

연구데이터 공유·활용체계 구축 사업 예비타당성조사 기획 연구

(A study on preliminary feasibility planning on
establishment and operation of research data sharing system)

연 구 기 관 : 한국과학기술정책플랫폼협동조합
연구책임자 : 박미영

2019. 2. 20

과 학 기 술 정 보 통 신 부

연구데이터 공유·활용체계 구축 사업 예비타당성조사 기획 연구

(A study on preliminary feasibility planning on
establishment and operation of research data sharing system)

연 구 기 관 : 한국과학기술정책플랫폼협동조합

연구책임자 : 박미영

2019. 2. 20

과 학 기 술 정 보 통 신 부

본 연구보고서에 기재된 내용들은 연구책임자의
개인적 견해이며 과학기술정보통신부의 공식견해가
아님을 알려드립니다.

과학기술정보통신부 장관 유 영 민

제 출 문

과 학 기 술 정 보 통 신 부 장 관 귀하

본 보고서를 “ 연구데이터 공유·활용체계 구축 사업 예비타당성조사에 관한 연구 ”의
최종보고서로 제출합니다.

2019 . 2 . 20 .

연구기관명 : 한국과학기술정책플랫폼협동조합

연구책임자 : 박 미 영

연 구 원 : 김 준 모

연 구 원 : 이 희 영

연 구 원 : 이 주 영

연 구 원 : 남 승 주

연 구 원 : 유 승 원

연 구 원 : 신 혜 령

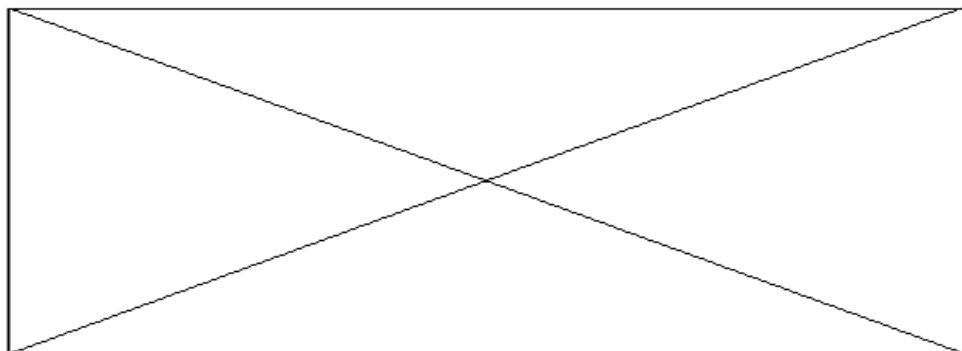


요약문

국가 연구데이터 공유·활용 체제
구축 사업 기획보고서

I. 사업개요

1. 추진 배경

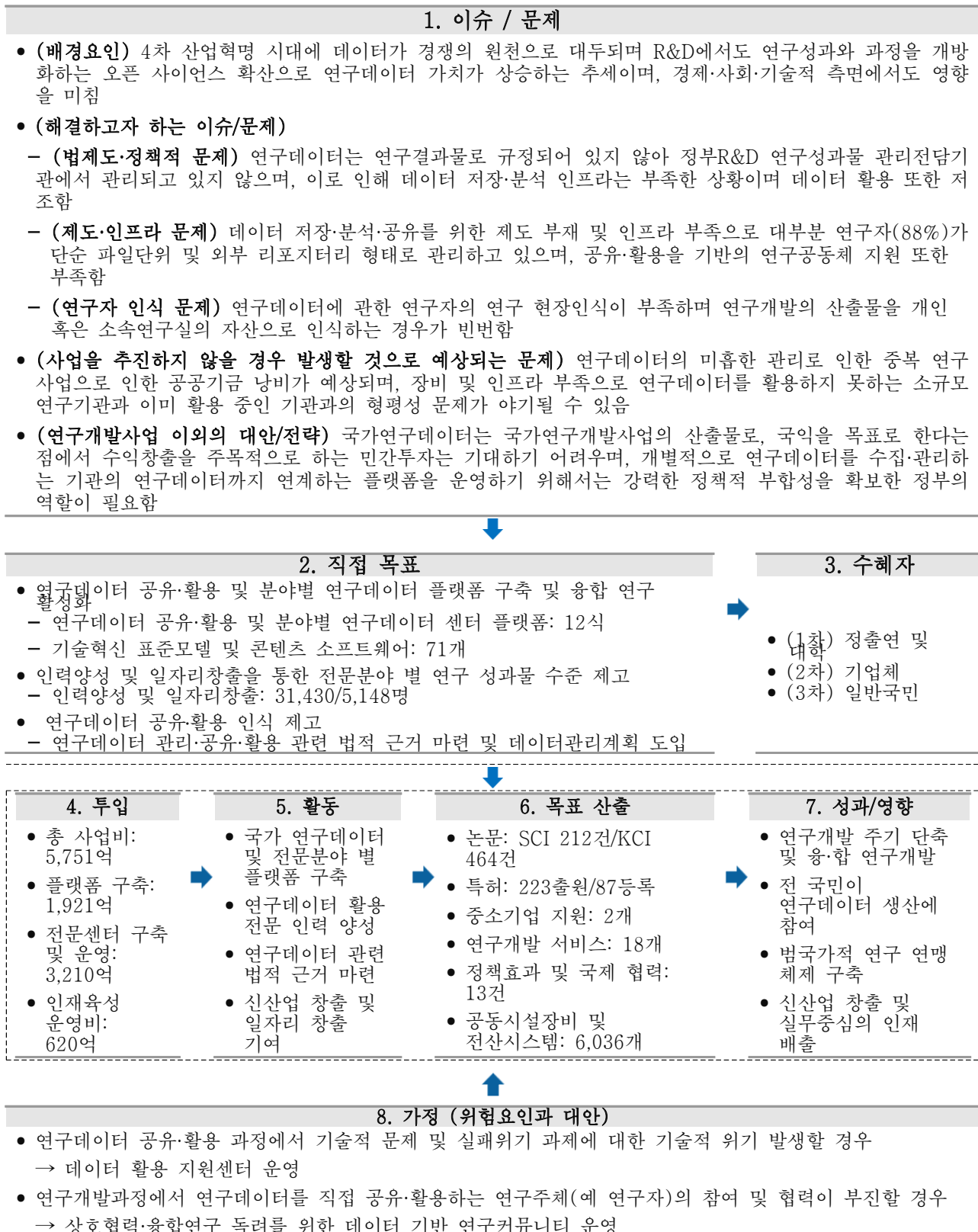


- (4차 산업혁명과 데이터의 중요성) 데이터의 활용이 과학기술 경쟁력의 핵심으로 대두
 - (기술의 융합과 데이터 혁명) 다양한 기술의 개방과 융합을 핵심으로 하는 4차 산업혁명이 산업구조 및 시장경제 모델에 가져올 파괴적인 혁신은 데이터 혁명에 기반을 둔(세계경제포럼, 2016)
 - (데이터 기술과 과학기술 경쟁력) 데이터 분석 및 활용의 중요성이 증가함에 따라 데이터 활용 역량이 4차 산업시대 과학기술 경쟁력의 핵심으로 대두
- (연구개발 패러다임의 변화) 디지털기술의 발전 등으로 연구개발 패러다임도 데이터 중심으로 변화
 - (디지털 기술 확산) 디지털기술은 연구 성과를 확산시키는 방식뿐만 아니라 연구를 추진하는 방식 자체를 변화시킴(Crouzier, 2015; Bosman and Kramer, 2015)
 - (데이터 중심 연구개발) 연구개발 패러다임이 '데이터의 생성, 처리, 분석 등 데이터 중심으로 전환
- (오픈사이언스의 확산) 연구개발 과정 및 성과의 개방화를 통해 연구개발 가치, 목표, 전략의 재정립 확산
 - (과학계의 개방화 추세) 오픈 액세스 저널과 셀프아카이빙 플랫폼을 통한 온라인 논문 공개의 증가 등 개방·공유 사례 증가(OECD, 2015b; 신은정, 2015에서 재인용; EC, 2016)
 - (연구개발 정책의 변화) 연구개발 과정과 성과의 개방과 공유를 국가적 차원에서 주도하는 오픈사이언스 정책이 확산되고 있음
- (국내 현황) 제도·인력·인프라의 미비로 연구개발 분야 전반적으로 관리·활용 저조
 - (제도·인력·인프라) 일부 국가연구개발사업의 성과물 관리제도 운영되고 있으나, 연구데이터의 전반을 관리하는 관련 법·제도 및 전담조직·정보시스템 등 인프라 미흡
 - (관리·활용) 대학 연구자 131명 중 연구데이터 관리에 대한 지침/내규/정책 등이 있는 연구자는 45명(34.3%), 연구데이터를 모든 사람들이 활용할 수 있도록 공유하는 연구자는 7명(5.3%)에 불과

2. 필요성

- (R&D 효율성 증대) 연구데이터 재사용을 향상 및 중복연구 최소화를 통한 R&D 투자효율성 증대
 - (연구데이터 재사용을 향상) 기 수행된 연구를 통해 확보된 연구데이터를 다른 연구에 활용할 수 있는 기반을 마련하여 재사용을 제고
 - － 대부분 연구데이터가 개방적 접근, 체계적 보존(생성~폐기), 적극적 활용(재사용)을 통해 부가가치를 높이지 못하고 단순 저장 상태로만 보존 중(과학기술정책연구원, 2012)
 - － 본 사업을 통해 기 관측된 자료를 재사용함으로써 비용, 시간, 대형 장비를 같은 목적으로 재사용하는 비효율을 예방할 수 있음(한국과학기술기획평가원, 2012)
 - (중복 연구 최소화) 연구데이터 개방과 공유를 통하여 중복 연구를 최소화함으로써 투자효율성 증대
 - － 연구데이터 공유·활용체제 구축에 앞선 주요국에서는 중복 연구 최소화와 투자효율성 증대를 통한 경제적 이득을 크게 기대하고 있음(Beagrie, N. and Houghton J.W., 2016)
- (연구개발 경쟁력 향상) 연구 재현성 및 신뢰성 제고를 통한 국가 연구개발 경쟁력 향상
 - (연구 재현성 및 투명성) 연구데이터 및 소재의 공개를 통해 연구 재현성 및 투명성을 향상시키고 결과적으로 연구 신뢰도를 향상시킴
 - － 연구개발 경쟁력 제고를 위해 현재의 낮은 연구 재현성과 투명성을 개선하여 연구 신뢰도를 높일 수 있음
 - (융합 및 공동 연구) 연구데이터의 개방과 공유로 인해 학제 간 융합 및 공동 연구를 활성화시킴
- (신산업 및 일자리 창출 효과 기대) 신산업 및 양질의 새로운 일자리 창출 효과 기대
 - (산업창출 효과) 연구데이터의 적극적인 공유 및 활용을 통해 신기술 및 미래 신산업 창출 기대
 - － 연구데이터의 적극적인 공유·활용체제 구축은 지능정보기술을 활용한 연구개발을 촉진시켜 새로운 가치와 미래 신산업 창출에 기여할 것으로 예측
 - ※ 실험 없이 최소한의 예산으로 데이터분석만으로 신약후보 물질 발굴 가능해짐(네이처, 2018)
 - (일자리 창출) 연구데이터 활용 확산 시 신규 일자리 창출기대
 - － 연구데이터 공유 및 활용을 위한 데이터 전문직 신규 수요 창출
 - ※ ‘정보통신기술 융합 실증 프로젝트’ 실시로 신규 일자리 2,683건을 창출(과기정통부, 2018)
 - ※ 국가 과학데이터센터 구축 시 취업자유효효과는 63,453명(국가 과학데이터센터를 구축하였을 경우 생산유발효과에서 1.5배의 경제적 효과가 발생함)(박성욱, 한선화, 2011)

3. 사업논리 모형



Ⅱ. 연구데이터 공유·활용 체제 현황분석

1. 해외 현황분석

□ (국가별 연구데이터 관리체제 현황 종합 분석) 한국과 일본은 연구데이터 공유·활용체제 구축에 있어서 미국, 영국, 호주, 중국 등에 비해 상당히 뒤쳐진 상태임

국가			미국	영국	호주	중국	일본	한국
법·제도	기본법		●	●	●	◐	○	○
	적용 범위	출판물	●	●	●	●	○	●
		데이터	●	●	◐	●	○	◐
정책	DMP		●	◐	◐	●	○	○
	엠바고		●	◐	◐	●	○	○
	공유		◐	◐	◐	●	○	◐
	소유권		●	●	◐	●	○	◐
	보상		◐	○	◐	◐	○	○
	보안		●	◐	◐	●	○	◐
	개인정보보호		●	○	◐	●	○	●
	공유시기		●	●	◐	●	○	○
	품질		◐	◐	◐	●	○	○
추진 체계	R&D 전담조직		●	●	●	●	●	●
	데이터관리 지원조직		●	●	●	◐	○	○
인프라	정부지원 리포지터리		●	●	◐	●	○	◐
	대학지원 리포지터리		●	●	◐	●	○	○

□ (시사점) 국내외 현황을 고려할 때, 한국은 국가주도의 연구데이터 공유·활용체제의 구축을 시급히 추진할 필요가 있음

- 기존의 연구 환경을 바탕으로 국가 중심의 연구데이터 공유·활용체제 구축을 위한 효과적인 모델로서 영국과 호주의 사례를 벤치마킹 대상으로 삼을 수 있을 것으로 판단됨
- 호주의 연구데이터 공유·활용체제는 정책의 수립과 인프라 구축이 균형적으로 추진된 모델이며 ANDS는 한국의 연구데이터 인프라 및 서비스 구축 계획 시 참고할만한 가장 이상적인 모델로 판단됨

2. 국내 현황분석

- 공공기관이 보유하고 있는 공공데이터와 일부 연구데이터에 대한 공유·활용 제도*가 마련
 - 연구데이터를 국가연구개발 사업의 결과물로 인정하고 국가차원의 공유·활용을 위한 법적 기반부재
 - * 「과학기술기본법 제11조의3」의 연구성과물의 소유권 규정, 「국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률 제2조」, 「국가연구개발사업의 관리 등에 관한 규정 제3장」 연구개발결과의 등록대상과 기탁대상
- (국가과학기술 기본계획) 연구데이터 공유·활용을 위한 중장기 계획 및 종합 대책을 수립하는 초기 단계
 - 「제3차 연구성과 관리·활용 기본계획」에서 국가연구개발사업의 성과를 연구 종료 시의 성과물에서 연구개발 전주기 정보로 성과 공개 범위 확대
- (부처별 정책현황) 과학기술정보통신부가 연구데이터 공유·활용체계 구축에 있어 중추적 역할
 - (과학기술정보통신부) 국가 R&D 시스템 혁신을 전략과제로 연구데이터 공유·활용 체계 구축을 통한 연구성과 확산을 세부 추진과제로 마련
 - (산업통상자원부) 에너지 빅데이터 플랫폼구축('18)을 시작으로 산업별 데이터 플랫폼 확산추진
 - (보건복지부) 공공기관 보건의료 빅데이터 플랫폼 구축('18~'20) 예정
 - (해양수산부) 해양수산 빅데이터 구축('21년까지 283종 통합)과 민간개방을 위한 정보제공시스템 구축·운영('18)
 - (행정안전부) 공공데이터 소재 및 연관관계를 시각화한 '국가데이터맵' 구축·공개
 - (농림축산식품부) 농식품 빅데이터 플랫폼 구축 등 과학역농 기반 구축
- (출연연 규정) 연구데이터 활용도가 높은 일부 출연연에서 연구데이터 관리규정 운영
 - 연구데이터 생산을 미션으로 하고 연구활용도가 높은 일부 정부출연연구소의 경우 자체적인 연구데이터 관리규정을 수립하여 실행 중
 - (지질자원연구원) 2016년부터 데이터관리계획(DMP)을 주요 사업계획서에 포함하도록 연구데이터 관리규정에 명시하고 데이터 소유권, 생산 및 보존, 공개원칙, 활용에 따른 의무, 관리현황보고 등 세부 정책 수립
 - (극지연구소) 한국극지데이터센터를 설립하고 극지데이터관리정책에 의해 데이터를 관리하고 있으며 해당 정책에는 DMP 제출의무, 소유권, 관리대상 범위, 관리 및 저장, 공개 및 활용, 인용, 보고 등에 대한 세부내용 포함

III. 경쟁력 진단 및 추진방향 설정

1. SWOT 분석

- 연구데이터의 공유·활용체제 구축에 대한 정부의 강력한 의지 - ICT 인프라 기술선도 경험 보유		- 연구데이터 공유·활용을 위한 연계 및 추진체제 미비 - 연구데이터의 공유·활용에 대한 연구현장 인식 및 전문역량 부족	
기술선도 경험과 정부의 강력한 추진의지	S (강점, Strength)	W (약점, Weakness)	추진체제 및 인식·역량 부족
	내부 역량		

데이터 기반 R&D 혁신	외부 환경		주요국과의 경쟁심화
	O (기회, Opportunity)	T (위협, Threat)	
- 데이터 사이언스의 발전과 더불어 데이터 기반 연구개발 문화 및 기술혁신 노력 확산 - 연구데이터의 개방과 공유를 위한 국가차원의 인프라 및 지원체제 구축 필요성 증가	- 법·제도 및 정책 미비로 인한 주요국과의 격차 심화 우려 - 글로벌 경쟁자들의 플랫폼 장악에 따른 기술 종속화		

□ 기회(Opportunities)

- 데이터 사이언스의 비약적 발전으로 인한 연구개발 방식의 변화와 급속한 기술혁신 가속화
- 연구개발 활동에 있어서의 데이터의 활용과 다양한 연구자간의 융합·공동연구 추진을 위한 연구데이터 공유체제의 필요성이 점증하는 추세

□ 위협(Threats)

- 연구데이터 관련 법·제도 및 정책 미비로 인한 주요국과의 격차 심화 가능성
- 글로벌 경쟁자들의 관련 기술 및 플랫폼 장악에 따른 기술 종속화 우려

□ 강점(Strengths)

- 연구데이터 관리·활용을 위한 정부의 강력한 추진 의지로 연구데이터의 공유 및 활용의 필요성이 널리 인식되면서 정부 차원에서 체제 구축을 위한 기반을 마련
- 한국은 전 세계 ICT 인프라 기술을 선도했던 경험을 보유하고 있음

□ 약점(Weaknesses)

- 연구데이터 공유·활용을 촉진하는 법·제도의 미비
- 연구데이터의 관리·공유에 대한 연구현장 인식 및 전문역량 부족
 - － 관리·공유에 대한 연구현장 인식이 부족(79%가 개인자산으로 인식)
 - － 데이터 집약형 연구 확산에 따라 특정 분야 차원에서 일부 데이터를 수집관리 중이나, 활용은 저조

2. 대응전략

	강점(S)	약점(W)
기 회 요 인 (O)	SO 전략: 강점과 기회요인의 시너지 전략 [SO ①] 국가 연구데이터 플랫폼·인프라 구축·운영 [SO ②] 대·소분야 전문센터 플랫폼·인프라 지원	WO 전략: 약점 극복 전략 [WO ①] 연구데이터 공유·활용 위한 추진체계 구축 [WO ②] 연구데이터 공유·활용의 인식제고 및 활성화를 위한 시범사업 추진
위 협 요 인 (T)	ST 전략: 위협 극복 전략 [ST ①] 선진국과의 격차를 줄이기 위한 기술개발 [ST ②] 연구데이터 활용 기반 조성을 위한 법·제도 마련	WT 전략: 위협 회피 전략 [WT ①] 연구데이터 활용을 위한 전문인력 양성 [WT ②] 연구데이터 경쟁력 강화를 위한 대한 중장기·집약적 투자

<그림 3-1> SWOT 매트릭스

- [SO 전략] 연구데이터 공유·활용체계 전반의 연구데이터를 원스톱으로 검색하고 활용할 수 있는 국가연구데이터 플랫폼 및 대·소분야 전문센터 구축·운영
 - 각종 연구데이터에 대한 색인·검색 기능과 연구자가 데이터 융합 및 분석에 활용할 수 있는 분석도구 및 가상작업환경을 제공하는 국가연구데이터 플랫폼 구축
 - 연구데이터 공유·활용 우선순위 도출을 통해 중점분야를 선정하고 분야별 대·소분야 전문센터 구축
- [WO 전략] 추진체계 구축 및 시범사업 추진
 - 다양한 분야에서 개별적으로 생성되는 연구데이터를 체계적으로 취합·관리·공유할 수 있도록 계층적 공유·활용 체계를 구축·운영할 추진체계 구축
 - 시범사업 실시를 통해 연구데이터 공유·활용의 인식제고 및 활성화
- [ST 전략] 기술개발 및 법·제도 정비
 - 연구데이터 공유·활용 인프라 및 관련 기술을 선점한 선진국과의 격차 해소를 위한 기술개발
 - 연구데이터 활용 기반 조성을 위한 법·제도 정비 및 연구개발혁신을 위한 규제 완화
- [WT 전략] 전문인력 양성 및 투자 강화
 - 데이터 기반 집단연구를 통해 각 분야의 문제해결이 이루어지는 각 분야 전문센터에서 R&D와 인력양성 동시 추진으로 고급인재 육성
 - 연구데이터 경쟁력 강화를 위한 대한 중장기·집약적 투자

IV. 사업의 추진전략

1. 사업의 비전과 목표

비 전
연구데이터의 지식자산화·빅데이터화로 << 국가 R&D성과혁신 달성 >>

목 표
- 연구데이터의 공유·활용 활성화로 데이터기반 R&D 정착 -
국가 연구데이터의 공유·활용 생태계 조성 전문분야별 연구데이터의 집약과 공동연구 기반확보 빅데이터, AI기술 등 데이터를 활용하는 R&D 확산

전 략
① 연구데이터 공유·활용 원칙 및 제도 등을 반영한 플랫폼 설계 ② 융합·공동연구 역량 강화를 위해 분야별 연구공동체 형성 ③ 연구데이터의 검색·분석 Tool 등의 환경제공으로 연구자 부담경감 ④ 데이터 집약형 연구 분야부터 우선 추진, 단계적 확장 ⑤ 플랫폼의 운영과 빅데이터 활용에 필요한 기술력의 체계적 확보

	구분	세부과제
추진과제	국가 연구데이터 플랫폼 구축 및 운영	○ 연구데이터 수집·연계 및 공유 기술 ○ 연구데이터 저장·보안 기술 ○ 연구데이터 품질·표준화 기술 ○ 지능형 연구데이터 고속 검색 기술 ○ 연구데이터 융합·분석 환경 구축 기술 ○ 연구데이터 활용·플랫폼기술
	대·소 분야 전문센터 구축 및 운영지원	○ 바이오 분야 전문센터 구축 및 운영지원 ○ 소재 분야 전문센터 구축 및 운영지원 ○ 인공지능 분야 전문센터 구축 및 운영지원 ○ 기타 9개분야 전문센터 구축 및 운영지원
	데이터 및 컴퓨팅 활용 기반의 R&D 인재육성	○ 연구데이터 관리를 위한 ICT기반 기술교육 ○ 연구데이터의 R&D 응용을 위한 분야별 전문활용 교육 ○ 연구데이터활용 인재육성을 위한 교육용DataBank 및 서비스 구축

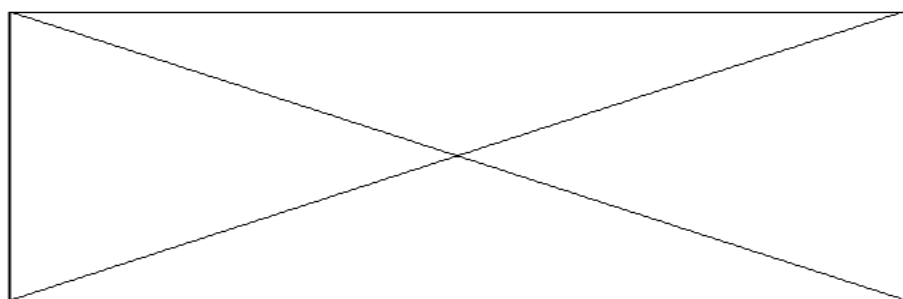
2. 사업전략

□ 연구데이터 공유·활용 원칙 및 제도 등을 반영한 플랫폼 설계

- 「국가연구개발사업 공동관리 규정」 개정을 통해 DMP를 통한 법적 정의 및 관리제도 마련
 - * DMP(Data Management Plan): 데이터의 양,종류,품질,포맷,표준 및 메타데이터,수집 방법 등의 관리 계획
- 단계별 법제 정비 및 개선의 내용이 관리 되도록 DMP정보 공유 등의 플랫폼 설계
 - * DMP작성 템플릿과 작성 가이드 등을 제공하여 연구자가 DMP를 용이하게 작성하도록 서비스
- 연구데이터를 공유·활용할 수 있도록 국가연구데이터 관리를 플랫폼화 하고 각종 데이터 분석 SW 개발과 운영을 통해 공동연구를 할 수 있는 협업환경을 구축
 - 빅데이터 시뮬레이션, 인공지능을 접목한 분석 등 데이터 기반 R&D 협업공동체 형성 촉진 및 확산
 - * 연구 분야별 전문센터를 구축하여 연구역량을 제고하는 프로그램제공, 데이터관리 기능 등의 제공

□ 연구데이터의 검색·분석 Tool 등의 환경제공으로 연구자 부담경감

- 분산파일시스템, NoSQL, 비디스크 기반 DBMS에대한 원스탑 검색 및 활용통합 검색도구 제공
- 데이터의 특징, 연관관계, 인과관계 등을 도출, 종합적 분석 처리할 수 있는 분석도구·방법론·S/W제공



- 데이터 입출력, 시각적 변환, 상호작용적인 렌더링 등의 시각화 도구를 공개 소프트웨어 형태로 연구자에게 지원

□ 데이터 집약형 연구 분야부터 우선 추진, 단계적 확장

- 전문센터 구축을 3단계로 추진하되, 1단계는 바이오·미래소재·인공지능 분야를 우선구축
 - 바이오: 중장기적 집단연구 및 기존 연구커뮤니티와의 연계를 추진, 연구데이터 관리 고도화 및 성공적 활용모델 창출

- 미래소재: 다양하고 소량으로 구성된 데이터 특징을 고려한 응용분야별 연구데이터 수집 및 활용체계 구축
- 인공지능: 국가 R&D 과정에서 연구수행기관(출연연·대학·기업 등)이 축적한 기계학습 데이터를 취합·공개·활용

□ 플랫폼의 운영과 빅데이터 활용에 필요한 기술력의 체계적 확보

- 대학생·대학원생 및 현직 연구자들을 대상으로 전문분야에서의 데이터·컴퓨팅 활용을 위한 교육제공
 - 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
- 데이터 기반 집단연구 시 문제해결 과정과 인력양성을 동시 추진하여 고급인재를 육성

3. 사업구성

□ 3대 추진과제의 구성

- 사업전략을 통해 플랫폼구축, 전문센터구축, 인재육성의 3대 추진과제와 각 추진과제별 이행방안 수립과정을 통해 20대 세부과제를 도출
- **(국가 연구데이터 플랫폼 구축 및 운영)** 연구데이터 공유·활용체계 전반의 연구데이터를 원스탑으로 검색하고 활용할 수 있는 통합 플랫폼 구축 및 운영
 - * 6대 세부과제: 연구데이터 수집·연계 및 공유기술, 연구데이터 저장·보안기술, 연구데이터 품질·표준화기술, 지능형 연구데이터 고속검색 기술, 연구데이터 융합·분석 환경 구축기술, 연구데이터 활용·플랫폼 기술
- **(대·소분야 전문센터 구축 및 운영지원)** 클라우드 형태로 분야별 컴퓨팅·데이터 인프라를 서비스하고, 개별 연구자에도 데이터관리 클라우드 제공
 - * 11대 세부과제: 생명·보건의료, 소재·나노, ICT·S/W, 기계·제조, 에너지·자원, 농림수산·식품, 우주·항공·해양, 건설·교통, 환경·기상, 원자력·핵융합, 재난·안전
- **(데이터 및 컴퓨팅 활용기반의 R&D 인재육성)** 대학·연구소 등 연구자의 데이터 분석·활용능력 제고를 위해 빅데이터 등의 기반기술과 도메인별 R&D응용을 위한 활용과정 운영
 - * 3대 세부과제: 연구데이터 관리를 위한 ICT기반 기술교육, 연구데이터의 R&D 응용을 위한 분야별 전문 활용교육, 연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스구축

□ 3대 추진과제의 목표 및 기대효과

3대 추진과제	목표	기대 효과
국가 연구데이터 플랫폼 구축 및 운영	연구데이터를 효과적으로 공유·활용할 수 있는 환경과 기술 지원을 통해 국가 과학기술의 글로벌 경쟁력 제고	- 연구에 필요한 시간과 비용을 절감하여 국가 R&D 효율성을 제고 - 다양한 주제 분야들을 결합한 융합 연구 활성화 - 연구의 효율성 및 투명성과 연구결과물의 수준 제고
대·소분야 전문센터 구축 및 운영지원	데이터기반 연구역량 제고를 위한 분야별 연구데이터 공유·활용 전문데이터센터 구축	- 분야별 정보의 효율적인 활용을 통한 각 분야별 산업발전에 기여 - 빅데이터 활용으로 인해 데이터 엔지니어링, 빅데이터 분석기술제고 - 분야별 연구데이터 현황분석 및 관리를 위한 다양한 활용 인프라 구축
데이터 및 컴퓨팅 활용 기반의 R&D 인재육성	연구데이터 활용 R&D 인재의 지속적 육성을 위한 실무중심 교육체계 구축	- 데이터 전문과학자 양성으로 국가 데이터 산업육성과 발전에 기여 - 다양한 데이터 분석기술 습득 및 연구에 활용 - 실무 적용 및 활용 중심의 빅데이터 교육을 실시함으로써 현장 전문 인력 양성 및 배출

4. 성과목표(지표)

□ 성과지표 선정근거

- (과학적 성과)국가연구개발 사업임에 따라 성과물의 과학적 수준을 나타내는 논문의 게재 학술지 수준, 영향력 지수 등의 성과측정
- (기술적 성과)연구데이터를 집약하는 사업의 특성상 빅데이터화, 클라우드기반 서비스, 고속의 컴퓨팅 환경제공 등 기술적 혁신이 필요하고 이를 지원하는 특화된 S/W 개발 등의 성과측정
- (경제적 성과)플랫폼, 전문센터의 분석인프라 활용, 중소기업의 인재양성 교육과정지원, 연구데이터를 활용한 연구개발 지원 등의 성과측정
- (사회적 성과)인재육성 교육을 통한 인력양성 및 일자리창출 등의 성과측정

□ 사업 성과목표

구분		1단계			2단계				3단계			계
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	
과학적 성과	논문 (SCI)	14	26	24	27	22	19	20	20	20	20	212건
	논문 (KCI)	34	57	54	60	47	39	43	44	43	43	464건
기술적 성과	특허	출원 (19)	출원 (26)	출원 (25)	출원 (28)	출원 (22)	출원 (19)	출원 (21)	출원 (21)	출원 (21)	출원 (21)	출원 (223)

구분		1단계			2단계				3단계			계
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	
		등록 (7)	등록 (11)	등록 (10)	등록 (11)	등록 (9)	등록 (7)	등록 (8)	등록 (8)	등록 (8)	등록 (8)	등록 (87)
	기술혁신	표준 모델(6)	표준 모델(6)	표준 모델(8)	표준 모델(7)	표준 모델(7)	표준 모델(8)	표준 모델(4)	표준 모델(4)	표준 모델(4)	표준 모델(4)	58개
	콘텐츠 S/W		4	6		1		2				13건
경제적 성과	기술활용 효과					5조					5조	10조
	중소기업 지원					기업 양성(1)					기업 양성(1)	2개
	연구개발 서비스	분야 과제 (1)	분야 과제 (1)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	18개
사회적 성과	인력양성	2,605	2,747	2,843	2,943	3,049	3,208	3,323	3,443	3,569	3,700	31,430명
	일자리 창출	421	439	459	479	500	522	545	569	594	620	5,148 명
	정책효과	10%	10%	20%	20%	30%	30%	40%	40%	50%	50%	10건
	홍보확산	1	1	1	1	1	1	1	1	1	1	10건
인프라 성과	국제협력			1				1			1	3건
	공동활용 시설장비	데이터 set (250)	데이터 set (250)	데이터 set (500)	데이터 set (500)	데이터 set (500)	데이터 set (1000)	데이터 set (500)	데이터 set (500)	데이터 set (1000)	데이터 set (1000)	6,000개
		교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (2)	교육용 DB구축 (1)	교육용 DB구축 (1)	11개
			서버장 비(6)	서버장 비(9)				서버장 비(3)				18개
	전산 시스템	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	30Core 1PB
		시스템 (1)	세부별 (2)	세부별 (3)				세부별 (1)				7개

5. 결과물 공유 및 부처간 협력

□ 결과물 공유

－(정책적 측면) 데이터 관리계획(DMP)에 따라 데이터에 대한 공개범위와 시기가 의무적으로 제공되어 연구데이터 공유·활용의 문화 정착과 연구데이터를 활용

* 연구데이터 검색을 통한 활용, 연구데이터 활용 연구결과에 대한 객관성 확보, 연구데이터 인재육성 프로세스 참여 등

－(운영 측면) 계층적으로 운영되는 연구데이터 관리 구조에서 플랫폼, 표준기술, 상용기술 등 계층구조에 맞는 연구데이터를 관리 기술과 운영기술 활용

* 플랫폼, 표준기술 등의 가이드라인 활용, PaaS, 분석기술 등 연구자측면의 다양한 서비스 활용

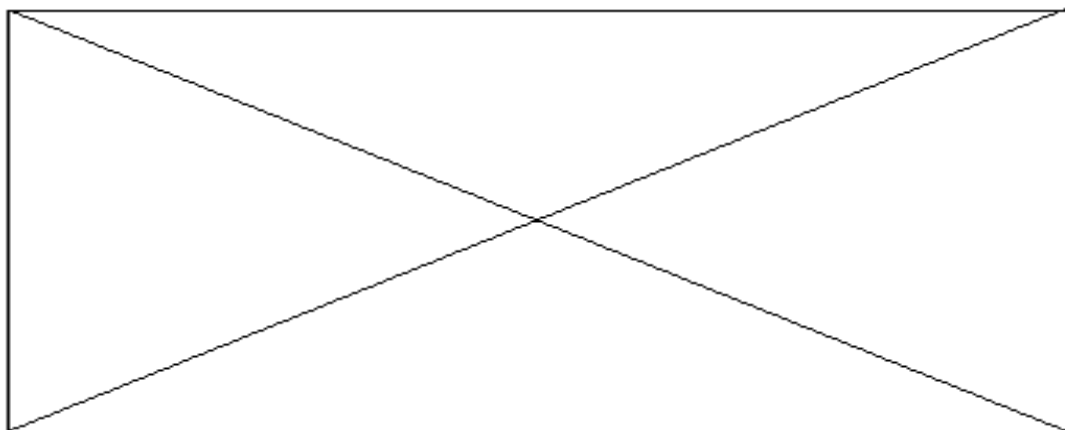
－(기술 측면) 빅데이터 기술, 인공지능기술, 머신러닝 기술 등 다양한 4차 산업혁명의 기반이 되는 핵심기술 활용하고 플랫폼 및 클라우드 기술을 활용

* 빅데이터 처리기술, 플랫폼 기술, 클라우드 기술과 연구데이터 분석 SW 등의 R&D 활용

－(데이터 측면) 관찰, 센싱 등 대용량 연구데이터, 비정형, 반정형의 다양한 연구데이터, 연구결과를 검증하기 위한 가공 연구데이터 등을 DMP 범위 내에서 활용

□ 부처간 협력

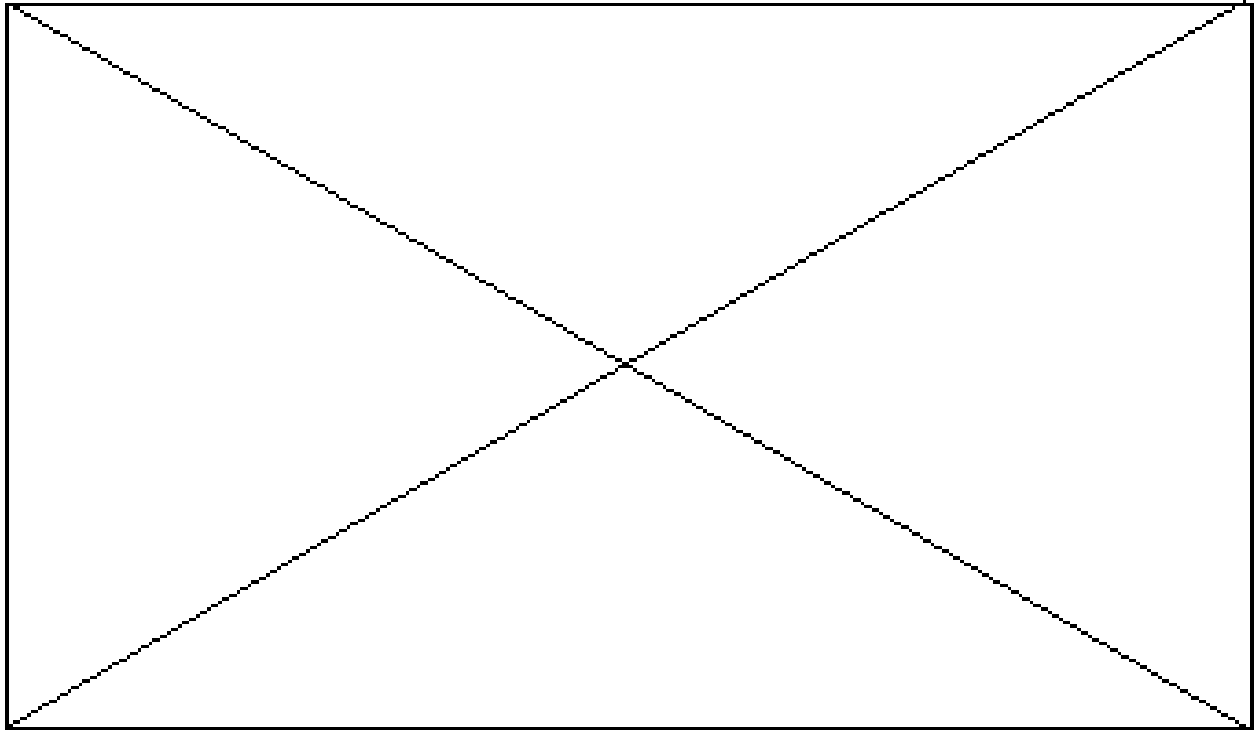
- 과기 정통부는 산·학·연·관이 필요로 하는 4차 산업혁명 관련 원천기술을 개발·공유
 - * 빅데이터, 인공지능, 클라우드 등에 대한 기술을 농업·물류·제조 등 타 분야로 공급, 새로운 비즈니스 모델 및 혁신기업 육성을 위한 토대 마련
 - * 혁신기술 개발에 관련 S/W, 기초기술 등이 규제로 인해 장애가 되지 않도록 관계부처와 협력



- 정부부처의 적극적인 참여 및 지원이 유지되도록 범부처 협력체계를 운영
 - 범부처 협의체는 사업의 종료시점까지 운영하고 정기적으로 범부처 협의체 회의를 개최
 - 연간 사업계획에 따른 사업 추진경과 점검 및 성과 공유·확산 방안, 사업추진에 중요하다고 판단하는 사항에 대한 범부처 의사결정
- 정부의 협의기구를 통해 국가 R&D 사업의 혁신적인 변화를 위한 주요 의사결정 창구로 활용
 - * 4차산업혁명 위원회, 정보통신전략위원회, 국가과학기술심의회(운영위원회, 특별위원회), 생명공학 종합정책심의회 등

V. 국가연구데이터 플랫폼 구축 및 운영

1. 목표 및 추진전략



□ 전략1. 국가연구데이터플랫폼의 기능적 역할 정의 및 요소기술 선정

- 수집·연계, 분석 및 활용지원 등을 위한 기능구성,기능요건,상세기능정의, 필요 요소기술과 세부기술 선정

* 수집·연계 프레임워크, 서비스 프레임워크, 보안 프레임워크, 시스템 관리 프레임워크 4개 영역구성

□ 전략2. 데이터의 상호운용성 확보를 위한 표준화 및 품질 관리

- 상호운용성 확보를 위한 분류체계 및 코딩표준, 데이터베이스 구축 및 공유를 위한 표준, 서비스 표준 등 관리

* 메타데이터 등록과 인증을 통해 표준화된 메타데이터 관리 및 공유할 수 있는 레지스트리 구축

□ 전략3. 연구데이터 수집·연계를 위한 국내·외 연구기관 및 협력 네트워크와 협업

- 국내·외 R&D 사업으로부터 발생하는 연구데이터의 수집·연계를 위한 국내·외 관련기관 협업추진
 - * 분야별 전문센터, 개별 연구데이터센터와 협업 및 연계 API 보급
 - * 연계대상 : 전문분야 데이터센터, 연구과제관리시스템, 해외연구데이터 플랫폼, 분석환경시스템, 개별연구관리 시스템

□ 전략4. 연구자가 데이터 융합 및 분석에 활용할 수 있는 다양한 환경 지원

- * 데이터 분석지원을 위한 분석도구 및 가상작업 환경 구축, 슈퍼컴퓨터(KISTI)와 연계

□ 전략5. 안전하고 투명한 데이터 유통체계 확보

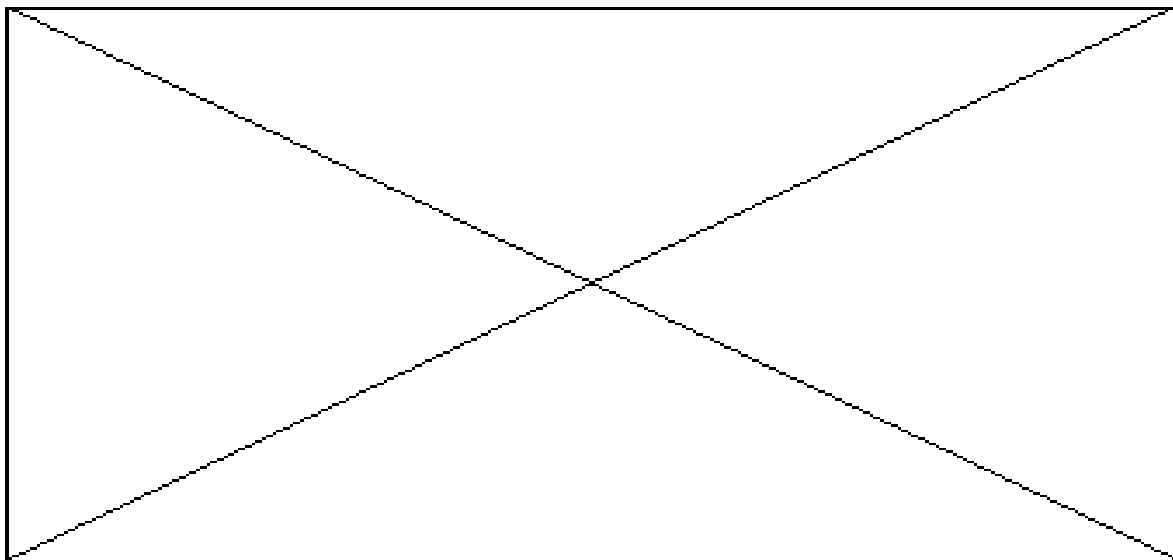
- 데이터 보안기술 적용, 생산자의 권리보호를 위한 엠바고, 인용표기, 이용조건 및 라이선스 관리

□ 전략6. 범 국가 차원의 연구데이터 공유·활용 체계 활성화 지원

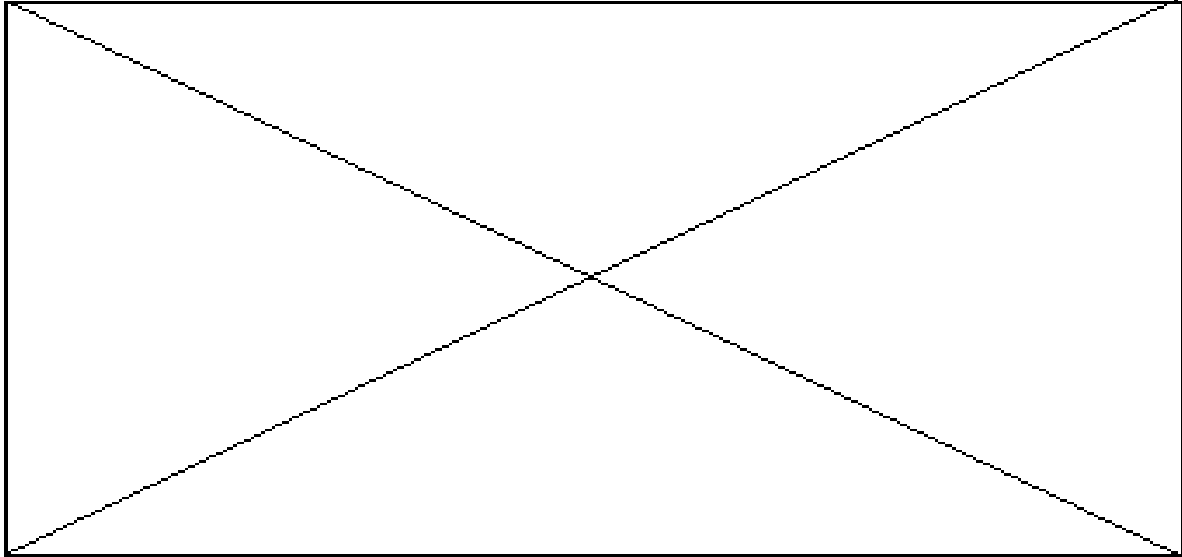
- 대·소분야 전문센터, 연구기관이나 연구자에게 클라우드서비스(IaaS, PasS) 제공

