정보은행'과 연계를 통한 산학연 활용도 제고

□ 추진 내용

- (미래소재 전문센터) 소재 전분야의 R&D 연구데이터 허브 구축
 - 소분야 전문센터 구축·지정 우선순위 설정, 수집 기획 및 사후관리
 - 소재 전 분야 데이터의 표준화 큐레이션 시스템 구축, 소재정보은행과 연계
 - 기존과제(글로벌 프론티어 등)에서 생성된 연구데이터 수집·관리체계 구축

- (소분야 전문센터) 분야별 소재연구데이터 구축 및 활용기술 개발
 - 열전소재, 촉매 등 응용분야별 연구데이터 수집(국내·외) 및 활용 연구
 - * 웹기반 계산과학 플랫폼 시범사업('16~), 열전소재 연구데이터 플랫폼 시범 구축('17~)
 - 미래소재 전문센터와 소분야 전문센터의 데이터 연계 시스템 구축·활용

□ 향후 계획

- (미래소재 전문센터) 소재연구데이터 수집 및 활용 인프라 구축('19~)
 - * '18년 응용분야별 데이터 수집 우선순위, 지적재산권 보호방안 등 미래소재 전문센터의 역할 연구 추진
 - * (일) MI2사업(Materials Information Integration Initiative, 2015): 5년간 총 21억엔 지원
- (소분야 전문센터) 응용분야별 데이터 수집·활용과제 추진('17~)
 - * '17년 시범과제(열전소재, 촉매분야) 추진, '18년 시범과제(2개 내외) 신규지원 예정
 - * '19년부터 3~4개 분야 내외 본격 지원(15억원/연)

< 2018년도 시범사업 추진현황 >

- ▶ 열전소재 분야 연구데이터 플랫폼 구축 시범 연구('17~'21)
 - (목적) 소재 개발 효율성 제고를 위한 열전소재 분야 빅데이터 플랫폼 구축 및 활용 기술개발 시범 추진
 - (내용) 웹기반 열전소재 물성 데이터(실험, 계산 및 공개된 데이터 등) 자동 수집기술 개발 등 빅데이터 플랫폼을 구축하고, 이를 이용한 기계학습 기반의 소재 탐색 최적화 기술 개발
- ▶ (출연연 기관고유) 촉매분야 데이터 기반 R&D 환경 구축(KIST/'17~'18년)
 - (목적) R&D 빅데이터 및 정보 기술을 이용할 수 있는 인프라를 구축함으로써 데이터 기반의 연구개발 환경을 제공
 - (내용) ① KIST 내부 R&D 데이터의 수집 인프라로서, 연구자의 통상적 연구 활동이 DB화 될 수 있는 'R&D 플랫폼 구축, ② 데이터 기반 실시간 연구활동 지원을 위한 '비정형 데이터의 수집, 가공 및 저장 인프라 구축, ③ 내외부 데이터의 연동 및 R&D 인포메틱스 플랫폼 구축 및 서비스

□ 데이터관리 현황

- 해외 대형장비(CERN, LIGO, KEK) 및 일부 국내 대형장비의 연구 데이터 공유·분석 환경 지원 중*

* 기초연구실험데이터글로벌허브(기초연구기반구축사업내, KISTI 운영('10년~))

