



보도일시	2018. 1. 19.(금) 조간(온라인 1. 18. 12:00)부터 보도해 주시기 바랍니다.		
배포일시	2018. 1. 18.(목) 09:00	담당부서	HPC팀
담당과장	조경옥(02-2110-2360)	담당자	장완익 사무관(02-2110-2363)

서랍 속 연구데이터 함께 쓰는 빅데이터로 새롭게 거듭난다

- 「혁신성장 촉진을 위한 연구데이터 공유·활용 전략」 수립 -

□ 과학기술정보통신부(장관 유영민, 이하 과기정통부)는 **혁신성장** 촉진을 위한 「**연구데이터 공유·활용 전략**」을 추진한다.

○ 「**연구데이터 공유·활용 전략**」은 국가 R&D 추진 과정에서 축적되는 연구데이터를 체계적으로 관리·공유하고 연구자가 새로운 지식과 가치 창출에 연구데이터를 적극 활용토록 지원하기 위해 마련됐다.

○ 과기정통부는 이번 전략을 1월 19일(금) **국가과학기술심의회 운영위원회**에 보고하고, 전략 추진에 본격 나설 계획이다.

□ 세계적으로 데이터 중심 R&D 활성화*되어 데이터 분석이 연구개발의 주요 추진 동력으로 부각되고, 이를 활용한 다분야융합·공동연구가 활성화됨에 따라 연구데이터 공유·활용체계의 필요성이 점증하는 추세다.

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

○ 주요국은 이와 같은 변화에 주목하고 **연구성과와 과정을 개방화하는 오픈사이언스 정책***을 본격 추진하고 있으나, 국내는 연구데이터를 성과물로 관리하지 않는 등 관련 제도·인프라가 미비**한 실정이다

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

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

□ 이에 과기정통부는 **연구데이터 공유·활용체계** 구축을 목표로 **업무 혁신 TF** 중 하나인 '**모아서 새롭게 TF**'를 구성하고, '17년 7월부터 12월까지 총 26회 이상의 TF회의를 개최하는 등 정책방안을 강구했으며,

○ 연구현장과 일반 국민 대상의 공청회('17.1.18.)와 관계부처 협의 등 의견수렴 절차를 거쳐 '**연구데이터 공유·활용 전략**'을 마련하였다.

□ 과기정통부는 '**연구데이터 공유·활용 전략**'을 통해 ▲ 연구데이터 관리체계 구축 및 커뮤니티 형성 촉진, ▲ 국가연구데이터 플랫폼 구축 및 서비스 제공, ▲ 데이터 및 컴퓨팅 활용 인재 성장 지원, ▲ 연구데이터 공유·활용에 관한 법·제도 마련, ▲ 연구데이터의 산업적 활용 및 일자리 창출 등 과제를 추진할 계획이다. 【별첨 참조】

□ 과기정통부는 전략의 효과적 이행과 신속한 성과 창출을 위해 '**18년부터 바이오·미래소재·대형연구장비·인공지능** 등 데이터 기반 R&D가 활발한 분야에 대한 시범사업에 즉시 착수*하고,

○ 관련 **법제 개선안** 및 **지원 사업 계획**을 마련해나갈 예정이다.

* 신약개발 플랫폼 구축, 열전소재 연구데이터 플랫폼 구축 등('18년 총 41억원 규모)

□ 과기정통부 정병선 연구개발정책실장은 “‘연구데이터 공유·활용 전략’은 세계적인 **오픈사이언스 확산 추세**에 발맞추어 나가기 위한 **핵심 정책과제**”라며,

○ “연구데이터 공유 문화가 **연구공동체 내에 신속하게 정착**될 수 있도록 **시범사업 성과의 조기 가시화**를 위해 노력하겠다”고 말했다.

○ “다만, 연구데이터의 공유·활용 제도는 연구현장에 큰 변화를 미칠 수 있는 사안인 만큼 시범사업 성과와 연구현장의 목소리 등을 종합적으로 검토하여 신중하게 확산시켜 나갈 것”이라고 말했다.



이 자료에 대하여 더욱 자세한 내용을 원하시면
과학기술정보통신부 장완익 사무관 ☎ 02-2110-2363)에게 연락주시기 바랍니다.