2. H/W, S/W 구성방안

□ 국가연구데이터플랫폼 H/W 구성 방안



□ 국가연구데이터플랫폼 S/W 구성 방안



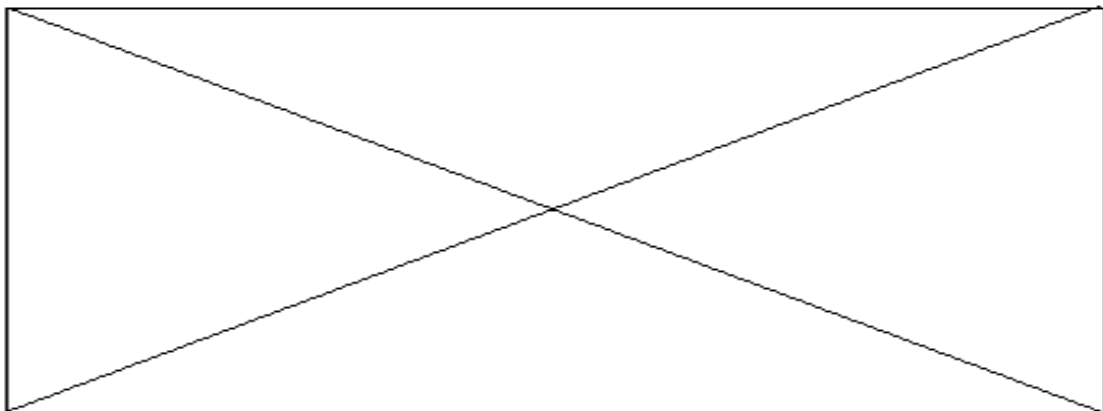
3. 세부과제 내용

① 연구데이터 수집·연계 및 공유 기술 개발

□ 핵심기술 구성

- 데이터 수집·연계 관리, 연구과제 정보 및 DMP 연계, 분석환경 연계 기능개발 및 연계 모니터링 개발

☞ 수집·연계 기술로 활용되고 있는 REST, XML-RPC, SOAP, FTP, OAI-PMH 등의 기술 중 적합한 기술을 이용하여 연구 데이터/메타데이터를 효과적으로 수집 전송을 위한 표준 기술 개발



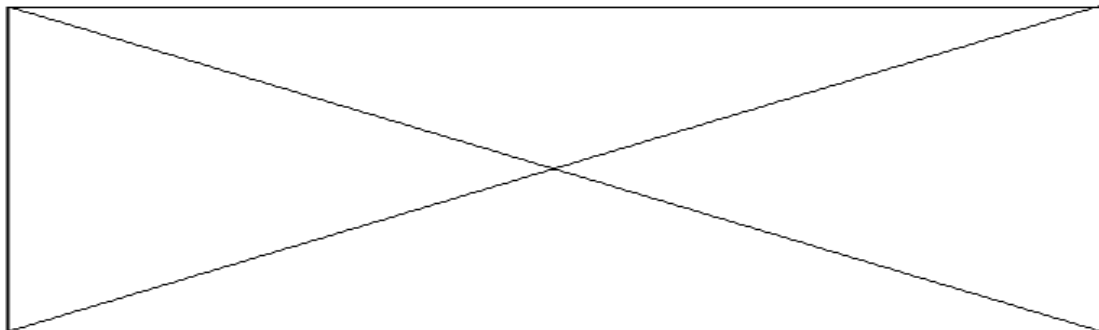
□ 기술확보 전략

- 연구데이터/메타데이터 연계 : REST API, OAI-PMH 표준을 준수하여 자체 개발
- 연구데이터 연계기술 개발 : FTP, Cloud환경의 데이터 공유 기술을 활용하여 자체 개발

② 연구 데이터 저장·보안 기술 개발

□ 핵심기술 구성

- 대용량 데이터 저장 클라우드 기반 분산, 병렬 저장기술과 데이터 연계 및 저장 시 보안 기술개발
 - ☞ 빅데이터 저장 및 분석에 활용되고 있는 분산/병렬 파일시스템과 NoSQL 중심으로 구현 기술을 도출, 보안 프레임워크는 위변조 방지, 암호·복호화, 전송 및 접근제어 등 데이터 보안 및 개인정보 관리 기술개발



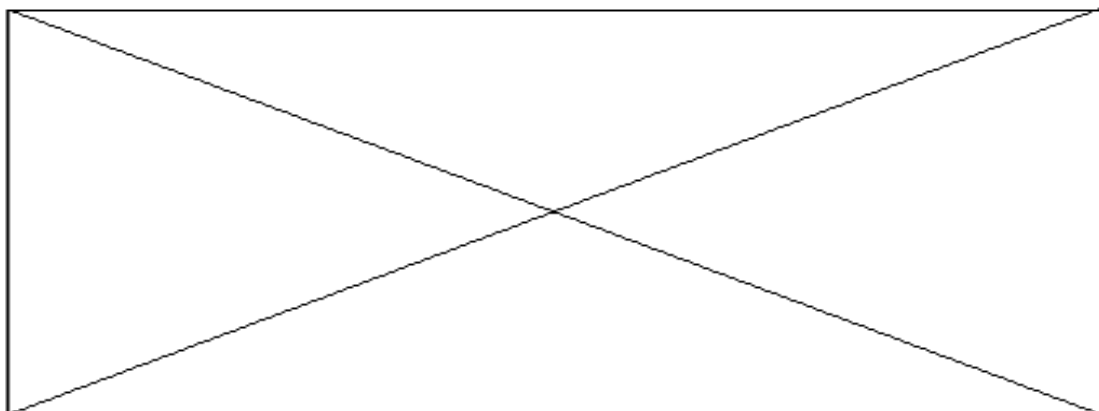
□ 기술 확보전략

- 분산, 병렬 저장 기술 : Hadoop 기술을 활용하여 자체 개발
- 보안기술 : SSO, 암호화 솔루션을 도입, 통합계정관리는 자체 개발

③ 연구 데이터 품질·표준화 기술 개발

□ 핵심기술 구성

- 연구데이터의 품질 관리를 위한 표준 정의, 관리기술 및 연구데이터 출처 및 저작권 관리
 - ☞ 연구과제정보와 DMP정보를 기반으로 메타데이터 관리를 위한 스키마 및 항목에 대한 표준 수립과 메타데이터 수집 시 메타데이터 표준(안)이 반영된 룰 기반의 자동 품질검증 기술개발



□ 기술 확보전략

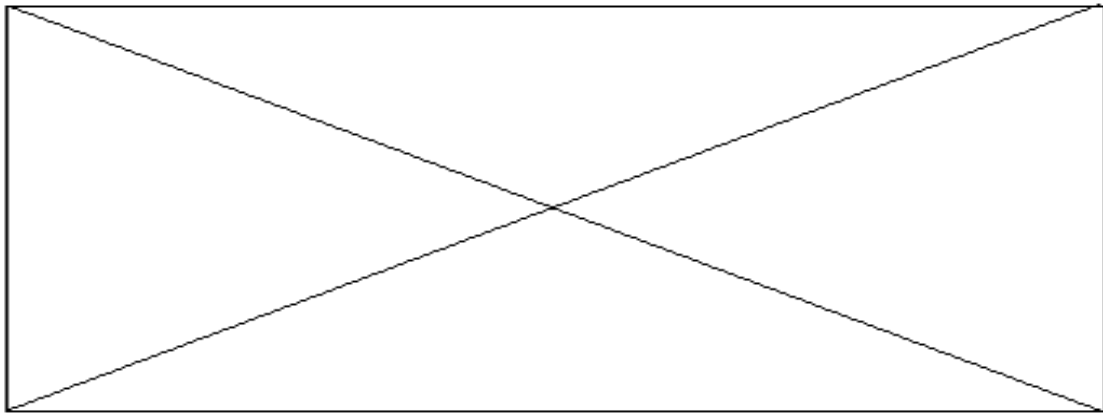
- 표준수립은 연구과제정보, DMP표준을 기반으로 표준 정의, 품질관리는 품질관리 솔루션을 도입, 저작권관리는 오픈소스 기반 자체 개발

④ 지능형 연구 데이터 검색 기술 개발

□ 핵심기술 구성

- 메타데이터 기반의 검색 및 클라우드 서비스 항목(S/W, 분석도구)에 대한 검색 기술 개발

- ☞ 사용자의 등록 연구데이터들에 편리하고 쉽게 조회할 수 있도록 하기 위해 등록된 메타데이터 기반으로 효율적인 검색 방안 구축
- ☞ 연구데이터 플랫폼은 등록된 연구데이터 검색을 위한 검색 엔진 도입



□ 기술 확보전략

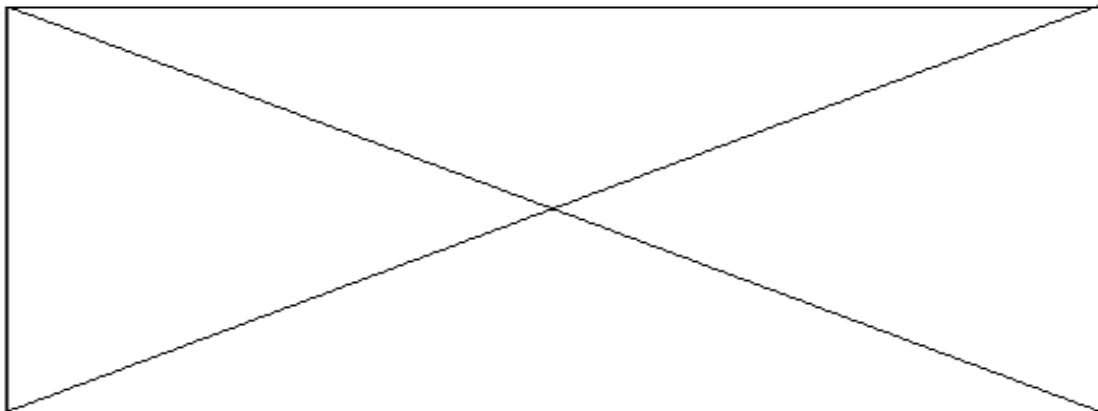
- 오픈소스 분산 검색엔진 기반의 자체 개발

⑤ 연구데이터 융합·분석 환경 구축 기술

□ 핵심기술 구성

- BI, AI, ML, DL솔루션을 연구플랫폼에 탑재하고 분석 솔루션을 workflow기반으로 활용할 수 있도록 구축

- ☞ 클라우드 환경에서 연구데이터를 실시간분석, 데이터 분석, 지능형 분석 및 시각화 기술을 기반으로 분석 프레임워크 구성



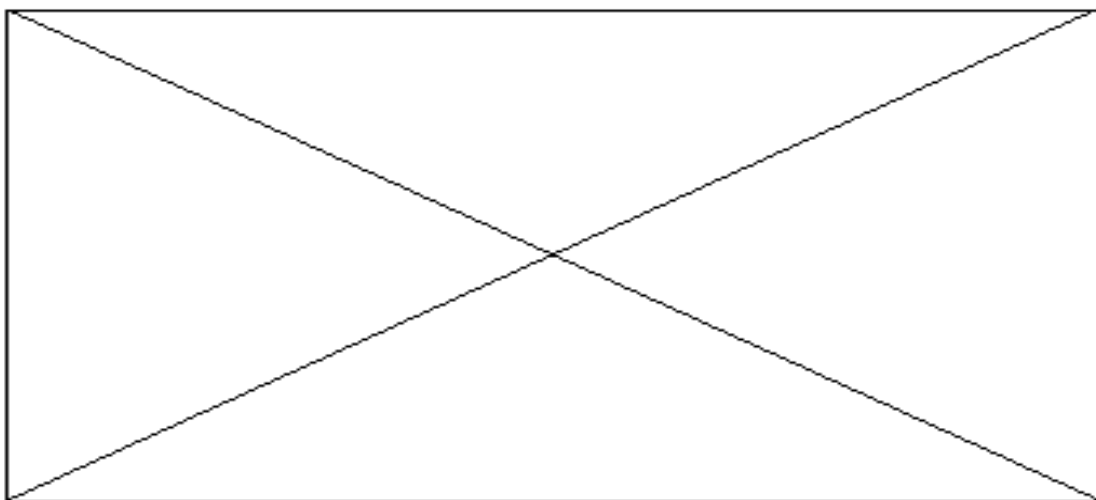
□ 기술 확보전략

- 오픈소스 분석 기술 기반의 자체 개발

⑥ 연구데이터 활용·플랫폼 기술

□ 핵심기술 구성

- 연구자간 협업 및 커뮤니티 기능 및 클라우드 서비스 운영을 위한 기반기술 개발
 - ☞ 물리 자원 및 가상 자원 관리를 위한 기능 기술과 실행 및 운용을 위한 가상 자원관리 기능기술 개발
 - ☞ 관리환경은 플랫폼에서 제공하는 기능을 각각의 제공 서비스에 배포 및 관리하기 위한 모듈로 구성

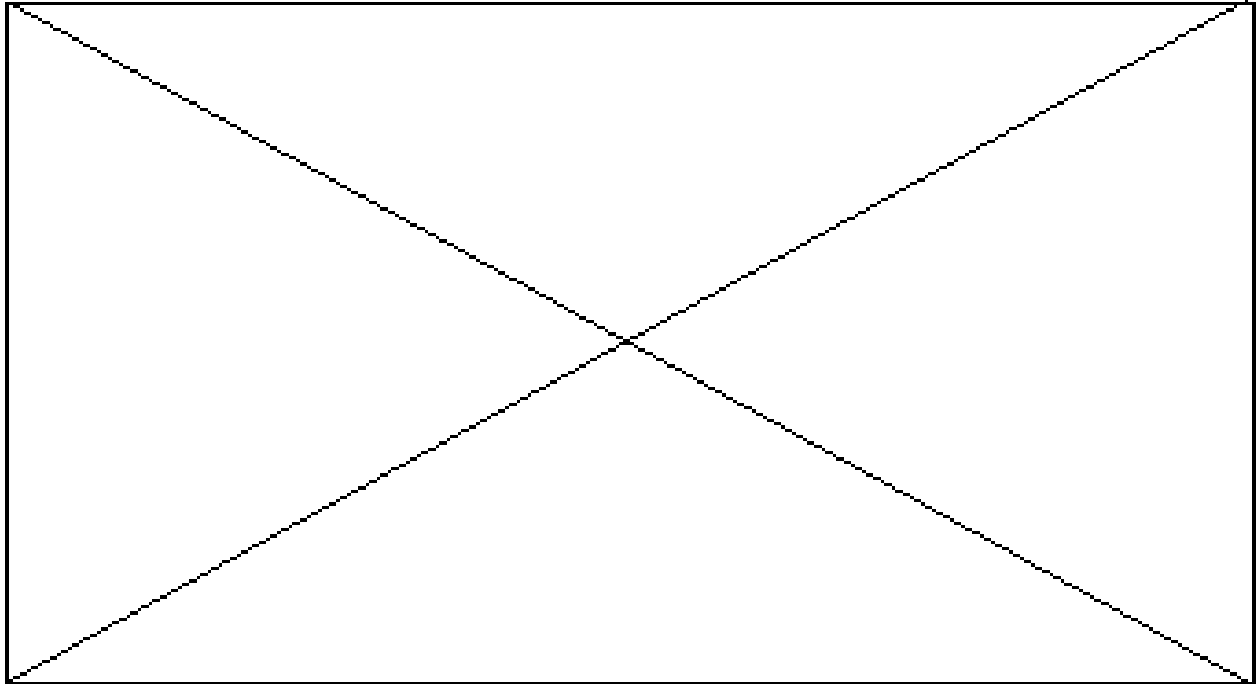


□ 기술 확보전략

- 협업서비스, 커뮤니티 생성 기술은 자체 기술개발을 통해 구현, 클라우드 서비스 관리 기술은 오픈소스 솔루션을 기반으로 기술 개발

VI. 대·소분야 전문센터 구축 및 운영지원

1. 목표 및 추진전략



□ 전략1. 현황분석 및 설문을 통한 전문센터 우선분야 선정

- 정부의 연구개발 중점분야*, 국가의 중점과학기술 분야* 등을 통해 연구개발 분야도출
 - * 2019년도 정부연구개발투자 방향 및 기준: ICT·S/W, 생명·보건의료, 기초연구 등 11개 분야
 - * 제4차 과학기술기본계획(2018~2022): ICT·S/W, 생명·보건의료, 재난안전 등 11개 분야
 - * 국가융합기술 시행계획(2017): 빅데이터,차세대반도체,융합형 콘텐츠,스마트 자동차 등 15개 분야
- 평가지표, 평가기준 등의 개발 및 정량적 평가 등을 통한 우선분야 도출 및 순위선정

순위	1	2	3	4	5	6	7	8	9	10	11
분야명	ICT·SW	생명·보건의료	소재·나노	농림수산·식품	환경·기상	기계·제조	우주·항공·해양	원자력·핵융합	에너지·자원	건설·교통	재난안전

□ 전략2. 전문센터의 체계적 선정평가를 통해 데이터 집약형 분야부터 단계적 구축

- 우선분야 선정 결과에 따라 단계별로 평가절차에 따라 대·소분야 전문센터를 선정하여 구축

－ 평가단계를 선정평가 → 중간평가 → 최종평가로 구분 및 정의하여 순차적으로 평가

□ 전략3. 데이터 기반 연구공동체 지원환경 구축

- 분석도구, 연구 및 협업 지원, 연구자료 공유 등 분야별로 가장 필요한 기본적인 연구지원 서비스 제공
- 검색된 데이터를 분석하고, 협업연구 할 수 있는 연구자 지원 서비스

2. 세부과제 내용

① 바이오분야 전문센터 구축 및 운영지원

□ 추진 목표

바이오 연구데이터 확보 · 관리 · 활용 체계 구축		
목표1: 바이오 연구데이터 연계체계 구축 국가연구데이터 플랫폼과 바이오 연구데이터 간 메타정보 연계 ['22년 연계체계 구축 완료]	목표2: 바이오 연구데이터 수집 과기정통부 생명연구자원 정보 확보 강화 [생명연구자원 '18년 2,668,777건 → '28년 4,350,000건] 생명정보 연구성과물 확보 강화 ['18년 52 TB → '28년 80TB]	목표3: 바이오 연구데이터 활용시스템 구축 생명연구자원 정보기반 멀티오믹스 분석시스템 구축 ['23년 오픈 계획] 정보연계 및 분석 지원을 위한 인프라 확보 CPU: '18년 700 core → '28년 1,000 core 저장공간: '18년 2 PB → '28년 3 PB

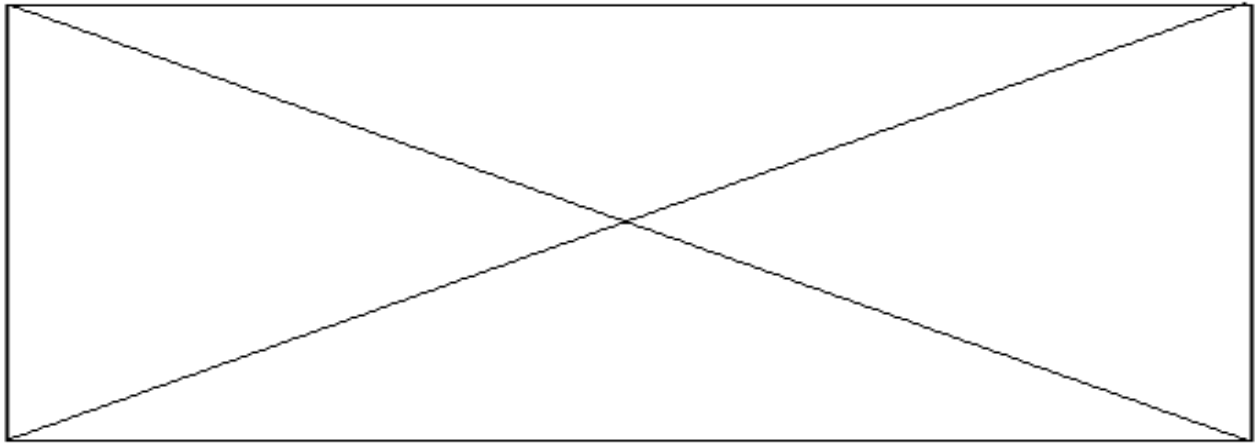
□ 추진 내용

- 바이오 연구데이터 연계체계 구축
 - － 바이오 연구데이터 전문센터와 국가연구데이터 플랫폼 간 메타정보 연계를 위한 기반 구축 및 연계
- 바이오 연구데이터 수집
 - － (과기정통부 생명연구자원 정보 확보) 바이오 분야의 연구에서 산출된 바이오 연구데이터를 분야별 연구내용, 연구성과 발표 타입별, 생산된 연구데이터 타입별로 나누어 등록·관리하고, 제공하는 정보의 다변화
 - － (생명정보 연구성과물 확보) 연구성과물 등록시스템 기반의 연계 체계를 활용하여, 바이오 분야의 연구 방법론, 결과물 등을 수집하고 정제 강화
- 바이오 연구데이터 활용시스템 구축 및 지원을 위한 인프라 구축

- 생명연구자원 정보기반 멀티 오믹스* 분석시스템 구축(개발 중)
 - * 유전체학, 전사체학, 단백질체학 등 생명정보학 연구에 활용되는 집합체 전부
- 정보연계 및 분석 지원을 위한 인프라 확보

② 소재분야 전문센터 구축 및 운영지원

□ 추진 목표



□ 추진 내용

- 소재분야 연구데이터 대·소전문센터 구축
 - 빅데이터 구축에 필요한 표준화, 소재 온톨로지 개발 및 머신러닝 활용을 통한 신물질 개발을 체계적으로 수행하기 위한 응용분야별 소분야 전문센터 구축
- 소재·정보·데이터 융복합 혁신기업 양성
 - 데이터 구동형 소재설계(데이터, 인공지능), 소재 데이터 분석·해석 기반 엔지니어링 및 데이터 패키지를 제공하는 혁신기업 양성*
 - * 계산과학·머신러닝 서비스 제공 혁신기업 양성 (1개)
 - * 실험데이터 기반 데이터베이스 서비스 제공 혁신기업 양성 (1개)
- 소재 데이터 활용 연구개발에 의한 경제적 효과 창출
 - 미래소재 연구개발 비용 및 개발기간 단축에 의한 경제적 효과 약 10조 창출*
 - * 미국 MGI 경제적 파급효과(RTI社, NIST 용역) : 138~300 조원/년 경제적 효과
 - * 한국 경제규모 고려(GDP 미국의 1/20) → 7~15조원/년 예상
- 정량적 연구개발 성과 목표
 - (인력양성) 현재 국내 데이터과학 활용 가능 소재전문가는 수십여명 이내로 소재 연구데이터 전문센터 사업 수행 및 교육프로그램 운용으로 소재데이터 융합기술 보유 소재전문가 양성 (700명/10년)

- (학술논문) 응용분야별 소분야 전문센터에서 학술지 논문 90편 이상 게재
- (특허) 대분야 및 소분야 전문센터에서 관련 특허 50편 이상 출원
- (소프트웨어) 대분야 전문센터 및 프레임워크 소프트웨어, 소분야 전문분야 개발 소프트웨어 5개 이상
- (신물질 개발) 응용분야별 소분야 전문센터 세세부과제 수행으로 미래소재 신물질 개발 30개 이상

③ 인공지능분야 전문센터 구축 및 운영지원

□ 추진 목표

“연구공유형 AI데이터 허브 구축·활성화”을 통하여 국가 AI경쟁력 혁신

1단계 (2019~2021)	2단계 (2022~2024)	3단계 (2025~2028)
• 학습데이터셋 : 1,000개 • 허브 수용율 : 20% • 데이터 표준모델 : 8종(기준모델) • 인력양성 : 1,500명	• 학습데이터셋 : 2,000개 • 허브 수용율 : 40% • 데이터 표준모델 : 2종(보완모델) • 인력양성 : 3,000명(누적)	• 학습데이터셋 : 2,000개 • 허브 수용율 : 80% • 데이터 표준모델 : 5종(통합모델) • 인력양성 : 5,000명(누적)

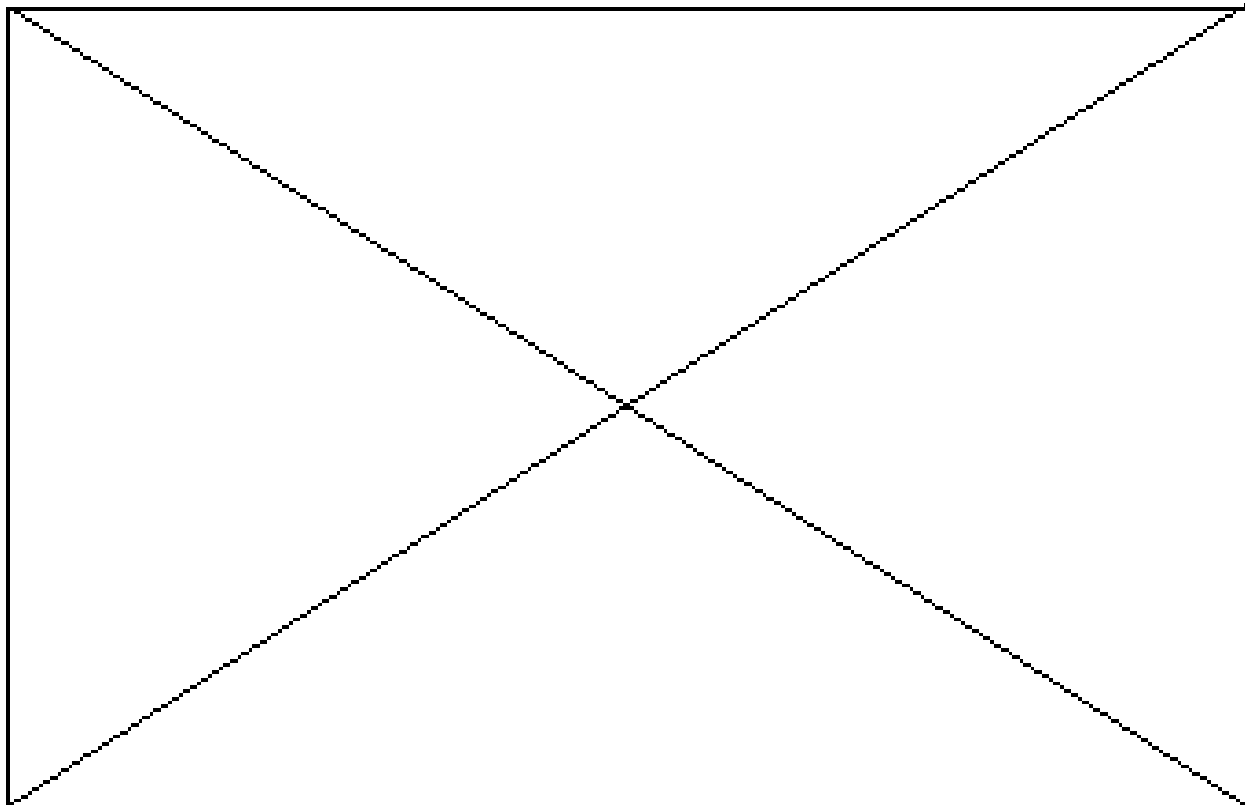
□ 추진 내용

- 인공지능 학습용 데이터셋 6,000개 제공
 - 1단계(1~3차년도): 2억개의 데이터를 포함한 분야별 데이터셋 1,000개
 - 2단계(4~6차년도): 5억개의 데이터를 포함한 분야별 데이터셋 2,000개
 - 3단계(7~10차년도): 10억개의 데이터를 포함한 분야별 데이터셋 3,000개
 - * 정부는 기업들이 필요로 하는 인공지능 학습용 데이터를 2022년까지 1억 6천건을 구축 할 계획
- 인공지능 관련 연구데이터의 클라우드 허브 플랫폼 수용률 80% 달성
 - 1단계(1~3차년도): IT분야 중심으로 20% 달성
 - 2단계(4~6차년도): 전체산업응용 중심으로 40% 달성
 - 3단계(7~10차년도): 일반연구분야 확산을 통해 80% 달성
- 타분야 연구데이터가 참고가능한 데이터 표준모델 10종류 수립
 - 1단계(1~3차년도): 지능분야 별 데이터 기준모델 2종류씩 총 8종류의 표준모델 구축

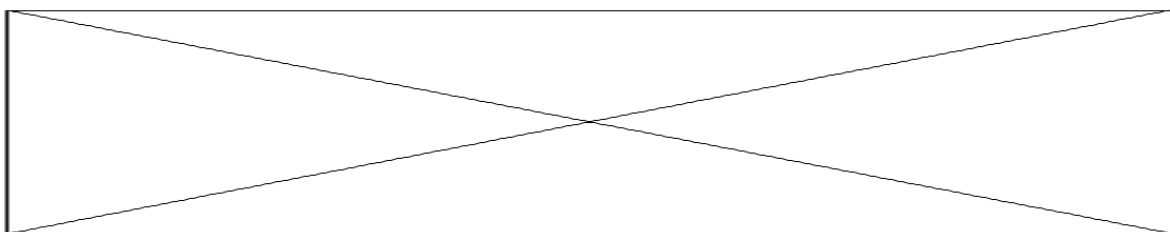
- 2단계(4~6차년도): 1단계 표준모델의 단점 보완하여 두 종류의 모델 추가, 누적 10종류의 표준모델 구축
- 3단계(7~10차년도): 2단계의 표준모델 중 중복되는 모델을 통합하여 5종류로 단순화
- AI기술 이론과 실데이터 경험을 보유한 AI데이터 전문가 5,000명 양성
 - 1단계(1~3차년도): 단기 교육(6개월)과 표준화 경험 중심의 연 500명 전문인력 양성
 - 2단계(4~6차년도): 단기 교육(6개월)과 표준화 경험 인력 및 장기 교육(1년)과 500명의 전문인력 양성
 - 3단계(7~10차년도): 장기 교육(1년)과 연구 경험 중심의 연 500명의 전문인력 양성

VII. 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성

1. 목표 및 추진전략



- 전략1. 연구데이터 활용 및 운영 과정의 통합적 관점 설계 및 데이터와 컴퓨팅 활용 직무역량 정의
- 연구데이터와 (빅)데이터의 활용절차를 통합관점에서 7개의 연구데이터 및 컴퓨팅 R&D Task를 설계



* 시뮬레이션, 실험, 관찰 등은 데이터 생성단계로 생성이후 수집단계부터 연구데이터 활용단계

□ 전략2. 연구 분야별 분석·활용 교육은 전문 교육인프라를 활용

- 연구데이터 전처리, 연구데이터 분석, 분석결과 시각화 영역의 교육
 - * KISTI 과학데이터 스쿨(KSDS) 교육 인프라 및 UST 교육 인프라 활용

□ 전략3. ICT기반 데이터관리 기술교육은 전문교육기관 활용 위탁교육

- 데이터 수집, 데이터 저장, 데이터 운영 영역의 교육
 - * 전문교육기관 교육인프라 활용, 국내 15개 대학 빅데이터/데이터 교과과정 중 R&D와 부합하는 교과 과정·활용

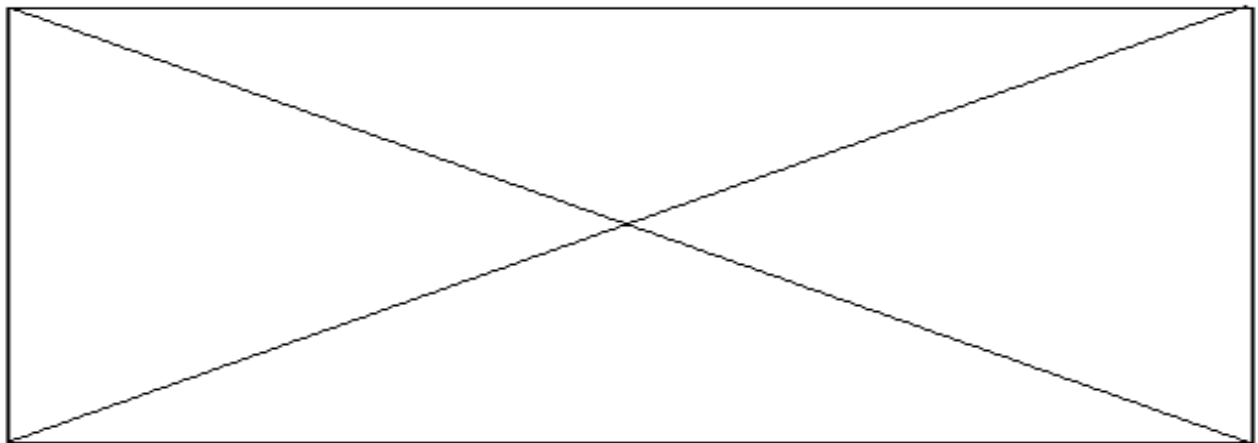
□ 전략4. 계산 과학공학 플랫폼인 EDISON의 교보재 체계 및 인프라 활용

- 언제, 어디서나 기계학습을 활용하여 데이터를 직접 분석·예측·학습할 수 있는 플랫폼
 - * 계산 + 데이터 + 인공지능을 융합 및 활용
 - * 웹 기반 EDISON 플랫폼을 활용하여 6개 계산과학공학 분야(전산열유체, 나노물리, 계산화학, 구조동력학, 전산설계, 전산의학)

2. 세부과제 내용

① 연구데이터 관리를 위한 ICT 기반기술 교육

□ 추진 목표

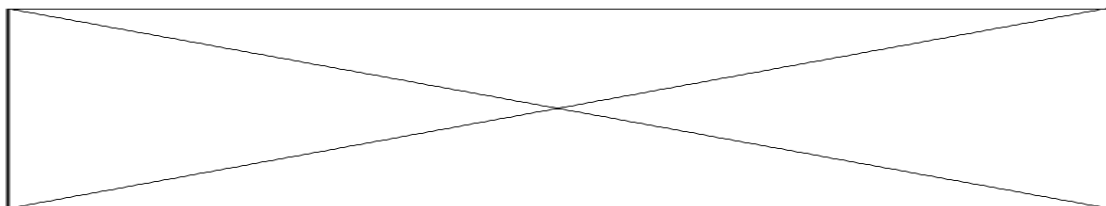


□ 교육대상

- 국가 연구기관 및 대학등 학계의 초·중급 연구원, 기업등 산업계의 사립기술 개발인력
공공분야 연구인력을 대상으로 초·중급 데이터 관리 및 운영 인재육성

국가 연구기관	대학 등 학계	기업 등 산업계
석·박사급 고급인력	석·박사급 고급인력	산업기술 개발인력
데이터처리 중급연구원	데이터 처리 중급연구원	공공분야 연구인력
데이터 처리 초급연구원	데이터 처리 초급연구원	

□ 교육범위



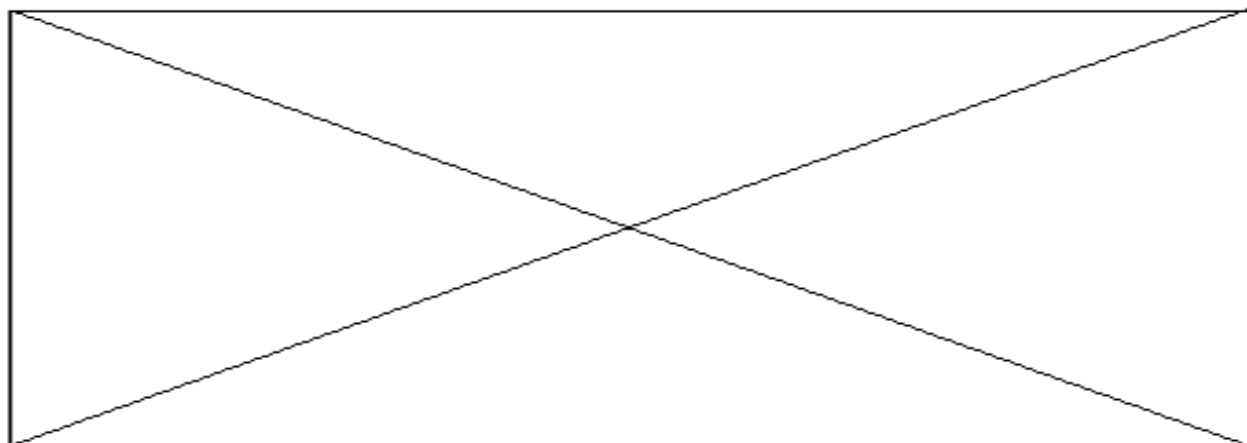
- 연구데이터 및 컴퓨팅 활용 R&D Task 중 데이터 관리영역인 연구데이터 수집과 연구 데이터 저장, 데이터 운영영역인 연구데이터 운영 Task에 대한 교육

□ 교육방법

- 데이터분석을 위한 기본 지식(데이터 수집, 저장, 운영 등)과 스킬에 대해서는 전문교육 기관의 교육과정을 이용할 수 있도록 전문기관을 활용한 위탁교육을 수행
- 지원 내용 : 1인당 연간 최대 100만원 이내 데이터 교육 강좌 수강료 지원
- 지원 방식 : 쿠폰 지급 또는 전자 바우처 등

② 연구데이터의 R&D 응용을 위한 분야별 전문활용 교육

□ 추진 목표

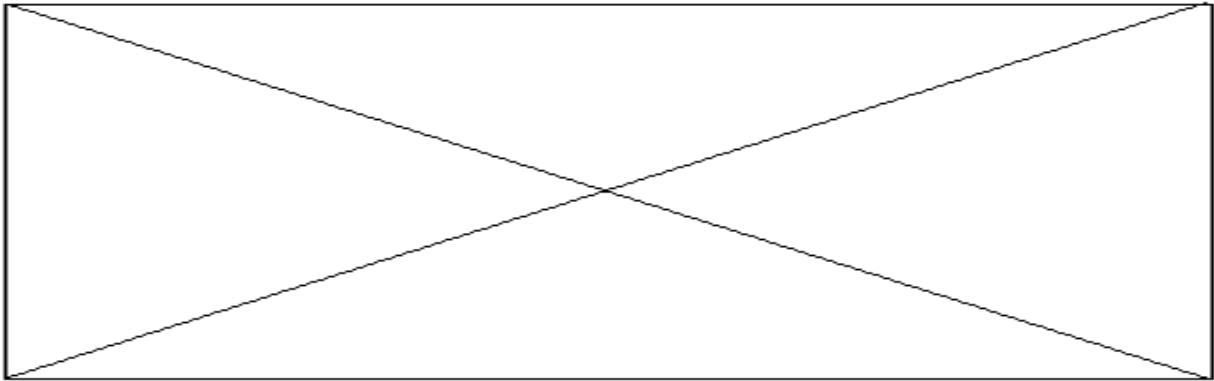


□ 교육대상

- 국가 연구기관 및 대학 등 학계의 석·박사급 고급인력을 대상으로 분야별로 고급 인재육성데이터 관리 및 운영, 기업 등 산업계에 대해서는 필요 시 고급인력 교육 수행

국가 연구기관	대학 등 학계	기업 등 산업계
석·박사급 고급인력	석·박사급 고급인력	산업기술 개발인력
데이터처리 중급연구원	데이터 처리 중급연구원	공공분야 연구인력
데이터 처리 초급연구원	데이터 처리 초급연구원	

☐ 교육범위



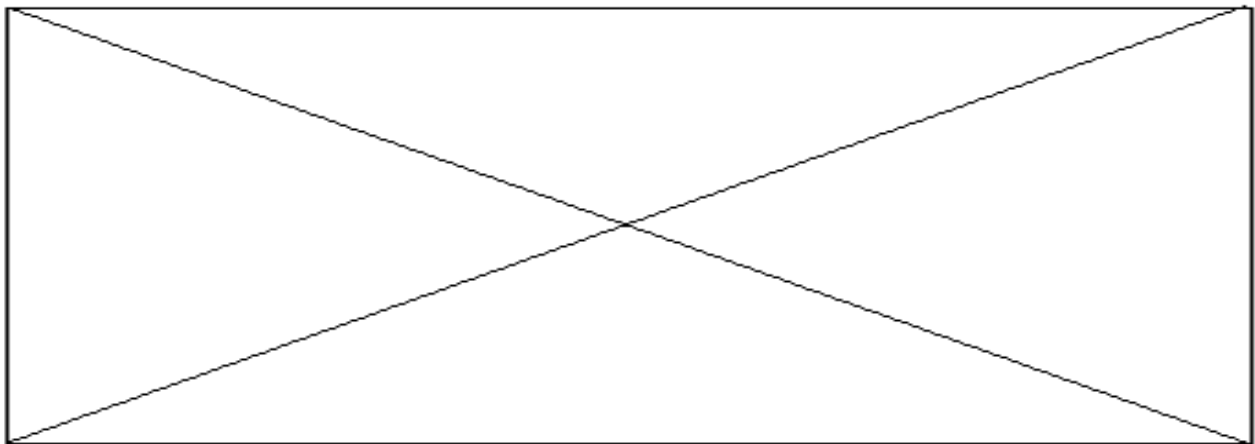
- 연구데이터 및 컴퓨팅 활용 R&D Task 중 데이터 관리영역인 연구데이터 전처리, 연구데이터 분석, 분석결과 시각화 Task에 대한 교육

☐ 교육방법

- KISTI KSDS의 분야별 연구현장의 수요를 기반으로 실무중심의 커리큘럼과 HPC 실습을 위한 상시 교육체계 활용
- EDISON 교육플랫폼과 연계한 UST의 연구분야별 전문교육 시스템 활용

③ 연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축

☐ 추진 목표



☐ 주요 구축내용

- 전문분야 데이터셋 구축 및 데이터 기반 교육 콘텐츠 개발
 - － 플랫폼 기반으로 전문분야 교육을 위한 분석/추론 교육 콘텐츠(context)* 개발 및 교

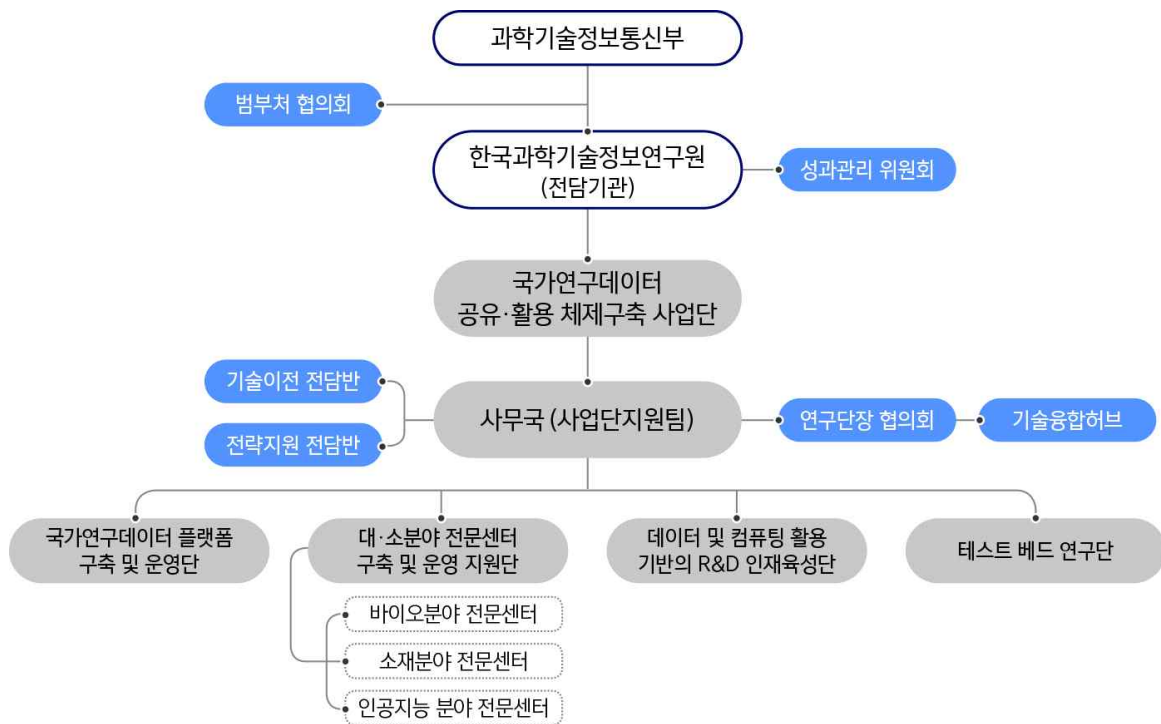
육용 Dataset 구축

* 콘텍스트(context): 데이터 혹은 콘텐츠의 단순집합을 넘어서 그것들의 활용 맥락을 의미

- 전문분야별 데이터 기반 연구교육 체계 수립
- 교육용 데이터셋 구축
 - 실험 및 계산 등의 데이터, 수집 및 생산을 통한 체계적 구축
- CX-Learning 도구개발 및 콘텐츠 교재제작
 - 고정된 공간의 데스크톱뿐만 아니라 노트북/스마트폰 등 언제 어디서나 학습을 할 수 있는 서비스로, 웹브라우저에서 반응형 컴퓨팅 교육과 동영상/전자문서/전자책 등의 다양한 매체를 융합한 강의 서비스구축
 - 교육환경 시대적 변화에 대응한 새로운 교육 개념인 CX-Learning* 도입
 - 교육 수요자의 요구에 기반한 CX-Learning 교육 도구 기반 교육 콘텐츠 제작
 - 워크벤치 기반 HTML5 이론교재 및 계산과학 시뮬레이션 연동 실습 도구
 - 멀티 학습요소(Learning element) 기반 온라인 학습도구 및 저작도구를 활용한 CX-Learning 콘텐츠 제작

VIII. 사업운영계획 및 소요예산

1. 사업추진체계



□ (조직구성) 본 사업 수행을 위한 조직은 크게 관리조직, 지원조직, 연구개발 조직으로 구분

- (관리조직) 한국과학기술정보연구원, 성과관리 위원회, 국가연구데이터 공유·활용 체제 구축 사업단, 연구단장 협의회로 구성되며 주로 연구단 연구프로세스 모니터링, 성과관리, 목표관리, 지원조직 관리, 테스트베드 운영관리 등의 역할 수행
- (지원조직) 기술이전 전담반, 변화관리 지원단, 기술융합 허브로 구성되며 사업수행에 요구되는 행정업무를 수행하고 연구단의 연구개발 과정에서 요구되는 기술, 변화관리, 기술융합 등을 지원하여 사업성과를 극대화하고 확산하는 역할 수행
- (연구개발조직) 연구단은 세부과제를 수행하는 연구주체들의 컨소시엄 형태와 테스트베드를 운영하는 조직으로 구성. 국가연구데이터 플랫폼 구축·운영, 대소분야 전문센터 구축·운영, 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 등의 역할을 수행

2. 사업관리

□ 사업의 추진절차

추진절차	추진내용	일정
국가전략 프로젝트 심의	사전기획내용 검토 및 기술성 평가(2월) 사업추진여부 확정(과학기술 전략회의, 2월)	'18년 7~8월
예비타당성 조사	예비타당성조사	'18년 11~
사업단 구성	국가연구데이터 공유·활용 체제구축 사업단 및 사무국 구성	'20년 02~
사업기획	과제기획서 및 RFP 작성(사업단, 관계기관 합동)	'20년 03~
사업심의위원회	사업의 기획, 평가 등 조정(과기부 장관) ※ 관계부처 담당 국장 사업심의위원회 참여	'20년 03~
사업시행 공고	세부추진계획 및 신규사업 공고	'20년 04~
신규과제 접수	신규과제 사업계획서 접수(사업단)	'20년 04~
과제선정·평가	신규과제 평가(사업단) ※ 사전검토-서면평가-발표평가-종합심의 신규과제 확정(과기부 장관)	'20년 04~
협약체결	주관기관과의 협약 체결(사업단)	'20년 12~
평가(연차, 단계, 최종)	연차별 실적 및 계획서 평가(사업단) 자문위원회 의견 및 평가위원회 심의	'20년 12~
사업비 정산	연간 사업비 회계 정산	'21년 02~
성과활용 보고/ 평가, 사업화	연구활용 성과 조사 및 성과관리	'21년 02~