□ 추진방향

- 연구데이터 공유·활용 우선 분야, 대형연구장비 보유 분야, 연구 데이터 생산규모 등을 고려한 전략적인 확대 추진

※ (1단계) 가속기, 생명의료 분야 ⇒ (2단계) 천문, 우주, 원자력·핵융합 등 분야 확대 ⇒ (3단계) 대형연구장비 전분야

- 장비 특성(범용, 특수목적)에 따라 추진체계를 설계하고 연구커뮤니티 수요를 기반으로 우선순위 설정

□ 추진내용 및 단계별 추진계획

- 국가연구데이터 플랫폼에 대형연구장비를 포괄하는 로드맵을 구성하되, ICT와 연계되었을 때 시너지가 큰 분야로 점진적 확대



- **(1단계)** 대용량 연구데이터 공유·활용 수요가 있는 커뮤니티를 중심으로 시범사업을 실시
 - **R&D 프로세스 시간 단축**(50%이상) 및 **데이터 융합 성공 사례 도출** (포항방사광가속기, 전자현미경)
 - ※ 연구데이터가 일시적으로 활용되고 소멸되는 구조를 개선하고, 대용량 데이터의 잦은 복사 및 저장으로 인한 비효율성을 개선하는 R&D 효율화
- **(2단계)** 대상 분야를 확대(천문, 우주, 원자력)하고, 장비별 특성*을 고려한 연구데이터 공동 활용 체계 구축
 - * 범용 연구장비는 실험커뮤니티를 육성하여 공유활용 체계를 구축(국가연구데이터 센터에서 지원), 특수목적 장비는 보유기관 중심으로 공유활용 체계 육성(소분야전문센터 지정)
- **(3단계)** 연구데이터 중 최종 정제되어 재활용이 가능한 데이터(참조데이터)를 배포하여 **R&D 장비데이터의 활용 확대**

< 2018년도 시범사업 추진계획 >

▶ 이기종 대형연구장비 데이터 융합 분석 사업('18년, 5.5억)

- **(목적)** 개별 관리중인 대형연구장비의 데이터를 융합하여 시너지 창출
- **(내용)** 연구장비(포항방사광가속기, 전자현미경)의 특성에 따라 생산되는 데이터를 융합하여 구조를 정밀분석
 - ※ 다양한 연구장비에서 생산되는 다양한 연구데이터를 통합적으로 분석 및 활용하는 통합 구조생물학(Integrative Structural Biology) 연구모형 확립
- (대형연구시설장비) 실험을 통해 원천데이터 생성 및 전송 전처리 체계 구축
- (데이터수집 및 분석) 대형연구시설장비의 데이터를 고속 전송할 수 있는 환경, 이기종 데이터를 융합할 수 있는 분석환경 구축
- (데이터융합 4개 프로그램) 이 기종 대형장비의 데이터를 융합하여 정밀한 분석을 할 수 있는 성공사례 도출
- (인력양성 프로그램) 대형연구시설장비, 데이터분석환경, 데이터융합 4개 프로그램이 연계한 데이터 공유·융합 인재양성 교육프로그램 구성

□ 데이터관리 현황

- 인공지능 관련 국가R&D사업을 통해 구축된 기계학습용 데이터셋*을 “공공 인공지능 오픈API/데이터 포털”을 통해 개방·공유 중

* 한국어 QA 데이터셋, 한국어 언어분석 말뭉치, 한국어 형태·의미 말뭉치, 한국어 시공간 말뭉치, 한국형 이미지 등

□ 추진방향

- 인공지능 국가 R&D 과정에서 연구수행기관(출연연·대학·기업 등)이 축적한 기계학습 데이터를 취합·공개·활용

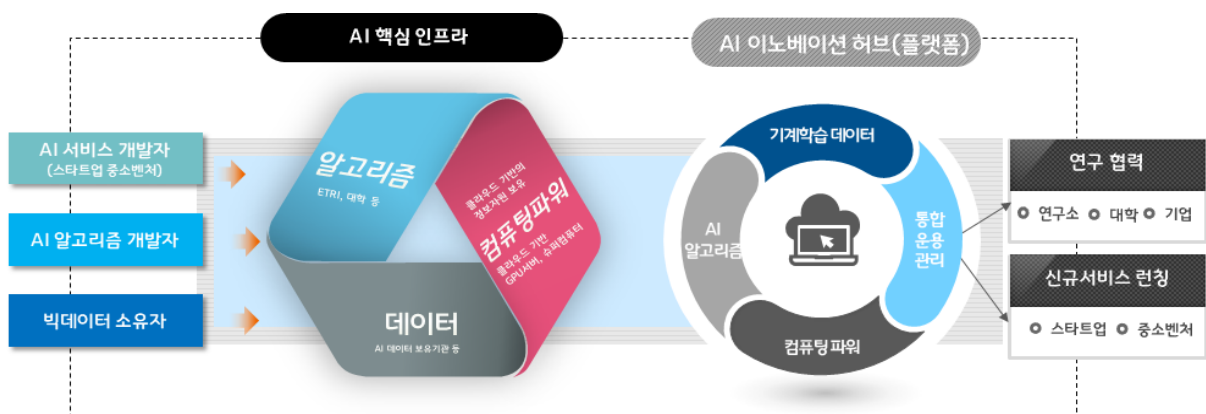
* ETRI(엑소브레인: 언어지능, 딥뷰:시각지능), KIST(안면인식) 등 주요 AI연구기관을 포함한 국가 인공지능 R&D 사업 전반의 기계학습 데이터를 통합관리·운영