붙임1 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

- (관리체계) 연구데이터의 체계적인 관리를 위해 국가연구데이터센터 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 담당, 분야별 전문센터는 데이터 취합·관리 및 활용촉진 등 담당

< 연구데이터 관리 체계 >



- (커뮤니티 형성 촉진) 대규모 연구데이터를 구축·활용하는 집단연구 과제를 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 플랫폼을 구축

- 통합플랫폼을 통해 사용자에게 연구 데이터 융합 및 분석을 실시할 수 있는 가상 작업환경과 데이터 분석도구*를 제공

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

- 데이터 유통이력의 투명한 관리와 위변조 방지 등을 위해 블록 체인 기술을 접목하는 방안 검토

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

③ 데이터 및 컴퓨팅 활용 R&D 인재 성장 지원

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

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

- 온라인 빅데이터 교육 프로그램 운영을 위해 교육용 데이터·컴퓨팅 플랫폼을 개발*하고 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*를 마련하고, 연구데이터 관리제도 도입, 데이터에 대한 연구자의 권리보장 등을 위한 법령정비**를 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 데이터 활용 API를 제공하고, 활용 과정에서 발생하는 애로사항 해소를 돕는 데이터활용 지원센터*를 운영

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

- (연구데이터 사업화) 연구데이터를 활용한 사업화 아이디어 경진대회, 연구데이터 사업화를 위한 후속 R&D 지원 사업 등 추진

- (연구데이터 서비스업 육성) 민간의 연구데이터 분석 SW 개발을 지원하고, 우수 분석 SW는 선제적으로 조달·도입, 연구데이터 컨설팅 전문 기업 육성 등 추진

붙임2

'18년 시범사업 추진계획(안)

□ 바이오

- (유전체 빅데이터 경진대회) 국가 R&D 과정에서 축적된 유전체 연구 데이터(약 25만건)를 분석하는 우수SW 발굴·육성을 위한 경진대회 개최*

* 우수 아이디어 포상(2팀 내외) 및 후속 R&D 지원 연계(과제당 1.25억원 내외)

- (AI·빅데이터 신약개발 플랫폼) 국가 R&D를 통해 축적된 화합물 데이터(약 50만건)에 인공지능을 접목하여, 신약 후보물질에 대한 효능·독성 예측 시스템 개발('18년 10억원)

□ 미래소재

- (열전소재 연구데이터 수집·분석 플랫폼) 열전소재 물성 빅데이터 플랫폼* 구축 및 기계학습 기반의 소재 탐색 최적화 기술 개발

* 열전소재 데이터(실험, 계산 및 공개된 데이터 등) 자동 수집 등 빅데이터 기술 접목

- (촉매 연구데이터 공유 플랫폼 구축) 연구노트 등 비정형 데이터에 대한 수집·가공·저장 인프라를 구축하고 데이터 분석도구 개발·제공

※ 내·외부 데이터 간 상호운용성 확보하고 사용자에게 R&D 인포매틱스 서비스 제공

□ 대형연구장비

- (이기종 대형장비 데이터 융합) 포항방사광가속기와 기초과학원 전자현미경의 대규모 연구데이터를 공유·교차 분석하는 시스템을 구축하고, 이를 활용한 단백질 구조분석 연구 추진('18년 5.5억원)

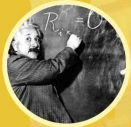
□ 인공지능

- (인공지능 오픈이노베이션 허브) 국내 연구자 및 기업 등에 기계학습 데이터와 컴퓨팅·SW 등 인공지능 핵심 인프라*를 원스톱 서비스

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

붙임3 '연구데이터 공유·활용 전략' 주요 개념 알아보기

①	②
“오래된 서랍을 열어보고 당황하신 적 없나요?” 	기억하지 못하던 수많은 아이템들이 쏟아져 나올 때, “여기 있었는데 또 싶어...” 
③	④
실패하거나 진행중인 사업의 수많은 연구 데이터가 서랍속에서 나와 모아진다면? 모아서 새롭게 TF의 출범! 	87% 1914년 이후 노벨상 수상 연구의 87%가 대형장비의 대용량 데이터 분석 연구에서 나왔습니다  [LIGO 연구단] 중력파 검출로 노벨상 수상 (2017)  [인간 지능 프로젝트] 인간 30 억개 DNA 염기서열 분석 (~2003)

⑤	⑥
이제는 한명의 천재에 의한 연구가 아니라, 잠자는 데이터를 모아 빅데이터화하여 국제 공동연구의 방향으로 바뀌어야 합니다  과거  현재	지금까지는 . 국가 R&D에서 연구데이터는 관리해오지 않았음 . 대형 연구장비에 대한 데이터 저장 인프라 부족 
⑦	⑧
이제부터는 . 연구데이터 관리 및 공유활동 제도화 . 분야별 데이터 공유활용 공동체 형성 . 데이터 관련 인재 성장 지원 . 국가 연구데이터 플랫폼 구축 	누군가에겐 서랍속 기록 누군가에겐 성공의 열쇠  모아서 새롭게  세금 속 연구데이터, 모아서 가치 창출