3. 소요예산 및 인력

□ 세부사업별 투자규모

○ 총 5,751억원 투자(정부 출연금 100%)

－ 본 사업은 3개분야 20개 세부과제로 구성되어 있으며, 각 추진과제별 투자규모는 아래 참조

(단위: 백만 원)

분야별		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계		32,075	41,195	57,887	61,332	69,388	64,176	65,151	63,534	60,598	59,850	575,187
국가 연구 데이터 플랫폼 구축 및 운영	소계	15,595	23,778	22,418	25,315	19,309	15,579	17,283	17,738	17,569	17,581	192,165
	연구데이터수집· 연계및공유기술	1,804	2,522	2,405	2,739	2,066	1,671	1,861	1,909	1,898	1,899	20,774
	연구데이터저장· 보안기술	3,315	5,324	4,984	5,544	4,186	3,385	3,771	3,869	3,846	3,848	42,072
	연구데이터품질· 표준화기술	2,828	3,953	3,769	4,292	3,238	2,618	2,916	2,992	2,975	2,976	32,556
	지능형연구데이터 고속 검색 기술	1,634	2,624	2,457	2,733	2,063	1,669	1,859	1,907	1,896	1,897	20,738
	연구데이터융합· 분석환경구축기술	2,612	4,029	3,854	4,500	3,566	2,851	3,104	3,190	3,106	3,111	33,923
	연구데이터활용· 플랫폼기술	3,402	5,326	4,948	5,508	4,190	3,387	3,772	3,871	3,848	3,850	42,101
대·소 분야 전문 센터 구축 및 운영 지원	소계	9,500	10,900	28,850	29,300	43,350	43,050	42,200	40,200	37,300	36,400	321,050
	바이오분야 전문센터	1,150	1,200	1,200	2,400	2,400	2,400	2,950	2,950	2,950	2,950	22,550
	소재분야 전문센터	2,850	4,200	5,400	5,150	4,700	4,400	4,000	4,000	3,100	2,200	40,000
	인공지능분야 전문센터	5,500	5,500	6,250	5,750	4,250	4,250	3,250	3,250	3,250	3,250	44,500
	농림수산·식품			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	환경·기상			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	기계·제조			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	우주·항공·해양			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	원자력·핵융합					4,000	4,000	4,000	4,000	3,500	3,500	23,000
	에너지·자원					4,000	4,000	4,000	4,000	3,500	3,500	23,000
	건설·교통					4,000	4,000	4,000	4,000	3,500	3,500	23,000
	재난안전					4,000	4,000	4,000	4,000	3,500	3,500	23,000
데이터 및 컴퓨팅 활용 기반의 R&D 인재 육성	소계	6,980	6,517	6,619	6,717	6,729	5,546	5,669	5,596	5,730	5,869	61,971
	ICT기반 기술교육	1,802	1,887	1,976	2,070	2,168	2,271	2,378	2,491	2,609	2,732	22,384
	분야별 전문 활용교육	1,678	1,129	1,143	1,147	1,061	1,076	1,090	906	921	937	11,088
	교육용 데이터뱅크 및 서비스 구축	3,500	3,500	3,500	3,500	3,500	2,200	2,200	2,200	2,200	2,200	28,500

□ 비목별 투자규모

(단위: 백만 원)

연도		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인건비	소계	8,010	9,410	17,409	18,794	25,054	23,806	23,797	22,906	21,614	21,213	192,013
	투입인력(명)	259	295	321	344	432	356	349	339	313	308	3,318
	인건비	31	32	54	55	58	67	68	68	69	69	570
직접비	소계	21,384	28,080	35,128	36,030	36,226	32,745	33,523	33,012	31,749	31,498	319,375
	연구장비/재료비	12,645	17,794	23,334	21,947	20,095	17,838	18,075	17,888	17,062	16,784	183,462
	연구활동비	1,883	2,583	3,225	4,052	4,781	4,354	4,531	4,437	4,249	4,207	38,303
	연구과제추진비	4,255	4,104	4,722	5,172	5,683	5,670	5,855	5,715	5,731	5,840	52,747
	연구수당	1,277	2,029	1,734	2,694	2,734	2,394	2,655	2,661	2,609	2,612	23,400
	위탁연구개발비	1,323	1,569	2,113	2,166	2,934	2,489	2,406	2,311	2,097	2,055	21,463
간접비		2,682	3,705	5,350	6,508	8,108	7,624	7,831	7,617	7,236	7,140	63,800
연구비총액		32,075	41,195	57,887	61,332	69,388	64,176	65,151	63,534	60,598	59,850	575,187

□ 소요인력

- 2021년부터 2030년까지 전체 3,318명 인력 투입 예상
 - － 국가연구데이터플랫폼 1,114명, 대소분야 전문센터 1,811명, 인재육성 393명

IX. 사전타당성 분석

1. 분석 개요

- 대규모 신규 사업에 대한 예산 편성 시 합리적인 의사결정과 평가항목 간 중복 및 편향성 방지
 - * 「국가재정법」 제38조 및 동법 시행령 제13조의 규정
- － 정책적 타당성 분석 : 정책의 일관성 및 추진체계, 사업추진상의 위험요인
- － 과학기술적 타당성 분석 : 과학기술 개발계획의 적절성, 과학기술개발 성공가능성, 기존 사업과의 중복성
- － 경제적 분석 : 사업추진 목적달성에 따른 편익, 경제적 파급효과 및 투자의 적절성 등

2. 정책적 타당성

- 상위계획과의 부합성
 - － 과학기술분야의 최상위계획인 과학기술기본계획에 부합 사업, 문재인 정부 100대 국정 과제, 연구성과 관리활용 기본계획 등에 부합
- 사업 추진체계
 - － 한국과학기술정보연구원을 전문기관으로 하여 사업 추진에 필요한 조직과 협조, 사업에 참여하는 이해관계자의 연구데이터 공유·활용에 관한 역할분담 및 협조가 기대되는 사업
- 사업 추진의지
 - － 국가 연구데이터 공유·활용체제 구축사업은 동 사업 사전에 결성된 TF와 20여 건의 회의 등으로 보아 주관부처의 사업의지가 확고
 - － 이해관련자들이 연구개발과정에서 생산되는 연구데이터 활용의 어려움과 그 필요성에 관해 인지하고 있으며, 사업 참여예정 주체와 이해관계자의 사업에 관한 의지를 객관화할 수 있는 공식적인 문서와 구체적인 근거를 확보
- 재원조달 가능성
 - － 국가 연구데이터 공유·활용체제 구축사업은 총사업비 3,808억 원이 100% 정부출연금으로 추진되는 사업으로, 정부의 재정지원 능력 및 사업 추진 의사가 확보되어 사업 추진 중 예상되는 재정적 위험수준이 양호한 사업
- 법·제도적 위험요인
 - － 국가 연구데이터 공유·활용체제 구축사업은 「과학기술기본법」 과 「정보통신 진흥 및 융합 활성화 등에 관한 특별법」 등 적절한 법적 근거를 기반으로 추진된 사업이며,
 - － 동 사업 수행 중 연구데이터 공유·활용과 마찰을 일으키는 법·제도적 제약이 없으나 연구자의 연구개발 생산물인 연구데이터의 활용에 대해서 이해관계 간 협의가 필요

3. 기술적 타당성

- 기획과정의 적절성
 - 산·학·연의 전문가로 구성된 기획 참여 집단은 과학기술별 전문성과 전문가가 속한 기관 간 균형성을 확보
 - 과학기술 전문분야 센터의 우선순위 선정을 위해 실시한 평가에는 사업의 특성을 반영하고 적절한 평가항목 및 척도를 산정하여 적용
 - * 항목별 정량적 평가와 전문가 설문조사를 통해 공정하고 합리적인 항목별 가중치 적용 등의 기획과정 수행
- 사업목표의 적절성
 - PEST분석을 통해 사업이 직면한 문제 및 이슈를 해결할 수 있는 구체적이고 타당한 근거를 가진 사업목표를 시간의 관점에 따라 단계별로 제시
 - 사업목표와 함께 제시된 성과지표는 미래창조과학부의 지침서에 기반하여 산정된 것으로 사업을 관리함에 있어 체계적이고 객관적이며 이슈/문제와 유기적이고 논리적으로 연결
- 구성 및 내용의 적절성
 - 세부과제는 사업목표와 유기적으로 연계, 다수의 세부과제를 제시하였으나 서로 부족한 점을 보완하고, 하나의 목표를 제시했을 때보다 효과적으로 목표를 달성할 수 있도록 기획
 - 세부과제를 관리하기 위한 성과지표를 연차별로 제시하여 세부사업 간 선·후 관계에 있어 동 사업의 목표 달성의 효율성을 제고
- 기술추세 분석
 - 연구데이터 공유활용 관련 기술은 세계시장 주요 국가에서 기술 성장기에 속하여 연구개발이 활발하고 국가 간 경쟁이 치열해지는 시기임
 - 세계시장을 목표로 하는 기술개발을 위해 합리적인 연구개발 투자가 필요하며, 이에 따라 대규모의 재원투입 정당성이 확보되는 사업
- 과학기술 수준 분석
 - 특허 활동도, 기술 경쟁력 그리고 질적 수준에 있어 세계시장에서 우위를 차지하지 못하고 있으며 최근에는 저조한 활동을 이어나가고 있음으로 세계 주요 국가와의 격차 감소를 위한 투자가 시급함
 - 연구데이터 공유·활용체제 구축의 6대 핵심기술 개발은 기술 별 서로 다른 부분에서 우위를 확보하며 향후 기술개발 보강을 통해 전략적·정책적 목표달성이 가능한 사업임
- 기존 사업과의 중복성
 - 기존 사업과의 중복을 지양하고 사업시행의 효과가 유사성 및 동일성 여부를 분석하여 국가기관의 특정분야에 대한 과잉투자를 방지
 - 한정된 국가재정을 효과적으로 사용하고 정부의 연구개발투자가 균형을 이루도록 함

4. 경제적 타당성

○ 비용/편익추정

- 국가 연구데이터 공유·활용체제 구축사업은 3개의 세부 사업으로 총사업비는 5,751억 원으로 동 사업의 경제성 분석을 위한 총비용은 5,404억 원으로 산정
- 국가 연구데이터플랫폼 및 전문센터 구축 사업은 새롭게 생산된 산출물의 상용화로 인해 부가가치의 증가분을 바탕으로 편익을 추정하는 시장수요접근법으로 추정
- R&D 인재육성 사업의 경우, 교육수요, 월 평균 임금 및 육에 따른 임금 상승효과를 편익으로 산정

(단위: 억 원)

구분		비용					편익			
		플랫폼 구축 및 운영	대·소 분야전문센 터구축및운 영지원	R&D 인재육성	합계		부가가치 편익	교육효과 편익	합계	
					경상가치	현재가치			경상가치	현재가치
사업기간	2021	142	90	66	299	262		29	29	25
	2022	218	104	62	383	321		30	30	25
	2023	205	274	63	542	435		31	31	25
	2024	233	278	64	575	441		33	33	25
	2025	178	412	64	654	480		35	35	25
	2026	144	409	52	605	425		36	36	25
	2027	160	401	53	614	413		38	38	25
	2028	164	382	53	598	385		40	40	26
	2029	163	354	54	571	352		42	42	26
	2030	163	346	55	563	332		43	43	26
회임기간	2031				0	0			0	0
	2032				0	0			0	0
편익발생 기간	2033				0	0	805		805	416
	2034				0	0	889		889	440
	2035				0	0	982		982	465
	2036				0	0	1,084		1,084	491
	2037				0	0	1,198		1,198	519
	2038				0	0	1,323		1,323	548
	2039				0	0	1,461		1,461	580
	2040				0	0	1,614		1,614	613
합계		1,769	3,049	586	5,404	3,847	9,355	356	9,711	4,325

○ 경제성 분석

- 비용과 편익을 종합하여 경제적 타당성을 분석한 결과, 비용편익 비율(B/C Ratio)은 1.12, 순현재가치(NPV)는 477억원, 내부수익률(IRR)은 5.67%로 나타나 사업 추진의 경제적 타당성을 확보하는 것으로 분석됨

(단위: 억 원, %)

편익(현재가치)	비용(현재가치)	B/C	NPV	IRR
4,325	3,847	1.12	477	5.67%

○ 민감도 분석

- 사회적 할인율의 경우 $\pm 1\%p$ 단위로, 편익 및 비용은 $\pm 10\%p$ 단위의 변경요인이 발생하였을 때 경제성 판단 지표들이 어떻게 변화하는지를 검토함
- 할인율을 1%p씩 증감시켜 민감도 분석을 수행한 결과, 할인율이 1.3%p 증가하면, B/C가 0.99로 경제적 타당성을 확보하지 못하는 것으로 나타남
- 편익을 10%씩 증감시켜 민감도 분석을 수행한 결과, 편익이 12% 감소하면, B/C가 0.99으로 경제적 타당성을 확보하지 못하는 것으로 나타남
- 비용을 10%씩 증감시켜 민감도 분석을 수행한 결과, 비용이 13% 증가하면, B/C가 0.99로 경제적 타당성을 확보하지 못하는 것으로 나타남

5. 기대효과 및 파급효과

○ 경제적 기대효과

- 기존에 수행된 연구개발의 성과물을 활용하여 연구데이터의 불필요한 연구데이터 재생산 및 중복연구를 방지하여 국가연구개발사업의 효율성을 제고
- 연구자는 국가 연구데이터 플랫폼을 통해 수집된 연구데이터를 공유·활용하여 새로운 산업분야를 개척하고 고부가가치 지식을 창출
- 체계적으로 관리되는 연구데이터 및 연구개발 대한 정보를 제공함으로써 연구개발사업의 산업화 과정을 단축하고 효율성을 제고하여 신산업 창출에 기여
- 연구데이터의 다양성·복잡성으로 인해 시간이 많이 소요되었던 연구개발 및 실용화 기간의 획기적 단축
- 교육현장과 연구, 산업현장 간 공간적·시간적 단절을 극복하고 교육에서 사회로 진입하는 기간을 단축하여 전문 인력을 조기에 확보

○ 과학기술적 기대효과

- 국가 연구데이터 플랫폼은 연구자 커뮤니티를 형성하여 연구자 간 교류 및 융합연구를 실현할 수 있도록 함
- 연구데이터가 다수의 연구자에게 공유될 경우, 연구자의 자유로운 참여와 전문성으로 연구 결과 및 연구데이터의 품질 향상과 투명성 제고에 기여
- 통합하여 수집·관리하는 전문센터의 기능으로 연구데이터의 유실이 방지되며, 공유·활용으로 연구데이터를 지식·자산화 함
- 빅데이터 기술 발전의 현상과 연구데이터의 공유·활용 환경이 상호작용하여 연구개발 생태계를 구축하는 시너지효과를 일으킴
- 연구데이터의 지속적인 수집·관리를 통해 개별 연구기관이 자원별·기관별로 용이하게 연구데이터를 활용할 수 있도록 지원

○ 파급효과

- 연구데이터 공유·활용은 연구개발의 주기를 획기적으로 단축시키며, 이를 통해 정부 및 민간 투자를 통한 타 연구개발사업 진출에 기여
- 다양한 분야와 연령층에서 슈퍼컴퓨팅을 활용할 수 있는 기회를 제공하여 전 국민이 연구데이터 생산에 참여하는 잠재적 연구자로서의 자질을 키움
- 연구데이터의 일괄 수집·관리를 통해 분산적인 정보유통을 중앙 집중화하고, 양질의 데이터를 신속하게 제공·연계하여 범국가적인 연구 연맹 체제를 구축
- 대학과 출연연구소 센터를 연계하여 실무 적용 및 활용 중심의 빅데이터 교육을 실시하고 현장 전문 인력을 신산업 창출과 동시에, 실무중심의 인재를 사회 전반에 필요한 적재적소에 배치



본 문

국가 연구데이터 공유·활용 체제
구축 사업 기획보고서



목차

CONTENTS

제1장 사업 개요	1
1.1. 추진 배경 및 필요성	3
1.1.1. 사업 추진 배경	3
1.1.2. 사업의 필요성	6
1.1.3. 정부지원 필요성	9
1.2. 사업 추진 경과 및 체계	10
1.2.1. 사업 추진 경과	10
1.2.2. 사업기획 추진체계	18
1.2.3. 사업논리모형	21
1.3. 연구데이터 공유·활용 체제의 정의	25
1.3.1. 연구데이터의 정의	25
1.3.2. 연구데이터 공유·활용 체제의 정의	27
제2장 연구데이터 공유·활용 체제 현황 분석	29
2.1. 국외 현황 분석	31
2.1.1. 미국	31
2.1.2. 영국	35
2.1.3. 호주	39
2.1.4. 중국	42
2.1.5. 일본	45
2.1.6. 분석 및 시사점	47
2.2. 국내 현황 분석	48
2.2.1. 정책적 현황	48
2.2.2. 연구데이터 관련 기존 유사사업 분석	51
2.2.3. 인력양성 현황	52

2.3. 수요 분석	54
2.3.1. 수요 분석 개요	54
2.3.2. 연구데이터 1차 수요조사 결과	55
2.3.3. 연구데이터 2차 수요조사 결과	64
2.3.4. 수요조사 주요 시사점	70
 제3장 경쟁력 진단 및 추진방향 설정	 72
3.1. 국내 R&D 역량 분석	74
3.1.1. R&D 투자 현황	74
3.1.2. 기술역량 분석	77
3.1.3. R&D 인프라 현황	82
3.1.4. 연구데이터 관련 특허 분석	96
3.1.5. 연구데이터 관련 논문 분석	101
3.2. PEST 분석 및 종합	108
3.2.1. 분석 개요	108
3.2.2. 정책적 환경 분석	110
3.2.3. 경제적 환경 분석	111
3.2.4. 사회적 환경 분석	112
3.2.5. 기술적 환경 분석	113
3.3. SWOT 분석 및 대응전략 도출	114
3.3.1. SWOT분석	114
3.3.2. 대응전략	116
 제4장 사업의 추진전략	 118
4.1. 목적과 범위	120
4.1.1. 사업의 비전	120
4.1.2. 사업 추진 목표	121
4.1.3. 사업의 범위	124
4.2. 사업의 추진방향	127
4.2.1. 사업전략	127
4.2.2. 사업의 구성	134
4.2.3. 세부과제별 개요	137
4.2.4. 세부과제별 성과목표(지표)	144

4.2.5. 과제별 연계성 강화방안	153
4.2.6. 결과물 공유 및 부처 간 협력방안	157
제5장 국가연구데이터플랫폼 구축 및 운영	159
5.1. 과제개요	162
5.1.1. 추진배경	162
5.1.2. 필요성	162
5.2. 추진전략	165
5.2.1. 추진방향 및 목표	165
5.2.2. 추진전략 및 세부과제	167
5.3. 세부과제별 추진내용	193
5.3.1. 연구데이터 수집·연계 및 공유 기술 개발	193
5.3.2. 연구데이터 저장·보안 기술 개발	201
5.3.3. 연구데이터 품질·표준화 기술 개발	209
5.3.4. 지능형 연구데이터 검색 기술 개발	214
5.3.5. 연구데이터 융합·분석 환경 구축 기술	219
5.3.6. 연구데이터 활용·플랫폼관리 기술	226
5.4. 기대효과	235
5.5. 연구개발 대상 기술 선정과정	237
제6장 대소분야 전문센터 구축·운영 지원	239
6.1. 과제개요	242
6.1.1. 추진배경	242
6.1.2. 필요성	243
6.2. 추진전략	244
6.2.1. 추진방향 및 목표	244
6.2.2. 추진전략 및 세부과제	245
6.3. 바이오 전문센터 구축	250
6.3.1. 사업개요	250
6.3.2. 환경분석	253
6.3.3. 추진전략	262
6.3.4. 사업운영계획	274
6.3.5. 기대효과	278

6.4. 소재 전문센터 구축	279
6.4.1. 사업개요	279
6.4.2. 환경분석	279
6.4.3. 추진전략	309
6.4.4. 사업운영계획	346
6.4.5. 기대효과	354
6.5. 인공지능 전문센터 구축	355
6.5.1. 사업개요	355
6.5.2. 환경분석	359
6.5.3. 추진전략	366
6.5.4. 사업운영계획	387
6.5.5. 기대효과	390
6.6. 전문센터 우선분야 선정과정	392
6.6.1 우선분야 선정개요	392
6.6.2 대상 분야 선정	392
6.6.3 평가지표 선정 및 정량평가	393
6.6.4 평가지표별 중요도 산정	395
6.6.5 분야별 우선순위 도출	396
 제7장 데이터 및 컴퓨팅활용 기반의 R&D 인재육성	 398
7.1 과제 개요	400
7.2. 환경분석	403
7.2.1. 연구데이터 활용 인재양성 프로그램	403
7.2.2. (빅)데이터활용 인재양성 프로그램	404
7.2.3. 해외선진 인재양성 프로그램	406
7.2.4 연구데이터와 (빅)데이터 활용의 절차적 특징	408
7.3. 추진전략	414
7.3.1 추진방향 및 목표	414
7.3.2 추진전략 및 세부과제	417
7.4 세부과제별 추진 내용	422
7.4.1 연구데이터 관리를 위한 ICT 기반기술 교육	422
7.4.2 연구데이터의 R&D 응용을 위한 분야별 전문활용 교육	427
7.4.3 연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축	440
7.5 기대 효과	450

제8장 사업운영계획 및 소요예산	452
8.1. 사업추진체계 및 역할분담	454
8.1.1. 사업추진체계	454
8.1.2. 사업관리 방안	459
8.1.3. 성과활용 방안	469
8.2. 과제 · 소요예산 및 재원조달 방안	474
8.2.1. 전체 과제 및 사업비	474
8.2.2. 세부 과제 및 사업비	476
8.2.3. 재원조달방안	480
8.2.4. 소요인력	480
8.2.5. 중장기 재정운영계획과의 연계	481
제9장 사전타당성 분석	483
9.1. 사전타당성 분석 방법론	485
9.1.1. 개요	485
9.1.2. 영역별 타당성 분석 수행 체계	488
9.2. 정책적 타당성 분석	495
9.2.1. 상위계획과의 부합성	495
9.2.2. 사업 추진체제 및 추진의지	503
9.2.3. 사업 추진 상의 위험요인 및 대응방안	512
9.3. 기술적 타당성 분석	521
9.3.1. 사업 기획의 적절성	521
9.3.2. 기술개발 성공가능성	540
9.4. 경제적 타당성 분석	550
9.4.1. 비용 추정	550
9.4.2. 편익 추정	551
9.4.3. 경제성 분석	572
9.5. 기대효과 및 파급효과	576
9.5.1 기대효과	576
9.5.2. 파급효과	577

사업 개요

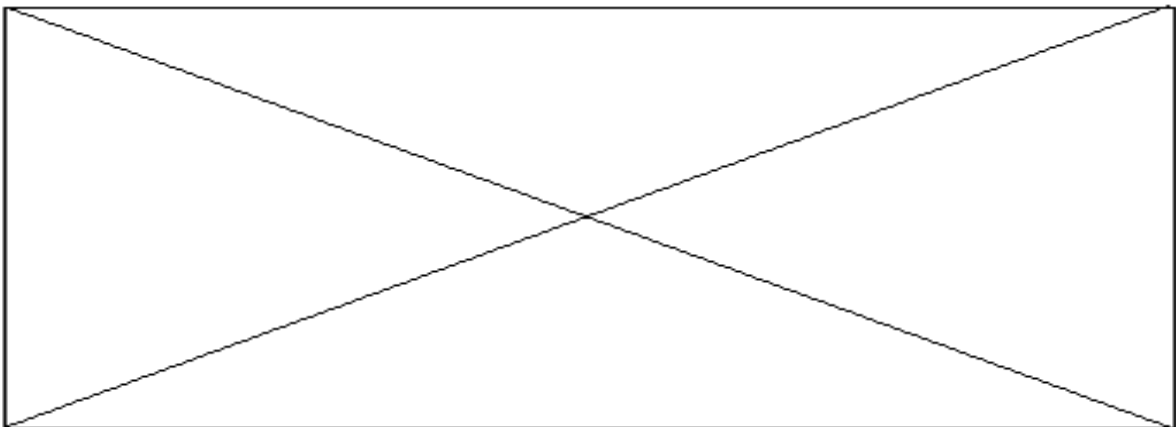
- 1.1. 추진 배경 및 필요성
- 1.2. 사업 추진 경과 및 체제
- 1.3. 연구데이터 공유·활용 체제의 정의

1.1. 추진 배경 및 필요성



1.1.1. 사업 추진 배경

4차 산업혁명을 통해 데이터 분석의 중요성이 증가함에 따라 연구개발의 패러다임도 데이터 분석 중심으로 변화하고 있으며, 선진국을 중심으로 연구의 성과와 과정을 개방화하는 오픈 사이언스 정책이 확산되고 있음. 그러나 국내에는 연구데이터의 공유·활용을 위한 제도·인력·인프라가 전반적으로 부족



<그림 1-1> 사업 추진 배경

- (4차 산업혁명과 데이터의 중요성) 4차 산업혁명과 함께 데이터의 활용이 과학기술 경쟁력의 핵심으로 대두
 - (기술의 융합과 데이터 혁명) 다양한 기술의 개방과 융합을 핵심으로 하는 4차 산업혁명이 산업구조 및 시장경제 모델에 가져올 파괴적인 혁신은 데이터 혁명에 기반을 둔 (세계경제포럼, 2016)
 - － 21세기의 원유로 비유되는 데이터 기술이 미래 경쟁우위 확보 여부를 결정(가트너, 2011)하며 새로운 비즈니스 가치 창출의 핵심 열쇠임(PwC, 2010)
 - － 데이터는 사물인터넷(IoT), 클라우드 기술, 인공지능(AI) 등 핵심기술의 기반으로 4차 산업혁명의 변화 동인(한국정보산업연합회, 2016)

- **(데이터 기술과 과학기술 경쟁력)** 데이터 분석 및 활용의 중요성이 증가함에 따라 데이터 활용 역량이 4차 산업시대 과학기술 경쟁력의 핵심으로 대두
 - － 데이터의 활용을 통한 고급 분석, 미래 예측 능력은 기업의 경쟁력, 국가의 경쟁력에 직결됨(한국정보산업연합회, 2016)
- **(연구개발 패러다임의 변화)** 디지털기술의 발전과 데이터 혁신으로 인해 연구개발 패러다임도 데이터를 중심으로 변화함
 - **(디지털 기술 확산)** 디지털기술은 연구 성과를 확산시키는 방식뿐만 아니라 연구를 추진하는 방식 자체를 변화시킴(Crouzier, 2015; Bosman and Kramer, 2015)
 - **(데이터 중심 연구개발)** 연구개발 패러다임이 ‘데이터의 생성, 처리, 분석 등 데이터 중심으로 전환¹⁾

<표 1-1> 4세대 연구환경의 분류 및 내용

1세대 연구	2세대 연구	3세대 연구	4세대 연구
경험 중심 연구	이론 중심 연구	컴퓨팅 자원 중심 연구	데이터 중심 연구
관측이나 실험을 통해 데이터를 수집-생산한 후 이를 기반으로 연구를 진행	수백 년 전부터 행해지던 연구 방식으로 모델링과 일반화 방법을 통해 연구를 진행	지난 십수년 전부터 가능해진 방법으로 컴퓨팅 자원을 활용해 복잡한 현상을 시뮬레이션하는 연구	방대한 데이터를 연구의 중심 도구로 사용하는 연구

- **(오픈사이언스의 확산)** 주요국을 중심으로 연구개발 과정 및 성과의 개방화를 통해 연구개발의 가치, 목표, 전략을 재정립하려는 움직임이 확산되고 있음
 - **(과학계의 개방화 추세)** 연구개발 과정에서의 디지털 기술 활용이 늘고 연구의 추진 및 성과 확산 과정에서 새로운 개방화의 추세가 증가하고 있음
 - － 오픈 액세스 저널과 셀프아카이빙 플랫폼을 통한 온라인 논문 공개의 증가 등 과학계 전반에 걸쳐 개방·공유 사례 증가(OECD, 2015b; 신은정, 2015에서 재인용; EC, 2016)
 - **(연구개발 정책의 변화)** 연구개발 과정과 성과의 개방과 공유를 국가적 차원에서 주도하는 오픈사이언스²⁾ 정책이 확산되고 있음

1) The Forth Paradigm, Data Intensive Scientific Discovery

2) 디지털 기술의 확산과 함께 연구의 성과와 과정이 점차 개방화되고 있으며, 이러한 추세를 오픈사이언스로 개념화함(OECD, 2016b; EC, 2016)

- 디지털 기술을 통해 연구의 전 과정을 보다 개방적으로 전환하고자 하는 일련의 움직임
을 지칭하는 오픈사이언스 정책이 미국, 영국, 호주, 중국 등 주요국을 중심으로 확산
(신은정, 2017)

□ (국내 현황) 일부분야(바이오·소재 등) 연구데이터 공유·활용 추진 중이나 제도·인력·인프라 미비하여 연구개발 분야 전반적으로 관리·활용 저조

○ (제도·인력·인프라) 일부 국가연구개발사업의 성과물 관리제도가 운영되고 KISTI에서 과학데이터스쿨 운영, 계산과학공학 EDISON 플랫폼 개발 등 교육인프라를 구축하였으나, 연구데이터의 전반을 관리하는 관련 법·제도 및 전담조직·정보시스템 등 인프라 미흡³⁾

○ (관리·활용) 대학 연구자 131명 중 연구데이터 관리에 대한 지침/내규/정책 등이 있는 연구자는 45명(34.3%), 연구데이터를 모든 사람들이 활용할 수 있도록 공유하는 연구자는 7명(5.3%)에 불과⁴⁾

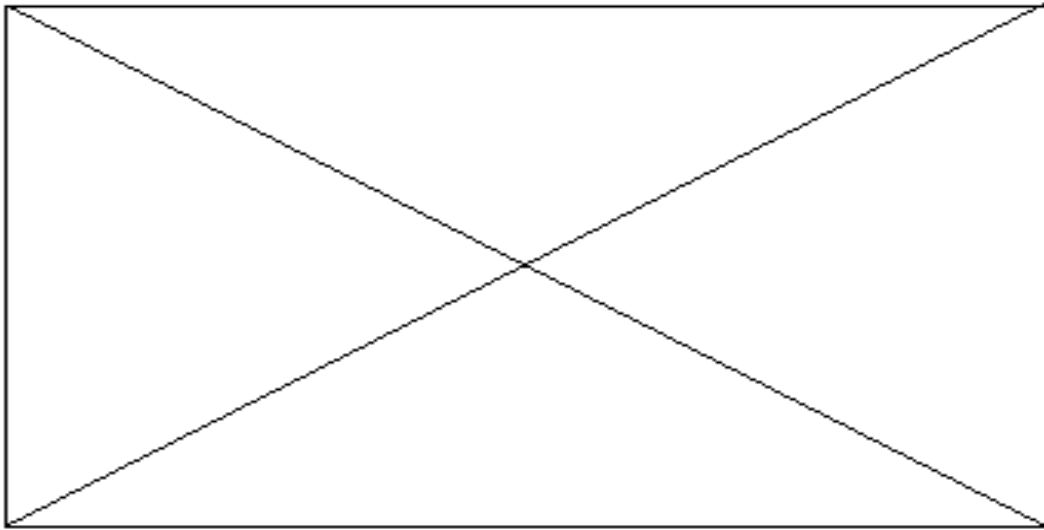
- 국내 출연연의 45%는 연구데이터 등록 시 제도적/기술적 지원 필요하고, 출연연의 54.5%는 연구데이터 등록/활용 지원 서비스 필요(KISTI, 2016)

3) 윤종민(2012), 국가과학데이터의 효율적 관리 및 활용을 위한 법제도 기본연구

4) 김지현(2012), 대학 내 연구자들의 연구데이터 관리에 관한 연구

1.1.2. 사업의 필요성

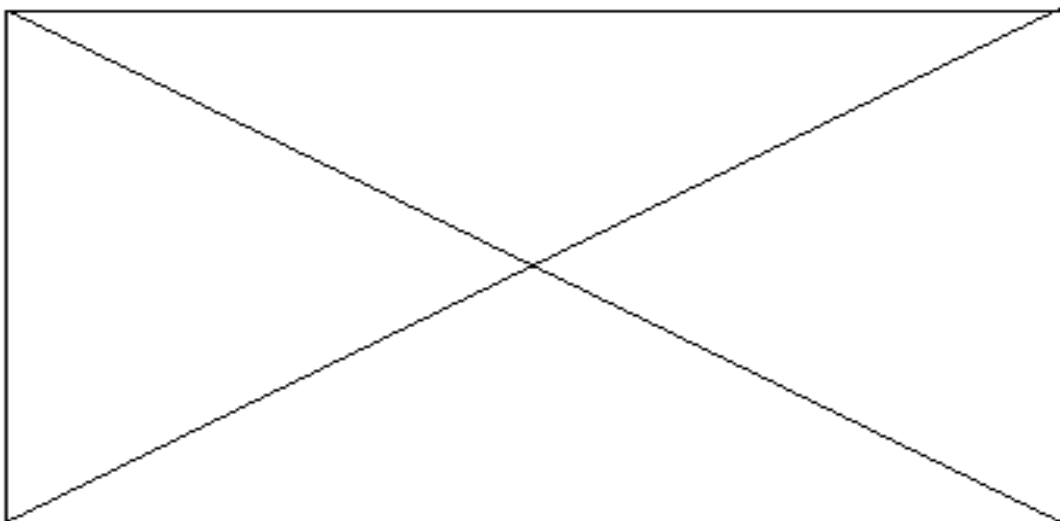
국가 연구데이터 공유·활용체제 구축은 **연구 재현성 및 투명성**을 제고하고 **융합 및 공동 연구**를 **활성화**시켜 국가 **연구개발 경쟁력**을 향상시킬 뿐 아니라, 연구데이터의 **재사용률**을 향상시키고 **중복연구**를 **최소화**함으로써 **R&D 효율성**을 증대시키고, 결과적으로 **신산업 및 양질의 일자리 창출**효과를 기대할 수 있음.



<그림 1-2> 사업 추진 필요성

- (R&D 효율성 증대) 연구데이터 재사용률 향상 및 중복연구 최소화를 통한 R&D 투자효율성 증대
 - (연구데이터 재사용률 향상) 기 수행된 연구를 통해 확보된 연구데이터를 다른 연구에 활용할 수 있는 기반을 마련하여 재사용률 제고
 - － 대부분 연구데이터가 개방적 접근, 체계적 보존(생성~폐기), 적극적 활용(재사용)을 통해 부가가치를 높이지 못하고 단순 저장 상태로만 보존 중(과학기술정책연구원, 2012)
 - － 본 사업을 통해 기 관측된 자료를 재사용함으로써 비용, 시간, 대형 장비를 같은 목적으로 재사용하는 비효율을 예방할 수 있음(한국과학기술기획평가원, 2012)
 - (중복 연구 최소화) 연구데이터 개방과 공유를 통하여 중복 연구를 최소화함으로써 투자효율성 증대
 - － 연구데이터 공유·활용체제 구축에 앞선 주요국에서는 중복 연구 최소화와 투자효율성 증대를 통한 경제적 이득을 크게 기대하고 있음(Beagrie, N. and Houghton J.W., 2016)

- (연구개발 경쟁력 향상) 연구 재현성 및 신뢰성 제고를 통한 국가 연구개발 경쟁력 향상
 - (연구 재현성 및 투명성) 연구데이터 및 소재의 공개를 통해 연구 재현성 및 투명성을 향상시키고 결과적으로 연구 신뢰도를 향상시킴
 - － 연구개발 경쟁력 제고를 위해 현재의 낮은 연구 재현성⁵⁾과 투명성을 개선하여 연구 신뢰도를 높일 수 있음⁶⁾
 - (융합 및 공동 연구) 연구데이터의 개방과 공유로 인해 학제 간 융합 및 공동 연구를 활성화시킴
 - － 과학기술분야 연구 활동에서 생산되는 연구데이터에 대한 효과적인 저장·분석·응용·적용을 위한 투자는 융합 및 공동연구 활성화를 위한 기반이 됨(한국과학기술기획평가원, 2012)
 - (신산업 및 일자리 창출 효과 기대) 연구데이터 공유·활용체계 구축을 통한 신산업 및 양질의 새로운 일자리 창출 효과 기대
 - (산업창출 효과) 연구데이터의 적극적인 공유 및 활용을 통해 신기술 및 미래 신산업 창출 기대
 - － 연구데이터의 적극적인 공유·활용체계 구축은 지능정보기술을 활용한 연구개발을 촉진시켜 새로운 가치와 미래 신산업 창출에 기여할 것으로 예측
- ※ 실험 없이 최소한의 예산으로 데이터분석만으로 신약후보 물질 발굴 가능해짐(네이처, 2018)



5) 현재 진행되는 상당수의 유망한 연구가 재현 불가능한 상황임(C. Glenn Begley & Lee M. Ellis, 2012)

6) 영국의 경우, UK Open Research Data Forum을 통해 연구관리기관, 대학, 연구자 간의 공감대 형성을 위한 네트워크 활동을 수행하였으며, 연구성과 평가보다는 연구 신뢰성 및 데이터 활용성 향상이 목적(KISTI, 2015)

\ <그림 1-3> 지능정보기술로 인한 신산업 창출 및 산업구조 변화 전망
(한국사물인터넷 협회, 2015)

- (일자리 창출) 국가연구데이터 공유·활용 체제 구축에 따른 연구데이터 활용 확산 시 신규 일자리 창출 기대

－ 연구데이터 공유 및 활용을 위한 데이터 전문직 신규 수요 창출

- ※ ‘정보통신기술 융합 실증 프로젝트’ 실시로 신규 일자리 2,683건을 창출(과기정통부, 2018)
- ※ 미국의 경우 2018년까지 14만~18만 명의 전문가 및 150만 명 정도의 데이터 관리자와 분석 인력 필요 예상(McKinsey(2011))
- ※ 국가 과학데이터센터 구축 시 취업자유발효과는 63,453명(국가 과학데이터센터를 구축하였을 경우 생산유발효과에서 1.5배의 경제적 효과가 발생함)(박성욱, 한선화, 2011)

1.1.3. 정부지원 필요성

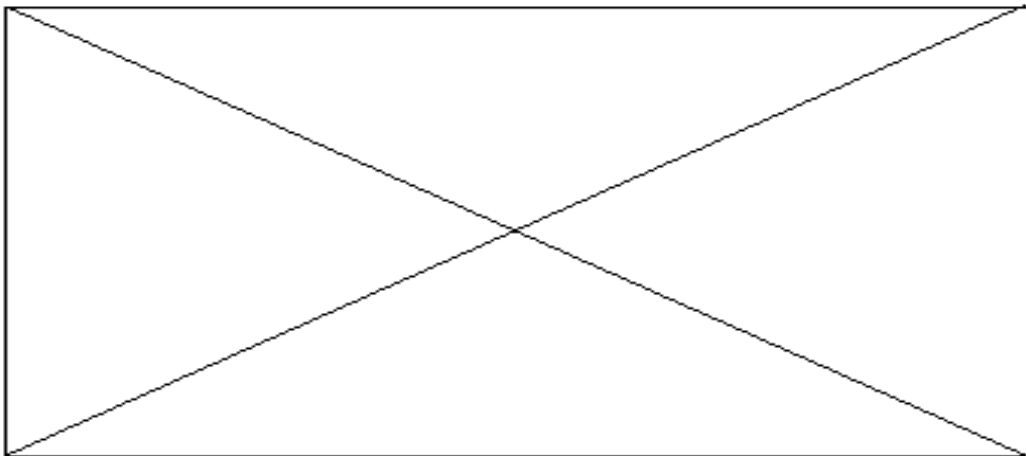
연구데이터의 공유·활용에 대한 수요는 있으나 연구성과 개방·공유에 대한 의지가 부족하여 정부 주도의 법·제도 정비를 통한 개방·공유의 촉진이 필요하며, 연구데이터의 체계적이고 종합적인 관리의 필요성을 고려할 때 연구기관별 사업 추진보다는 정부지원을 통해 국가연구개발사업의 데이터 관리전담기관을 지정하여 국가 차원의 연구데이터 공유·활용체계의 구축 필요

- (국가 차원의 관리체계 구축 필요성) 연구데이터 공유·활용을 위한 국가차원의 종합적 관리체계 구축 및 확산을 위해 정부주도 플랫폼 구축이 절대적으로 필요
 - (종합적 관리체계 필요성) 연구데이터를 개별 플랫폼 차원에서 관리하는 것에 그치지 않고 종합적·체계적으로 관리하고 공유하기 위해서는 국가적 차원의 전략과 지원 필요
 - － 연구데이터 기탁·관리·활용을 위한 구체적인 절차와 실현을 위해 법·제도를 정비하고 인프라 및 종합적 관리체계를 구축하기 위해서는 국가 차원의 전략 수립과 정부의 지원이 필수적임(이상환, 심원식, 2009)
 - (대학 및 소규모 연구기관 참여 유도) 자체적으로 연구데이터 공유·활용 플랫폼을 구축하여 운용하기 어려운 대학이나 소규모 연구기관들의 연구데이터 공유·활용 생태계 참여 유도를 위한 정부주도 플랫폼 구축 필요
 - － 대학과 소규모 연구기관은 과학데이터에 대한 관리 경험, 인식, 준비수준이 모두 낮은 상황(KISTI, 2016)에서 연구자와 기관의 인식을 개선하고 공유·활용 생태계를 활성화시키기 위해서는 정부의 역할이 매우 중요함
- (사업의 시급성과 정부의 역할) 연구데이터의 공유·활용체계 관련 주요국과의 격차 해소의 시급성을 고려할 때 정부의 역할이 매우 중요
 - (주요국과의 격차) 우리나라는 연구데이터 관련 법제도, 관련 기술수준 등에서 주요국과 격차를 보이고 있음
 - － 규제 중심의 법체계는 ICT 융합연구 활성화를 저해하고 주요국과의 격차를 보이는 주요 원인이며(KISDI, 2017), 빅데이터와 인공지능 등 관련 기술 수준도 미국 등과 격차 보임
 - (정부 지원의 중요성) 주요국과 벌어진 격차를 해소하기 위해서는 정부의 강력한 추진 의지와 신속한 지원이 매우 중요함
 - － 미국이 2013년을 기점으로 연방정부 차원의 오픈사이언스 정책을 본격적으로 추진하는 등 주요국의 사례를 통해서 정부지원의 중요성을 엿볼 수 있음(신은정, 2015 재인용)

1.2. 사업 추진 경과 및 체계

1.2.1. 사업 추진 경과

- 혁신성장 촉진을 위한 연구데이터 공유·활용 전략(안) 수립('18.1, 국가과학기술심의회)
 - 과학기술정보통신부는 모아서 새롭게 TF'를 구성·운영하여 「연구데이터 공유·활용 전략(안)」을 수립함
 - －(추진 배경) R&D 전과정의 성공과 실패를 포함한 연구데이터를 빅데이터화하여 인공지능을 적용·활용함으로써 R&D 성과 제고
 - －(구성) 학연 전문가 및 연구관리 전문기관, 과학기술정보통신부 관계자(6개 부서) 등으로 4개 분과별(총괄·제도화, 바이오, 소재, 대형연구장비) TF 구성



<그림 1-4> 모아서새롭게 TF 조직도

- 총괄·제도화 TF

<표 1-2> 총괄·제도화 TF 명단

No.	성명	소속	직급	비고
1	정병선	연구개발정책실	실장	팀 장
2	조경옥	HPC팀	팀장	활용·육성 체계
3	이현정	HPC팀	서기관	
4	이형진	KISTI 정책연구실	실장	
5	노서영	KISTI 대용량데이터허브실	실장	
6	이광렬	KIST 기술정책연구소	소장	
7	이재훈	연구제도혁신과	과장	제도화
8	임나리	연구제도혁신과	사무관	
9	황영희	과학기술정보과	사무관	
10	류영대	한국연구재단 기초연구총괄실	실장	
11	이원근	한국연구재단 원천연구사업실	실장	
12	장경수	한국연구재단 정책연구혁신센터	센터장	
13	김래수	기초연구진흥과	사무관	분야별 방안
14	이호준	생명기술과	사무관	
15	이선미	융합기술과	사무관	

- 대형연구장비 TF

<표 1-3> 대형연구장비 TF 명단

No.	성명	소속	직급	비고
1	윤성훈	과학기술정보통신부 기초연구진흥과	과장	-
2	노서영	KISTI GSDC	실장	데이터센터
3	윤진희	인하대 물리학과	교수	CERN(ALICE)
4	이형목	서울대 물리학과	교수	LIGO(중력파)
5	권영준	연세대 물리학과	교수	Belle II(중간자)
6	김수봉	서울대 물리학과	교수	RENO(중성미자)
7	구태영	포항가속기연구소 빔라인부	부장	포항방사광가속기(4세대)
8	추 용	국가핵융합연구소	부장	초전도핵융합연구장치
9	송충환	IBS 정책기획본부	본부장	중이온가속기
10	손봉원	한국천문연구원	책임연구원	한국우주전파관측망
11	권희석	KBSI 전자현미경연구부	부장	전자현미경
12	김형래	이화여대 의과대학	교수	유전체
13	정성진	한국뇌연구원 뇌연구정책센터	센터장	뇌지도, 뇌신경망 등
14	송지준	KAIST 생명과학과	교수	구조생물학
15	김래수	과학기술정보통신부 기초연구진흥과	사무관	

\ • 바이오·신약개발 TF

<표 1-4> 바이오·신약개발 TF 명단

No.	성명	소속	직급	비고
1	이석래	과학기술정보통신부 생명기술과	과장	
2	이도현	유전자동의보감사업단	단장	
3	여재천	신약개발조합	전무	
4	김 철	신약개발지원센터(오송)	팀장	
5	최환근	신약개발지원센터(대구)	팀장	
6	이주연	화학연구원 화합물은행	선임연구원	
7	김선규	한국생명공학연구원	연구원	
8	김성훈	의약바이오컨버전스연구단	단장	
9	임병권	한국파스퇴르연구소	본부장	
10	이병욱	국가생명연구자원정보센터	실장	
11	박장환	한국연구재단	PM	
12	김홍열	생명공학정책연구센터	센터장	
13	문성훈	생명공학정책연구센터	실장	
14	이호준	과학기술정보통신부 생명기술과	사무관	

• 소재분야 TF

<표 1-5> 소재분야 TF 명단

No.	성명	소속	직급	비고
1	최미정	융합기술과장	과장	
2	김선재	세종대	교수	원천소재 확보전략 기획과제 책임자
3	이병주	포항공대	교수	
4	장현주	화학연구소	PM	소재 빅데이터 플랫폼 시범구축 과제 책임자
5	이상관	재료연구소 전략기획부장	부장	원천소재 확보전략 기획과제 참여
6	이광렬	KIST 기술정책연구소장	소장	계산과학 플랫폼 기능 고도화 과제 책임자
7	김주선	한국연구재단 나노소재 PM	PM	—
8	김태연	국가나노기술정책센터		—
9	이선미	과학기술정보통신부 융합기술과	사무관	—