- 인공지능 R&D를 위한 기계학습 데이터의 효율적 연계·관리·활용을 위해 “AI 오픈 이노베이션 허브” 구축

- 국내 연구자에 기계학습 데이터를 중심으로 컴퓨팅·개발환경 등 인프라*를 통합 제공하여 국내 인공지능 연구 장벽을 완화

* AI 핵심 인프라: ①기계학습 데이터, ②인공지능 알고리즘, ③고성능 컴퓨팅 자원

< AI 오픈이노베이션 허브 개념 >



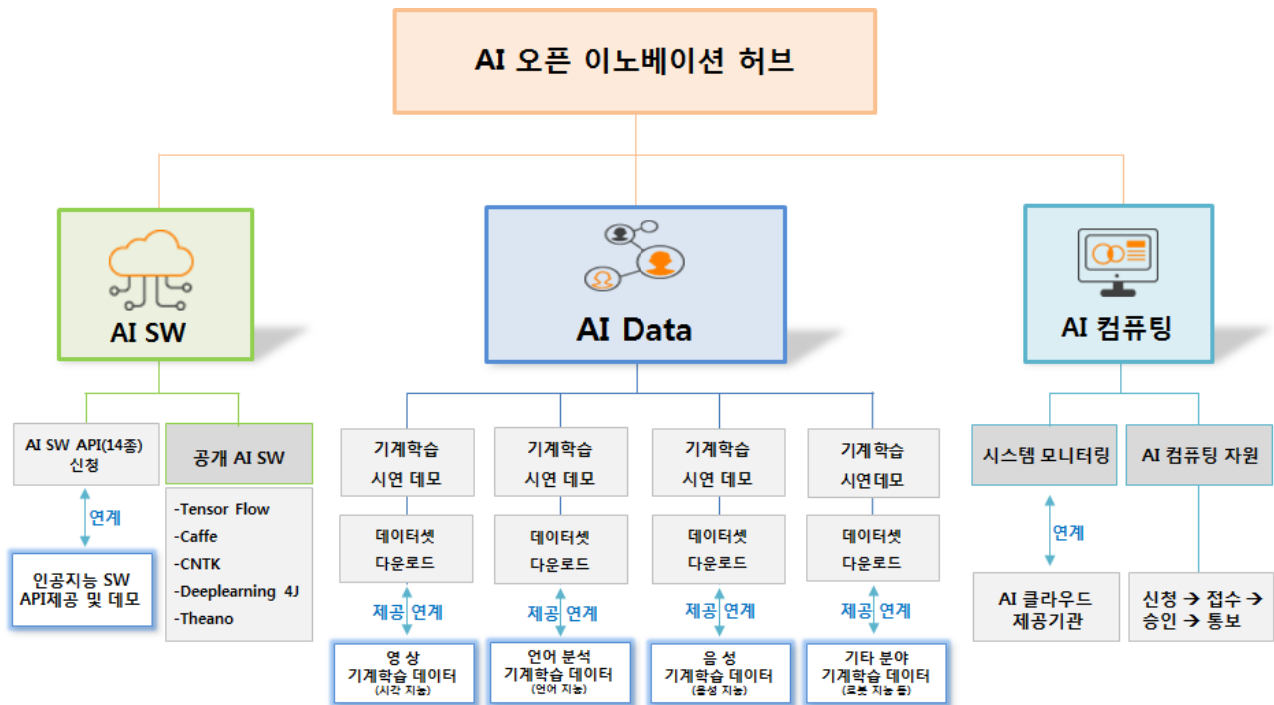
□ 추진내용

- 인공지능 분야 신규 R&D 과제 추진 시 데이터관리계획(Data Management Plan)*을 제출받아 통합 관리

* AI DMP: 기계학습 데이터의 형식·규격, 저장형태, 연계/제공 방법, 공개규모/범위 등

- 전담기관은 데이터관리계획에 대한 검토·평가의견을 연구기관 선정 및 단계평가 등에 반영하고 데이터 활용성과를 관리
- DMP에 따라 수행기관이 “AI 오픈 이노베이션 허브”에 기계학습 데이터를 제출·등록·연계하도록 조치

< AI 오픈 이노베이션 허브 시범구축·운영안('18년) >



* '18년은 노인심리검사 데이터(이미지), 한국 자동차 데이터(이미지) 등 구축·제공 예정

□ 향후 추진 계획

- ('19년) 기계학습(딥러닝) 전문 연구기관·기업(스타트업 등) 협의체 구축을 통한 데이터 활용 촉진기반 마련
- ('20년)전문 분야별 데이터간 융합형 기계학습모델 테스트 플랫폼 구축

□ 데이터관리 현황

- 글로벌프런티어사업('10~'23) : 10개 연구단, 170여 연구기관 참여
 - 연구단과 연구기관 세부과제(약 170개) 참여 연구자별 연구데이터 관리