— (주요 논의사항) 연구데이터 공개 제도 및 수집·관리 범위, 단계별 인프라 확보 및 지원 방안 등 논의

<표 1-6> ‘모아서 새롭게 TF’ 주요 논의사항 및 결정사항

주요 논의사항	결정사항
연구데이터 공개 의무 제도와	과학기술기본법령 개정, 연구규정 개선
데이터 수집대상·범위, 우선순위, 단계별 목표, 전체 계획기간	단계별 목표, 계획기간, 우선추진 시범사업 설정
데이터 수집관리 의무 범위, 데이터관리계획 수준	대상사업과제 기준, 공개시점 결정, 기관평가 규정 개정
단계별 인프라 확보 및 지원방안	소요예산, 인프라 지원담당 기관
데이터 신뢰성 확보 및 디지털컨텐츠 식별자 부여 등 데이터 공유 필요조치	신뢰성 확보 프로세스 도출, 데이터 접근권 부여 범위(국내, 해외) 등
연구분야별 데이터 권라책임 보유 커뮤니티 육성방안	출연연 및 대학의 역할, 프로그램 규모(예산) 등

－ (진행 현황 및 내용) 총괄 TF 총 10회(총괄 TF 확대회의 2회)개최 , 분과별 TF 총 24회 개최됨

• 총괄 TF

<표 1-7> 모아서 새롭게 총괄 TF 추진 현황 및 주요 논의 내용

일 자	구분	주요 논의내용
'17.08.25	총괄 TF	• 연구데이터 공개 의무 제도와, 공유·활용 대상 분야 및 우선순위, 단계별 목표
'17.08.31	총괄 TF	• 연구데이터 공유·활용 체계 구축 추진전략로드맵, 분야별 시범사업 후보 선정논리 마련
'17.09.04	총괄 TF	• 바이오·소재 분야에서 수집 가치가 있는 데이터, 공개 시점, 출연연 역할 등 논의
'17.09.07	총괄 TF	• 연구데이터 공유·활용 체계 및 커뮤니티 육성방안, 데이터 관리 및 공유·활용 제도화 방안, 공유·활용 인프라 클라우드 구축 방안, 바이오·소재·거대실험장비 분야 추진방안 논의
'17.09.15	총괄 TF	• 연구데이터 관리 및 공유·활용 법·제도화안 및 대안, 시범분야 추진 방안, 데이터·컴퓨팅 활용 R&D 인력양성 논의
'17.09.21	총괄 TF (분야별 전문가 초청 확대회의)	• DMP 제출·이행 및 이행 점검에 있어 일반적인 원칙 및 기준, 데이터 및 컴퓨팅 활용 R&D 인력양성 방안 논의
'17.09.28	총괄 TF (전문가 간담회)	• TF에서 논의·마련한 연구데이터 관리·활용체계 구축 방안 관련 전문가 의견수렴
'17.10.19	총괄 TF (분야별 전문가 초청 확대회의)	• 클라우드 방식의 데이터 정리용 플랫폼을 통한 과거 및 진행 중인 과제 데이터 유실방지 방안, 적용 분야, 제공 필요서비스, 운영 방식·원칙 등 논의

일 자	구분	주요 논의내용
'17.10.26	총괄 TF 확대회의	- 총괄 및 분과 TF에서 논의·마련한 연구데이터 관리·활용 방안 집중 토론 ① 연구데이터 관리 및 공유·활용 제도화 ② 분야별 데이터 공유·활용 연구공동체 형성 ③ 데이터 및 컴퓨팅 활용 R&D 인력양성 ④ 국가 연구데이터플랫폼 구축 등
'17.11.29	총괄 TF 확대회의	- 연구데이터 관리·활용 체계 구축 방안 ① 빅데이터 네트워크 구축 사업과 연계 방안 ② 국가연구데이터플랫폼 구축에 있어서 KISTI 역할 ③ 2018년 시범사업 계획 ④ 연구데이터에 대한 연구자 권리 부여 방안 ⑤ 의료정보 비식별화(동향 파악, 개인정보 보호 이슈) ⑥ 뇌연구 분야 연구데이터 관리·활용 방안 ⑦ 연구데이터 정의, 범위·범주

－ 분야별 TF

• 제도화

<표 1-8> 모아서 새롭게 제도화 분야 TF 추진 현황 및 주요 논의 내용

일시	주요 논의내용
'17.08.28	• 미국, 영국, 호주 등 해외 선진국들의 연구데이터 제도화 사례 및 시사점 분석
'17.09.04	• 연구데이터 제도화 시 고려사항 논의(기본방향, 연구현장 의견 등)
'17.09.11	• 연구데이터 관리 및 공유·활용, 지원을 위한 단계적 입법 추진, 연구데이터를 R&D 성과물로 지정·관리
'17.09.18	• 과제 전주기(공고, 선정, 평가 등)에 걸쳐 DMP 제도화 적용 여부 및 해외사례(NSF) 분석 등

• 바이오

<표 1-9> 모아서 새롭게 바이오 분야 TF 추진 현황 및 주요 논의 내용

일시	주요 논의내용
'17.07.26	• 바이오분야 연구데이터 연구내용 현황
'17.08.10	• 인공지능 빅데이터를 활용한 신약 개발 방안
'17.09.13	• 인공지능 빅데이터를 활용해 신약을 개발하는 벤처 기업 사례 공유 및 빅데이터 활용 신약 개발 방안
'17.09.27	• 1~3차 회의 및 총괄 TF 논의사항 공유 및 향후 운영계획

- 소재

<표 1-10> 모아서 새롭게 소재 분야 TF 추진 현황 및 주요 논의 내용

일시	주요 논의내용
'17.08.10	• 소재 분야 빅데이터 필요성·정의, 국내 소재정보은행·참조표준센터 및 미국 NIST 연구데이터 수집 현황 등
'17.08.18	• 소재 분야 빅데이터 수집·관리·활용을 위한 전문가 설문 검토 및 분류체계 필요성 논의(소재전문가 설문조사 실시('17.08.18-23, 250명 대상): 빅데이터 정의, 소재 분류체계 개발 필요성, 빅데이터 필요성 등)
'17.08.25	• 소재 분야 연구데이터 플랫폼 구축 연구자 설문조사 결과 분석, 국내외 연구데이터 수집·활용 사례 발표 및 논의 - 딥러닝을 이용한 인포머틱스 구현' 사례 발표 세미나 실시
'17.09.13	• 소재분야 Tier-1, Tier-2 역할 및 범위 논의, Tier-1,2 예산 및 단계별 계획 정리

- 대형연구장비

<표 1-11> 모아서 새롭게 대형연구장비 분야 TF 추진 현황 및 주요 논의 내용

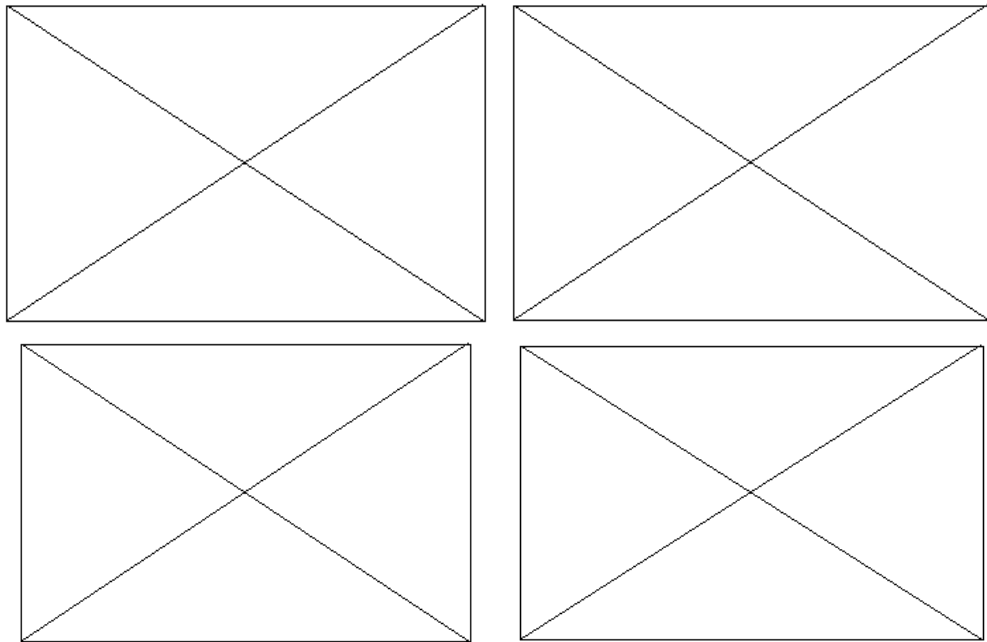
일시	주요 논의내용
'17.08.29	• 대형연구장비 데이터 저장, 공유, 활용 현황 및 구조 생물학 분야에서의 대형연구장비 데이터 분석, 활용 현황 등
'17.09.12	• 대형연구장비별 데이터 저장, 공유, 활용 현황 등
'17.09.12	• IBS 연구단 연구데이터 관리현황 및 향후 공유·활용 방안
'17.09.12	• 글로벌 프런티어 연구단별 연구데이터 관리현황 파악, 향후 공유·활용 방안

□ (공청회) 모아서 새롭게 확대 TF

- (목적) 장관 주재로 『연구데이터 빅데이터화를 위한 ‘모아서 새롭게’ TF』*에서 논의·마련한 연구데이터 관리·활용 방안 집중 토론

※ 추진 경과 : TF Kick Off(7.25) → 총괄 TF 및 분과별 TF(24회) → 초안 마련

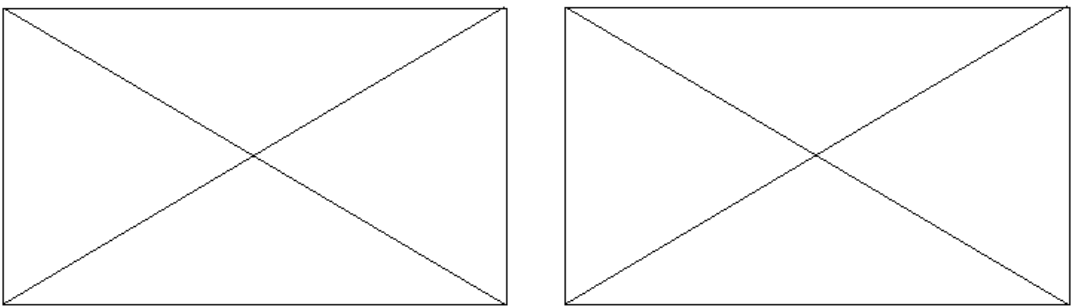
- (일시/장소) '17.10.26.(목) 14:00~15:50 / 서울 중앙우체국(21층 국제회의실)
- (회의내용) 총괄 및 분과별 TF에서 논의된 연구데이터 관리·활용 방안 등
 - ①연구데이터 관리 및 공유·활용 제도화, ②분야별 데이터 공유·활용 연구공동체 형성, ③데이터 및 컴퓨팅 활용 R&D 인력양성, ④국가 연구데이터플랫폼 구축 등
- (참석자) 과기정통부 장관, 연구개발정책실장, 성과평가정책국장, 바이오·소재·천문·물리·우주 등 데이터 활용 연구 전문가, 대용량데이터·컴퓨팅 전문가 등(25명)



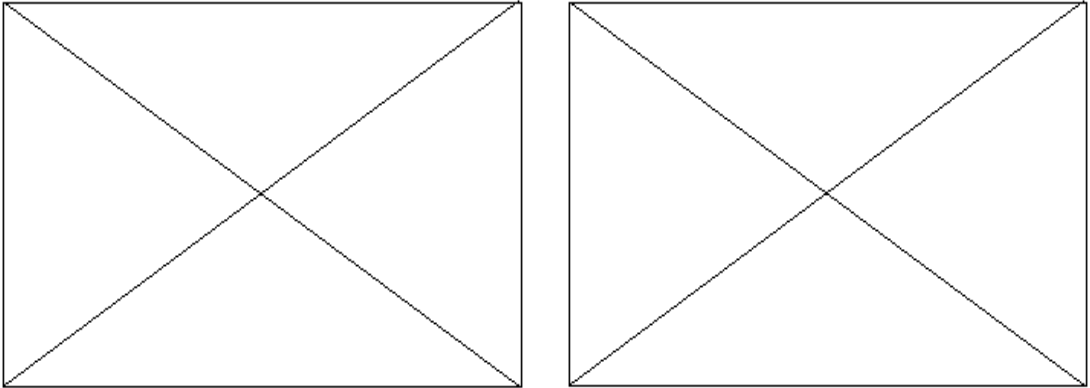
<그림 1-5> 모아서 새롭게 확대 TF

□ 국가 연구데이터플랫폼 공유·활용 체제 기획위원회 구성·운영('18.06~)

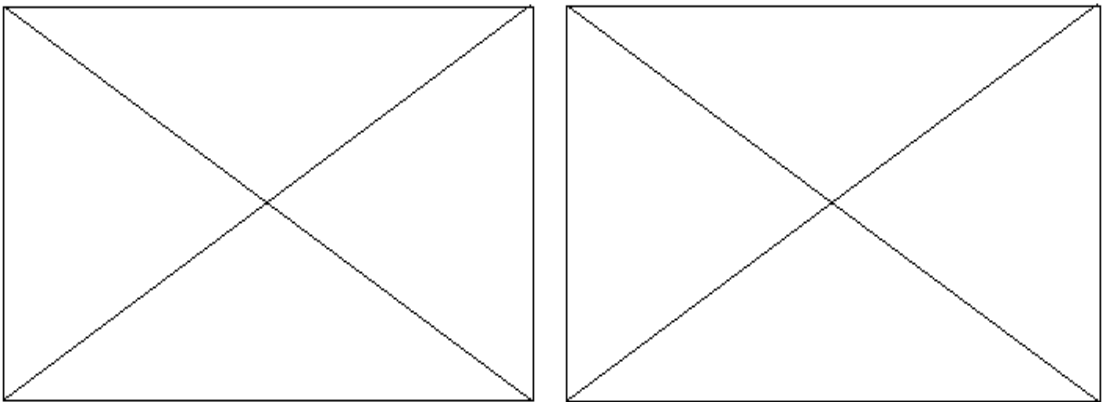
- 3개 분과(플랫폼, 전문센터, 인력양성)별 산·학·연 전문가 중심으로 구성하여 기능 및 기술체제(안) 마련
 - － 핵심 기능을 포함한 플랫폼 도출을 위해 플랫폼 분과위원회와 전문센터 분과위원회를 구성하였고, 연구데이터 활용을 위한 인력교육을 위해 인력양성 분과를 구성하여 추진
 - 플랫폼 분과위원회

구분	분과위원회	
회의사진		
일시	• 1차: '18.6.8.(15:00~18:00) • 2차: '18.8.16.(14:00~16:30) • 3차: '18.9.20.(14:00~16:00)	
회의내용	• 연구데이터 플랫폼의 개발 및 운영, 수요 및 요소기술 관련 의견 수렴 • 국가연구데이터 플랫폼 기능에 대한 타당성 및 적합성 검토 • 국가연구데이터플랫폼 세부 설계(안) 검토	

• 전문센터 분과위원회

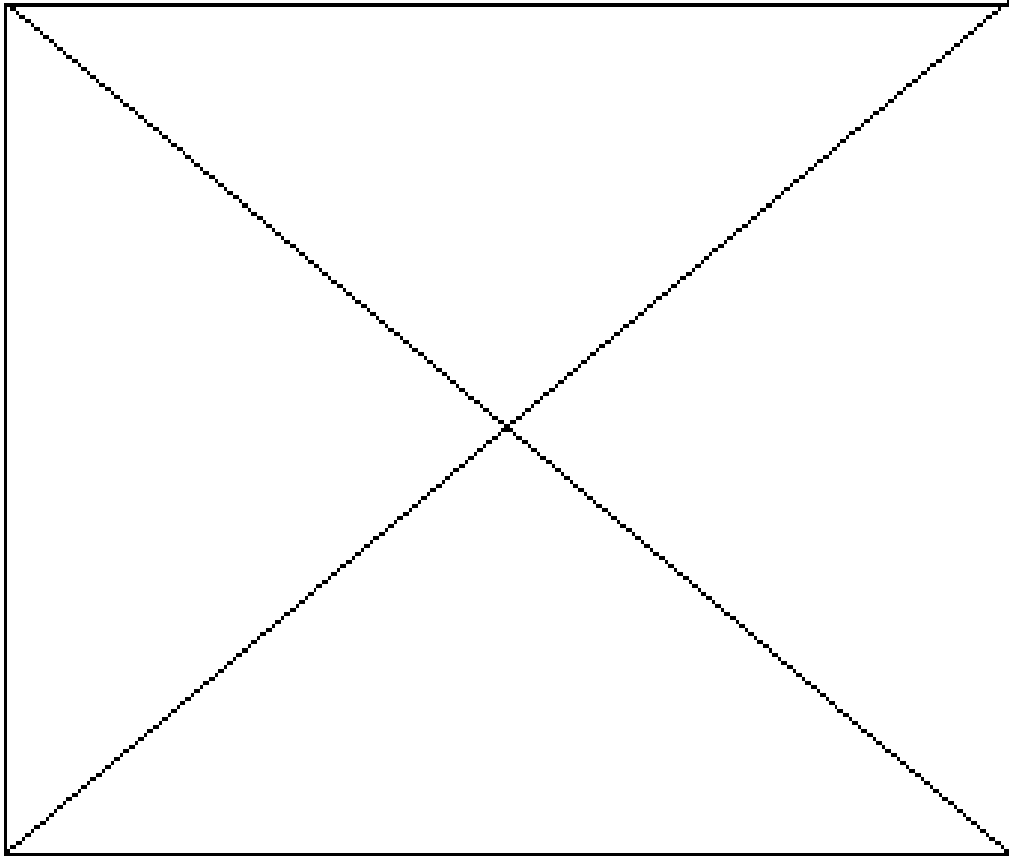
구분	분과위원회	
회의사진		
일시	• 1차: '18.5.03. (10:00~12:30) • 2차: '18.9.20.(10:00~12:00)	
회의내용	• 국가연구데이터플랫폼 시범구축 계획 논의 • 국내외 전문센터 현황 및 전문센터 구축 방향성 검토, 전문센터 선정 기준(안) 검토	

• 인력양성 분과위원회

구분	분과위원회	
회의사진		
일시	• 1차: '18.9.13(10:00~12:00)	
회의내용	• 국내외 교육 및 인력양성 현황, 연구데이터 인력 양성 방향성 검토 및 시사점 도출	

1.2.2. 사업기획 추진체계

- 동 사업의 추진체계는 과학기술정보통신부가 총괄부처 역할을 수행하고, 한국과학기술정보연구원 과학데이터연구센터가 참여하는 방식으로 구성



<그림 1-6> 사업추진체계

- (주관부처) 과학기술정보통신부 주관의 국가 연구데이터 공유·활용 체제 관련 기본계획 수립부터 예비타당성 조사 기획까지 총괄
 - 예비타당성조사 기획과 관련된 중요 의사결정(기술기획위원회의 산·학·연 전문가 추천, 전반적인 사업의 방향 등) 수행
- (사업기획 추진단) 한국과학기술정책플랫폼협동조합, 한국과학기술정보연구원, 투이컨설팅이 각각 사업기획, 기술기획, 과제기획을 담당하여 사업기획 실시
 - (사업기획) 국가 연구데이터 공유·활용 체제 관련 주요 선진국의 정책적·경제적·사회적·기술적 동향 및 기술개발 트렌드, 국내 R&D 동향 등을 고려하여 국가 정책 차원에서의

사업전략 도출

- (과제기획) 연구데이터를 생산·공유·활용하는 국내 연구자를 대상으로 연구데이터 활용의
향 수요조사 및 세부 과제 선정
- (기술기획) 국내 산·학·연 전문가를 대상으로 연구데이터 활용의 현재 또는 미래에 요구
되는 기술수요를 파악함
- (기술기획위원회) 분야별 분과위원회를 국내 산·학·연 전문가로 구성하여 아이디어 발굴,
목표 설정, 전략 수립 등 기획보고서 작성의 기술적 부분 담당
 - (목적) 분과위원회를 통해 연구방법 선정 및 사업설계 상세내용 검토 등 주요 진행사항
에 대한 모니터링을 실시하여, 기획과정의 공정성 및 투명성을 제고함
 - (구성) 연구데이터 관련 각 분야의 전문가들의 다양한 의견을 수렴할 수 있도록 플랫폼
분과위원회, 전문센터 분과위원회, 인력양성 분과위원회로 구성함

<표 1-12> 플랫폼 분과위원회 소속 전문가 명단

No.	성명	소속	직급
1	유재수	충북대학교	교수
2	김영국	충남대학교	교수
3	이문용	한국과학기술원	교수
4	이종후	국가보안연구소	책임
5	최재훈	한국전자통신연구원	책임
6	예상준	한국한의학연구원	팀장
7	김동욱	한국화학연구원	팀장
8	송중호	국가핵융합연구소	선임
9	진태은	국가생명연구자지원정보센터	실장
10	이영기	정보화진흥원	수석연구원
11	임성준	데이터베이스진흥원	차장
12	이민호	한국과학기술정보연구원	센터장
13	오상화	SK텔레콤	Manager
14	서완석	NVIDIA	상무
15	박정혜	한국정보통신기술협회	책임
16	장성우	오라클	본부장
17	박상욱	아이티센	상무
18	김종태	솔리데오	상무

〈표 1-13〉 전문센터 분과위원회 소속 전문가 명단

No.	성명	소속	직급
1	이상환	한국과학기술정보연구원	센터장
2	이선경	한국화학연구원	책임
3	진태은	국가생명연구자원정보센터	실장
4	조완섭	충북대학교	교수
5	김경수	코아아이티	부사장
6	유호진	한국정보화진흥원	수석
7	한종규	한국지질자원연구원	책임연구원
8	이영진	기초과학연구원	팀장
9	노서영	한국과학기술정보연구원	센터장
10	오창석	재료연구소	책임연구원
11	서덕록	한국과학기술연구원	팀장
12	안재평	한국과학기술연구원	센터장
13	한승우	서울대학교	교수
14	이민호	버츄얼랩	사장
15	한상혁	한국항공우주연구원	책임연구원
16	김종수	한국천문연구원	본부장
17	김현우	한국과학기술정보연구원	선임

〈표 1-14〉 인력양성 분과위원회 소속 전문가 명단

No.	성명	소속	직급
1	김화종	강원대학교	교수
2	이종숙	한국과학기술정보연구원	센터장
3	김평	전주교육대학교	교수
4	박수형	전국대학교	교수
5	성원경	과학기술연합대학원대학교	학과장
6	최규술	환경아카데미	부장
7	최종성	한국데이터진흥원	실장
8	이인구	인텔코리아	전무
9	안부영	한국과학기술정보연구원	실장

1.2.3. 사업논리모형

- (사업목표의 도출) 본 사업의 목적이 연구데이터 공유·활용 체계 구축과 이를 통한 국가 연구개발 효율성 및 경쟁력 강화라는 점을 고려하여 국가 연구 데이터 공유·활용 활성화를 주요 목표 설정
 - (국가연구데이터 공유·활용 생태계 조성) 연구데이터 공유·활용 및 분야별 연구데이터 플랫폼 구축 및 융합 연구 활성화
 - － 연구데이터 공유·활용 및 분야별 연구데이터 센터 플랫폼: 12식
 - － 기술혁신 표준모델 및 콘텐츠 소프트웨어: 71개
 - (연구 성과물 수준 제고) 인력양성 및 일자리창출을 통한 전문분야 별 연구 성과물 수준 제고 : 인력양성 및 일자리창출: 31,430/5,148명
 - (연구데이터 공유·활용 인식 제고) 연구데이터 관리·공유·활용 관련 법적 근거 마련 및 데이터관리계획 도입
- (투입) 본 사업의 추진 과제인 국가연구데이터 플랫폼 구축 및 운영, 대·소분야 전문센터 구축 및 운영 지원, 데이터 및 컴퓨터 활용 기반의 R&D 인재육성 별로 인건비, 직접비 그리고 간접비를 투입요소로 고려하여 산정
 - 총 사업비 : 5,751억원
 - － 국가 연구데이터 플랫폼 구축 : 1,921억원
 - － 대·소분야 전문센터 구축 및 운영 지원 : 3,210억원
 - － 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 : 620억원
- (활동) 본 사업의 주요 활동인 국가 연구데이터 및 전문분야 별 플랫폼 구축, 연구데이터 활용 전문 인력 양성, 연구데이터 관련 법적 근거 마련, 신산업 창출 및 일자리 창출 기여로 구성
 - 본 사업이 10년의 장기 과제이고 국가 연구 데이터 공유·활용 활성화를 목적으로 한다는 점에서 전문 분야별 플랫폼 구축을 포함시켰으며,
 - 국가 연구데이터 공유·활용체계 기반은 연구자간 공유 네트워크가 지속적이고 안정적으로 구축·확대 되어야 한다는 점에서 연구데이터 활용 전문 인력 양성도 본 사업 활동으로 구성

- (산출) 본 사업은 장기간(10년) 동안 진행되는, 사업유형이 기초연구, 중장기산업 기술 개발, 인력양성, 시설장비 구축 성격의 혼합 사업으로 성과지표 분류체계상 5대 성과지표인 과학적성과, 기술적성과, 경제적성과, 사회적성과, 인프라성과 지표를 모두 적용함
 - 과학기술적 성과 : 논문(SCI) 수, 논문(KCI) 수
 - 기술적 성과 : 특허 출원 및 등록 건수, 기술혁신 표준모델 개발 수, 콘텐츠 S/W 수
 - 논문 및 특허 관련 성과지표의 설정기준을 마련하기 위해 기 추진한 유사사업의 최근 3년간 논문 및 국제특허의 출원/등록 수를 분석한 결과 SCI급 논문은 10억 원당 1.15건, 국제 특허는 10억 원당 1.11건으로 조사(SW컴퓨팅원천기술개발 및 ICT융합원천기술개발사업 등 2개 사업)하여 목표를 설정하였으며, 콘텐츠S/W는 세부과제별 성과목표별로 개발 기술에 대한 S/W 출원수로 설정
 - 경제적 성과 : 중소기업 지원 및 연구개발 서비스 수
 - 사회적 성과 : 정책효과 및 국제 협력 건수
 - 정책효과는 DMP 정책에 따른 연구데이터 활용 R&D 집행을 10%씩 증감, 홍보확산은 매년 1회 성과에 대한 홍보활동, 국제협력은 단계별 해외 세미나 참석 및 홍보수를 목표로 설정
 - 인프라 성과 : 공동시설장비 및 전산시스템 개수
 - 공동활용시설 장비는 세부과제별로 Web/Was 서버, DB서버, Computing 서버 책정, 전산시스템은 세부과제별 1개로 교육용 DB 구축을 목표로 설정목표를 산정
- (성과) 본 사업은 혼합사업으로 기초 연구부터 인력 육성 까지를 지원한다는 점에서 매출, 사업화, 시장점유율 등 시장관련 지표 보다는 연구데이터를 공유하고 활용적인 측면에서 설정
 - 국가연구데이터 공유·활용 생태계 조성을 통해 연구개발 주기 단축 및 융·합 연구개발
 - 사용자 중심의 서비스 제공(데이터 융합 및 분석에 활용할 수 있는 다양한 환경 지원)을 통해 전 국민이 연구데이터 생산에 참여
 - 연구자간 협업연구를 위한 커뮤니티 활동 지원을 통한 범국가적 연구 연맹 체제 구축
 - 연구데이터의 R&D 응용을 위한 분야별 전문활용 교육을 통해 실무중심의 인재 배출

성과 유형		성과지표	비고
과학적 성과	논문	논문(SCI급)	논문 건수
		논문(KCI)	논문 건수
기술적 성과	지식재산	특허	특허 건수
	비지식재산	기술혁신	공정혁신, 표준획득
	성장동력창출	콘텐츠 S/W	S/W 등록
경제적 성과	간접성과	기술활용효과	분석환경, 클라우드
		중소기업지원	교육, 분석환경
	연구개발서비스	연구개발서비스	표준화, 연구개발지원
사회적 성과	인적자원·고용	인력양성	인력수
		일자리창출	일자리 창출 수
	공공복지	정책효과	
	과학대중화	홍보, 확산	
	국제협력	국제협력	
인프라 성과	연구인프라	공동활용시설장비	
		전산시스템	응용시스템

□ (수혜자 설정) 본 사업 연구데이터가 활용 및 효과가 파급될 수 있는 범위 내 주체들을 파급단계에 따라 1차, 2차, 3차 수혜자로 설정

- 본 사업이 연구데이터의 생산과 축적을 1차 목적으로 한다는 점에서 정부출연연구소·대학이 1차적 수혜자가 되고,
- 개발된 플랫폼의 데이터를 기업에서 활용 할 수 있다는 점에서 기업은 2차 수혜자가 되며,
- 연구데이터가 민간 및 공공시장에서 활용됨으로써 국민적 편익을 제고한다는 점에서 국민이 3차 수혜자가 됨

[논리모형]

1. 이슈 / 문제

- **(배경요인)** 4차 산업혁명 시대에 데이터가 경쟁의 원천으로 대두되며 R&D에서도 연구성과와 과정을 개방화하는 오픈 사이언스 확산으로 연구데이터 가치가 상승하는 추세이며, 경제·사회·기술적 측면에서도 영향을 미침
- **(해결하고자 하는 이슈/문제)**
 - **(법제도·정책적 문제)** 연구데이터는 연구결과물로 규정되어 있지 않아 정부R&D 연구성과물 관리전담기관에서 관리되고 있지 않으며, 이로 인해 데이터 저장·분석 인프라는 부족한 상황이며 데이터 활용 또한 저조함
 - **(제도·인프라 문제)** 데이터 저장·분석·공유를 위한 제도 부재 및 인프라 부족으로 대부분 연구자(88%)가 단순 파일단위 및 외부 리포지터리 형태로 관리하고 있으며, 공유·활용을 기반의 연구공동체 지원 또한 부족함
 - **(연구자 인식 문제)** 연구데이터에 관한 연구자의 연구 현장인식이 부족하며 연구개발의 산출물을 개인 혹은 소속연구실의 자산으로 인식하는 경우가 빈번함
- **(사업을 추진하지 않을 경우 발생할 것으로 예상되는 문제)** 연구데이터의 미흡한 관리로 인한 중복 연구사업으로 인한 공공자금 낭비가 예상되며, 장비 및 인프라 부족으로 연구데이터를 활용하지 못하는 소규모 연구기관과 이미 활용 중인 기관과의 형평성 문제가 야기될 수 있음
- **(연구개발사업 이외의 대안/전략)** 국가연구데이터는 국가연구개발사업의 산출물로, 국익을 목표로 한다는 점에서 수익창출을 주목적으로 하는 민간투자는 기대하기 어려우며, 개별적으로 연구데이터를 수집·관리하는 기관의 연구데이터까지 연계하는 플랫폼을 운영하기 위해서는 강력한 정책적 부합성을 확보한 정부의 역할이 필요함



2. 직접 목표

- 연구데이터 공유·활용 및 분야별 연구데이터 플랫폼 구축 및 융합 연구 활성화
 - 연구데이터 공유·활용 및 분야별 연구데이터 센터 플랫폼: 12식
 - 기술혁신 표준모델 및 콘텐츠 소프트웨어: 71개
- 인력양성 및 일자리창출을 통한 전문분야 별 연구 성과물 수준 제고
 - 인력양성 및 일자리창출: 31,430/5,148명
- 연구데이터 공유·활용 인식 제고
 - 연구데이터 관리·공유·활용 관련 법적 근거 마련 및 데이터관리계획 도입

3. 수혜자

- (1차) 정출연 및 대학
- (2차) 기업체
- (3차) 일반국민



4. 투입	5. 활동	6. 목표 산출	7. 성과/영향
• 총 사업비: 5,751억 • 플랫폼 구축: 1,921억 • 전문센터 구축 및 운영: 3,210억 • 인재육성 운영비: 620억	• 국가 연구데이터 및 전문분야 별 플랫폼 구축 • 연구데이터 활용 전문 인력 양성 • 연구데이터 관련 법적 근거 마련 • 신산업 창출 및 일자리 창출 기여	• 논문: SCI 212건/KCI 464건 • 특허: 223출원/87등록 • 중소기업 지원: 2개 • 연구개발 서비스: 18개 • 정책효과 및 국제 협력: 13건 • 공동시설장비 및 전산시스템: 6,036개	• 연구개발 주기 단축 및 융·합 연구개발 • 전 국민이 연구데이터 생산에 참여 • 범국가적 연구 연맹 체제 구축 • 신산업 창출 및 실무중심의 인재 배출



8. 가정 (위험요인과 대안)

- 연구데이터 공유·활용 과정에서 기술적 문제 및 실패위기 과제에 대한 기술적 위기 발생할 경우
→ 데이터 활용 지원센터 운영
- 연구개발과정에서 연구데이터를 직접 공유·활용하는 연구주체(예 연구자)의 참여 및 협력이 부진할 경우
→ 상호협력·융합연구 독려를 위한 데이터 기반 연구커뮤니티 운영

1.3. 연구데이터 공유·활용 체제의 정의

1.3.1. 연구데이터의 정의

- (연구데이터의 정의) 「연구데이터 공유·활용 전략」에서 연구데이터를 ‘국가연구개발사업의 각종 실험, 관찰, 조사, 분석 등을 통하여 산출된 자료로 연구성과의 재현에 필수적이고 객관적인 사실데이터’로 정의함⁷⁾
- (연구데이터의 종류 및 형태) 연구 데이터는 R&D 활동의 과정·결과를 통해 얻은 데이터로 아래와 같은 범위를 포함

<표 1-15> 연구데이터의 종류와 형태

구분	실험데이터	관측데이터	시뮬레이션 데이터	파생데이터	참조데이터	조사데이터
정의	실험장비에서 생산되는 데이터로 많은 비용 소요	관측장비를 통해 실시간 관측되어 생산되는 데이터	모델링을 통해 생산된 데이터	원천데이터로부터 재생산된 데이터	평가를 거쳐 신뢰성이 공인된 데이터 (다른 데이터의 기준이 되는 데이터)	설문조사 등을 통해 생산된 데이터
예시	가속기, 화학/바이오 실험 장비 등	망원경, 전자현미경, 인공위성 등	기후모델링, 경제전망 모델링	텍스트마이닝, 3D모델링	유전자배열, 플라즈마 물성 표준, 뇌MRI영상 등 참조 표준데이터	시장조사, 연구과제정보, 예측조사

형태	수치(numerical)	공간(spatial)	도표(graphical)	문서(text)

7) 국가과학기술심의회(2018), 연구데이터 공유·활용 전략(안)

□ (주요 국가별 정의) 주요 선진국들은 연구데이터를 아래와 같이 정의함

- (OECD) 수치, 문자, 이미지나 음성 등의 사실적 데이터이며, 주로 과학 연구의 주요한 원천으로 사용하고, 과학분야 학술자 사이에서 연구결과를 검증하기 위해 필수적인 것⁸⁾
- (미국) 연구 전 주기과정에서 발생하는 데이터(미국정보표준화기구(NISO))⁹⁾
 - 연구계획 단계의 연구데이터는 단순히 연구방법을 보충하는 것이 아닌 연구자가 수집할 정보와 수집 방법, 정보 처리, 분석 계획 등을 포함¹⁰⁾
- (영국) 연구데이터 관리에 관한 지침서를 통해 데이터 관리차원에서 연구의 전 주기에서 생성되는 데이터(영국정보시스템합동위원회(JISC))¹¹⁾
- (독일) 연구를 수행하는 과정에서 기초·원천연구, 실험, 측정, 조사 또는 설문을 통해 생성되는 데이터로서 디지털 또는 전자데이터로 저장할 수 있는 데이터(독일연구재단(DFG))
- (호주) 사실, 관찰, 이미지, 컴퓨터 프로그램 결과, 기록, 측량 또는 (논거, 이론, 테스트 또는 가설 또는 기타 연구물에 기초한) 경험의 형태에서 생성되는 데이터(호주국가데이터서비스(ANDS))

□ 국외 주요 국가들은 연구데이터를 연구 과정 중에 발생되고, 수치, 문자 등 기록 가능한 데이터로 정의 중

<표 1-16> 국가별 연구데이터 유형 및 범위 비교

구분	OECD	미국	영국	독일	호주
연구데이터 형태	수치, 문자, 이미지나 음성 등의 사실적 데이터	인정된, 기록이 된 사실적 자료 (예비분석, 초안 등 제외)	—	디지털 또는 전자데이터로 저장할 수 있는 데이터	수치, 기술, 시각, 청각
연구데이터 범위	—	연구의 전 주기	연구의 전 주기	연구를 수행하는 과정	경험의 형태에서 생성

8) Dirk Pilat, Yukiko Fukasaku, OECD(2007), OECD principles and guidelines for access to research data from public funding

9) NISO, Research Data Management, 2015.01.01

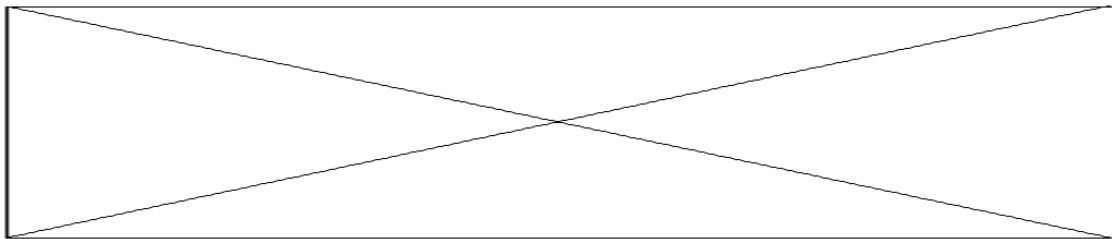
10) NISO, Research Data Management, 2015.01.01

11) JISC, How and why you should manage your research data, 2016.01.07

1.3.2. 연구데이터 공유·활용 체제의 정의

□ 국가 연구데이터 플랫폼

- **(국가 연구데이터 플랫폼)** 연구데이터 공유·활용체계 전반의 연구데이터를 원스탑으로 검색하고 활용할 수 있는 국가연구데이터 플랫폼 구축·운영
 - ※ 해외 데이터 공유·활용 플랫폼과 연계, 해외 데이터 리포지터리와 상호운용성 확보
- 독립된 기능을 수행하는 대·소분야 전문센터의 데이터와 인프라 서비스를 유기적으로 통합하는 허브
- 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 클라우드 지원)** 인프라를 보유하고 있지 않은 대·소분야 전문센터 임무수행 지원을 위한 컴퓨팅·스토리지·네트워크 인프라 구축*
 - ※ 연구현장 수요 변화에 따른 탄력적 운용성 확보 등을 위해 민간 클라우드 적극 활용

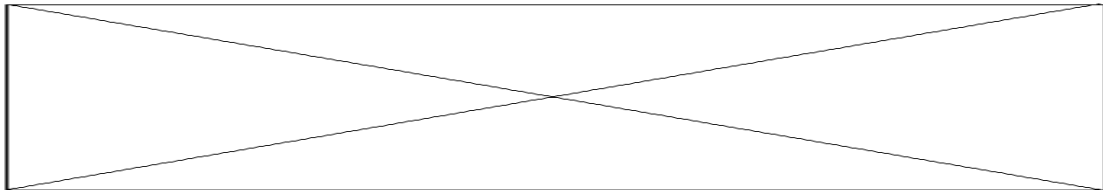


<그림 1-7> 국가연구데이터플랫폼 개념도

□ 대·소분야 전문센터

- **(대분야 전문센터)** 대분야(바이오, 소재 등) 또는 중분야에 따라 1개씩 두되, 여러 소분야 전문센터와 연계하여 시너지가 나도록 구성
- (역할) 대분야 내 필요 소분야 우선순위 제시 및 공통적으로 사용되는 데이터 관리 및 분석 SW 등 개발 담당
 - ※ 소분야 전문센터가 없는 분야의 개인·집단 과제 또는 대형연구개발사업에서 수집된 데이터의 1차 아카이빙(Archiving)도 대분야 전문센터가 담당

- \ ○ **(소분야 전문센터)** 연구데이터가 실질적으로 공유되어 활용됨으로써 새로운 부가가치를 창출할 수 있는 **소분야 연구 공동체** 인프라
 - － (역할) 연구데이터 활용 방향을 설정하고, 데이터 공유를 위한 표준화·큐레이팅, 분석 SW 개발 등 데이터 관리 책임 보유
 - － 소분야 연구 공동체(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



<그림 1-8> 대·소분야 전문센터 체계 개념도

연구데이터 공유·활용 체제 현황 분석

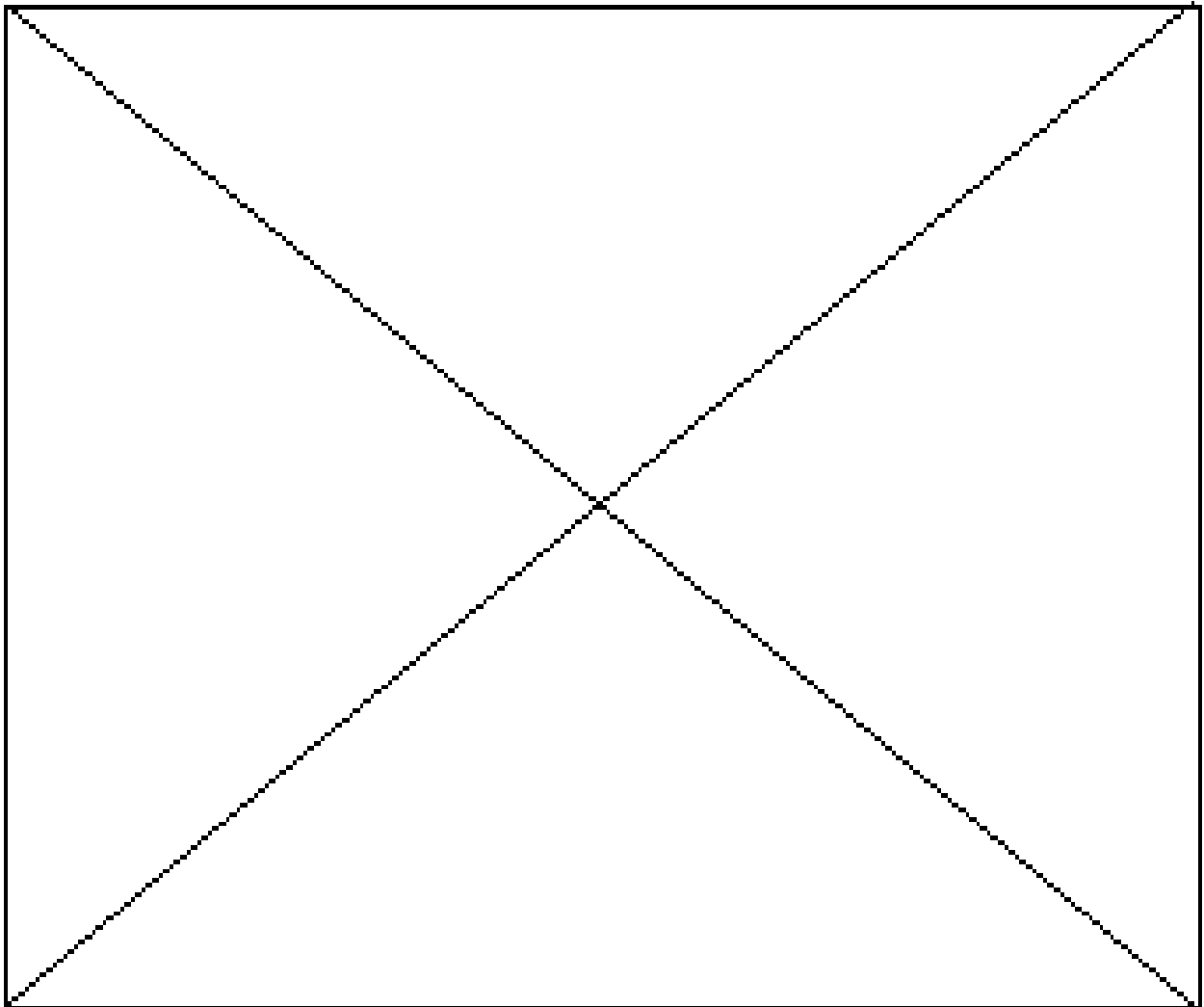
- 1.1. 추진 배경 및 필요성
- 1.2. 사업 추진 경과 및 체제
- 1.3. 연구데이터 공유·활용 체제의 정의

2.1. 국외 현황 분석



2.1.1. 미국

국가 차원 보다는 자율적인 데이터 관리 및 공유 시스템이 기관이나 분야별로 구축되어 국제규모의 분야별 연구 커뮤니티를 주도해오다가 최근 국제적인 흐름에 따라 연방정부 차원에서의 데이터 공유 정책을 수립하여 구현하기 시작함



<그림 2-1> 미국 연구데이터 거버넌스 현황(한국과학기술정보연구원, 2016)

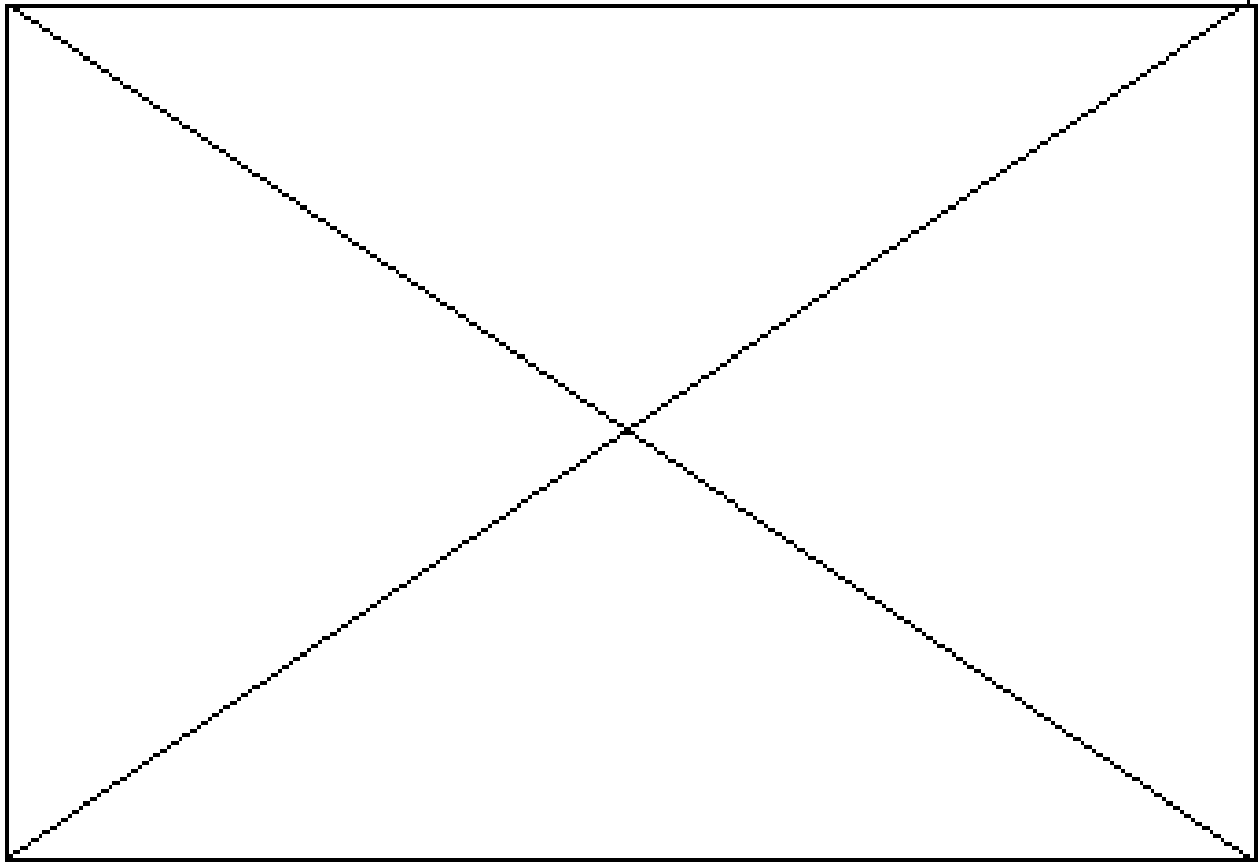
- (공유·활용 법제도) 정부자유법(1966), 미국경쟁력강화법(2010)을 통해 공공 연구결과물인 연구데이터의 공개와 접근을 위한 법적 근거 마련
 - (정보자유법) 정부가 소유하고 있는 정보에 대한 국민의 알권리 및 통제할 권리를 보장하며, 두 차례 개정을 통해 정보공개제도와 절차를 강화
 - 백악관 예산정책국(OMB)은 정보자유법에 근거하여 요청이 있는 경우 공공연구비 지원을 받아 연구를 수행한 기관에 데이터 공개를 의무화할 수 있다고 규정¹²⁾
 - (미국경쟁력강화법) 연구데이터에 대한 관리원칙을 천명하며, 연방 지원 연구 프로그램 성과물(논문 및 연구데이터)에 대한 개방적 접근(public access)과 이용에 대한 법적 근거 마련
 - 과학기술정책국(OSTP)이 중심이 되어 연구데이터를 수집하되, 연방연구 프로그램을 통해 생산되는 연구데이터의 구체적 생산, 관리, 공개 및 보존 등에 관한 사항은 미국과학재단(NSF)을 통해 규율
- (공유·활용 정책) 백악관 산하 과학기술정책국(OSTP)에서 연구관리기관의 연구데이터 공유 원칙 제시
 - ‘연방정부 지원을 통해 수행된 과학연구 결과의 접근 증대 지침(Memorandum 2013)’을 발표하여 연간 1억 달러 이상의 정부 연구개발비를 운용하는 연방기관들에게 공공 연구결과물의 공적접근계획(Public Access Plan) 제출 의무화
 - 공적접근계획의 대상인 공공연구 결과물은 연구출판물과 디지털 형태의 과학데이터이며 데이터관리계획(DMP) 제출 조항이 필수적으로 포함되어야 함(OSTP, 2013)
 - 해당 연방기관들은 공적접근계획을 통해 데이터관리계획(Data Management Plan, DMP) 제출을 의무화하고 DMP 타당성 및 이행여부 평가결과를 과제선정 및 연구비 지급 여부에 반영
 - NSF는 Peer Review를 통해 DMP 타당성 평가 수행, NOAA/NASA 등은 DMP 불이행 시 연구비가 지급되지 않을 수 있으며, 차기 과제 선정에 반영
 - 주요 연구관리기관의 연구데이터 공유·활용 정책에는 DMP 제출 이외에도 데이터 소유권, 품질, 정보보호, 보상, 출판 및 엠바고, DMP 평가 등에 대한 규정 제시

12) 백악관 예산정책국(1999). The Office of Management and Budget (OMB) Circular A-110

<표 2-1> 미국 주요 연구관리기관별 데이터 공유·활용 정책

주요 정책	국립과학재단 (NSF)	국립보건원 (NIH)	국립항공우주국(NASA)	에너지부 (DOE)	국립해양대기청(NOAA)
DMP	○	○	○	○	○
소유권	○	○	○		
품질/문서화	○	○	○		○
공유예외 기준/정보보호	○	○	○	○	○
공유시기	○	○	○	○	○
보상	○	○	○		
DMP 평가	○		○		○

- (공유·활용 추진체계) 과학기술정책국(OSTP), 예산정책국(OMB) 등이 연방 차원의 법제도 및 원칙을 수립하고, 이를 기반으로 연구관리기관들이 세부적인 과학데이터 공유·활용 계획 수립
- (정책추진조직) 연간 1억 달러 이상 연구개발비를 운용하는 모든 연방 기관*이 연구데이터 공유·활용 추진 조직 역할 수행
 - 국립과학재단(NSF), 국립보건원(NIH), 에너지부(DOE) 등 연구관리기관들은 분야별 특성을 반영한 연구데이터 공유·활용 계획 수립 및 실행
 - 연방 기관(Federal Agency)은 정부부처 산하 주요 연구관리기관(예, NIH, NOAA 등) 및 독립 연구관리기관(예, NSF, NASA 등)과 연구개발 미션을 가진 연방부처(예, DOD, DOE, DOC 등)로 구성
 - (관리/활용 지원조직) 대학 도서관이 대학 소속 연구자들의 DMP 제출 및 데이터 관리/공유를 지원하는 역할 수행
 - 연구관리기관별 다양한 데이터 공유 및 관리 지침으로 인한 연구자들의 어려움을 해결하기 위해 대학 도서관 내 전담부서를 설치하고 DMP 관련 컨설팅, 데이터 리포지터리 서비스, 데이터 관리 및 공유 관련 교육 제공

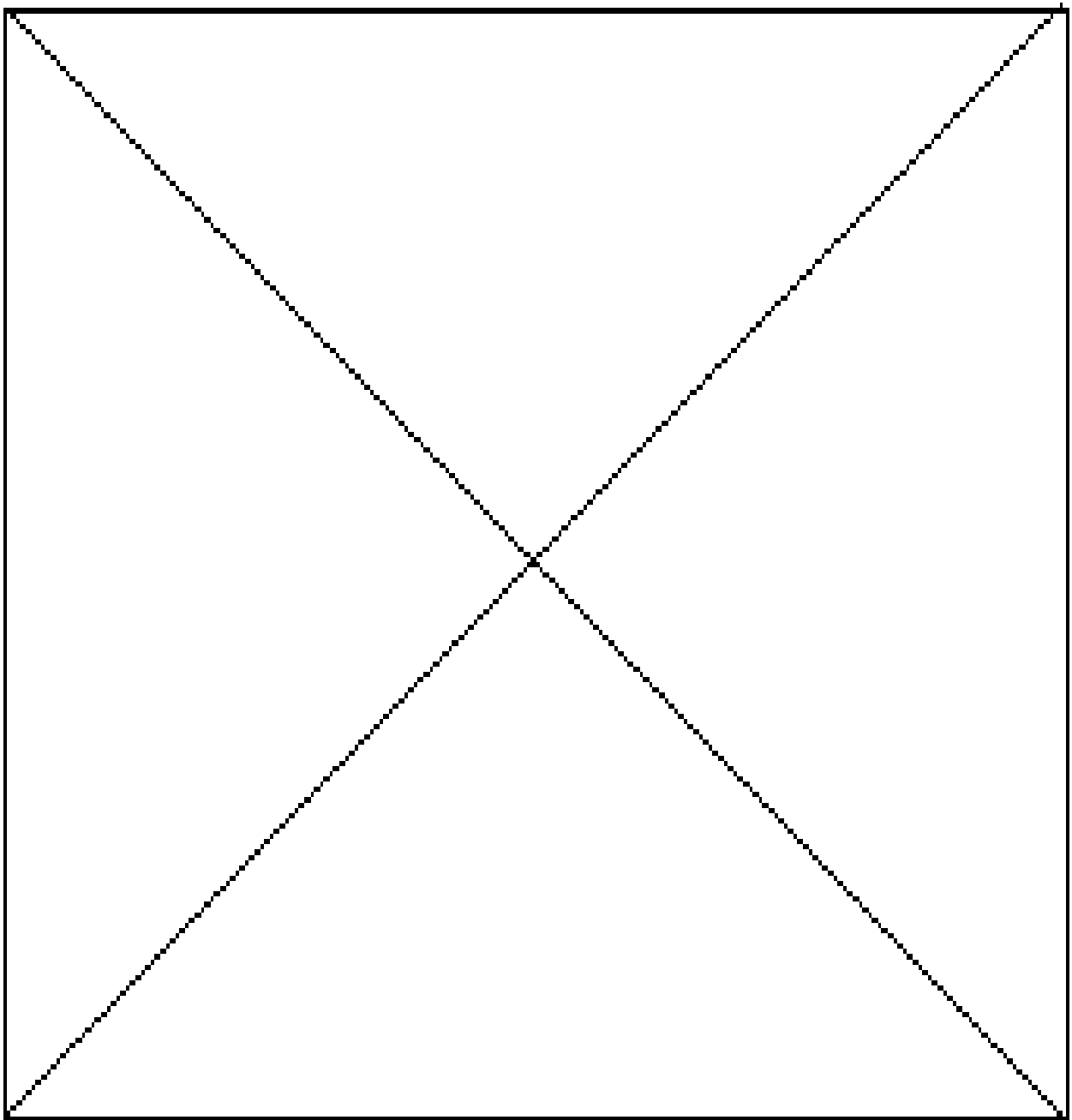


<그림 2-2> 미국 연구기관 및 대학의 연구데이터 공유·활용 추진체계

- (공유·활용 인프라) 주요 연방기관 및 대학에서 연구데이터 관리 및 공유를 위한 데이터 센터와 리포지토리를 운영하고 있으나 각 시스템을 총괄하거나 연계하는 정책과 체계는 부재
 - 각 전문분야 국가 연구기관(NASA-우주항공, NOAA-해양, NIH-바이오 등)에서 데이터 센터*를 구축하여 내·외부 연구과제에서 생성되는 데이터 수집, 관리 및 공개하고 있으나 분야별 연구데이터를 연계하는 공통 플랫폼은 부재
 - NOAA소속 해양데이터센터(National Oceanographic Data Center, NODC), NASA 우주과학데이터 공동 아카이브(NASA Space Science Data Coordinated Archive, NSSDCA), NIH 소속 국가바이오정보센터(National Center for Biotechnology Information, NCBI) 등
 - 다수 연구기관 및 대학에서는 데이터 리포지토리를 통해 연구데이터의 보존과 공유 지원
 - 하버드(Harvard Dataverse), 스탠포드(Stanford Digital Repository), 위스콘신 메디슨(MINDS@UW) 등

2.1.2. 영국

영국의 연구데이터는 7개의 분야별 연구회를 중심으로 자율적으로 운영되며 이들의 협의체인 UKRI의 데이터정책공동원칙(2011)과 EPSRC의 Publication Scheme을 기반으로 국가 데이터관리 정책 수립



<그림 2-3> 영국 연구데이터 거버넌스 현황

- (공유·활용 법제도) 정보자유법은 예외사항을 제외한 모든 공공정보 공개가 원칙이며 국가연구기관 및 대학에서 보유한 정보를 재사용 가능한 구체적 데이터셋 형태로 공개하기 위한 법적 근거 제시
 - 정보자유법(2000)은 영국연구혁신청(UKRI)를 포함하여 정부 지원을 받는 대학과 연구소, 민간 기업 등이 적용 대상이며 각 기관이 보유한 연구데이터도 공개 대상
 - 개정(2013)을 통해 재사용 가능한 형태의 데이터셋 제공과 제3자가 사용할 수 있는 라이선스 허용 조항 포함(ICO, 2015)
- (공유·활용 정책) 데이터 정책에 관한 공동원칙을 기반으로 분야별 연구위원회에서 세부적인 규정 제시
 - 영국연구혁신청(UKRI)는 2011년 데이터 정책에 관한 7가지 공동 원칙(Common Principles on Data Policy)을 발표하여 분야별 데이터 정책의 상위 프레임워크 제공¹³⁾

<표 2-2> 영국연구혁신청의 데이터 정책에 관한 7가지 공동원칙

구분	내용
원칙 1	공적자금으로 수행된 연구데이터는 최소한의 제한으로 적시에 활용 가능
원칙 2	연구특성을 반영한 데이터 관리정책과 계획 시행 및 장기적 가치를 지닌 데이터의 보존과 접근성 유지
원칙 3	타 연구자들의 데이터 활용을 위한 충분한 메타데이터 공개
원칙 4	연구데이터 품질을 위해 모든 연구단계에서 관련 정책 고려
원칙 5	연구자가 일정 기간 독점적으로 데이터 사용
원칙 6	데이터 출처 표기 및 이용조건 준응
원칙 7	데이터 공유를 위한 공적자금 사용

- 공동원칙을 기반으로 각 연구위원회가 독립적으로 분야의 특성을 반영한 세부적인 데이터 관리 및 공유 정책, 큐레이션 규정 등을 제시¹⁴⁾
 - 연구비 지원기관의 대표적인 데이터 관리 계획(DMP) 의무제출 규정은 7개 연구위원회 중 6개가 시행 중
 - 대부분의 연구관리기관은 연구결과에 대한 제한 없이 접근 가능한 오픈액세스를 추구하며 이를 위해 리포지터리, 데이터센터 서비스 제공

13) RCUK 홈페이지, www.rcuk.ac.uk

14) Jones, S. (2009). A report on the range of policies required for and related to digital curation, version 1.2, (DCC, Glasgow)

<표 2-3> 영국 연구위원회별 데이터 공유·활용 정책

주요 정책		예술인문 AHRC	생명과학 BBSRC	공학물리 EPSRC	경제사회 ESRC	의학 MRC	자연환경 NERC	연구장비 STFC
정책범 위	출판물	○	○	○	○	○	○	○
	데이터	○	○		○	○	○	
큐레이 션규정	기한	○	○	○	○	○	○	○
	DMP	○	○		○	○	○	○
	접근/공유	○	○	○	○	○	○	○
	장기큐레이션	○	○		○	○	○	
	모니터링		○		○		○	
지원	가이드	○	○		○	○	○	
	리포지터리		○		○	○	○	○
	데이터센터	○	○	○	○	○	○	○

□ (공유·활용 추진체계) 분야별 연구위원회가 공공 연구과제에서 생성된 데이터와 성과물을 관리 및 공유하고 있으며, 연구자를 지원하기 위한 디지털 큐레이션 센터 운영

- (정책추진조직) 7개 분야별 연구위원회가 독립적으로 R&D 연구수행 및 지원, 산출물 관리 및 공유정책을 수립 및 추진하며, 각 연구위원회의 전략적 파트너십과 공동활동을 촉진하는 협의체로 영국연구위원회(RCUK) 운영

－ 영국연구혁신청(UKRI)는 분야별 연구위원회의 협의체로 전체 연구위원회의 최고 정책 결정기구로서의 역할을 하며 분야별 연구위원회의 공동 활동 촉진 활동을 수행

- (관리/활용 지원조직) 디지털 큐레이션 센터(Digital Curation Center, DCC)에서 DMP 컨설팅, 데이터 관리자 교육 등 정책지원 수행, 정보시스템합동위원회(Joint Information System Committee, JISC)에서 연구데이터 공유 인프라 지원

□ (공유·활용 인프라) 정부 지원을 통해 수행된 연구의 다양한 산출물을 수집·공유하는 국가 차원의 통합 플랫폼(ResearchFish)을 운영하고 있으며 개별 연구위원회 및 대학도 데이터 공유를 위한 데이터 센터와 리포지터리 운영

- ResearchFish는 연구관리기관, 연구기관, 연구자를 대상으로 연구결과물 수집 및 공유 서비스를 통해 공동연구의 기회를 제공하며 연구결과물 등록 채널 일원화로 데이터 공유·활용을 위한 연구자들의 추가적인 노력 감소

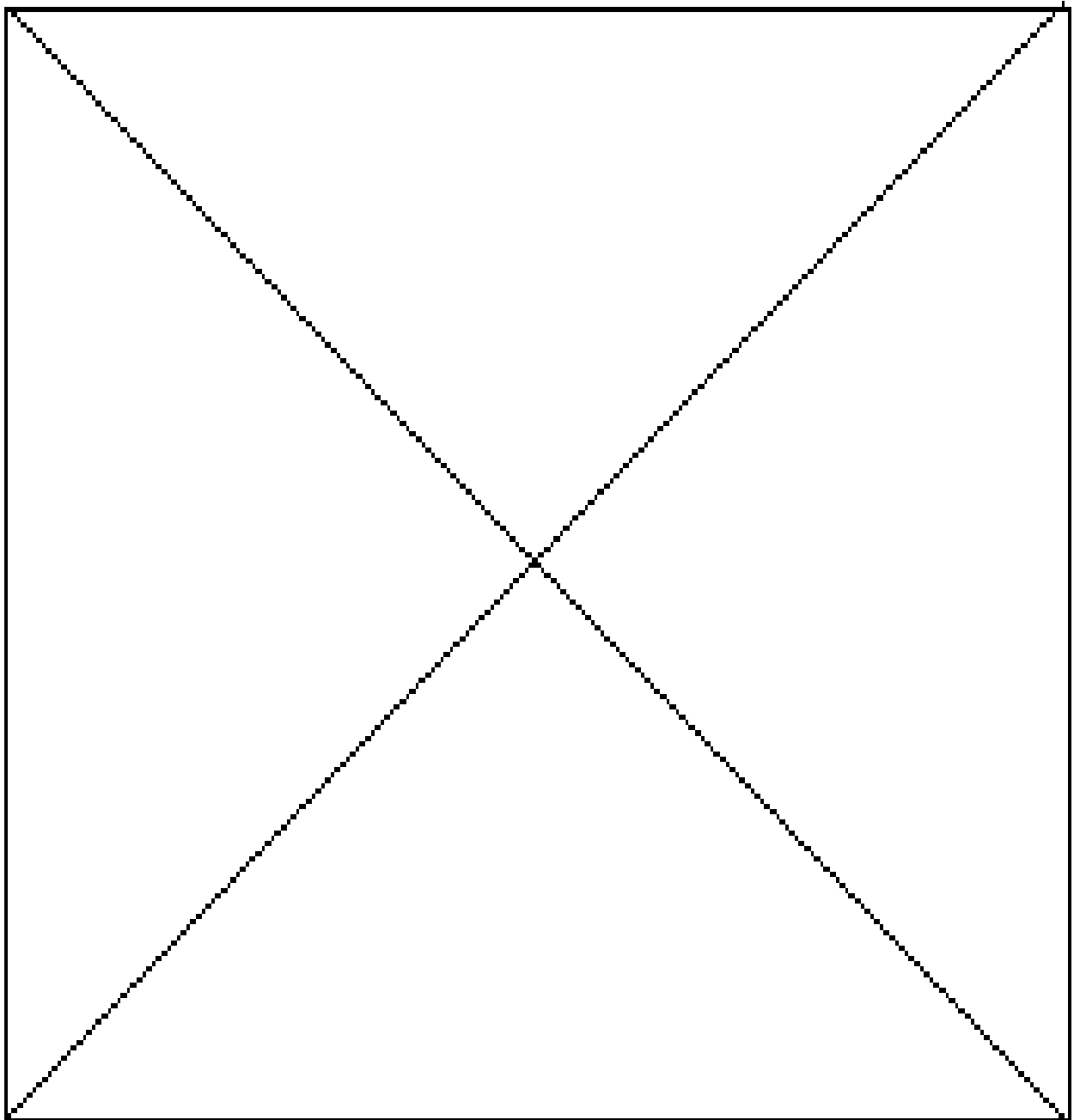
－ ResearchFish는 연구에서 발생한 모든 산출물(최종보고서, 논문, 데이터 등)을 수집하는 플랫폼으로 현재 약 200만개 이상의 연구 성과물 보유¹⁵⁾

- \ ○ DCC에서 제공하는 온라인 DMP 작성 도구인 DMPOnline은 영국의 다양한 연구관리기관 및 연구 커뮤니티의 데이터 정책을 분석하여 연구자들이 다양한 요구에 대응할 수 있도록 체크리스트 등의 서비스 제공
- JISC는 JISC 공유데이터센터를 구축(2014)하여 연구커뮤니티 및 대학의 연구데이터 공유를 지원하는 클라우드 기반 인프라 제공
 - JISC 공유데이터센터를 통해 이용자들은 멀리 떨어져 있는 데이터 스토리지와 서비스에 접근할 수 있으며 대량의 연구데이터 공유 가능
- 정부와 연구위원회의 데이터 공유 정책을 준수하도록 연구자를 지원하기 위해 분야별 연구위원회 및 대학에서 개별 데이터 센터 및 리포지터리 운영
 - 7개 연구위원회 중 NERC(지구해양환경 분야)와 ESRC(경제사회 분야)는 직접 데이터 센터를 운영하고, SFRC(연구시설·장비 분야)는 EU 및 글로벌 데이터 리포지터리와 협업하여 세계적으로 데이터 공유
 - NERC-국가 지구과학 데이터 센터(National Geoscience Data Centre, NGDC), 환경정보 데이터 센터(Environmental Information Data Centre, EIDC)/ESRC-UK Data Archive 등
 - 캠브리지 대학-Apollo, 옥스퍼드 대학-ORA-Data, 셰필드 할람 대학-SHURDA 등

15) ResearchFish 홈페이지, www.researchfish.net

2.1.3. 호주

정보자유법과 책임 있는 연구수행을 위한 국가지침(the Code)을 중심으로 한 제도적 기반 위에서 호주국가데이터서비스(ANDS)를 통해 국가 차원의 연구데이터 관리 및 공유를 지원하고 대학 기관별 지원은 도서관을 중심으로 이루어짐



<그림 2-4> 호주 연구데이터 거버넌스 현황

□ (공유·활용 법제도) 정보자유법(1982)을 통해 공공정보에 대한 접근 보장

- (정보자유법) 개인이 정부 정책 관련 정보에 대한 접근권을 보장하는 법으로 일반적으로 정부기관 문서를 대상으로 하지만 정부주도 과학연구 사업 데이터 공유에 법적 기반 제공

□ (공유·활용 정책) ‘책임 있는 연구수행을 위한 국가 지침(이하 the Code, 2007)’을 통해 연구에 대한 책임과 진실성 증진을 위해 연구단계별 우수사례 제공 및 연구데이터 생산, 관리, 공유 원칙 제시

- (the Code) 모든 연구 활동으로부터 생산되는 사실, 관찰, 이미지, 컴퓨터 프로그램 기록 등 모든 데이터와 자료를 대상으로 체계적인 보존과 관리 의무 부과
 - the Code는 대표 연구관리기관인 호주연구위원회(ARC)와 보건의료연구위원회(NHMRC)의 협업으로 개발되어 ARC와 호주대학연합의 모든 연구프로그램의 필수 규정
- 호주연구위원회(ARC)는 2014년부터 ARC의 연구비를 지원 받는 연구기관에 대하여 의무적으로 DMP를 제출할 것을 규정하였으며 DMP 세부 지침 및 템플릿 제공 관련 책임은 개별 연구기관 및 대학에 부여

□ (공유·활용 추진체계) 연구관리기관이 제시한 지침을 기반으로 개별 연구수행기관들이 자체 규정을 수립 및 실행하고 호주대학연합(UA)과 호주국가데이터서비스(ANDS)가 연구데이터 관리 및 공유 정책 준수를 위한 지원 활동 수행

- (정책추진조직) 호주연구위원회(ARC)와 보건의료연구위원회(NHMRC)가 연구결과물에 대한 관리와 공유 방향을 제시하는 기본 지침(the Code)을 수립하고, 연구기관 및 대학과 같은 연구수행기관이 연구데이터 관리 및 공유에 관한 세부 규정을 자체적으로 수립하여 시행
- (관리/활용 지원조직) 데이터 관리 및 공유 인프라를 지원하기 위해 호주국가데이터서비스(Australian National Data Service, ANDS)를 운영하고 호주대학연합은 각 대학의 연구데이터 정책 수립 및 지원 수행

□ (공유·활용 인프라) 국가차원에서 연구데이터의 수집, 공유, 관리 등 전주기에 대한 인프라가 체계적으로 구축되어 있으며, 대학에서도 개별적인 플랫폼 운영

- Data.gov.au는 호주연방정부 및 주정부가 보유한 공공 및 연구데이터에 대한 검색, 접근 및 다운로드, 재사용 서비스 제공

- Data.gov.au는 약 50개 정부 기관의 데이터 정보를 제공하며 주로 국가 소유의 연구데이터(환경 및 지리정보, 기상데이터 등)를 확인할 수 있으며 데이터셋을 직접 다운로드할 수 있도록 기능 제공¹⁶⁾
- ANDS의 Research Data Australia는 시스템에 모든 데이터를 등록·보관·관리하는 형태가 아닌 개별 연구기관 및 대학에서 관리하는 연구데이터 플랫폼 정보 및 개별 연구에 대한 메타데이터 제공 서비스가 중심이며 데이터 소유 연구기관과의 연계를 통해 데이터 설명이나 링크 등 제공
- ANDS와 연구기관과의 연계를 통해 연구기관에서 생산하는 데이터가 관리된 형태로 전환되고 데이터 간 연계를 가능하게 할 뿐만 아니라 데이터의 재사용도 가능
- CSIRO Data Access Portal은 연방 과학산업 연구기관인 CSIRO에서 운영하는 연구데이터 관리 포털로 보유한 분야별 연구결과물의 공유 및 재사용이 목적이며 공유가 허용되지 않는 데이터의 경우 관련 메타데이터 제공
- 대학에서 운영하고 있는 연구데이터 공유플랫폼으로 Monash University Data Repository, ANU(Australian National University) Data Commons, UQ(University of Queensland) eSpace 등이 있음

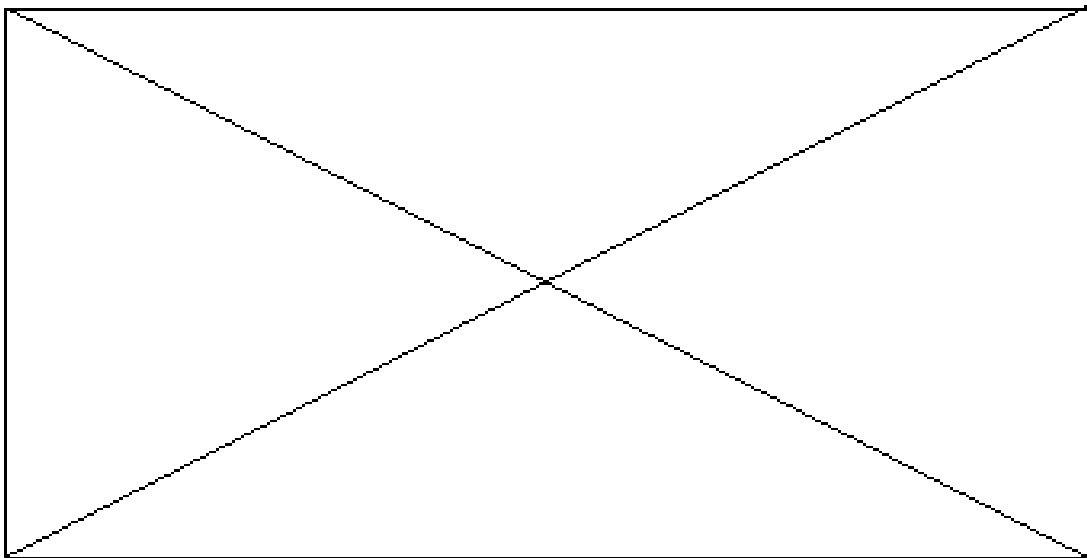
16) Data.gov.au 홈페이지, <http://data.gov.au>

2.1.4. 중국

2001년부터 장기적인 계획 하에 2020년까지 전반적인 연구데이터 공유·활용체제의 완성을 목표로 연구데이터 전반을 규율하는 법체계 수립과 인프라 구축을 추진 중임

□ (공유·활용 법제도) 연구데이터 전반을 규율하는 법체계 완성을 목표로 입법 추진 중

- ‘중화인민공화국 과학기술 진보법’과 ‘중화인민공화국 과학기술성과전환 촉진법’에서 연구데이터를 포함한 연구성과 관리 및 공유 인프라 구축 규정 보유
- ‘국가과학기술계획 항목 과학데이터제출입시법’에서 연구데이터 종류와 범위, 데이터제출 계획(DMP), 제출의무자의 권리와 의무 등 연구데이터 공유와 관련된 전반 내용 규율¹⁷⁾
 - － 과학데이터 공유에 대한 법적 근거 마련 및 구체화를 위해 과학데이터 공유법률(안)과 하위 세부 행정법규인 과학데이터공유조례(안)을 통해 법률 구체화
 - － 과학데이터 공유 법체계 구조는 상위 법률에 의거하여 행정법규인 조례에서 데이터 생산, 제출, 보관, 이용 및 관리로 세분화하여 전반의 내용 규율



<그림 2-5> 중국의 과학데이터 공유 법체계 구조

□ (공유·활용 정책) 연구데이터 공유 사업을 국가혁신 및 국가연구개발사업의 중요 부분으

17) 윤종빈, 김규빈(2011). 과학데이터에 관한 입법례와 관리정책 그리고 대응방안 : 호주, 미국, 중국을 중심으로. 한국기술혁신학회지. 161(1). 87.

로 인식하여, 2001년부터 장기적 계획 하에 기반 환경 구축

- 2001년 제10차 5개년 계획에 ‘국가과학기술 기초조건 인프라 건설’을 포함시켜 처음으로 국가정책화 하였고 2011년 제12차 5개년 계획에 ‘기초연구와 선도기술연구 우선추진’을 포함시켜 오픈사이언스 정책 추진
- 과학기술부(MOST) 주도로 국가차원의 과학공유 기반시설 및 데이터공유사업(China Scientific Infrastructure, Science Data Sharing Projects, SDSP)이 3단계에 걸쳐 체계적으로 추진되고 있으며 2020년에는 전반적 공유체계가 완성될 예정
 - 1단계(‘02-‘05)에서는 국가 차원의 시범사업 운영
 - 2단계(‘06-‘10)에서는 연구데이터 공유 법률 제정 및 데이터 공유 기반시설(6개 연구 분야*별 데이터센터) 확충
 - * 인구 및 보건, 지구과학, 산림, 기상 및 대기과학, 농업, 지진과학
 - 3단계(‘11-‘20)에서는 오픈 액세스 활성화, 연구데이터 평가제도 및 관리기구 설치

□ (공유·활용 추진체계) 2002년 이후 과학기술부(MOST)가 연구데이터 공유 사업을 추진하고 있으며 환경, 농업, 인구, 보건 등 24개 관련 부처가 참여

- (정책추진조직) 과기부(MOST) 주도로 2002년 이후 18개 연구데이터 공유 사업 추진
- (관리/활용 지원조직) 중국과학원(CAS)은 연구데이터위원회(China Committee of Codata)와 중국데이터센터(World Data Center China Centers)를 설립하여 연구데이터 공유·활용 지원
 - 중국과학원은 과학데이터공유공정*을 통해 과학데이터를 통합적이고 체계적으로 관리·활용하기 위한 데이터베이스, 데이터센터, 관련 포털 사이트 구축
 - * 과학데이터공유공정은 중국과학원이 대형 컴퓨터 개발 공정으로 중국과학원 산하 연구소 연구자들이 쉽게 연구에 필요한 각종 데이터를 공유할 수 있도록 구축한 플랫폼 공정

□ (공유·활용 인프라) 중국과학원(CAS)은 기관 리포지터리 네트워크 시스템인 CAS IR GRID를 구축 및 운영하며 데이터를 포함한 연구 성과의 접근성 및 활용성 제고¹⁸⁾

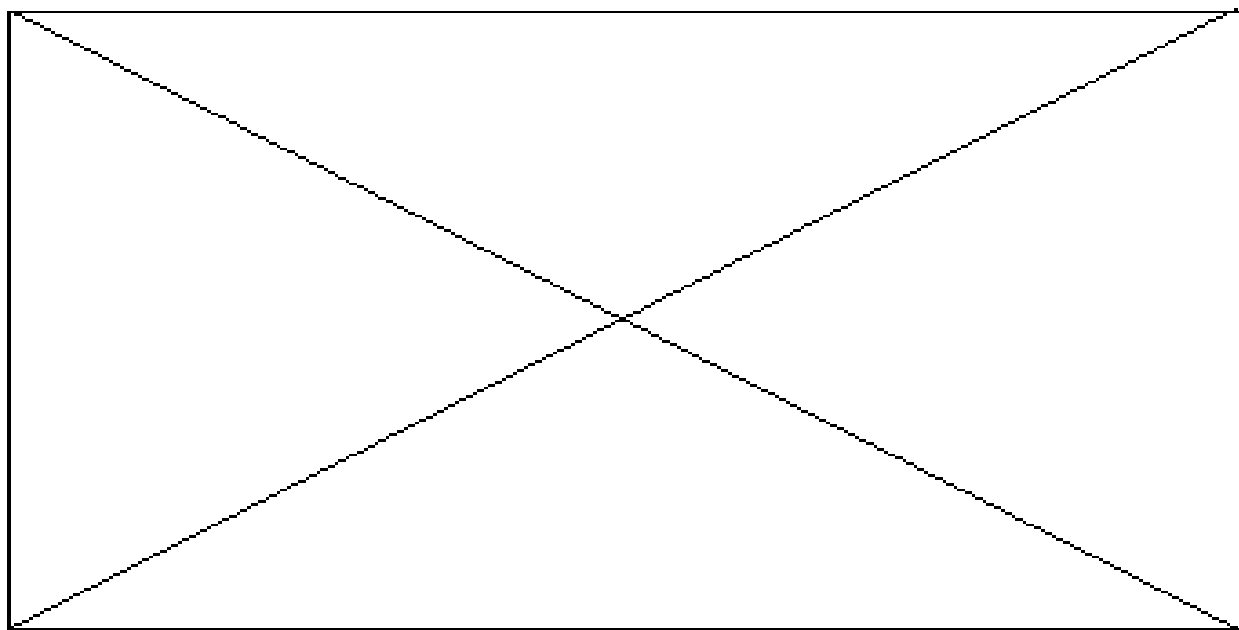
- (CAS IR GRID) 연구 성과 기탁 지원, 연구 성과의 배포 및 권한관리, 저작권, 검색 서비스, 통계서비스, 연구자 프로필 등 서비스 제공
 - 112개 리포지터리가 통합되어 있으며 CAS IR GRID를 통해 공개된 콘텐츠는 논문, 특허, 단행본, 학위논문, 연구보고서, 데이터셋 등을 포함하여 약 26만 건

18) CAS IR GRID 홈페이지. <http://www.irgrid.ac.cn/>

- \ ○ (데이터센터) 6개 분야(삼림, 지구과학, 인구 및 보건, 농업, 기상, 지구과학)별 데이터 시스템을 구축해 과학기술 데이터의 수집·관리·공유 기반으로 활용

2.1.5. 일본

일본은 현재 문부과학성 산하 과학기술정보위원회와 학술정보위원회를 중심으로 오픈사이언스 관련 정책 수립을 위해 논의 중임



<그림 2-6> 일본 연구데이터 거버넌스 현황

- (공유·활용 법제도) 국가차원의 연구데이터 공유·활용을 규정하는 법률은 부재
- (공유·활용 정책) 내각부와 종합과학기술혁신회의의 주도로 오픈 사이언스를 실현하기 위한 계획과 정책을 수립하고자 논의하고 있는 단계
 - 문부과학성은 연구비 지원기관에게 자율적으로 공적자금으로 진행된 연구 성과를 5차 과학기술기본계획 기간(2020년) 내에 공개하는 계획을 수립하도록 요구
 - － 과학기술기본계획에서 연구기관의 기관 리포지토리를 구축하고 논문, 실험 데이터 등을 전자화하여 오픈 액세스를 추진하는 방안을 발표하고 진행 중
- (공유·활용 추진체계) 문부과학성 산하 과학기술정보위원회와 학술정보위원회에서 연구성과물 및 데이터 공유에 관한 논의가 진행 중이며, DOI 및 리포지토리 관련 시험 활동 수행 중

- \ ○ 과학기술진흥기구(JST)는 연구데이터 공유를 위해 과학기술정보위원회를 설치하여 연구데이터 공유 관련 정책, 인프라, 인력양성, 성과평가, 연구 커뮤니티 지원 방안 등 논의('14)
 - 과학기술정보위원회는 문부과학성 산하 독립행정법인으로 학계, 산업계 등 연구개발과 관련된 전반적인 프로젝트 관리
- 학술정보위원회는 2015년부터 공적 연구 자금으로 진행된 논문 및 연구데이터 등 학술연구 성과 공개와 데이터 관리계획 제출 등을 포함한 학술정보 공개 추진에 관한 논의 진행
 - 문부과학성 과학기술학술심의회 내 학술정보위원회는 연구기관과 대학 등 학술 관리를 전담으로 하는 정부부처
- DOI 등록기관인 재팬링크센터(JaLC)는 DOI 등록 구조를 구축 및 관련 정책 수립을 목표로 DOI 액세스의 지속성 보장, DOI 활용방법 등 연구자들의 다양한 요구사항을 검토하기 위한 연구데이터에 DOI를 부여하는 시험 프로젝트 진행 중

2.1.6. 분석 및 시사점

□ (국가별 연구데이터 관리체계 현황 종합 분석) 한국과 일본은 연구데이터 공유·활용체제 구축에 있어서 미국, 영국, 호주, 중국 등에 비해 상당히 뒤쳐진 상태임

<표 2-4> 국가별 연구데이터 관리체계 현황

국가			미국	영국	호주	중국	일본	한국
법·제도	기본법		●	●	●	◐	○	○
	적용 범위	출판물	●	●	●	●	○	●
		데이터	●	●	◐	●	○	◐
정책	DMP		●	◐	◐	●	○	○
	엠바고		●	◐	◐	●	○	○
	공유		◐	◐	◐	●	○	◐
	소유권		●	●	◐	●	○	◐
	보상		◐	○	◐	◐	○	○
	보안		●	◐	◐	●	○	◐
	개인정보보호		●	○	◐	●	○	●
	공유시기		●	●	◐	●	○	○
	품질		◐	◐	◐	●	○	○
추진 체계	R&D 전담조직		●	●	●	●	●	●
	데이터관리 지원조직		●	●	●	◐	○	○
인프라	정부지원 리포지터리		●	●	◐	●	○	◐
	대학지원 리포지터리		●	●	◐	●	○	○

□ (시사점) 국내외 현황을 고려할 때, 한국은 국가주도의 연구데이터 공유·활용체제의 구축을 시급히 추진할 필요가 있음

- 기존의 연구 환경을 바탕으로 국가 중심의 연구데이터 공유·활용체제 구축을 위한 효과적인 모델로서 영국과 호주의 사례를 벤치마킹 대상으로 삼을 수 있을 것으로 판단됨
- 호주의 연구데이터 공유·활용체제는 정책의 수립과 인프라 구축이 균형적으로 추진된 모델이며 ANDS는 한국의 연구데이터 인프라 및 서비스 구축 계획 시 참고할만한 가장 이상적인 모델로 판단됨

2.2. 국내 현황 분석



2.2.1. 정책적 현황

- 우리나라는 공공기관이 보유하고 있는 공공데이터와 일부 연구데이터에 대한 공유·활용 제도가 마련되어 있으나, 연구데이터를 국가연구개발 사업의 결과물로 인정하고 국가차원에서 공유·활용하기 위한 법적 기반 부재
 - 「과학기술기본법 제11조의3」의 연구성과물의 소유권 규정에서는 국가연구개발사업의 성과를 연구형태와 비중, 개발성과 유형 등을 고려하여 대통령령으로 정하는 바에 따라 연구기관 등의 소유로 함
 - 「국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률 제2조」는 연구성과를 ‘연구개발을 통하여 창출되는 특허·논문 등 과학기술적 성과와 그 밖에 유·무형의 경제·사회·문화적 성과’를 모두 말하는 것으로 널리 규정 중
 - 「국가연구개발사업의 관리 등에 관한 규정 제3장」은 국가연구개발사업을 통하여 창출된 연구성과 중에서 연구개발결과의 등록대상과 기탁대상을 정의함

<표 2-5> 국가연구개발사업 연구성과의 등록 및 기탁 범위

등록대상		기탁대상	
논문	국내외 학술단체에서 발간하는 학술지 및 학술대회지에 수록된 학술논문	생물자원 (생명자원)	미생물자원(세균, 곰팡이, 바이러스 등), 동물자원(사람·동물세포, 수정란 등), 식물자원(식물세포, 종자 등), 유전체자원(DNA, RNA, 플라스미드 등) 및 관련 정보
특허	국내외에 출원 또는 등록된 특허정보		
보고서 원문	연구개발 종료 시 제출하는 최종보고서 및 연차보고서(전자자원문 포함)		
연구시설·장비	국가연구개발사업을 통하여 취득한 3천만원 이상 연구시설·장비 또는 3천만원 미만이라도 공동활용이 가능한 연구시설·장비		
기술요약 정보	기초·응용개발단계 등의 최종보고 및 연차보고가 완료된 결과물의 기술정보를 요약하여 공유·활용(기술이전·사업화 등)할 수 있도록 작성된 기록정보	화합물	합성 또는 천연물에서 추출한 유기화합물 및 관련 정보
생명정보 (생명자원)	유전체정보(서열, 발현정보 등), 단백질체정보(서열, 구조, 상호작용 등), 발현체정보(유전자(DNA)칩, 단백질칩 등) 및 그 밖의 관련 정보		
소프트웨어	창작된 소프트웨어 및 등록에 필요한 관련 정보		

□ (국가과학기술 기본계획) 연구데이터 공유·활용을 위한 국가 차원의 중장기 계획 및 종합 대책을 수립하는 초기 단계

- 「제3차 연구성과 관리·활용 기본계획」에서 국가연구개발사업의 성과를 연구 종료 시의 성과물에서 연구개발 전주기 정보로 성과 공개 범위 확대¹⁹⁾

<표 2-6> 「연구성과 관리·활용 기본계획」 내 연구성과물 공개정책 변화 과정

	제1차 기본계획('06~'10)	제2차 기본계획('11~'15)	제3차 기본계획('16~'20)(안)
핵심 전략	(전략 1) 연구 성과 종합관리 체계 구축	(전략 6) 연구 성과 관리·활용 시스템의 고도화	(전략 4) 성과의 수집·관리 체계 개선, 정부 3.0 제고
주요 과제	(1-1) 연구 성과 유형별 분류체계 마련 및 관리 • 연구 성과 유형별 분류체계 마련 및 표준관리양식 마련 - 4가지 유형 (1) 기록물(논문, 특허, 보고서) (2) 성과물(생명자원, 화합물, 소프트웨어) 및 성과물정보 (3) 기술정보 (4) 성과통계(논문, 특허, 기술료, 인력양성, 기타 파급효과 통계)	(6-2) 범부처 성과관리 연계, 협업체계 구축 • NTIS를 통해 15개 부처·청 국가 R&D정보의 상시 연계·수집 체계 정착 • 부처별 산재한 연구 성과 시스템 특성을 고려한 연계·협업 추진	(4-2) 연구개발 정부 3.0 제고 • NTIS의 연구개발 전주기 정보구축 및 공개 확대, 활용 중심의 정보접근성 및 이용편의성 제고 • 연구 성과 정보를 미래기술마당, NTB, 테크브리지 등 유관 온·오프라인 성과활용 관련 기관들에 제공 확대
연구 성과 범위	논문, 특허, 보고서원문, 기술요약정보, 생명자원, 화합물, 소프트웨어, 연구기자재	NTIS-TBN, NTIS-NDSL 등 시스템간의 연계·협업으로 성과정보 상세 제공	연구개발 전주기 정보구축 및 활용

□ (2018년 부처별 데이터 관련 정책 현황) 과학기술정보통신부가 연구데이터 공유·활용체계 구축에 있어 중추적 역할, 각 부처별 해당 분야 관련 빅데이터 플랫폼 구축을 통한 기반 조성

- (과학기술정보통신부) ‘과학기술과 정보통신기술(ICT)로 열어가는 사람 중심의 4차 산업혁명 구현’을 위해 국가 R&D 시스템 혁신을 전략과제로 연구데이터 공유·활용 체계 구축을 통한 연구성과 확산을 세부 추진과제로 마련
- (산업통상자원부) 미래 신산업 성과 조기 창출 및 주력산업 구조 혁신(전략과제) 실현을 위한 실천과제로 혁신성장 핵심역량 확충, 세부 추진과제로 에너지 빅데이터 플랫폼

19) 국가과학기술위원회(2006; 2010), 국가과학기술심의회(2016)

Ⅱ 구축('18)을 시작으로 산업별 데이터 플랫폼 확산 추진

<표 2-7> 산업 데이터 플랫폼(예시)

차량 데이터 분야	바이오헬스 분야	에너지 분야
차량 성능·위차·운행 → 법인 사업자의 차량운영 효율화 서비스	임상 정보·생체신호 → 맞춤형 건강관리 서비스	에너지 공급·사용 → 신재생E 공급량 예보 서비스

- (보건복지부) “국민의 삶의 수준 제고와 모든 국민을 포용하는 복지 구현”을 위해 혁신 주도를 추진 방향으로 삼고 혁신성장을 견인하기 위한 세부 추진과제로 공공기관 보건 의료 빅데이터 플랫폼 구축('18~'20) 예정
 - (해양수산부) “해양수산업 혁신산업화”를 위한 주요 정책과제로 해양공간 통합관리 체계 구축을 선정하고 해양수산 빅데이터 구축('21년까지 283종 통합)과 민간개방을 위한 정보제공시스템 구축·운영('18)
 - (행정안전부) 지능형 정부 실현을 위해 세부 추진과제로 고부가가치 공공데이터 개방 확대를 위한 공공데이터 접근성을 제고하고자 공공데이터 소재 및 연관관계를 시각화한 ‘국가데이터맵’ 구축·공개
 - (농림축산식품부) 농식품산업의 지속가능한 성장토대 마련을 위한 농식품 빅데이터 플랫폼 구축 등 과학역농 기반 구축
- (출연연 규정) 연구관리기관 및 연구기관의 개별 규정은 전무한 실정으로 연구데이터 활용도가 높은 일부 출연연에서 연구데이터 관리규정 운영
- 연구데이터 생산을 미션으로 하고 연구활용도가 높은 일부 정부출연연구소의 경우 자체적인 연구데이터 관리규정을 수립하여 실행 중
 - － (지질자원연구원) 2016년부터 데이터관리계획(DMP)을 주요 사업계획서에 포함하도록 연구데이터 관리규정에 명시하고 데이터 소유권, 생산 및 보존, 공개원칙, 활용에 따른 의무, 관리현황보고 등 세부 정책 수립²⁰⁾
 - － (극지연구소) 한국극지데이터센터를 설립하여 극지관련 연구데이터를 장기적으로 보관하고 관리하며 국내외에 공유 중
 - 극지연구소는 자체적으로 극지데이터관리정책에 의해 데이터를 관리하고 있으며 해당 정책에는 DMP 제출의무, 소유권, 관리대상 범위, 관리 및 저장, 공개 및 활용, 인용,