 - * 분야 : 바이오(2개), 소재(4개), 나노/소재(2개), IT(2개)
- IBS : 28개 연구단(본원5, 캠퍼스14, 외부9) 별 연구데이터 관리
 - * 분야 : 바이오(10개), 소재(7개), 나노(4개), 수학/이론물리(3개), 기타(4개)
 - * 컴퓨팅 계산수요가 큰 데이터 집약형 연구단은 12개(본원, 캠퍼스)

□ 추진방향

- 대형 연구단은 특정분야가 아닌 여러 분야에 설치되어 있으므로, 분야별 전문센터 생성 주도 보다는 선제적 참여 원칙
 - 다만, 외부에 커뮤니티가 없는 일부 분야는 전문센터 생성 주도

□ 추진내용

- (글로벌프런티어) 연차실적계획서에 DMP(데이터관리)*를 포함하여, 연차점검/단계평가 시 전문가 점검을 통해 데이터 공유 활성화
 - * 실험데이터 양, 공유할 데이터 범위/형태, 데이터의 규격화, 공개 시기 등
 - 종료('19.8~ 순차)가 임박한 연구단은 데이터를 국가연구데이터센터에 우선 보관
 - 메타물질 연구데이터 공유 플랫폼 시범구축(과동에너지, '17년/500건~)
- (IBS) 연간실적보고서에 DMP를 포함, 연구데이터 관리실적 점검
 - 대형장비 연구단(3개)*과 기후물리연구단은 IBS에 전문센터(1~3개) 신설
 - * 지하실험연구단, 초강력레이저과학연구단, 엑시온및극한상호작용연구단

□ 단계별 추진계획

- 단기 : 바이오, 소재 분야 시범사업 우선 참여(프런티어 8, IBS 17)
- 장기 : IBS 자체 소분야 전문센터(4개 연구단) 신설 등 전 분야로 확대

과학기술정보통신부 기초원천연구정책관 HPC팀	
담당자	장완익 사무관
연락처	전 화 : 02-2110-2326 E-mail : wanik.jang@korea.kr