20) 한종규 외(2015). KIGAM 지질자원정보 통합관리체계 구축을 위한 중장기 추진전략 수립. 한국지질자원연구원

보고 등에 대한 세부내용 포함²¹⁾)

2.2.2. 연구데이터 관련 기존 유사사업 분석

- (유사과제 검색) 국가과학기술지식정보서비스(NTIS) 검색 결과 우리나라 국가연구개발사업 중 동 사업과 유사한이 없는 것으로 나타나, 정부의 동 사업 투자는 타당한 것으로 확인함
- 국가과학기술지식정보서비스에서 기수행과제, 타인등록과제, 공공R&D과제를 대상으로 유사과제를 검색함

<표 2-8> NTIS 유사과제 검색결과

항목	내용	
검색일시	2018.07.19.	
검색범위	기 수행과제 + 타인등록과제 + 공공R&D과제	
기준유사도	60	
검색결과 요약	등록과제 수	유사과제여부
	1건	0건

21) 한국국지데이터센터 홈페이지, <http://kpdc/kopri.re.kr>

2.2.3. 인력양성 현황

- 정출연 및 전문교육기관에서 데이터과학 전문인력 양성을 목표로 데이터 중심 교육 프로그램이 마련되어 있으나, 해당 교육은 슈퍼컴퓨팅 활용, 산업계 중심 빅데이터 인력 양성에 초점이 맞춰져 있어 국가 R&D 연구데이터 중점 교육 필요
 - KISTI는 교육·연구용 계산과학공학 시뮬레이션 플랫폼인 EDISON을 개발하고 과학데이터스쿨을 운영하여 정보이용 및 슈퍼컴퓨팅 교육 프로그램을 운영 중
 - 한국데이터진흥원(KDATA)는 빅데이터 아카데미를 운영하며 시장수요에 대응한 빅데이터 인력을 양성하고 산업 활성화 도모를 위해 산업계 재직자 대상의 빅데이터 전문인력 양성 프로그램 운영 중
 - 과기정통부, 교육부는 각각 빅데이터 청년인재 일자리 연계 사업, 빅데이터 사이언티스트 인력양성 사업 등을 통해 산업 수요에 부응하는 실무중심의 빅데이터 전문교육 시행 중
 - 서울대, KAIST, 연세대, 고려대, 성균관대 등 주요 대학에서도 빅데이터 교육 과정을 신설하여 운영 중
- (EDISON) KISTI는 이공계 교육·연구용 계산과학공학 시뮬레이션 플랫폼인 EDISON을 개발하여 국내 SW 실용화 기반 마련
 - 교육·연구과정에서 도출되는 데이터 공유체제 구축, 우수 SW 활용성 증대를 위한 전후 처리기 개발, 강의활용 수준의 계산과학공학용 CX-learning 콘텐츠 개발, 유지보수를 통한 선순환 지원체제 구축 등을 주요 과제로 추진
 - 현재 7개 전문분야 40여개 대학 연구팀에서 개발한 1200여종의 계산과학공학 응용 SW 및 콘텐츠를 탑재한 사이버 교육·연구 환경으로 연간 1만명 이상 활용 중
- (KISTI 과학데이터스쿨) 과학기술분야 연구데이터 및 HPC 활용 전문 인력 양성을 위한 교육 전담 조직으로서 KISTI과학데이터스쿨 설치 및 운영 개시('18.03)
 - 빅데이터, 사물인터넷, 딥러닝 등 교육 프로그램을 개발하고 연구데이터·컴퓨팅 활용 온라인 교육 환경(KSDSweb) 및 콘텐츠(KSDSbank) 구축
 - 2018년 기준 4차에 걸쳐 출연연 인재양성 과정을 통해 22개 기관 대상 34개 과정을 운영하여 800명 양성

- (빅데이터 아카데미) 빅데이터 프로젝트 진행 또는 예정인 산업계 재직자 대상의 빅데이터 전문인력 양성 프로그램으로 과기정통부에서 주관하고 한국데이터진흥원에서 운영 중
 - 현업 DB전문가를 대상으로 체계적인 교육을 제공하자고 직무별(빅데이터 기술/분석/기획)·산업별(통신/제조/유통/금융/융합) 빅데이터 교육 프로그램 설계
- (빅데이터 인력양성 사업) 정부는 빅데이터 청년인재 양성을 목표로 미취업 청년 대상의 빅데이터 인력양성 사업을 진행 중
 - 과기정통부는 빅데이터 청년인재 일자리 연계 사업(전담기관: 한국데이터진흥원) 및 혁신성장 청년인재 집중양성 사업(전담기관: 정보통신기술진흥센터)을 통해 산업 맞춤형 실무중심의 빅데이터 전문교육을 실시하고 청년과 수요기업간 일자리 연계 추진
 - 교육부는 빅데이터 사이언티스트 인력양성 사업(수행기관: 고려대학교)을 통해 실무 중심의 글로벌 프로젝트 교육 시스템을 통해 빅데이터 사이언티스트 인력 양성과정 추진
- (대학 교육과정) 산업적 요구에 부응하여 주요 대학에서 빅데이터 교육과정 운영 중
 - KAIST, 연세대, 고려대, 성균관대는 대학원 석사과정에 빅데이터 교육과정 신설
 - 서울대는 빅데이터 현장 인력 양성을 위해 빅데이터 연구원(SNU Big Data Academy)을 설립하고 학내 대학원 대상 BIG CAMP를 실시, 전국 대학(원)생 대상 산업체와의 연계 교육 프로그램을 운영 중

2.3. 수요 분석

2.3.1. 수요 분석 개요

□ 연구데이터 관리실태 1차 설문조사 개요

- (조사 목적) 연구과제 수행 시 생산되는 연구데이터 생산 현황 및 외부 공유 의견 파악
- (조사 기간) 2018. 04. 09. ~ 2018. 04. 20.
- (조사 방법) 온라인 설문조사
- (조사 대상) 국가연구개발사업 종료 후 2년 이내 과제(현재 진행 중인 과제 포함)
 - － 회수 현황: 총 3,205건
- (조사 내용) 국가 연구개발사업에서 유발되는 연구데이터 생산·관리·공유 현황

□ 연구데이터 관리실태 2차 설문조사 개요

- (조사 목적) 연구데이터 관리·활용 현황 및 애로사항 조사
- (조사 기간) 2018. 08. 14. ~ 2018. 09. 07.
- (조사 방법) 온라인 설문조사
- (조사 대상) 국가연구개발사업 종료 후 2년 이내 과제(현재 진행 중인 과제 포함)
 - － 회수 현황: 총 1,731건
- (조사 내용) 연구데이터 생산·관리·활용 현황 및 애로사항

2.3.2. 연구데이터 1차 수요조사 결과

가. 현황

- (응답자 현황) 수요조사 응답과제는 총 3,205개이고, 이 중 BT분야 과제가 42.59%(1,365개)로 가장 큰 비중을 차지
- 이후 IT(20.16%), 기타(13.76%), NT(11.20%), ST(8.02%), ET(2.75%), CT(1.53%)로 나타남

<표 2-9> 6T분야별 응답 현황

(단위: 개, %)

구분	빈도	비중
BT	1,365	42.59%
CT	49	1.53%
ET	88	2.75%
IT	646	20.16%
NT	359	11.20%
ST	257	8.02%
기타	441	13.76%
계	3,205	100.00%

- (연구데이터 생산 현황) 수요조사 대상과제 중 64.93%는 국가 R&D사업 수행으로 연구데이터를 생산하였다고 응답하였고, 35.07%는 그렇지 않았다고 응답함
- 동 수요조사 분석은 국가 R&D사업 수행으로 생산된 연구데이터가 있는 과제 2,081개를 대상으로 실시함

<표 2-10> 응답과제의 연구데이터 보유 여부

(단위: 개, %)

있음		없음		계	
빈도	비중	빈도	비중	빈도	비중
2,081	64.93%	1,124	35.07%	3,205	100.00%

- 6T분야별 연구데이터를 생산한 과제 수는 BT분야가 981개로 가장 많고, 이후 IT(334개), NT(266개) 순임
- － 연구데이터가 생산된 과제의 비중 기준은 ET분야가 75.00%로 가장 많고, 이후

\ NT(74.09%), BT(71.87%) 순임

<표 2-11> 6T분야별 연구데이터 보유 여부

(단위: 개, %)

구분	있음		없음		계	
	빈도	비중	빈도	비중	빈도	비중
BT	981	71.87%	384	28.13%	1,365	100.00%
CT	28	57.14%	21	42.86%	49	100.00%
ET	66	75.00%	22	25.00%	88	100.00%
IT	334	51.70%	312	48.30%	646	100.00%
NT	266	74.09%	93	25.91%	359	100.00%
ST	168	65.37%	89	34.63%	257	100.00%
기타	238	53.97%	203	46.03%	441	100.00%
계	2,081	64.93%	1,124	35.07%	3,205	100.00%

나. 연구데이터 생산 및 관리 현황

□ (연구데이터의 유형) 국가 R&D사업 수행으로 생산된 연구데이터는 57.53%(1,604개)가 정형데이터이고, 28.80%(803개)는 비정형데이터, 13.67%(381개)는 반정형데이터임

<표 2-12> 연구데이터의 유형(중복응답)

(단위: 개, %)

구분	빈도	비중
정형	1,604	57.53%
반정형	381	13.67%
비정형	803	28.80%
계	2,788	

□ (연구데이터의 형태) 연구데이터의 형태는 43.36%(1,658개)가 숫자 형태이고, 26.05%(996개)가 이미지, 23.46%(897개)가 텍스트임

- 기타에 응답한 과제 연구데이터 형태는 GIS 자료, spectroscopy, 뇌파신호, 도표, 생물분자분석, 소스 코드 및 Binary 등, 신호, 영상 등임

<표 2-13> 연구데이터의 형태(중복응답)

(단위: 개, %)

구분	빈도	비중
숫자	1,658	43.36%
텍스트	897	23.46%
동영상	224	5.86%
이미지	996	26.05%
소리	38	0.99%
기타	11	0.29%
계	3,824	100.00%

□ (연구데이터의 전체 용량) 연구데이터가 생산되었다고 응답한 과제 2,081개의 연구데이터 총 용량은 19,414,023,790MB임

- 6T분야별로는 NT분야의 연구데이터가 10,101,127,270MB로 52.03%를 차지하고, 기타 분야가 43.06%, BT분야가 3.65%를 차지함

<표 2-14> 6T 분야별 연구데이터의 용량

(단위: MB, %)

구분	있음	비중
BT	708,578,070	3.65%
CT	3,610,620	0.02%
ET	13,762,530	0.07%
IT	168,673,140	0.87%
NT	10,101,127,270	52.03%
ST	58,070,840	0.30%
기타	8,360,201,320	43.06%
계	19,414,023,790	100.00%

□ (연구데이터의 전체 건 수) 국가 R&D사업 수행으로 생산된 연구데이터 총 23,831,511,943건 중 98.99%는 IT분야에 해당함(23,591,680,753건)

표 2-15 > 6T분야별 연구데이터 전체 건 수

(단위: 건, %)

구분	건 수	비중
BT	132,114,123	0.55%
CT	10,679	0.00%
ET	31,731	0.00%
IT	23,591,680,753	98.99%
NT	103,420,502	0.43%
ST	207,263	0.00%
기타	4,046,892	0.02%
계	23,831,511,943	100.00%

- (연구데이터 관리 주체) 연구데이터를 생산하는 과제 중 1,005개(48.29%)는 연구데이터를 연구실 내부 시스템을 사용하여 관리한다고 응답하였고, 824개(39.60%)는 서버에 저장하여 단순 관리 또는 파일 단위로 관리한다고 응답함

<표 2-16> 연구데이터 관리 주체 현황

(단위: 개, %)

구분	기관 관리시스템을 사용하여 관리	서버에 저장하여 단순 관리 (파일단위로 관리)	연구실 내부 시스템을 사용하여 관리	외부 리포지토리를 사용하여 관리	관리하고 있지 않음	기타	계
빈도	98	824	1,005	64	42	48	2,081
비중	4.71	39.60	48.29	3.08	2.02	2.31	100.00

- 6T분야 모두 연구실 내부 시스템을 사용하여 연구데이터를 관리하는 과제가 가장 많은 것으로 나타남

<표 2-17> 6T 분야별 연구데이터 관리 주체

(단위: 개, %)

구분	기관 관리시스템을 사용하여 관리		서버에 저장하여 단순 관리(파일단위 로 관리)		연구실 내부 시스템을 사용하여관리		외부 리포지토리를 사용하여 관리		관리하고 있지 않음		기타		계	
	빈도	비중	빈도	비중	빈도	비중	빈도	비중	빈도	비중	빈도	비중	빈도	비중
BT	38	3.87	404	41.18	469	47.81	30	3.06	18	1.83	22	2.24	981	100.00
CT	1	3.57	10	35.71	14	50.00	2	7.14	0	0	1	3.57	28	100.00
ET	4	6.06	23	34.85	36	54.55	1	1.52	2	3.03	0	0	66	100.00
IT	15	4.49	142	42.51	148	44.31	19	5.69	8	2.40	2	0.60	334	100.00
NT	8	3.01	96	36.09	146	54.89	1	0.38	5	1.88	10	3.76	266	100.00
ST	14	8.33	64	38.10	84	50.00	1	0.60	2	1.19	3	1.79	168	100.00
기타	18	7.56	85	35.71	108	45.38	10	4.20	7	2.94	10	4.20	238	100.00
계	98	4.71	824	39.60	1,005	48.29	64	3.08	42	2.02	48	2.31	2,081	100.00

- (연구데이터 관리 방식) 연구데이터를 생산하는 과제 중 1,254개(60.26%)는 연구데이터를 소속 연구실 중심으로 관리하고 있고, 626개(30.08%)는 연구자 개인이 관리 중

- 6T 분야별로 소속 연구실에서 연구데이터를 관리하는 과제 수가 가장 많은 분야는 BT분야(581개, 59.23%)이고, 이후 IT분야 212개(63.47%), NT분야 181개(68.05%) 순임

표 2-18> 6T분야별 연구데이터 관리 방식

(단위: 개, %)

구분	개인		소속 기관		소속 연구실		기타		계	
	빈도	비중	빈도	비중	빈도	비중	빈도	비중	빈도	비중
BT	314	32.01	70	7.14	581	59.23	16	1.63	981	100.00
CT	6	21.43	4	14.29	18	64.29	0	0.00	28	100.00
ET	23	34.85	6	9.09	37	56.06	0	0.00	66	100.00
IT	77	23.05	43	12.87	212	63.47	2	0.60	334	100.00
NT	73	27.44	11	4.14	181	68.05	1	0.38	266	100.00
ST	47	27.98	14	8.33	106	63.10	1	0.60	168	100.00
기타	86	36.13	22	9.24	119	50.00	11	4.62	238	100.00
계	626	30.08	170	8.17	1,254	60.26	31	1.49	2,081	100.00

□ (연구데이터 관리 정책 보유 여부) 6T분야 중 연구데이터 관리 정책을 보유하고 있는 과제가 가장 많은 분야는 BT분야(452개, 33.14%)임

- 6T분야는 모두 연구데이터 관리 정책을 보유하지 않은 비중이 가장 많음

<표 2-19> 6T분야별 연구데이터 관리 정책 보유 여부

(단위: 개, %)

구분	있음		있으나 잘 준수하고 있지 않음		없음		계	
	빈도	비중	빈도	비중	빈도	비중	빈도	비중
BT	452	33.14	15	1.10	514	37.68	1,364	100.00
CT	10	20.41	1	2.04	17	34.69	49	100.00
ET	28	31.82	0	0.00	38	43.18	88	100.00
IT	145	22.45	9	1.39	180	27.86	646	100.00
NT	105	29.25	6	1.67	155	43.18	359	100.00
ST	73	28.40	5	1.95	90	35.02	257	100.00
기타	93	21.09	3	0.68	142	32.20	441	100.00
계	906	28.28	39	1.22	1136	35.46	3204	100.00

다. 연구데이터 공유 현황

□ (연구데이터 외부 공유 여부) 연구데이터를 생산한 과제 2,081개 중 연구데이터를 외부에 공유할 의사가 있는 과제는 313개(15.04%), 향후 공개할 예정이 있는 과제는 570개(27.39%)임

- 연구데이터를 외부에 공개할 의사가 있거나 향후 예정인 분야는 BT가 44.95%로 가장 많고, 기타가 43.70%, ST가 42.26%, IT가 40.12%임

<표 2-20> 6T분야별 연구데이터 외부 공유 여부

(단위: 개, %)

구분	있음		향후예정		없음		계	
	빈도	비중	빈도	비중	빈도	비중	빈도	비중
BT	147	14.98	294	29.97	540	55.05	981	100.00
CT	2	7.14	8	28.57	18	64.29	28	100.00
ET	7	10.61	17	25.76	42	63.64	66	100.00
IT	37	11.08	97	29.04	200	59.88	334	100.00
NT	52	19.55	47	17.67	167	62.78	266	100.00
ST	20	11.90	51	30.36	97	57.74	168	100.00
기타	48	20.17	56	23.53	134	56.30	238	100.00
계	313	15.04	570	27.39	1198	57.57	2,081	100.00

□ (연구데이터 공유 방법) 연구데이터를 생산한 과제 중 연구데이터를 URL, 웹 등의 자체 시스템을 통해 공유하는 과제가 33.82%(372개)로 가장 많고, 이후 이동저장장치가 25.91%, E-mail/FTP가 25.82%임

- 6T별로는 BT, ET, NT, ST, 기타 분야는 자체 보유시스템을 통해 공유하는 경우가 각각 36.84%, 31.03%, 30.08%, 37.36%, 33.87%로 가장 많음
- CT와 IT분야는 E-mail/FTP 등으로 공유하는 과제가 각각 45.45%, 39.18%로 가장 많은 것으로 나타남

표 2-21> 6T분야별 연구데이터 공유 방법(중복응답)

(단위: 개, %)

구분	E-mail/FTP		이동저장장치 (디스크, USB,CD-ROM 등)		자체 보유시스템을 통해 공유 (URL, 웹 등)		기타		계	
	빈도	비중	빈도	비중	빈도	비중	빈도	비중	빈도	비중
BT	141	25.59	130	23.59	203	36.84	77	13.97	551	100.00
CT	5	45.45	2	18.18	1	9.09	3	27.27	11	100.00
ET	6	20.69	8	27.59	9	31.03	6	20.69	29	100.00
IT	67	39.18	44	25.73	46	26.90	14	8.19	171	100.00
NT	22	17.89	35	28.46	37	30.08	29	23.58	123	100.00
ST	18	19.78	33	36.26	34	37.36	6	6.59	91	100.00
기타	25	20.16	33	26.61	42	33.87	24	19.35	124	100.00
계	284	25.82	285	25.91	372	33.82	159	14.45	1,100	100.00

□ (연구데이터 공유 대상) 연구데이터를 생산한 과제는 연구자에게만 연구데이터를 공유한 과제가 558개(54.81%)로 가장 많고, 이후 전체 공개(외국인 포함)가 306개(30.06%), 산업체 공개가 114개(11.20%) 순임

○ CT분야를 제외한 전체 분야에서 연구데이터는 연구자에게만 공유하는 경우가 가장 많았음

표 2-22> 6T 분야별 연구데이터 공유 대상(중복응답)

(단위: 개, %)

구분	전체공개 (외국인 포함)		연구자		산업체		기타		계	
	빈도	비중	빈도	비중	빈도	비중	빈도	비중	빈도	비중
BT	161	32.59	273	55.26	42	8.50	18	3.64	494	100.00
CT	5	45.45	3	27.27	1	9.09	2	18.18	11	100.00
ET	6	24.00	17	68.00	0	0.00	2	8.00	25	100.00
IT	39	23.49	93	56.02	30	18.07	4	2.41	166	100.00
NT	40	35.09	61	53.51	12	10.53	1	0.88	114	100.00
ST	15	17.65	54	63.53	13	15.29	3	3.53	85	100.00
기타	40	32.52	57	46.34	16	13.01	10	8.13	123	100.00
계	306	30.06	558	54.81	114	11.20	40	3.93	1,018	100.00

- (연구데이터 공유 조건) 연구데이터를 공유할 때 연구 목적으로만 공개하는 경우가 686개(77.69%)로 가장 많았고, 이후 완전 공개(123개, 13.93%), 비공개(32개, 3.62%) 순임

<표 2-23> 6T 분야별 연구데이터 공유 조건

(단위: 개, %)

구분	완전 공개		연구목적 공개		비공개		유료		기타		계	
	빈도	비중	빈도	비중	빈도	비중	빈도	비중	빈도	비중	빈도	비중
BT	66	14.97	345	78.23	15	3.40	6	1.36	9	2.04	441	100.00
CT	2	20.00	5	50.00	0	—	1	10.00	2	20.00	10	100.00
ET	2	8.33	19	79.17	1	4.17	0	—	2	8.33	24	100.00
IT	10	7.46	109	81.34	5	3.73	6	4.48	4	2.99	134	100.00
NT	12	12.12	82	82.83	2	2.02	0	—	3	3.03	99	100.00
ST	6	8.45	60	84.51	4	5.63	0	—	1	1.41	71	100.00
기타	25	24.04	66	63.46	5	4.81	4	3.85	4	3.85	104	100.00
계	123	13.93	686	77.69	32	3.62	17	1.93	25	2.83	883	100.00

- (연구데이터 향후 공유 기간) 연구데이터를 생산한 과제의 연구책임자는 연구데이터를 외부에 공유할 경우 과제 종료 후 평균 3.91년 후에 공개하는 것이 적절하다고 응답함
- 과제 종료 후 공개하기까지의 기간이 가장 짧은 분야는 IT 분야(3.17년)이고, ET분야가 5.10년으로 가장 길

<표 2-24> 6T 분야별 연구데이터 향후 공유 기간

(단위: 년)

구분	BT	CT	ET	IT	NT	ST	기타	평균
평균	3.78	3.85	5.10	3.17	4.73	3.98	4.21	3.91

2.3.3. 연구데이터 2차 수요조사 결과

가. 현황

- (응답자 현황) 동 수요조사 응답과제는 총 1,731개이고, 이 중 연구데이터를 생산한 과제는 1,134(65.51%)개임

<표 2-25> 응답과제의 연구데이터 보유 여부

(단위: 개, %)

있음		없음		무응답		계	
빈도	비중	빈도	비중	빈도	비중	빈도	비중
1,134	65.51%	595	34.37%	2	0.12%	1,731	100.00%

나. 연구데이터 관리 현황

- (연구데이터 관리 현황) 연구데이터를 생산한 경험이 있는 과제 중 연구데이터를 관리하고 있는 과제는 1,058개(93.30%)임
- 연구데이터를 관리하지 않는 과제는 76개로 연구데이터를 생산한 경험이 있는 과제 1,134개 중 6.70%를 차지함

<표 2-26> 응답과제의 연구데이터 관리 여부

(단위: 개, %)

있음		없음		계	
빈도	비중	빈도	비중	빈도	비중
1,058	93.30%	76	6.70%	1,134	100.00%

- (연구데이터 관리 시스템 자원) 연구데이터를 관리하고 있는 과제 1,058개의 관리 서버 보유 수는 총 1,181,938대임
- 1개 과제 당 평균은 1,117.14대 이지만 중앙값은 1대로 과제 당 주로 1대의 연구데이터 관리를 위한 서버를 보유하고 있음

<표 2-27> 응답과제의 연구데이터 관리 시스템 보유 현황

(단위: 대)

합계	평균	중앙값
1,181,938	1,117.14	1.00

- 연구데이터를 관리하기 위한 서버를 1대 보유한 과제는 총 591개로 55.86%를 차지하고, 이후 서버가 없는 과제가 177개(16.73%), 2개 보유한 과제가 149개(14.08%), 3대 보유한 과제가 12개(1.13%)임

<표 2-28> 연구데이터 관리 서버 수 별 응답과제 현황

(단위: 개, %)

서버 대 수	빈도	비중
없음	177	16.73%
1대	591	55.86%
2대	149	14.08%
3대	58	5.48%
4대	12	1.13%
5대	30	2.84%
6대 이상	41	3.88%
계	1,058	100.00%

□ (연구데이터 관리 시 애로사항) 연구 시 연구데이터를 생산한 과제 중 460개의 연구책임자는 연구데이터를 관리하는 전담자 또는 부서의 부족으로 연구데이터를 관리하는데에 어려움을 겪고 있음

- 이 후 연구데이터 관리 시의 애로사항은 연구데이터를 관리하기 위한 예산이 부족하거나 없는 경우가 376개(23.04%), 연구데이터 관리를 위한 저장·보존 시스템이 없는 경우가 344개(21.08%), 연구데이터 관리를 위한 규정이나 제도가 없는 경우가 271개(16.61%)임

<표 2-29> 연구데이터 관리 애로사항(중복응답)

(단위: 개, %)

구분	빈도	비중
연구데이터 관리를 위한 저장·보존 시스템이 없음	344	21.08%
연구데이터를 관리하는 전담자 또는 부서가 없음	460	28.19%
연구데이터를 관리하기 위한 예산이 부족하거나 없음	376	23.04%
연구데이터 관리를 위한 규정이나 제도가 없음	271	16.61%
기타	181	11.09%
계	1,632	100.00%

□ (연구데이터 유실·손실 경험) 연구 시 생산된 연구데이터의 유실 또는 손실된 적이 있는 과제는 총 708개로 연구데이터가 생산된 과제 중 과반수 이상을 차지함

표 2-30 연구데이터 유실·손실 경험

(단위: 개, %)

있음		없음		계	
빈도	비중	빈도	비중	빈도	비중
708	62.43%	426	37.57%	1,134	100.00%

다. 연구데이터 활용 현황

- (타 연구데이터 필요 경험) 동 수요조사 응답과제 총 1,731개 중 연구 시 다른 연구자가 생산한 연구데이터가 필요한 적이 있는 과제는 1,360개임
- 타 연구자의 연구데이터가 가끔 필요한 과제가 568개(32.81%)로 가장 많았고, 보통인 과제가 367개(21.20%), 자주 필요한 과제가 300개(17.33%)임

<표 2-31> 타 연구자 연구데이터 필요 여부

(단위: 개, %)

구분	빈도	비중
매우 자주	125	7.22%
자주	300	17.33%
보통	367	21.20%
가끔	568	32.81%
없음	371	21.43%
계	1,731	100.00%

- (타 연구데이터 활용 애로사항) 다른 연구자(기관)의 연구데이터 활용할 때의 애로사항은 다른 연구자(기관)가 연구데이터를 공개하지 않은 경우가 879개(30.25%)로 가장 많음
- 이후 연구데이터의 존재와 보유 위치를 모르는 경우가 702개(24.16%), 연구데이터에 접근하기 어려운 경우가 550개(18.93%), 정책/제도/보안 상의 이류가 333개(11.46%), 공유방식(시스템 등)이 미흡한 경우가 294개(10.12%)임

<표 2-32> 타 연구자 연구데이터 활용 어려운 이유(중복응답)

(단위: 개, %)

구분	빈도	비중
연구데이터의 존재와 보유 위치를 몰라서	702	24.16%
연구데이터에 접근하기 어려워서	550	18.93%
연구데이터를 공개하지 않아서	879	30.25%
정책/제도/보안 상의 이유로 인해	333	11.46%
공유방식(시스템 등)이 미흡해서	294	10.12%
기타	95	3.27%
애로사항 없음	53	1.82%
계	2,906	100.00%

- 타 연구자(기관)의 연구데이터 활용 시 애로사항이 있다고 응답한 과제 중 애로사항 해결 시 연구데이터를 공유·활용하는 것에 참여할 의향이 매우 있거나 있다고 응답한 과제가 865개로 49.97%를 차지함
- 참여할 의향이 없거나 매우 없는 경우는 298개로 17.21%를 차지함

<표 2-33> 연구데이터 애로사항 해결 시 공유·활용 의향

(단위: 개, %)

구분	빈도	비중
매우 그렇다	216	12.48%
그렇다	649	37.49%
보통	568	32.81%
그렇지 않다	255	14.73%
매우 그렇지 않다	43	2.48%
계	1,731	100.00%

- (연구데이터 공유·활용 플랫폼 필요 기능) 연구데이터를 활용하기 위해 클라우드 기반의 연구환경을 제공한다면 필요한 기능은 연구데이터 저장·공유 공간이라고 응답한 과제는 1,230개(43.01%)로 가장 많음
- 이후 필요한 기능은 연구데이터 장기 보존 공간(848개), 데이터 처리 및 분석을 위한 컴퓨팅 리소스 제공(623개), AI 플랫폼 등 데이터 분석을 위한 소프트웨어 제공(41개) 순임

<표 2-34> 클라우드 기반 연구환경 시 필요 기능(중복응답)

(단위: 개, %)

구분	빈도	비중
연구데이터 저장·공유 공간(비공개 가능)	1,230	43.01%
연구데이터 장기 보존 공간	848	29.65%
데이터 처리 및 분석을 위한 컴퓨팅 리소스 제공	623	21.78%
AI 플랫폼 등 데이터 분석을 위한 소프트웨어 제공	41	1.43%
기타	90	3.15%
해당 없음	28	0.98%
계	2,860	100.00%

- (연구데이터 공유·활용 기대효과) 연구데이터를 생산한 과제의 연구책임자는 국가 또는 기관 차원에서 연구데이터를 공유·활용하는 체계를 구축할 경우 연구 생산성이 매우 증가하거나 증가할 것이라고 응답함(57.48%)

<표 2-35> 연구데이터 공유·활용 체제 구축 시 연구생산성 증가여부

(단위: 개, %)

구분	빈도	비중
매우 그렇다	302	17.45%
그렇다	693	40.03%
보통	493	28.48%
그렇지 않다	196	11.32%
매우 그렇지 않다	47	2.72%
계	1,731	100.00%

2.3.4. 수요조사 주요 시사점

- 연구데이터를 국가 R&D 공식 성과물로 제도화하고 관리해야 연구데이터의 생산과 사용이 활성화될 수 있음
 - (연구데이터 생산분야) 연구데이터가 생산된 과제 비중이 ET(75%), NT(74.09%), BT(71.87%)순임에 반해 연구데이터 생산 과제 수가 주로 BT분야(47%)에서 생산되고 있음은 생명정보가 국가 R&D의 공식 성과물로 규정되어 관련한 연구데이터 또한 타분야 대비 높은 비중으로 생산되고 있음을 반증함
 - (연구데이터 유형·형태) 생산된 연구데이터의 유형이 주로 정형데이터(57.53%)이고, 형태가 숫자, 이미지, 텍스트임은 연구데이터 관리를 통한 재활용가능성이 높고, 플랫폼 내에서의 관리가 가능함을 시사함
- 개인·연구실 단위에서 연구데이터가 사장되지 않도록 국가차원의 관리체계를 구축하고 연구데이터 저장·보존 시스템이 미비한 분야에 연구데이터 플랫폼을 통한 지원체계 구축 필요
 - (연구데이터 관리현황) 연구데이터를 기관차원이 아닌 연구실(60%) 또는 연구자 개인(30%)이 관리하고 연구실 내부 시스템(48%) 혹은 서버에 저장하는 단순 관리(39%)를 하고 있음은 연구데이터의 유실·소실 위험성이 크다는 것을 반증하며, 국가차원의 공동관리규정 마련 후 기관차원의 체계적인 관리가 뒷받침되어야 함을 시사함
 - 데이터 관리를 위해 과제 당 1대의 서버를 보유하는 열악한 수준이며, 애로사항으로 연구데이터 저장·보존 시스템이 없다고 21% 응답
 - 클라우드 기반 연구환경 제공 시 가장 필요한 기능으로 저장·공유 공간(43%), 장기 보존 공간(30%)을 꼽음
 - 연구데이터 유실·손실 경험도 62%로 과반 수 이상 차지
 - 연구데이터 관리가 어려운 이유로 전담인력·조직의 부재(28%)를 가장 큰 문제로 지적
- 연구데이터 공유·활용을 위한 커뮤니티 조성, 데이터기반 협업 연구문화 확산 필요
 - (연구데이터 필요 대비 공유) 타 연구데이터 필요 경험(79%)은 있으나, 정작 자신의 연구데이터의 외부 공개 의사는 15%로 매우 낮아(비공개가 58%, 향후 공개가 27%), 대부분 연구자들이 연구데이터 공개에 대해 방어적이고 미온적 태도를 보임
 - 타 연구데이터 활용 애로사항 해결 시 응답자 절반(50%)이 연구데이터 공유·활용 참여

의사를 보였으나, 연구목적의(78%) 연구자에게만 공개(55%)하는 등 제한적 공개조건을 보임에 따라 연구 활성화를 위해 데이터 커뮤니티 조성 등 문화 정착이 우선적으로 필요

- 연구데이터 공유·활용 플랫폼 구축을 통해 데이터에 대한 접근성 향상 및 연구효율성 제고 필요
 - 연구데이터 활용의 가장 큰 애로사항으로 연구데이터 비공개(30%), 연구데이터 존재와 보유 위치를 모르거나(24%), 접근이 어렵다는 점(19%)을 지적
 - 연구데이터 공유·활용 체계를 구축할 경우 응답자 과반수 이상이 연구 생산성이 증가할 것이라 기대함(57%)

경쟁력 진단 및 추진방향 설정

- 3.1. 국내 R&D 역량 분석
- 3.2. PEST 분석
- 3.3. SWOT 분석 및 대응전략
도출

3.1. 국내 R&D 역량 분석



3.1.1. R&D 투자 현황

정부의 연구데이터 관련 사업이 다수 추진되고 있으나, 연구데이터 플랫폼 개발과 관련된 방향성이 부족한 상태에서 과제 단위로 연구개발이 수행되어 과제간 유기성을 찾기 어려움

- '13~'17년 5년간 정부의 연구데이터 R&D에 총 996.2억 원을 투자하였고, '15년 이후 꾸준히 230억 원 이상씩 투자 중
- 정부의 연구데이터 관련 R&D 비용 중 75.7%는 과학기술정보통신부가 차지

가. 분석 개요

- 국가과학기술지식정보서비스(NTIS) 상 연구데이터 관련 키워드 검색을 통해 최근 5년('13~'17년)간 연구데이터 관련 부처별 R&D 사업 분석 실시
 - 연구데이터 관련 키워드는 연구데이터, 과학데이터, 빅데이터, 대용량데이터로 선정함
- 연구데이터와 관련된 부처별 주요 R&D 사업의 연구목표, 연구내용, 기대효과 등을 검토하여 유사과제 선정
 - NTIS 키워드 검색 결과 총 897개 과제 검색
- 본 분석에서는 897개 과제 중 동 사업의 목적 및 범위와 관련이 없는 과제를 제외하여 총 312개 과제를 선별하여 분석 실시

나. 분석 결과

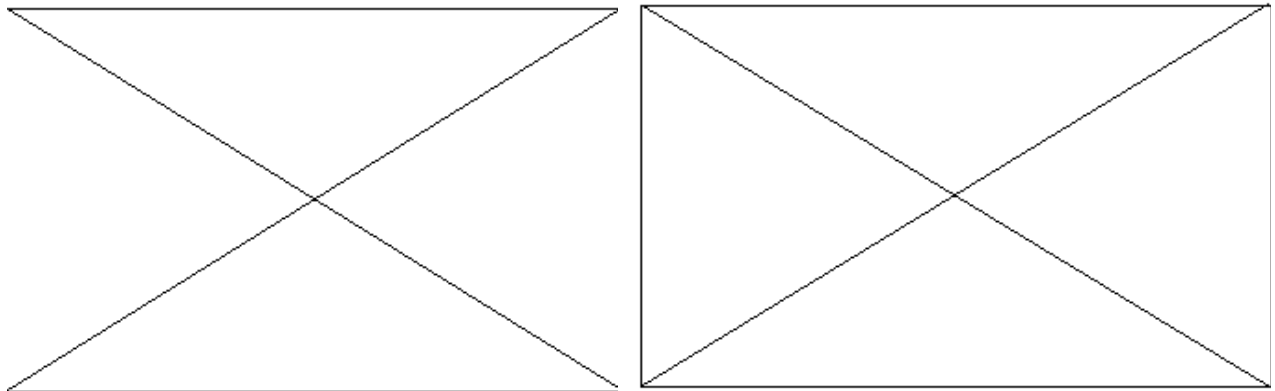
- (정부연구비 투자현황) 정부의 연구데이터 관련 R&D 투자 금액은 총 99,620백만 원이고, 연평균 증가율은 14.5%임('13~'17년 기준)
 - (연도별) '13~'17년 연구데이터 관련 R&D 투자 금액은 '16년까지 증가하다가 '17년에 소폭 하락함('16년 23,522백만 원, '17년 23,105백만 원)
 - (부처별) '13~'17년 연구데이터 관련 R&D 투자 금액은 과학기술정보통신부가 75.7%(75,428백만 원)로 가장 많은 비중을 차지함
 - 이후 중소벤처기업부 6.4%, 국토교통부 5.5%, 산업통상자원부 5.2% 순으로 투자함

<표 3-1> 연도별·부처별 연구데이터 관련 정부연구비 투자 현황

(단위: 백만 원, %)

부처명	2013년	2014년	2015년	2016년	2017년	계(평균)	비중
과학기술정보통신부	13,005	11,688	18,406	16,876	15,453	75,428	75.7
교육부	221	220	313	733	1,047	2,534	2.5
국토교통부	—	750	1,380	1,636	1,752	5,518	5.5
기상청	225	243	310	—	—	778	0.8
농촌진흥청	—	—	—	35	35	70	0.1
보건복지부	80	110	780	1,150	1,570	3,690	3.7
산업통상자원부	—	—	1,190	1,994	2,040	5,224	5.2
중소벤처기업부	994	2,031	1,049	1,097	1,208	6,378	6.4
계	14,524	15,042	23,427	23,522	23,105	99,620	100.0
증가율	—	3.6	55.7	0.4	-1.8	(14.5%)	

(단위: 백만 원)



<그림 3-2> 연도별·부처별 연구데이터 관련 정부연구비 투자 현황

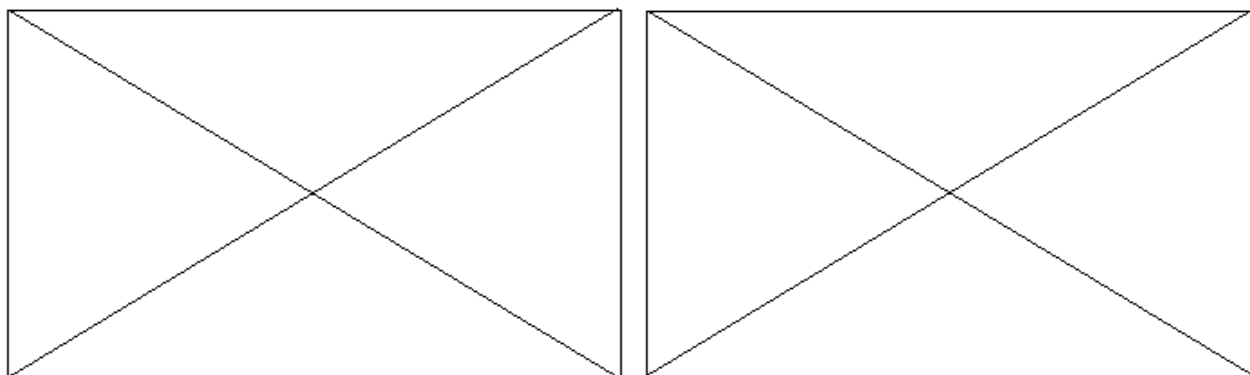
- (과제당 투자동향) '13~'17년 연구데이터 관련 과제 당 R&D 평균 투자액은 약 359.36백만 원임
- (연도별) '13~'17년 1개 과제당 R&D 투자액은 '15년에 414.36백만 원으로 가장 높고, 이후 감소 추세임
 - (부처별) 부처별 1개 과제당 R&D 투자액은 과학기술정보통신부가 873.01백만 원으로 가장 많고, 이후 국토교통부 689.75백만 원, 산업통상자원부 401.16백만 원, 보건복지부 260.17백만 원 순임

표 3-2> 연도별·부처별 과제 당 정부연구비 투자 현황

(단위: 백만 원)

부처명	2013년	2014년	2015년	2016년	2017년	평균
과학기술정보통신부	1,300.45	1,062.57	920.29	675.05	406.67	873.01
교육부	55.22	73.29	62.52	61.12	49.86	60.40
국토교통부	—	375.00	690.00	818.00	876.00	689.75
기상청	225.00	243.00	310.00	—	—	259.33
농촌진흥청	—	—	—	35.00	35.00	35.00
보건복지부	80.00	55.00	390.00	383.33	392.50	260.17
산업통상자원부	—	—	396.67	398.80	408.00	401.16
중소벤처기업부	141.99	184.61	131.07	137.19	150.94	149.16
평균	360.53	332.24	414.36	358.36	331.28	359.36

(단위: 억 원)



<그림 3-3> 연도별·부처별 연구데이터 관련 1개 과제 당 정부연구비 투자 현황

3.1.2. 기술역량 분석

국가연구데이터플랫폼은 전자정보·통신 분야의 ‘지식기반 데이터 활용기술’, ‘데이터 분산처리 시스템기술’, ‘융합서비스 플랫폼기술’과 연관성이 높음

동 기술의 우리나라 기술수준 그룹은 선도 그룹에 해당(최고기술국 대비 84.2%, 세계4위)하여 정부의 중장기적 투자를 통한 기술개발 시, 기술혁신 경쟁력 및 연구데이터 플랫폼 구축 성공을 이끌 수 있는 잠재력을 보유함

□ (전략기술 선정) 국가연구데이터플랫폼은 120개 국가전략기술 중 ‘전자정보·통신’ 분야의 ‘지식기반 빅데이터 활용기술’이 의미상 가장 부합되고, ‘데이터 분산처리 시스템기술’과 ‘융합서비스플랫폼기술’도 일부 부합함

- ‘지식기반 빅데이터 활용기술’은 대량의 정형 또는 비정형 데이터들간의 연관성을 분석, 발견하고 가치를 추출하여 새로운 정보·지식을 제공하는 측면이 동 사업의 국가연구데이터플랫폼에 부합함
- ‘데이터 분산처리 시스템기술’은 빅데이터의 입출력 처리를 가능하게 하는 부분이 부합하고, ‘융합서비스 플랫폼기술’은 서비스 이용자에게 지식 기반의 융합서비스를 제공하는 부분이 동 플랫폼에 부합함

<표 3-3> 연구데이터 공유·활용체계 관련 국가전략기술별 기술설명

분야	번호	전략기술명	전략기술 설명
전자·정보·통신	2	지식기반 빅데이터 활용기술	- 실생활에서 축적되는 <u>대용량 데이터들간의 연관성을 분석, 발견하고, 정보를 추출하여 새로운 지식을 생성, 관리하기 위한 정보의 지식화기술</u> - 기존 데이터베이스 관리도구의 데이터 수집·저장·관리·분석의 역량을 넘어서는 <u>대량의 정형 또는 비정형 데이터 세트 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하여 다변화된 현대 사회를 보다 정확하게 예측하여 효율적으로 작동케 하고, 개인화된 현대 사회 구성원마다 맞춤형 정보를 제공·관리·분석 가능케 하는 기술</u>
	5	데이터 분산처리 시스템기술	기하급수적으로 늘어나는 비정형 정보와 대규모 빅데이터를 분석하여 사용자의 상황에 맞는 유용한 정보를 즉시 제공하기 위해 실시간으로 데이터 저장 및 처리, 고성능 분산 데이터 처리, <u>고속 데이터 입출력 처리를 가능하게 하는 대용량 데이터 처리 기술</u>
	15	융합서비스 플랫폼기술	다양한 서로 다른 영역(예: 방송, 통신, 건설, 조선, 의료, 교육 등)의 지식들을 융합하여 체계적으로 구조화시키고 연계하여 지식의 전체상을 만들고, 이를 기반으로 단일 영역지식만으로 해결이 어려웠던 현실의 문제점을 해결하며, 새로운 지식을 창출함으로써 <u>서비스 이용자에게 지식 기반의 융합 서비스를 제공할 수 있도록 하기 위한 기반 기술</u>

□ (기술경쟁력 측면) 우리나라 전자정보·통신 분야의 기술수준 그룹은 선도그룹에 해당(최고기술국 대비 84.2%(세계 4위))하여 동 분야의 기술혁신 경쟁력을 보유함

- '14년과 '16년의 동 분야 최고기술 보유국은 미국이고, 우리나라는 선도 그룹에 해당함
 - － 우리나라의 기술수준은 '14년 83.2%에서 '16년 84.2%로 1.0%p 증가하였고, 기술격차는 '14년 2.7년에서 '16년 2.2년으로 0.5년 감소함

<표 3-4> 전자·정보·통신 분야 우리나라 기술수준 및 기술격차 변동('14~'16)

(단위: %, 년)

분야	최고기술국		기술수준 그룹		기술수준			기술격차		
	2014	2016	2014	2016	2014	2016	증감	2014	2016	증감
전자·정보·통신	미국	미국	선도	선도	83.2	84.2	1.0	2.7	2.2	-0.5

- － 기술수준은 기초연구가 82.0%, 응용개발 연구가 86.4%로 기초연구가 4.4%p 낮아 정부의 기초연구 지원으로 기술수준 향상이 필요함

<표 3-5> 전자·정보·통신 분야 우리나라 연구개발 단계별 기술수준

(단위: %, %p)

분야	기초연구 수준		응용·개발연구 수준		기초연구, 응용·개발 연구 수준 차
	기술수준 그룹	기술수준	기술수준 그룹	기술수준	
전자·정보·통신	선도	82.0	선도	86.4	4.4

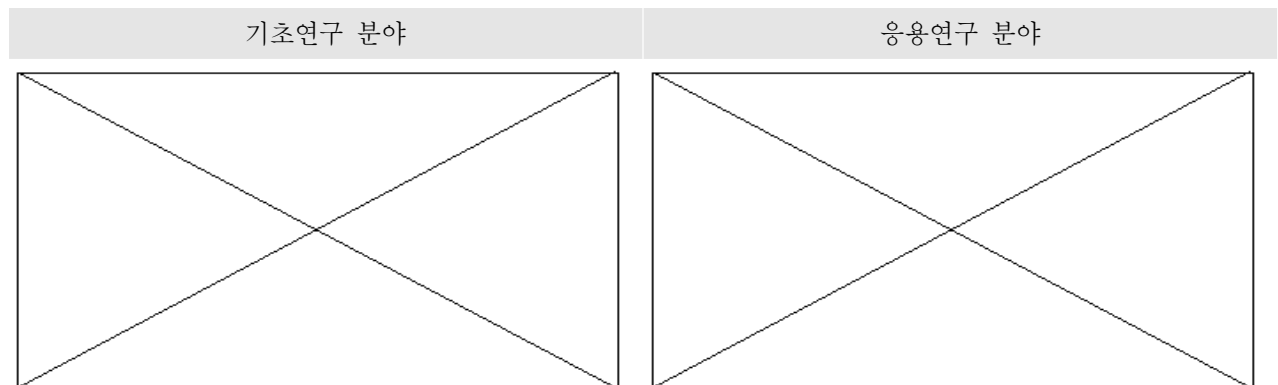
- '16년 국가전략기술별 기술수준은 융합서비스 플랫폼기술(80.7%), 데이터 분산처리 시스템 기술(79.9%), 지식기반 빅데이터 활용기술 순임(77.3%)
 - － '14년, '16년 최고기술 보유국은 미국이고, 우리나라는 융합서비스 플랫폼 기술이 '16년에 선도그룹에 진입함
 - － '14년 대비 '16년에 기술격차는 융합서비스 플랫폼기술이 0.4년, 지식기반 빅데이터 활용기술이 0.1년 감소함

<표 3-6> 전자·정보·통신 분야 국가전략기술별 우리나라 기술수준 및 기술격차 변동('14~'16)

(단위: %, 년)

분야	최고기술품		기술수준 그룹		기술수준			기술격차		
	2014	2016	2014	2016	2014	2016	증감	2014	2016	증감
지식기반 빅데이터 활용기술	미국	미국	추격	추격	77.9	77.3	-0.6	3.4	3.3	-0.1
데이터 분산처리 시스템기술	미국	미국	선도	추격	82.9	79.9	-3.0	2.3	2.5	0.2
융합서비스 플랫폼기술	미국	미국	추격	선도	79.0	80.7	1.7	2.7	2.3	-0.4

- 국가전략기술별 기초연구와 응용·개발연구의 기술수준 그룹은 추격~선도 그룹 수준이고, 특히 추격 그룹의 기술 중 융합서비스 플랫폼기술의 기술수준이 기초연구와 응용·개발연구 모두 '14년보다 향상하여 국가 연구데이터 공유·활용 체계 구축을 위한 기반기술이 구축됨
- 기초연구와 응용·개발연구의 수준 차는 전반적으로 응용·개발연구 수준이 기초연구 수준보다 최소 2.3%p에서 최대 5.7%p만큼 높음
- 국가전략기술별 기초연구와 응용·개발연구 수준 중 융합서비스 플랫폼기술의 '16년 응용·개발연구 수준이 81.8%로 가장 높고, 지식기반 빅데이터 활용기술의 '16년 기초연구 수준이 75.0%로 가장 낮음



<그림 3-4> 동 사업 관련 국가전략기술별 우리나라 연구개발 기술수준 및 기술격차 변동('14, '16)

\<표 3-7> 동 사업 관련 국가전략기술별 우리나라 연구개발 단계별 기술수준 및 기술격차 변동('14~'16)

(단위: %, %p)

국가전략기술명	기초연구 수준				응용·개발연구 수준				기초연구, 응용·개발 연구 수준 차	
	기술수준 그룹		기술수준		기술수준 그룹		기술수준			
	2014	2016	2014	2016	2014	2016	2014	2016	2014	2016
지식기반 빅데이터 활용기술	추격	추격	76.2	75.0	추격	추격	79.6	79.6	3.4	4.6
데이터 분산처리 시스템기술	선도	추격	82.1	78.6	선도	선도	83.7	81.1	1.6	2.5
융합서비스 플랫폼기술	추격	추격	78.0	79.5	추격	선도	80.0	81.8	2.0	2.3

- (기술역량 측면) 연구데이터 공유·활용 체계 구축 관련 기술분야인 전자·정보·통신 분야의 기술수준 그룹은 대기업, 중소기업, 연구계, 학계 모두 선도 그룹에 해당하여 연구데이터 플랫폼 구축 성공을 이끌 수 있는 잠재력 보유
- 동 사업의 주요 수혜대상인 연구계, 학계의 기술수준은 전자·정보·통신 분야에서 선도권에 위치하고 있어 중장기 대규모 투자가 집중적으로 이루어진다면 국내 연구의 효율성을 제고할 수 있는 원천기술 확보 잠재력 보유

<표 3-8> 전자·정보·통신 분야의 우리나라 연구주체별 기술수준

(단위: %)

분야	대기업			중소기업			연구계			학계		
	최고 기술 국	기술 수준 그룹	기술 수준	최고 기술 국	기술 수준 그룹	기술 수준	최고 기술 국	기술 수준 그룹	기술 수준	최고 기술 국	기술 수준 그룹	기술 수준
전자·정보·통신	미국	선도	85.4	미국	추격	75.0	미국	선도	81.7	미국	선도	80.5

- 인프라 구축수준 그룹도 선도권으로 기반기술의 기술혁신 인프라 측면에서도 연구데이터 공유·활용 체계 구축 성공 가능성을 제고하고 있음
- － 전자·정보·통신 분야의 연구개발 단계별 인프라 구축 수준은 기초연구(80.4%), 응용·개발연구(82.8%) 모두 선도그룹에 해당함
- － 기술수준과 인프라 구축 수준의 차이는 기초연구가 1.6%p, 응용·개발연구가 3.6%p로 타 분야 대비 수준 차이가 크지 않음

<표 3-9> 전자·정보·통신 분야의 우리나라 인프라 구축 수준

(단위: %, %p)

분야	기초연구 인프라		응용·개발연구 인프라		인프라		기초연구, 응용·개발연구 인프라 구축 수준 차
	구축수준 그룹	구축수준 (%)	구축수준 그룹	구축수준 (%)	구축수준 그룹	구축수준 (%)	
전자·정보·통신	선도	80.4	선도	82.8	선도	81.6	2.4

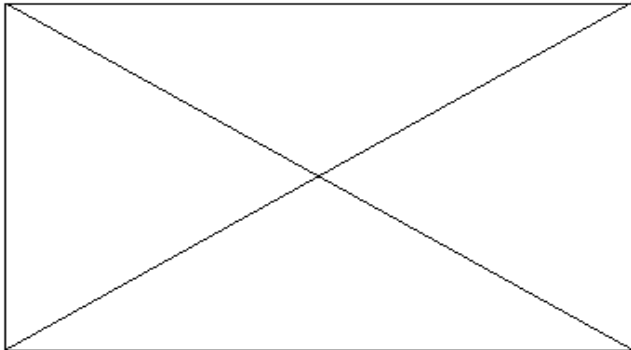
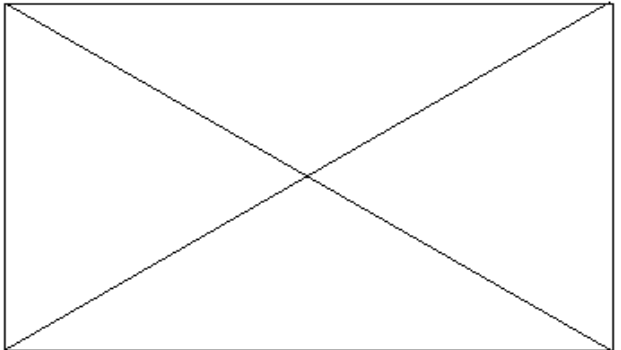
- － 특히, 전자·정보·통신 분야의 기술수준과 인프라 구축 수준이 비슷하여 연구데이터 공유·활용 체계 구축을 위한 기반기술을 보유하고 있는 상태임(기초연구 수준 차 1.6%p, 응용·개발연구 수준 차 3.6%p)

\<표 3-10> 전자·정보·통신 분야의 우리나라 연구개발 단계별 기술수준과 인프라 구축 수준 차

(단위: %, %p)

분야	기초연구			응용·개발연구			종합		
	기술 수준	인프라 구축 수준	기술 수준과 인프라 수준 차	기술 수준	인프라 구축 수준	기술 수준과 인프라 수준 차	기술 수준	인프라 구축 수준	기술 수준과 인프라 수준 차
전자정보·통신	82.0	80.4	1.6	86.4	82.8	3.6	84.2	81.6	2.6

- 반면, 최근 연구데이터 공유·활용체계 관련 인프라 기술이 다소 하향세를 보이고 있어, 빠른 시간 내에 정부의 지원이 이루어질 필요가 있음

기초연구 분야	응용연구 분야
	

<그림 3-5> 동 사업 관련 국가전략기술별 우리나라 연구개발 단계별 인프라 구축 수준('14~'16)

<표 3-11> 동 사업 관련 국가전략기술별 우리나라 연구개발 단계별 인프라 구축 수준('14~'16)

(단위: %, %p)

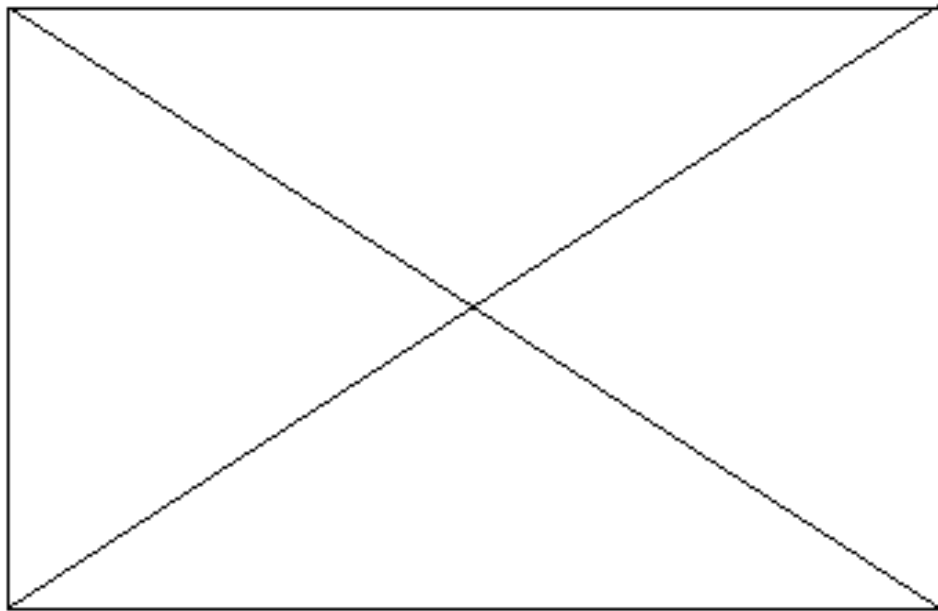
국가전략기술명	기초연구 인프라 구축 수준		응용·개발연구 인프라 구축 수준		인프라 구축 수준		기초-응용·개발 연구 인프라 구축 수준 차	
	2014	2016	2014	2016	2014	2016	2014	2016
지식기반 빅데이터 활용기술	80.0	77.2	80.2	80.0	80.1	78.6	0.2	2.8
데이터 분산처리 시스템기술	80.0	78.1	84.7	83.7	82.4	80.9	4.7	5.6
융합서비스 플랫폼기술	78.8	76.7	81.8	81.5	80.3	79.1	3.0	4.8

3.1.3. R&D 인프라 현황 (초기 R&D인프라 현황 자료로 대체함)

가. 과학기술분야 정부출연연구소 연구데이터 관리·공유 체계 현황

우리나라 정부출연연구소는 개별적으로 연구데이터 관리·공유를 위한 시스템을 구축하고, 일부 운영 중이나, 각 정부출연연구소의 연구데이터를 통합하여 관리할 수 있는 시스템은 부족함.
연구자의 공동융합 연구 활성화 및 연구 효율성 확대를 위해 정부 주도의 연구데이터를 통합 관리·공유 할 수 있는 연구데이터 플랫폼과 정부출연연구소의 연구데이터를 연계·활용 할 수 있는 공유 체계 마련이 필요함

- (한국과학기술연구원) 현재 R&D 빅데이터 및 정보학 기술을 이용할 수 있는 R&D 인프라를 구축 중으로 '18년 말 베타서비스 제공과 축매분야 데이터 수집 예정
- (예산) '17년 1,000백만 원, '18년 800백만 원
 - (인프라) 연구노트를 DB화하고, 기존의 비정형 데이터 및 외부 DB를 통합하여 데이터 통합관리 및 분석·시각화 서비스 제공 가능한 인프라 구축 중



<그림 3-6> 데이터기반 R&D 플랫폼의 구성도

- (수집·관리 현황) '18년 말 베타서비스 제공 후 구축 완료시점에 약 15만 건의 데이터 축적 예정(축매분야 논문 10만 건, 측정데이터 5만 건)
- 특성분석센터 측정 장비와 연계된 의한 물성측정 연계 데이터 구축 중

\ - 촉매분야 논문데이터 6만 건(58GB), 연말까지 약 6만 건 추가 예정

□ (한국과학기술정보연구원) 첨단 연구 장비에서 생산되는 대용량 데이터를 확보하고, 실험 데이터 분석을 위한 컴퓨팅 인프라 구축 중

○ (예산) '15년 2,829백만 원, '16년 2,565백만 원, '17년 2,138백만 원 투입('기초연구 실험데이터글로벌허브구축사업' 예산)

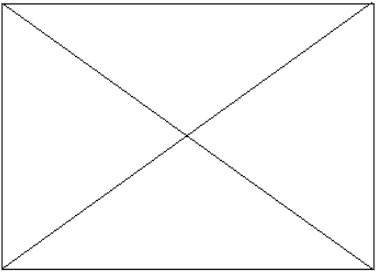
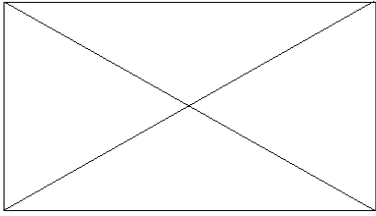
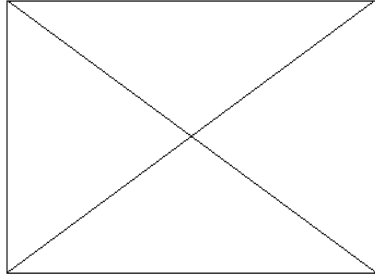
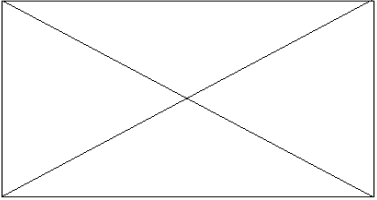
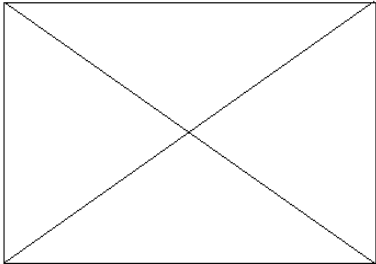
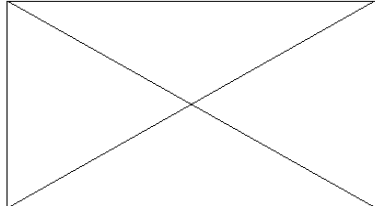
○ (인프라) 분야별 데이터 증가, 분석 수요 및 데이터 전송을 위한 인프라 구축

<그림 3-7> 대형장비 분야 인프라 구축 내용

실험데이터 증가 대응을 위한 인프라	데이터 분석 수요 대응을 위한 데이터공유·분석환경 인프라	안정적인 실험데이터 전송을 위한 10Gbps 데이터전용망
• '16년 현재까지 40.7 PB 데이터 생산(ALICE 6.5, ATLAS 14.5, CMS 13.1, LHCb 6.6) • '11년 15PB, '12년 23PB, '13년 27PB, '15년 50PB 생산) • '16년 8-9월 하루에 400TB 이상 데이터 생산(RUN1의 경우 하루 최대 데이터 생산 220TB) • '17년의 경우 작년대비 데이터 생산량 감소: '17년 10월까지 27PB 데이터 생산 (ALICE 3.2PB, ATLAS 13.1PB, CMS 7.2PB, LHCb 3.6PB)	• '15년 CPU 560코어, 2016년 684코어, 2017년 1,260코어 등 지속적으로 컴퓨팅 인프라를 확장 - '11년 이후 지속적으로 컴퓨팅인프라를 확장하여 구축하였으며, 2017년 현재 1만코어 규모의 시스템 운영 (관리노드, 서비스노드 포함) • '11년 이후 실험데이터 증가 대비 저장인프라를 확장구축, 특히 '16년 Disk 3PB, Tape 1.5PB 확장, '17년 Disk 1.5PB 확장 - 현재 디스크 10PB, 테이프 3.2PB, 총 저장규모 13.2PB	• '15년 KISTI에서 미국 시카고까지 기존 대역폭 2Gbps 구간을 10Gbps로 확장하고, 네덜란드 '서프넷(SURFnet)'과의 협약으로 시카고에서 제네바(CERN)까지 10Gbps 대역폭 확보 - 기존 전용망(2Gbps)을 사용할 때 (약 289일)보다 데이터 전송시간을 10분의 1 이하로 절감하는 효과 발생*(10Gbps의 가속기데이터 전용망 구축으로 1천TB 전송 시 23일 소요)

- 대형장비 분야는 데이터집약형 연구 지원을 위해 실험분야별 연구데이터 공유·활용 시스템 구성 중

<표 3-12> 대형장비 분야별 시스템 구성도

KIAF(ALICE)	중력파(LIGO/KAGRA)	Belle II
		
RENO	유전체	전자현미경
		

- 인프라 현황기초과학연구 선진화를 위한 데이터집약형 연구지원 인프라 구축 및 운영
 - 데이터 분석 수요 대응을 위한 데이터 공유분석 인프라 구축
 - (컴퓨팅 인프라 확장) '15년 CPU 560코어, '16년 684코어, '17년 1,260코어
 - (저장인프라 확장) '16년 Disk 3PB, Tape 1.5PB 확장, 2017년 Disk 1.5PB 확장
 - ※ 1.5PB의 CERN 가속기 원천데이터 확보 및 테이프시스템의 활용이 90%를 넘어, 추가적으로 1.5PB 테이프 저장인프라를 확장하여 3PB 규모의 원천데이터 확보 가능

<표 3-13> 저장인프라 전체 규모 및 사용량

구분	Disk		TAPE	합계
	SAN	NAS		
저장인프라 전체규모	5.5PB	4.5PB	3.2PB	13.2PB
저장인프라 사용량	3.8PB	4.5PB	3.2PB	11.5PB
가용량	1.7PB	—	—	1.7PB

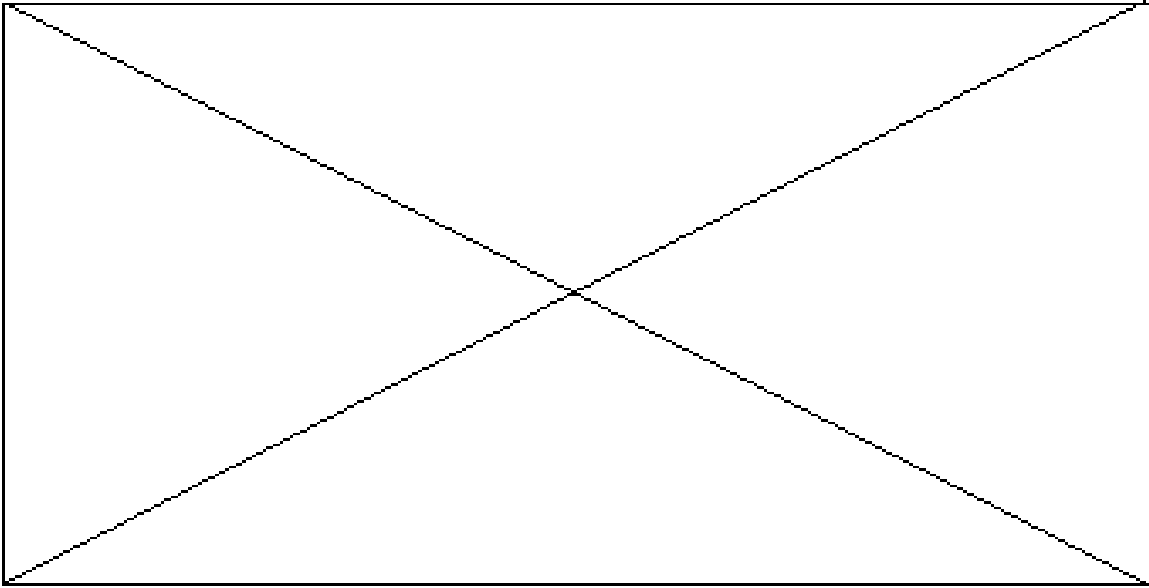
- (수집·관리 현황) 한국과학기술정보연구원은 ALICE, CMS 등 연구데이터 공유분석 환경을 제공하여 국내외 연구자 사용 중

표 3-14> 분야별 국내외 참여연구자 규모

실험명	참여연구자 규모
ALICE 실험(CERN)	• (국외) 37개국 151개 기관, 1,550명 참여 • (국내) 한국 ALICE 연구그룹(대표: 윤진희 인하대 교수) – 국내 6개 대학, 40여명(인하대, 부산대, 연세대 등)
CMS 실험(CERN)	• (국외) 40개국 172개 기관, 약 3,000명 참여 • (국내) 한국 CMS 연구그룹(대표: 양운기 서울대 교수) – 국내 8개 대학, 80여명(서울대, 고려대, 성균관대, 서울시립대, 경북대 등)
Belle II(일본 KEK)	• (국외) 20개국 67개 기관, 600여명 참여 • (국내) 한국 Belle 연구그룹(대표: 원은일 고려대 교수) – 국내 9개 대학 40여명(고려대, 숭실대 등)
중력파(LOGI/KAGRA)	• (국외) 16개국, 60개 기관, 1,000여명 • (국내) 한국중력파연구협력단(단장: 이형목 서울대 교수) 및 해외 공동연구자 – 10개 대학, 4개 연구소, 40여명 (서울대, 부산대, NIMS 등)
RENO	• (국내) 한국중성미자연구센터(대표: 김수봉 서울대 교수) – 국내 5개 대학 50여명(서울대, 전남대 등)
유전체	• (국내) 유전체 연구그룹(대표: 김형래 이화여대 교수) – 국내 40여명(이화여대, 서울대, 국립암센터, 삼성병원 등)
구조생물학(KBSI 전자현미경)	• (국내) 한국기초과학지원연구원(KBSI), KAIST 등 15개 연구그룹 50여명

□ (한국생명공학연구원) 생명연구자원통합정보시스템을 구축하여 생명연구 분야 연구데이터 수집·활용 지원 중

- (예산) ‘생명연구자원 통합시스템 활용기반 구축사업’으로 ’17년 25.8억 원 투입
- (인프라) CPU 3,100 cores, 저장 공간 10PB(페타바이트) 운영 중(’18.1.1. 기준)
 - 구축된 KOBIC 시스템 활용으로 생명연구자원 수집·관리 및 서비스지원 중



<그림 3-8> KOBIC 운영시스템 정보연계 추진체계

- (수집·관리 현황) 생명연구자원통합정보시스템 활용으로 5개 부처 유전체 정보 및 생명연구자원정보의 수집·관리 및 서비스지원 중

<표 3-15> 생명연구자원통합시스템의 생명연구자원 수집·관리 현황

(단위: 건, 점)

지표[지표유형]	2015		2016		2017		2018.04	
	목표	실적	목표	실적	목표	실적	목표	실적
생명연구자원정보 연계(누적)	—	3,166,417	3,260,000	2,336,451	3,360,000	3,199,049	3,406,667	4,837,155
생명정보 분석·활용 서비스 지원	—	50	55	55	60	63	21.67	27

— '16년부터 5개 부처 유전체정보 책임관리기관과 다부처 유전체 정보 연계 중('18년 4월 기준 14,568건, 82TB)

〈표 3-16〉 부처별 데이터의 KOBIC 연계 현황

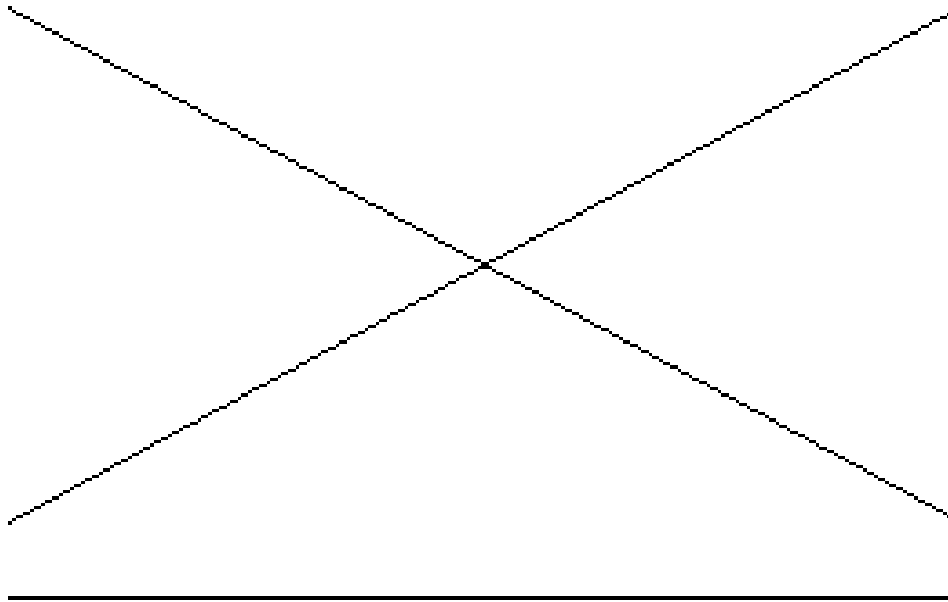
(단위: 개, G)

등록 년도	과기정통부 (국가생명연구 자원정보센터)		농식품부 (국립농업생명 공학정보센터)		해수부 (국가해양생명 자원센터)		보건복지부 (국립보건원 생명정보연구과)		산자부 (한국바이오안 전성정보센터)		소계	
	파일 수	용량	파일 수	용량	파일 수	용량	파일 수	용량	파일 수	용량	파일 수	용량
'16년	1,626	2,363	0	0	0	0	0	0	1	3	1,627	2,367
'17년	2,733	8,868	877	6,144	757	16,384	4,348	39,936	2,173	2,560	10,888	73,892
'18년*	1,087	2,194	966	3,912	0	0	0	0	0	0	2,053	6,107
소계	5,446	13,426	1,843	10,056	757	16,384	4,348	39,936	2,174	2,563	14,568	82,365

*부처별 '17년 생산된 유전체 데이터에 대한 수집 및 연계 작업은 진행중에 있으며, 5월 말까지 완료예정

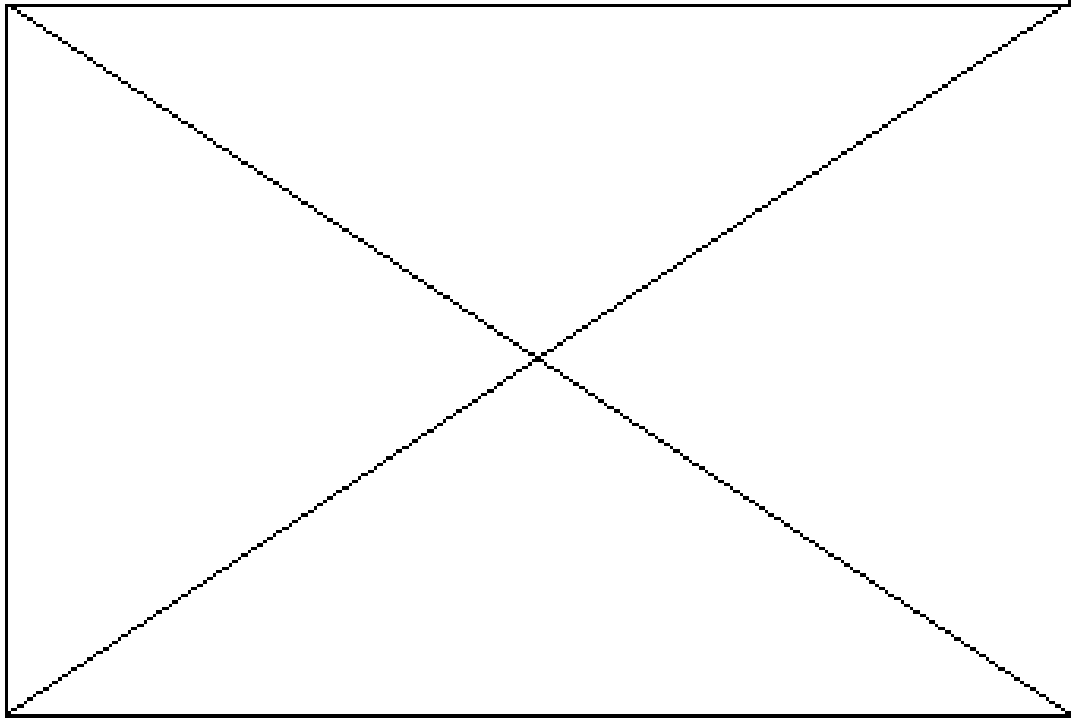
□ (한국지질자원연구원) 'KIGAM 지질자원정보 통합관리체계 구축을 위한 중장기 추진전략('15~'26)'에 따라 규정·조직 등 제도적 정비와 더불어 물리적/디지털 리포지토리를 구축하여 파일럿 서비스 중

- (예산) '15년 275백만 원, '16년 360백만 원, '17년 210백만 원 투입
- (인프라) 기관 자체적으로 DMP양식을 포함한 연구데이터 관리규정을 마련하여 시행 중이며, '16년 웹기반의 디지털 지질자원 연구데이터 리포지토리 시스템인 Geoscience Data Repository System(GDR)을 개발하여 현재 1차 파일럿 시스템을 서비스 중
- (수집·관리 현황) '18년까지 기반구축 단계로 디지털 리포지토리 시스템(소프트웨어) 개발 후 2단계('19~'22)부터 본격적으로 지질자원분야 연구데이터 수집 및 등록 예정



<그림 3-9> KIGAM 지질자원정보 통합관리체계

- (한국항공우주연구원) 저궤도 위성 다목적실용위성 시리즈와 정지궤도 위성 천리안 등 여러 관측위성을 통해 연간 약 10만 장의 위성영상을 확보하여 공공·민간부문에 위성데이터 서비스를 제공
 - (인프라 현황) ‘위성정보 3.0’ 추진에 따라 국가위성정보활용지원센터를 통해 국가 위성 정보의 통합적 관리를 위한 통합 데이터베이스 및 플랫폼 구축
 - － (국가위성정보활용지원센터 설립) 기존의 부처별 목적에 따른 위성영상 활용에서 범정부 차원의 위성정보 통합 관리·활용체제로 전환하고자 전문기구 설립(’15년)



<그림 3-10> 위성정보 통합플랫폼 제공 데이터

<표 3-17> 위성영상정보 관리·공급 및 활용체계

위성	수신	가공·공급(이전)		가공·공급(개편)		활용
다목적 실용위 성	과기정통부 (항우연)	항우연 (상용 공급은 대행업체)	→		→	23개 정부기관(위성정보활용 협의체) 지자체, 연구기관 등
						국내·외기업 등
천리안 위성	기상청 (국가기상위성센 터)	기상위성센터	→	국가위성정보활용 지원센터 (인공위성 통합 플랫폼)	→	공공·연구기관, 기업, 일반인 등
	해양수산부 (해양위성센터)	해양위성센터	→		→	
과학기 술 위성	KAIST (인공위성센터)	항우연/천문연	→		→	대학, 연구기관 등

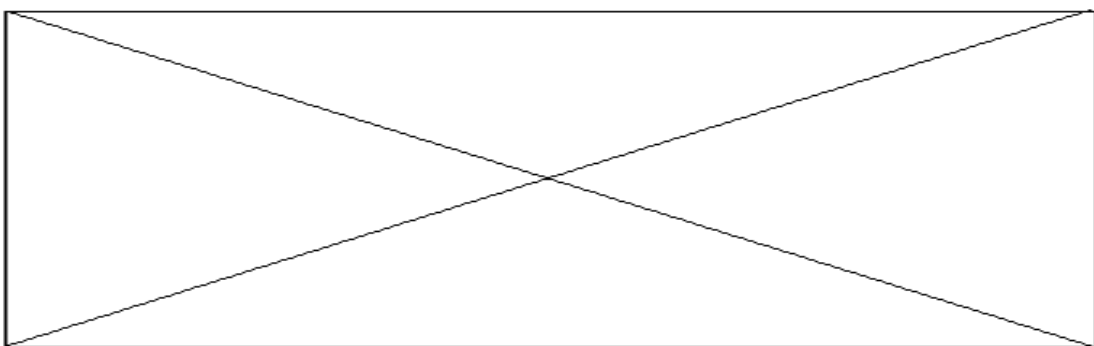
－(개방형 위성정보 통합플랫폼 구축) 국가 주요 공공재인 위성정보*의 통합(one-stop) 공급망이자 위성정보 기반의 새로운 가치창출의 장으로서 정보 플랫폼 구축 ('14~'17년)

- 인공위성을 이용하여 구축한 영상·음성·음향데이터 또는 이들의 조합으로 처리된 정보 (이를 가공·활용한 것을 포함(우주개발진흥법 제2조))

<표 3-18> 위성정보 통합플랫폼 인프라 구축내용

단계	추진 내용
1단계 (‘14~’15)	■ 위성정보 one-stop 유통시스템 구축 - 기존 전문가 중심의 다목적 위성영상 검색시스템을 개선, 다양한 수요에 부응하는 통합적 one-stop 검색환경 구축 - 통합플랫폼은 위성정보의 메타데이터(본 위성정보와 연결되는 기본 요약정보)만 제공하고 본 위성정보를 저장하고 있는 각 위성정보센터 DB와 연결하여 정보 공급
2단계 (‘16~’17)	■ 다중위성시대를 대비한 융복합형 통합플랫폼 구축 - 국내외 가용 위성정보의 통합 수신·저장 및 각 위성정보 활용센터와 유기적 연계시스템 구축(위성정보 유통 포맷 표준화 포함) ※ 신규 정지궤도 위성 및 차세대중형위성이 발사되는 '18년을 대비, 통합플랫폼 중심의 국가 위성정보 관리·활용체제 구축 - 과학기술 빅데이터 플랫폼('18년)과 연계, R&D 성과와 위성정보간 융복합을 통해 창조적 융합연구가 촉발·대동하는 창조 빅뱅플랫폼으로 진화토록 개발

- (데이터 현황) 800건의 위성영상데이터를 보유하고 있으며, 그 중 588건의 위성영상이 지리·해양·토양·재해·환경 등 다양한 부문에 활용됨
- － (데이터 유형) 원시데이터를 보정 처리한 중간데이터만 공급함으로써 중간데이터에는 기초보정데이터, 정밀보정데이터, 모자이크데이터를 포함함
 - * 기초보정데이터 : 기기·태양각 등에 따른 오차 제거 및 공간적 왜곡 보정 데이터
 - * 정밀보정데이터 : 지형의 고도 차이 등을 반영한 보정 데이터
 - * 모자이크데이터 : 지리적으로 인접한 정밀보정데이터를 이어 붙여 만든 데이터



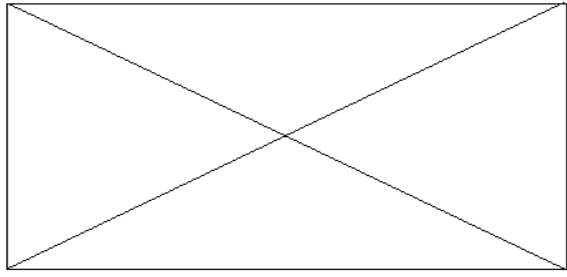
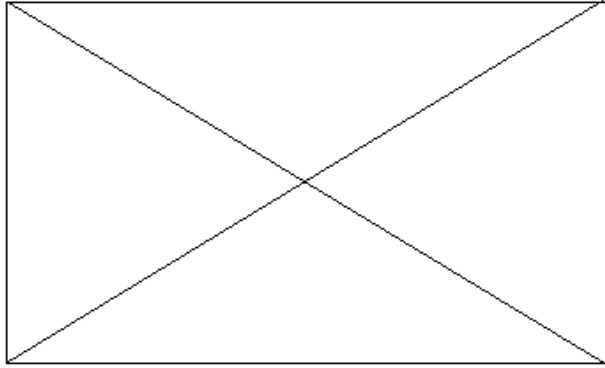
<그림 3-11> 위성정보 통합플랫폼 제공 데이터

- (서비스 현황) 위성정보 통합 DB 서비스 제공
 - － 위성정보 검색·주문·보급 단계별 자동처리 SW 및 대용량 위성정보 고속전송 기술, 다중 위성정보 융합처리 기술을 병행 개발, 정보 품질 및 공급 신속성 제고
 - － 위성정보를 지역·시간·센서·활용빈도 등의 기준으로 분류·테마화한 직관적 검색화면 제공

- \ - 한반도·지구전경 등의 기본적 위성정보와 사용자 눈높이 맞춤형 가공 SW 제공으로, 위성정보 활용 아이디어 구현 지원
- 위성정보별 부가정보(활용사례·관련 기술정보 등) 및 시계열·융복합* 정보 등 신규서비스 개발 중
 - * 해운정보서비스 = 해류해빙정보(관측위성) + 선박위치정보(항법위성) + 통신(통신위성)

- (한국표준과학연구원) '20년까지 신뢰성 있는 측정데이터 생산능력을 갖춘 연구데이터를 연구원 내 데이터센터(KRISS Data Center; KDC)를 5개까지 지정하는 것을 목표로 현재 시범사업 운영 중
 - (인프라) 한국표준과학연구원의 연구데이터 공유·활용 체계는 기획 단계로 현재 KDC 플랫폼의 구성 및 역할(안)까지 도출됨

<표 3-19>

KDC 플랫폼 구성도	KDC 플랫폼의 역할
	

- (수집·관리 현황) 현재 수집·관리 중인 연구데이터는 없고, '17년부터 1개 분야 선정하여 시범사업 수행 중(1단계)으로, 향후 인프라 개선 및 분야 확장(2단계) 실시로 총 5개 KDC 운영 목표 중
 - 현재 금속 박막 복소굴절률 측정 데이터를 시범사업으로 운영 중
 - 데이터마다 데이터의 특성에 따른 평가를 실시하여 고품질의 연구데이터 중심으로 KDC 구축 예정

<표 3-20> KDC 후보데이터 현황

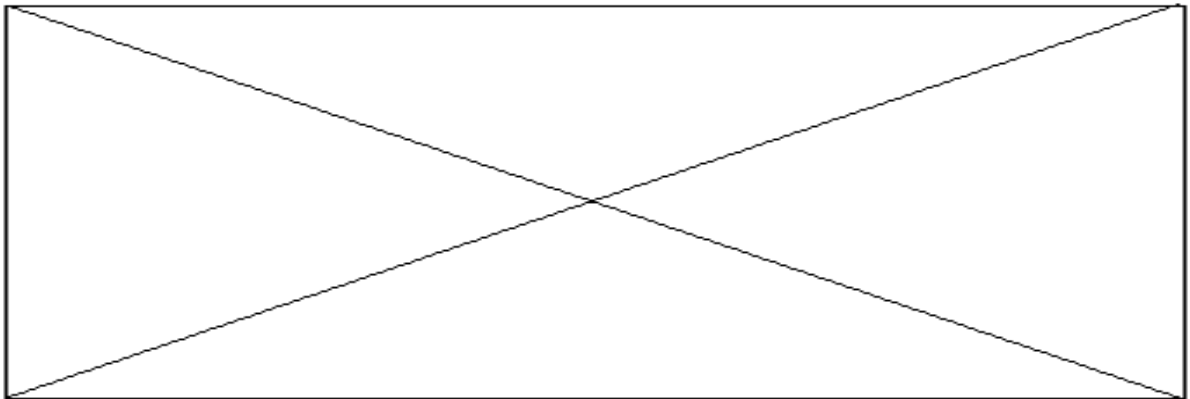
	데이터명	용도/활용처	산업분야
1	온실가스 온난화 지수	대기환경측정기준	환경
2	액체 열물성 데이터	열물성 측정 신뢰성 확보	첨단
3	복사물성 특성	재료공학	첨단
4	금속/기체 불순물 이온도 고정점에 미치는 영향 Matrix 개발	온도표준 및 불확도 향상	첨단
5	나노물질 안전 기준	국가의 나노 안전관리	첨단
6	나노 스케일 제백 계수	열전발전/냉각	첨단
7	진공재료물성	시험, 산업체 지원	첨단
8	식각공정데이터	식각공정 모델링, 산업체	첨단
9	의료 관련 데이터(신체, 장기 등)	의료, 복지	보건의료
10	한국인 성인 골밀도 데이터	골밀도 측정표준화	보건의료
11	성인 고혈압 환자 혈압 데이터	혈압계 측정 표준화 및 교정체계 구축	보건의료
12	0-4세, 영·유아 MRI 뇌지도(Brain Atlas) 개발	유아의 장애 조기진단 가능	보건의료
13	조직모사 초음파 재료의 음속 및 감쇠계수	팬텀재료/제조사	보건의료
14	방사선 의료장비 품질보증용 데이터	의료장비의 검증/병원	보건의료
15	기체 및 액체 방사성 물질의 방사능 측정용 데이터 (CRM)	측정장비의 교정 및 검증/원자력 발전소	보건의료
16	핵반응 데이터	신뢰성 있는 데이터 제공	보건의료
17	물질에 따른 이원 전자발생 효율 RBI 발생 효율, RBS 분산	금속재료 특성, 바이오 물질 특성	기간

- ☐ (한국한의학연구원) 현재 연구데이터 공유·활용 전략 수립을 위해 TFT를 구성하고 설문 및 자료조사를 실시하였고, 이후 한국과학기술정보연구원의 연구데이터 리포지시스템을 활용하여 연구원 환경에 맞도록 개발 예정임
- (예산) 연구개발적립금 사업(한의학 연구데이터 통합관리 체계 구축) 예산 투입 예정 (안)
 - (인프라) 한국과학기술정보연구원의 연구데이터 허브센터 및 연구데이터 플랫폼센터 등 구축 시 연계하여 연구데이터 리포지토리 시스템 설계 예정

☐ (극지연구소) ‘남극조약협의당사국’에 따라 ‘11년부터 연구데이터 수집·공유 중²²⁾

22) 극지연구소는 남극조약협의당사국(Antarctic Treaty Consultative Party, ATCP)에 따라 극지활동을 통해 수집되는 연구테

- \ ○ (수집·관리 현황) 연구소 내 사업 기획 시 DIF 포맷 사용의 DMP 작성 목표를 제시하고, 사업 수행 후 평가에 반영 중
- (수집 대상) 지상고층 기상, 전파레이더, 중력계, 지진계, GPS, 해수, 지형, 탄성과 등의 디지털 데이터와 빙하, 퇴적물, 운석, 암석, 화석, 해수, 동물 등 아날로그(실물) 데이터
- (수집 방식) 메타데이터는 DIF(Directory Interchange Format) 표준을 준수하고, 3가지 등급으로 구성하여 수집·공유 중
- 데이터관리계획서(DMP) 포맷



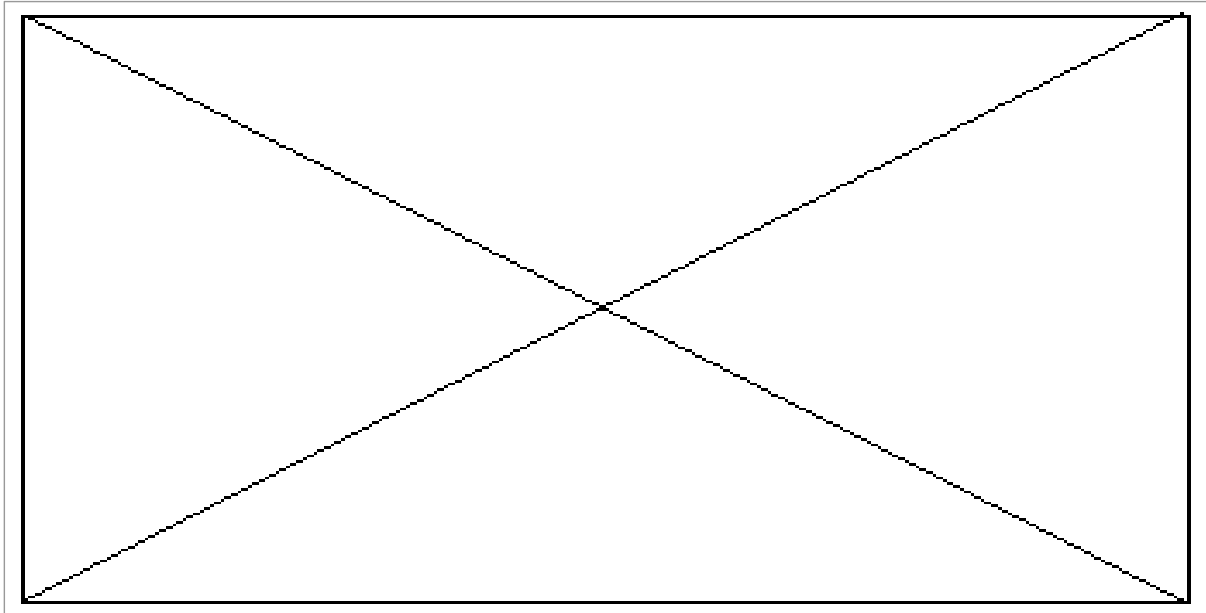
<그림 3-12> 극지연구소의 데이터관리계획서(DMP) 포맷

• 메타데이터 항목

<표 3-21> 메타데이터의 등급별 내용

Required	Highly Recommended	Recommended
• Entry ID • Entry Title • Science Keywords • ISO Topic Category • Organization • Summary • Related URL • Platform/Instrument • Temporal Coverage • Spatial Coverage • Project • Metadata Dates	• Dataset Citation • Personnel • Location • Data Resolution • Quality • Access Constraints • Use Constraints • Distribution Information • Dataset Progress • Dataset Language	• Originating Center • Multimedia Sample • Metadata Association • IDN Node • DIF Revision History • Version Description • Additional Attributes • Product Level Id • Collection Data Type • Extended Metadata • Ancillary Keyword • Publication/Reference

- (공유·활용 현황) 연평균 5회 정도(최대 10회) 연구데이터 공유 요청 중이고, KPDC 홈페이지에서 메타데이터 검색 가능
- '12년부터 시스템화하여 연구데이터 관리 중



<그림 3-13> KPDC 홈페이지 내 메타데이터 검색 화면

- 국가과학기술연구회 소속 25개 정부출연연구소 중 8개는 독립적으로 연구데이터 관리·공유 체계를 구축·운영 중으로 각 연구소의 연구데이터를 연계하고, 관리·공유 체계 미보유 연구소의 연구데이터를 수집하는 국가 차원의 연구데이터 수집·활용 체계 마련 필요
- (정부출연연구소) 국내 국가과학기술연구회 소관 25개 정부출연연구소 중 한국과학기술정보연구원, 한국생명공학연구원, 한국화학연구원 등 00개 연구소는 독립적으로 연구데이터 수집·활용을 위한 인프라 보유 중

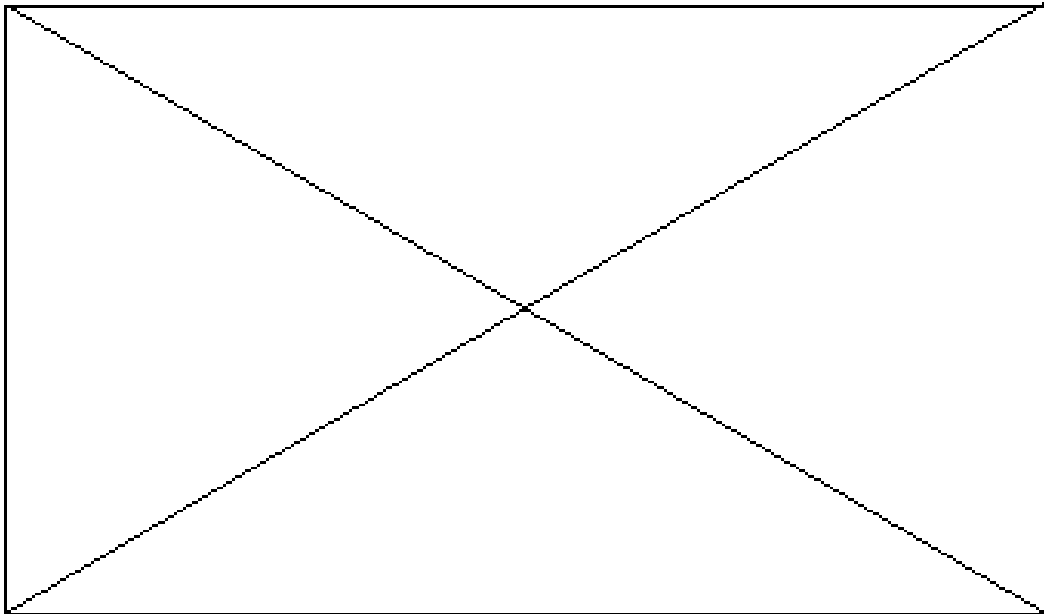
나. 분야별 전문기관의 연구데이터 관리·공유 체계 현황

- (정보화진흥원) 「인공지능 오픈이노베이션 허브 구축」 사업으로 AI기술을 개발·응용하는 주체(중소벤처, 연구소, 대학 등)들이 AI의 핵심자원을 원스톱 서비스 받을 수 있는 지원 플랫폼을 구축하여 연구데이터
- (예산) ‘지능정보산업인프라조성’ 사업으로 '17년 4,870백만 원 투입
- (인프라) 하드웨어는 비용을 절감하고 운영 효율을 위하여 x86급 민간 클라우드 자원을 임차하여 운영 중이고, 소프트웨어는 Open CMS인 Drupal²³⁾ 기반으로 구축·운영 중

<표 3-22> 정보화진흥원의 인공지능 분야 H/W·S/W 자원 현황

H/W	CPU	Memory		Disk
	4 Core 수준	8GB		500G
S/W	OS	DB	Web Server	CMS
	CentOS 6.9 (1식)	MariaDB 10.0.33 (1식)	Apache 2.2.15 (PHP 5.6.19) (1식)	Drupal 7.43 (1식)

- － (AI 컴퓨팅 자원관리)중소벤처, 스타트업 대상 AI용 컴퓨팅 자원(GPU 서버 등) 신청 접수, 자원할당, 사용 모니터링 체계 구축



<그림 3-14> 정보화진흥원의 AI 오픈 이노베이션 허브 시스템 개념도

- (수집·관리 현황) 회원 총 316명, 학습 데이터 파일 다운로드 총 440건 등의 성과 창출 ('18.4.10 기준)

<표 3-23> 정보화진흥원의 AIHub 가입자 현황

소속기관 분류	가입자 수	대표기관
중소기업	145	셀바스AI, 솔트룩스, ㈜웍스, 다나와 등
개인사용자	52	개인연구자 등

23) Open CMS(Contents Management System)인 Drupal로 제작·운영 중

소속기관 분류	가입자 수	대표기관
정부/공공기관	42	국회도서관, 소프트웨어정책연구소, 사회보장정보원 등
대학교	36	한국과학기술원, 고려대학교, 인하대학교 숭실대학교 등
대기업	24	현대자동차, 삼성전자, LG전자, KT, 현대오트모에버 등
기타	11	실로암장애인복지관, 경일관광고등학교 등
연구기관	6	지능정보기술연구원, 원주의료기기테크노벨리 등
합계	316	

－ 법률/특허/일반상식/한국형 이미지 분야별 데이터 구축 성과 창출

<표 3-24> 정보화진흥원의 AI 학습추론 데이터 구축·제공 현황

분야	데이터 종류	파일 형식	수량
법률	법령용어	json, N-triple	11,000건
	생활용어	json, N-triple	106,252건
	법률 지식베이스	json, N-triple	3,000건(표제어) 330,000건(트리플)
특허	특허 지식베이스	RDF	1,013,075건(표제어) 6천만건 이상(트리플)
	기술용어	RDF	120,687건
	도면태그정보	RDF	2,338만건
일반상식	일반상식 지식베이스	json	161,651건(표제어) 750,000건(트리플)
	일반상식 질문-답변-제시문	json	100,000건(MRC 데이터)
한국형 이미지	한국인 얼굴 이미지	jpg	6,480,000장
	한식 이미지	jpg	150,000장

－ 오픈 API 서비스 : 총 580 건 신청(호출 건수 기준으로 총 1,503,770 건 (일 평균 9,114 건))

표 3-25> 정보화진흥원의 오픈 API 서비스 신청 현황

소속기관 분류	신청자 수	대표기관
산업체	270	SKT, 삼성SDS, 삼성전자, 삼성증권, 엔씨소프트, 네이버 등
대학교	171	서울대, 연세대, 한양대, KAIST, 서강대, 충북대, 동국대 등
출연연	7	ETRI, KIST, 기계연, KISTI 등
기타(비영리)	29	소프트웨어정책연구소, 지능정보기술연구원 등
개인	103	개인연구자 등
합계	580	

3.1.4. 연구데이터 관련 특허 분석

가. 국가별 특허 경쟁력

국가 연구데이터 공유·활용체계 분야의 특허출원 분석결과, 모든 구간 및 모든 특허청에 걸쳐 최근 특허출원 활동이 성장하고 있어 우리나라의 연구개발 투자를 통한 기술력 확보가 필요한 것으로 판단

우리나라는 시장확보력 수준에서는 일정 부분 경쟁력이 있었으나, 특허출원의 양적 수준과 질적 수준이 높은 미국, 영국 등 국가 연구데이터 공유·활용체계 선도국 대비 양적·질적 수준에서 모두 열위하여 선도국과의 경쟁력 확보를 위해 해당 분야에 대한 집중 투자를 통한 정책적 육성이 필요

□ 연구데이터 공유·활용체계 기술은 기술의 양적·질적 수준에서 미국이 가장 선도적인 기술을 보유하고 있음

- 미국은 기술력(TS) 지수, 특허 인용도(CPP) 지수, 특허영향(PII) 지수에서 모두 일본, 독일, 한국, 영국 국가 대비 높은 수준을 나타냄
- 일본, 독일, 한국은 시장확보(PFS) 지수와 특허인용도(CPP) 지수가 기술력(TS) 지수와 특허영향(PII) 지수보다 높은 편
- 영국은 시장확보(PFS) 지수가 4.70으로 5개 국가 중 가장 높고, 기술력(TS) 지수, 특허 인용도(CPP) 지수가 미국 다음으로 높음
- 우리나라는 선도국 대비 시장확보 지수에서는 상대적으로 강점을 보유하고 있으나, 인용도 지수 및 특허영향 지수에서는 상대적으로 낮은 것이 기술의 질적 수준이 열위임

<표 3-26> 등록특허 상위 5대 국가 주요 특허지수 현황

국가	기술 경쟁력 판단 지수		기술 질적 수준 판단 지수	
	기술력(TS) 지수	시장확보(PFS) 지수	특허 인용도(CPP) 지수	특허영향(PII) 지수
미국	6973.35	2.42	7.79	1.17
일본	0.51	3.15	3.51	0.53
독일	0.37	2.76	3.43	0.52
한국	0.65	3.52	2.94	0.44
영국	4.61	4.70	4.33	0.65
기술력(TS) 지수			시장확보(PFS) 지수	
특허 인용도(CPP) 지수			특허영향(PII) 지수	

- 연구데이터 공유·활용 체제 구축 관련 핵심기술별로는 미국이 기술력(TS) 지수, 특허 인 용도(CPP) 지수, 특허영향(PII) 지수가 모두 1위임
 - 시장확보(PFS) 지수는 핵심기술별로 연구데이터 수집·연계 및 공유 기술과 품질·표준화 기술은 영국이, 저장·보안 기술과 고속검색 기술은 한국, 융합분석 환경 구축 기술은 일 본이 1위임

<표 3-27> 핵심기술별 상위 5대 국가 주요 특허지수 현황

핵심기술	국가	기술 경쟁력 판단 지수		기술 질적 수준 판단 지수	
		기술력(TS) 지수	시장확보(PFS) 지수	특허 인용도(CPP) 지수	특허영향(PII) 지수
연구데이터 수집·연계 및 공유 기술	미국	4987.56	2.43	7.86	1.14
	일본	0.51	3.14	3.62	0.53
	독일	0.37	2.79	3.39	0.49
	한국	0.65	3.11	3.10	0.45
	영국	4.61	4.22	5.93	0.86
연구데이터 저장보안 기술	미국	2384.81	2.32	8.37	1.18
	일본	0.51	3.08	3.71	0.52
	독일	0.37	2.91	4.80	0.68
	한국	0.65	4.21	3.93	0.56
	영국	4.61	2.98	3.87	0.55
연구데이터 품질·표준화 기술	미국	3441.16	2.44	6.87	1.22
	일본	0.51	3.03	2.94	0.52
	독일	0.37	2.83	2.24	0.40
	한국	0.65	3.68	2.67	0.47
	영국	4.61	4.85	6.38	1.13
지능형 연구데이터 고속 검색 기술	미국	38.21	2.31	7.21	0.98
	일본	0.51	3.90	1.40	0.19
	한국	0.37	4.33	1.33	0.18
	영국	0.65	4.00	3.00	0.41
	캐나다	4.61	2.50	1.50	0.20
연구데이터 융합분석 환경 구축 기술	미국	916.16	2.41	8.19	1.06
	일본	0.51	3.53	3.97	0.51
	독일	0.37	1.84	2.09	0.27
	한국	0.65	2.96	1.71	0.22
	아일랜드	4.61	2.15	5.20	0.67
연구데이터 활용·플랫폼 기술	미국	996.93	2.29	9.48	1.18
	일본	0.51	3.33	4.66	0.58
	독일	0.37	2.21	4.14	0.52
	한국	0.65	2.62	1.52	0.19
	캐나다	4.61	3.37	8.26	1.03

나. 국가별·기술분야별 특허동향

- 국가 연구데이터 공유·활용체제 6대 기술분야 중 ‘지능형 연구데이터 고속 검색 기술’분야를 제외한 5대 기술분야는 특허출원이 활발하게 진행 중
- 최근 특허출원 추세를 고려할 경우 「연구데이터 수집·연계 및 공유 기술」과 「연구데이터 저장·보안 기술」 및 「연구데이터 품질·표준화 기술」 분야가 국가 연구데이터 공유·활용체제 기술의 중점 분야로 당분간 본 기술분야의 성장을 주도할 것으로 전망
 - 「연구데이터 수집·연계 및 공유 기술」과 「연구데이터 저장·보안 기술」 및 「연구데이터 품질·표준화 기술」 등 3개 기술분야의 경우, 국가 연구데이터 공유·활용체제의 전체 특허 중 84%를 점유하고 있으며, 연도별 추세에서도 안정적인 성장세를 나타내며, 향후에도 국가 연구데이터 공유·활용체제 기술혁신을 주도할 것으로 전망됨
 - 「지능형 연구데이터 고속 검색 기술」 분야의 경우 최근 10년 간 출원된 특허건수가 230건에 불과, 현재까지 기술혁신은 더디게 진행 중인 것으로 판단됨

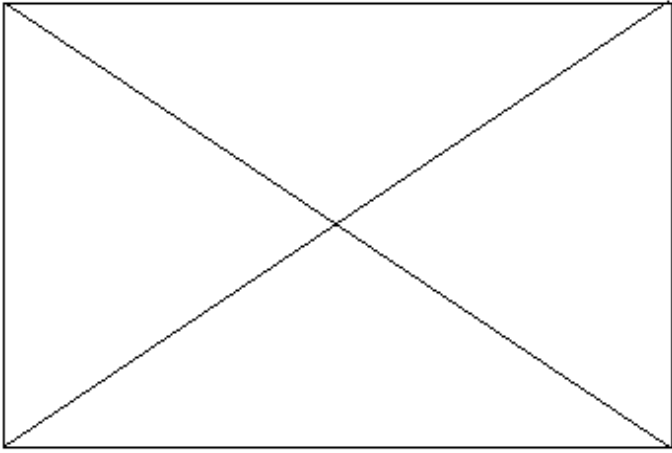
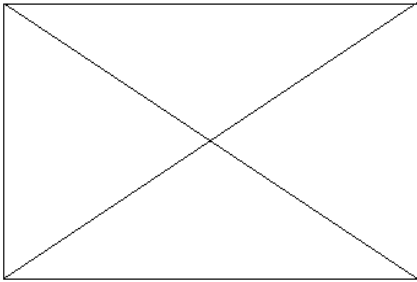
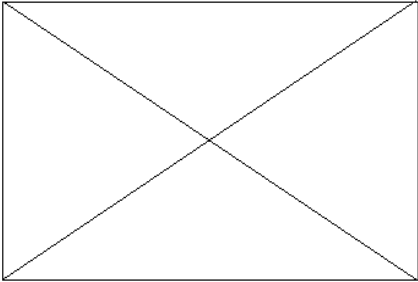
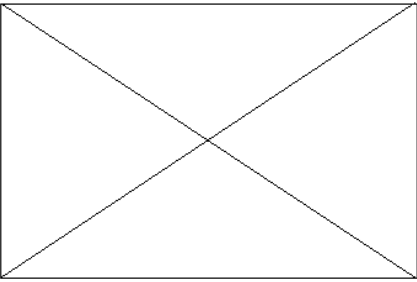
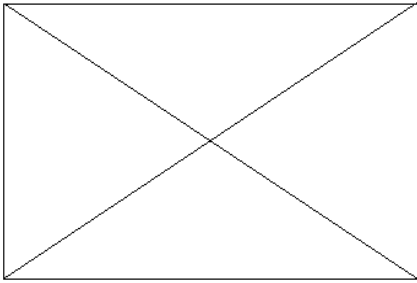
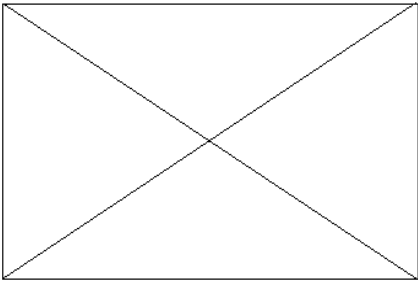
<표 3-28> 기술분야별 연도별 특허출원 현황 및 추이

구분	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	합계
연구데이터 수집·연계 및 공유 기술	1,317	1,237	1,300	1,487	1,573	1,996	2,248	3,059	2,473	2,590	19,280
연구데이터 저장·보안 기술	695	685	769	812	911	1,105	1,190	1,788	1,338	1,443	10,736
연구데이터 품질·표준화 기술	1,316	1,305	1,289	1,522	1,432	1,753	1,812	2,123	2,030	1,979	16,561
지능형 연구데이터 고속 검색 기술	22	13	9	22	14	30	32	26	33	29	230
연구데이터 융합·분석 환경구축 기술	253	249	298	313	327	417	507	682	726	805	4,577
연구데이터 활용·플랫폼 기술	390	358	366	385	384	499	483	567	509	642	4,583
합계(기술 간 중복특허제거)	2,234	2,167	2,313	2,568	2,622	3,293	3,576	4,755	4,047	4,174	31,749

기술분야별 특허출원 비율(2016년 누적 기준)	연도별 기술분야별 특허출원 추이

- 국가 연구데이터 공유·활용체계의 국가(특허청)별 포트폴리오 분석 결과 일본 특허청을 제외한 모든 특허청에서 ‘발전기’ 단계임
- 국가 연구데이터 공유·활용체계 기술분야는 최근 부각되고 있는 분야로서 일본특허청을 제외한 한국/미국/유럽/PCT 등 4대 주요 특허청에서 모두 기술개발 활동의 증가와 시장의 신규 진입자가 지속적으로 증가하는 ‘발전기’에 있는 것으로 분석
 - (출원건수) 기술개발의 활동정도 / (출원인수 증가) 시장의 신규 진입자의 증가를 의미
 - 일본특허청의 경우 비교적 활발한 특허활동에도 출원건수 및 출원인수가 증가와 감소를 반복하여 ‘발전기’ 단계에서 ‘성숙기’ 단계로 진입하는 양상을 보이는 것으로 판단

<표 3-29> 국가 연구데이터 공유·활용체계 기술분야 국가별 포트폴리오

			기술분야 전체
			1구간 : “07~”08 2구간 : “09~”10 3구간 : “11~”12 4구간 : “13~”14 5구간 : “15~”16
Total			
			
한국특허청(KIPO)	미국특허청(USPTO)	일본특허청(JPO)	
			
유럽특허청(EPO)	국제특허/PCT(WIPO)		

3.1.5. 연구데이터 관련 논문 분석

가. 국가별 논문 경쟁력

□ 논문영향력, 활용도, 집중도 3가지 측면에서 모두 우리나라는 일본과 비슷한 수준이나 가장 높은 수치인 미국과 큰 격차를 보이며 연구데이터 부문 국가 논문 경쟁력이 낮은 수준임

- (논문영향력) 특정국가 논문의 질적 수준을 다른 국가와 상대적으로 비교할 수 있는 정보 제공

논문영향력	$\frac{\text{특정국가 논문 인용 수}}{\text{총 논문 인용 수}}$

- (논문영향력 비교) 미국이 독보적으로 높은 수치를 보인 반면, 우리나라는 0.02로 논문 경쟁력에서 큰 격차를 보임
 - － (해외) 미국 : 0.34 / 영국 : 0.11 / 중국 : 0.07 / 호주 : 0.05 / 일본 : 0.02
 - － (국내) 연구데이터 논문영향력은 0.02 수준으로 일본과 비슷한 수준을 나타냄
- (논문활용도) 논문 게재 건수

논문활용도	$\frac{\text{특정국가 논문 수}}{\text{총 논문 수}}$

- (논문활용도 비교) 미국이 독보적으로 높은 수치를 보인 반면, 우리나라는 0.02로 논문 활용도에서 큰 격차를 보임
 - － (해외) 미국 : 0.28 / 중국 : 0.11 / 영국 : 0.08 / 호주 : 0.05 / 일본 : 0.02
 - － (국내) 연구데이터 논문활용도는 0.02 수준으로 일본과 비슷한 수준을 나타냄

<표 3-30> 연구데이터 논문영향력 및 논문활용도

구분	논문영향력	논문활용도
해외	0.98	0.98
미국	0.34	0.28
영국	0.11	0.08
호주	0.05	0.05
일본	0.02	0.02
중국	0.07	0.11
기타	0.39	0.45
국내	0.02	0.02
전체	1.00	1.00

주) 10년간 연구데이터 공유활용 관련 SCIE, SSCI 논문('08~'17년) 기준

- \ ○ (논문집중도) 특정국가가 다른 국가와 비교하여 상대적으로 집중하고 있는 기술 파악 가능

논문집중도	$\frac{(\text{특정국가 특정기술 논문 수} / \text{특정기술 전체 논문 수})}{(\text{특정국가 전체 논문 수} / \text{세계 전체 논문 수})}$

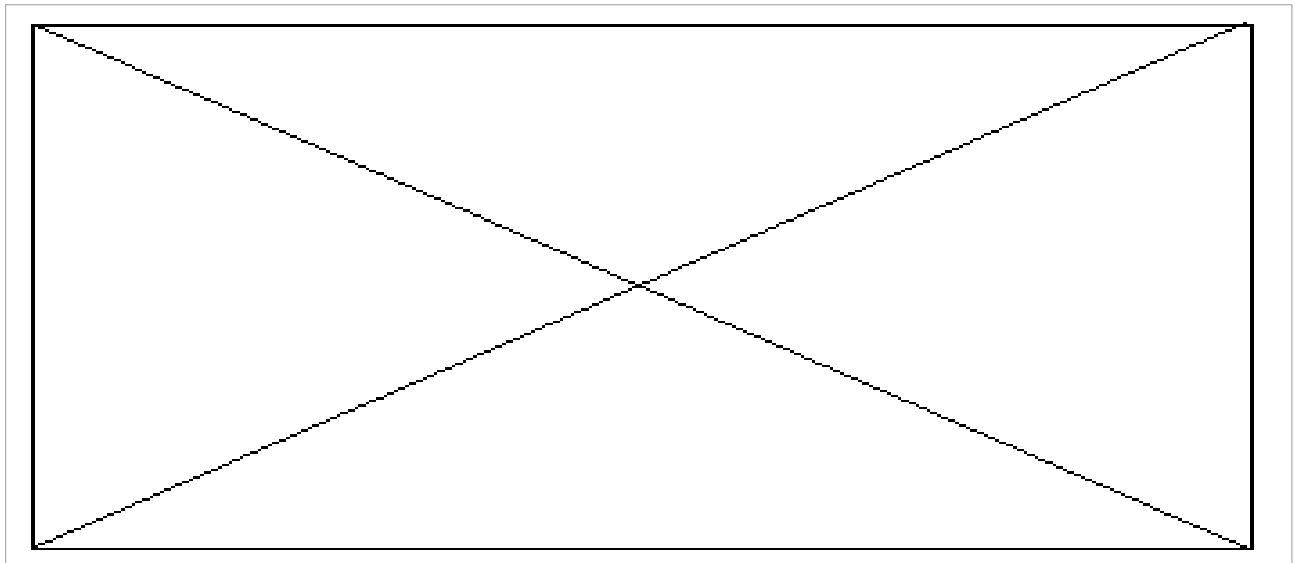
- (논문집중도 비교) 미국이 독보적으로 높은 수치를 보인 반면, 우리나라는 5개 분야에
서 0.01~0.04로 논문 집중도에서 큰 격차를 보임
- －(주요국) 6개 핵심기술 평균은 미국 : 0.27, 중국 : 0.12 / 영국 : 0.07 / 호주 : 0.05 / 일
본 : 0.02 순으로 나타남
- －(국내) 연구데이터 논문집중도 평균은 0.02 수준으로 일본과 비슷한 수준을 나타냈으
며, 연구데이터 저장·보안 기술 관련 집중도가 높은 것으로 나타남

<표 3-31> 연구데이터 논문집중도

구분	연구 데이터 수집·연계 및 공유 기술	연구 데이터 저장·보안 기술	연구 데이터 품질·표준화 기술	지능형 연구데이터 고속 검색 기술	연구데이터 융합·분석 환경 구축 기술	연구 데이터 활용·플랫폼 기술	평균
해외	0.98	0.96	0.99	0.97	0.98	0.98	0.98
미국	0.29	0.19	0.36	0.21	0.29	0.28	0.27
영국	0.09	0.06	0.03	0.09	0.06	0.09	0.07
호주	0.05	0.06	0.02	0.06	0.04	0.05	0.05
일본	0.02	0.02	0.01	0.03	0.03	0.02	0.02
중국	0.07	0.24	0.01	0.17	0.14	0.09	0.12
기타	0.45	0.40	0.56	0.42	0.42	0.44	0.45
국내	0.02	0.04	0.01	0.03	0.02	0.02	0.02
전체	1.00	1.00	1.00	1.00	1.00	1.00	1.00

나. 기술분야별 논문 동향

- 국내·외 모두 연구데이터 수집·연계·공유 기술 및 플랫폼 기술에 대한 연구가 활발하며, 국내의 경우 기술분야별 논문 빈도가 '14년부터 급격히 증가하며 연구데이터 관련 연구가 막 시작된 단계임
 - (해외) 기술분야별 논문 빈도를 살펴보면, 전체적으로 증가하는 추세를 나타냄
 - 지난 '08년~'17년까지 연구데이터 활용·플랫폼 기술 분야와 연구데이터 수집·연계 및 공유 기술에서 각각 9,874, 9,116건으로 가장 많았으며, 이어서 연구데이터 융합·분석 환경 구축 기술(5,738건), 지능형 연구데이터 고속 검색 기술(5,663건) 순으로 나타남



<그림 3-15> 연도별 기술분야별 해외 논문 추이

- (국내) 기술분야별 논문 빈도를 살펴보면, 전체적으로 증가하는 추세를 보였으며, 특히 '14년 이후 증가율이 두드러짐
 - 지난 '08년~'17년까지 연구데이터 수집·연계 및 공유 기술과 연구데이터 활용·플랫폼 기술 분야에서 206건으로 가장 많았으며, 이어서 지능형 연구데이터 고속 검색 기술(167건), 연구데이터 융합·분석 환경 구축 기술(141건) 순으로 나타남

<표 3-32> 해외 기술분야별 연도별('08~'17년) 논문 추이

(단위: 편, %)

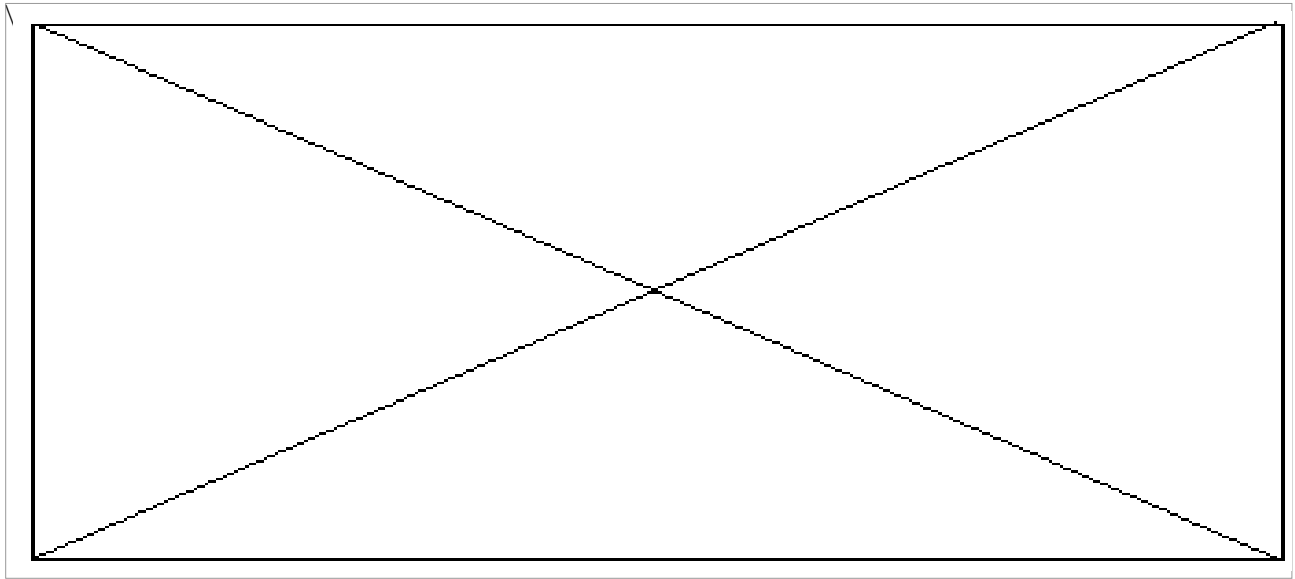
구분		'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	합계	CAGR	
													5년	10년
빈도	연구데이터 수집·연계 및 공유 기술	485	561	614	722	777	877	1,054	1,205	1,388	1,433	9,116	72.1	66.5
	연구데이터 저장·보안 기술	45	82	69	98	142	148	213	310	377	482	1,966	100.1	96.9
	연구데이터 품질·표준화 기술	250	300	337	415	450	459	565	581	711	715	4,783	63.7	77.5
	지능형 연구데이터 고속 검색 기술	366	374	394	398	512	565	605	782	801	866	5,663	57.3	71.5
	연구데이터 융합·분석 환경 구축 기술	249	282	333	395	477	588	674	753	910	1,077	5,738	71.8	67.3
	연구데이터 활용·플랫폼 기술	354	451	519	620	831	963	1,180	1,367	1,611	1,978	9,874	85.3	70.3
	합계	1,749	2,050	2,266	2,648	3,189	3,600	4,291	4,998	5,798	6,551	37,140	69.0	71.2
전년대비증감률	연구데이터 수집·연계 및 공유 기술	-	15.7	9.4	17.6	7.6	12.9	20.2	14.3	15.2	3.2	-	-	-
	연구데이터 저장·보안 기술	-	82.2	-15.9	42.0	44.9	4.2	43.9	45.5	21.6	27.9	-	-	-
	연구데이터 품질·표준화 기술	-	20.0	12.3	23.1	8.4	2.0	23.1	2.8	22.4	0.6	-	-	-
	지능형 연구데이터 고속 검색 기술	-	2.2	5.3	1.0	28.6	10.4	7.1	29.3	2.4	8.1	-	-	-
	연구데이터 융합·분석 환경 구축 기술	-	13.3	18.1	18.6	20.8	23.3	14.6	11.7	20.8	18.4	-	-	-
	연구데이터 활용·플랫폼 기술	-	27.4	15.1	19.5	34.0	15.9	22.5	15.8	17.8	22.8	-	-	-
	합계	-	17.2	10.5	16.9	20.4	12.9	19.2	16.5	16.0	13.0	-	-	-

<표 3-33> 국내 기술분야별 연도별('08~'17년) 논문 추이

(단위: 편, %)

구분		'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	합계	CAGR	
													5년	10년
빈도	연구데이터 수집·연계 및 공유 기술	16	10	12	15	16	20	15	35	32	35	206	62.2	78.1
	연구데이터	1	5	7	4	7	5	10	9	17	17	82	48.1	89.9

구분		'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	합계	CAGR	
													5년	10년
	저장·보안 기술													
	연구데이터 품질·표준화 기술	1	2	4	—	1	—	3	6	5	5	27	75.1	70.7
	지능형 연구데이터 고속 검색 기술	8	9	8	18	19	12	17	30	23	23	167	54.6	67.6
	연구데이터 융합·분석 환경 구축 기술	5	1	8	5	12	15	13	15	36	31	141	67.7	56.6
	연구데이터 활용·플랫폼 기술	13	10	8	12	14	16	23	29	35	46	206	72.7	62.0
	합계	44	37	47	54	69	68	81	124	148	157	829	62.2	65.3
전 년 대 비 증 감 률	연구데이터 수집·연계 및 공유 기술	—	−37.5	20.0	25.0	6.7	25.0	−25.0	133.3	−8.6	9.4	—	—	—
	연구데이터 저장·보안 기술	—	400.0	40.0	−42.9	75.0	−28.6	100.0	−10.0	88.9	0.0	—	—	—
	연구데이터 품질·표준화 기술	—	100.0	100.0	−100.0	—	−100.0	—	100.0	−16.7	0.0	—	—	—
	지능형 연구데이터 고속 검색 기술	—	12.5	−11.1	125.0	5.6	−36.8	41.7	76.5	−23.3	0.0	—	—	—
	연구데이터 융합·분석 환경 구축 기술	—	−80.0	700.0	−37.5	140.0	25.0	−13.3	15.4	140.0	−13.9	—	—	—
	연구데이터 활용·플랫폼 기술	—	−23.1	−20.0	50.0	16.7	14.3	43.8	26.1	20.7	31.4	—	—	—
	합계	—	−15.9	27.0	14.9	27.8	−1.4	19.1	53.1	19.4	6.1	—	—	—



<그림 3-16> 연도별 기술분야별 국내 논문 추이

3.2. PEST 분석 및 종합

3.2.1. 분석 개요

- (분석 목표와 방향) 데이터 관점에서의 PEST 분석을 통해 연구데이터 공유·활용체계 구축에 영향을 주는 외부환경과 그 시사점을 파악
 - (분석 목표) PEST 분석을 통해 역동적으로 변화하고 있는 **외부환경**과 그 시사점을 파악하여 연구데이터 공유·활용체계 **구축 전략**에 활용
 - (분석 방향) 데이터 관련 기술의 비약적 발전과 함께 **데이터 공유 및 활용**의 관점에서 거시적인 외부환경을 분석
- (분야별 세부항목) 외부환경을 정책적, 경제적, 사회적, 기술적 환경으로 구분하고 각각 아래와 같은 세부항목을 중심으로 분석
 - (정책적 환경 분석) 오픈사이언스 정책 추진 현황, 법·제도 등을 중심으로 분석
 - (경제적 환경 분석) 데이터를 중심으로 한 경제 패러다임의 전환 추세, 정부연구개발예산 집행 현황 등을 중심으로 분석
 - (사회적 환경 분석) 데이터 혁명이나 데이터 경제에 영향을 미치는 거시적 사회 이슈, 연구개발 문화의 변화, 연구자의 인식 변화 등을 중심으로 분석
 - (기술적 환경 분석) 데이터 연관 기술 발전 현황, 과학기술경쟁력 현황 등을 중심으로 분석

\<표 3-34> PEST 분석 개요

정책적 환경 분석 • 오픈사이언스 정책의 확산 • 4차 산업혁명 시대의 융·복합적 특징을 제대로 반영하지 못하는 현행 국내 법·제도 • 4차 산업혁명 대응, 혁신성장에 집중하는 정부 국가과학기술정책	경제적 환경 분석 • 데이터 경제로 전환 중인 세계경제 • 정부연구개발예산 증가율이 둔화되는 추세 속에서 연구개발 효율성을 제고할 필요성 증가
사회적 환경 분석 • 협업 및 융합연구 문화 확산 추세 • 오픈 사이언스에 대한 연구자 인식 확산 • 데이터 증가로 인한 에너지 소비 증가 추세와 데이터 기술 활용한 에너지 효율 증대 노력	기술적 환경 분석 • 데이터 기술 발전으로 인한 연구개발 및 생산공정 상의 기술혁신 가속화 • 국가경쟁력 순위에 비교해 높은 수준인 국내 과학기술 경쟁력 • 선진국에 비해 격차를 보이고 있는 국내 데이터 관련 기술수준

3.2.2. 정책적 환경 분석

- (오픈사이언스 정책의 확산) 디지털 기술의 발전과 함께 과학기술 연구의 성과와 과정을 개방화하는 정책이 확산되고 있음
 - (국제적 정책의제로서의 오픈사이언스) 연구개발과정의 개방에 대한 사회적 요청을 정책적으로 반영한 오픈사이언스 정책을 국제적 차원에서 논의
 - － 2002년 부다페스트 오픈엑세스 선언, 2013년 G8 오픈데이터 헌장, 2015년 세계과학기술관회의 대전선언문 등에서 오픈사이언스를 글로벌 정책의제로 제시
 - (국가정책으로서의 오픈사이언스) 주요 선진국을 중심으로 다양한 오픈사이언스 정책이 국가 차원에서 수립되어 이행되기 시작
 - － 2010년을 전후로 영국, 독일, 프랑스, 네덜란드, 핀란드 등 유럽의 주요 국가들과 유럽 연합 차원에서도 본격적인 오픈사이언스 정책을 추진함
 - － 미국은 2013년을 기점으로 연방정부 차원의 오픈사이언스 정책이 본격화됨
 - － 중국, 일본 등 아시아 국가들도 오픈사이언스 정책의 수립과 이행을 추진 중
- (국내 과학기술 관련 정책 환경) 과학기술 및 산업의 융·복합적 특징을 반영한 국가과학기술정책 수립을 통한 혁신성장을 지향하고 있음
 - (과학기술 관련 규제·제도 현황) 4차 산업혁명 시대의 융·복합적 특징을 제대로 반영하지 못하는 현행 법·제도
 - － 세계경제포럼(WEF) 발표 국가경쟁력순위에 의하면 한국의 정부규제 부담은 95위로 머물고 있음(국가경쟁력보고서 2017-2018)
 - － 법이 규정한 범위 밖의 사업이나 연구를 불허하는 ‘포지티브(Positive)’방식 법체계 등은 ICT 융합 연구 활성화를 저해(KISDI, 2017)
 - (문재인 정부 국가과학기술정책 방향) 4차 산업혁명 대응, 혁신성장에 집중하는 국가과학기술정책
 - － 대통령 직속 4차산업혁명위원회를 설치·운영하는 등 4차 산업혁명 대응에 중점을 둔 과학기술 혁신정책 추진
 - － 특정 기술과 특정 산업을 엄격히 구분하는 현행 규제·제도 대신 ‘선 허용, 후 규제’기조로의 전환을 통해 신기술·신사업 활성화 유도

3.2.3. 경제적 환경 분석

- (거시적 경제 환경) 세계 경제는 데이터 경제로 전환 중이며 데이터 시장 자체의 규모도 지속적으로 증가하는 추세
 - (세계 경제 패러다임) 세계 경제는 4차 산업혁명과 함께 데이터 경제로 전환 중임
 - － 세계 경제는 데이터가 모든 산업의 발전과 새로운 가치 창출의 촉매 역할을 하는 ‘데이터 경제’로의 패러다임이 전환 중(가트너, 2011)
 - － 4차 산업혁명은 2030년까지 세계 GDP에 15조 달러를 더하며 경제 전체에 영향을 미칠 가능성이 큰 것으로 예측 (GE 리포트)
 - (데이터 시장) 데이터 시장 자체 규모가 전 세계적으로 확대되는 추세
 - － 세계 데이터 시장 규모는 2020년 1,100억 달러로 성장할 것을 전망 (IDC 보고서, 2017)
- (국내 연구개발 투자 현황) 정부예산과 국가연구개발사업 예산은 유사한 수준의 증가율을 기록하고 있으나 2015년 이후 증가율이 둔화되는 추세
 - (국가연구개발사업 예산 증가율) 2015년 이후 증가율이 둔화되는 추세에 있음
 - － 1964년부터 2018년까지 국가연구개발사업 예산은 연평균 증가율은 18.6% 이나 2015년 이후 둔화되어 최근 3년간 증가율은 1%대 그침
 - (현 정부의 연개개발 투자 의지) 혁신성장에 대한 정부의 의지에 따라 2019년 국가연구개발 예산 증가율이 지난 3년에 비해 약간 상승함
 - － 2019년 국가연구개발사업 예산은 전년도에 비해 3.7% 상승한 20조 3,997억 원으로 편성
- (연구개발 효율성 제고의 필요성) 연구개발 투자 규모는 OECD 상위권에 해당하나 이에 비해 현저히 낮은 연구개발 효율성을 제고할 필요성이 꾸준히 제기되어 왔음(STEPI, 2009, KISTEP, 2011, 국회예산정책처 보고서, 2018)
 - (연구개발 투자 규모) OECD 국가 가운데 상위권을 유지하고 있는 연구개발 투자 규모
 - － 2016년 전체 연구개발 투자 규모는 GDP 대비 4.24%로 이스라엘에 이어 OECD 국가 중 2위에 해당하며 국가 연구개발 투자 역시 GDP 대비 0.885%로 4위를 기록함
 - (연구개발 투자 효율성) 국가 연구개발 사업의 효율성을 OECD 31개국과 비교·분석한 결과 현저히 낮은 것으로 나타남
 - － 2016년 국가 연구개발 사업의 효율성 비교·분석 결과, 28위에 불과한 것으로 나타남

3.2.4. 사회적 환경 분석

□ (연구개발 문화 변화) 4차 산업혁명의 핵심 기술혁신으로 인한 연구 문화의 변화

- (연구 환경 변화) 융·복합 연구 및 고품질·대용량 데이터 기반 연구 비율의 증가로 인해 연구 환경이 데이터 중심으로 변화하고 있음
 - － 융합 R&D의 SCI 논문과 정부연구개발 사업에서의 융합 R&D 비율이 전체적으로 증가 추세에 있음(KISTEP, 2015)
 - － 첨단 연구장비의 활용과 데이터 처리 기술의 발달로 인해 연구개발 패러다임이 “데이터 중심형”으로 전환됨(KISTI, 2016)
- (연구자 인식 변화) 연구성과물의 공유 및 활용을 연구개발의 효율성과 투명성을 제고해야 한다는 오픈 사이언스에 대한 인식 확산
 - － 개인의 자산으로 간주했던 연구데이터를 공공재로 인식하고 공유·활용을 통해 국가 경쟁력을 제고시켜야 한다는 인식의 확산
 - － 세계 주요국의 인식에 비해 국내 연구자들의 연구데이터 공유·활용 필요성에 대한 매우 인식이 낮은 편임

□ (데이터산업의 발전과 신기후 체제) 데이터산업 발전으로 에너지 소비 증가할 전망이며 데이터 관련 신기술 활용을 통해 에너지 절약 촉진할 필요 있음(국민경제자문회의, 2017)

- (신기후 체제와 에너지 감축의무) 신기후체제 시작에 따라 한국은 2030년까지 기준안(BAU) 대비 37%의 자발인 온실가스 감축 의무를 설정하였음
- (에너지 소비 증가) 컴퓨팅 자원 사용량의 폭발적 증가에 따른 에너지 소비 증가 전망과 데이터 관련 신기술 개발의 필요성
 - － 데이터 산업의 발전과 함께 컴퓨팅 자원 사용량과 에너지 소비가 급격히 증가할 것으로 전망됨에 따라 에너지 효율 제고를 위한 데이터 관련 신기술 개발 노력 필요성 증가함

□ (사회적 문제해결과 데이터) 재난관리, 교통정책 등 다양한 사회적 문제들에 대한 데이터 기반의 해결방안들이 증가하는 추세

- 주요 선진국을 중심으로 데이터 기반의 다양한 재난관리 시스템이 일찍부터 사용되고 있음(국가과학기술심의회, 2018)
- 세계 각국에서 다양한 교통문제 해결을 위해 데이터를 적극 활용하고 있음(서울연구원, 2016)

3.2.5. 기술적 환경 분석

- (데이터 관련 기술 혁신 현황) 인공지능, 사물인터넷, 빅데이터 등 관련기술의 발전으로 인한 연구개발 및 생산공정 상의 기술혁신 가속화
 - (연구개발혁신) 실험 과정 없이 연구데이터 공유를 통해 연구성과를 창출하는 등 연구개발혁신 가속화
 - － 정밀의료 디지털헬스케어, 신약후보물질 개발 등에 전통적인 실험 과정 대신 데이터 분석기술을 활용하여 성과를 창출(과학기술정보통신부, 2017)
 - (생산공정혁신) 연구개발 뿐만 아니라 생산공정 과정에서도 기간 단축 가능한 신기술 등장
 - － 인공지능 등 데이터 기술은 제조, 건설, 에너지 등 다양한 분야에 걸친 스마트화를 통해 생산공정혁신을 이루어 경쟁력 강화에 기여함
- (과학기술경쟁력) 2018년 기준 한국의 과학인프라 경쟁력은 7위, 기술인프라 경쟁력은 14위로 27위인 국가경쟁력에 비해 높은 수준임.
 - (과학인프라 경쟁력) 연구개발투자, 연구개발인력, 특허 관련 지표 등 정량적 지표가 강점이나, 노벨상 및 설문지표 등은 약점 영역
 - － 연구개발투자, 연구개발인력 관련 지표 중 일부는 순위가 소폭 하락했지만 지속적으로 강점 영역으로 유지
- (국내 데이터 관련 기술수준) 빅데이터, 인공지능 등 국내 데이터 관련 기술수준은 선진국과 격차가 발생하고 있음
 - (빅데이터 기술수준) 빅데이터 축적을 위한 인프라부터 분석 및 활용을 위한 소프트웨어 기술까지 외국에 비해 경쟁력이 높지 않은 상황
 - － 공급기업 관점 국내 기술수준 선진국 대비 65.7점, 기술수준 격차는 3.1년, 선진기술 도달시간은 3.4년이 걸릴 것으로 평가(2017 데이터산업 백서)
 - (인공지능 기술수준) 국내 인공지능 기술 수준은 최고 기술 보유국인 미국과 비요하여 70%로 2.4년의 격차 존재(한국정보화진흥원, 2017)
 - － 가장 영향력 있는 인공지능 연구자수, 인공지능 톱클래스 발표 논문과 특허, 인공지능 연구 역량, 기술 수준, 투자 금액 등 모든 부문에서 격차

3.3. SWOT 분석 및 대응전략 도출

3.3.1. SWOT분석

- 연구데이터의 공유·활용체계 구축에 대한 정부의 강력한 의지 - ICT 인프라 기술선도 경험 보유		- 연구데이터 공유·활용을 위한 연계 및 추진체계 미비 - 연구데이터의 공유·활용에 대한 연구현장 인식 및 전문역량 부족	
기술선도 경험과 정부의 강력한 추진의지	S (강점, Strength)	W (약점, Weakness)	추진체계 및 인식·역량 부족
	내부 역량		

데이터 기반 R&D 혁신	외부 환경		주요국과의 경쟁심화
	O (기회, Opportunity)	T (위협, Threat)	
- 데이터 사이언스의 발전과 더불어 데이터 기반 연구개발 문화 및 기술혁신 노력 확산 - 연구데이터의 개방과 공유를 위한 국가차원의 인프라 및 지원체계 구축 필요성 증가	- 법·제도 및 정책 미비로 인한 주요국과의 격차 심화 우려 - 글로벌 경쟁자들의 플랫폼 장악에 따른 기술 종속화		

가. 외부 환경 진단

□ 기회(Opportunities)

- 데이터 기반 연구개발 문화 및 기술혁신 확산 추세
 - － 데이터 사이언스의 비약적 발전으로 인한 연구개발 방식의 변화와 급속한 기술혁신 가속화
- 국가차원의 연구데이터 공유·활용체계 구축 필요성 증가
 - － 연구개발 활동에 있어서의 데이터의 활용과 다양한 연구자간의 융합·공동연구 추진을 위한 연구데이터 공유체계의 필요성이 점증하는 추세

□ 위협(Threats)

- 주요국과의 격차
 - － 연구데이터 관련 법·제도 및 정책 미비로 인한 주요국과의 격차 심화 가능성
- 기술 종속화 문제
 - － 글로벌 경쟁자들의 관련 기술 및 플랫폼 장악에 따른 기술 종속화 우려

나. 내부 역량 진단

□ 강점(Strengths)

- 연구데이터 관리·활용을 위한 정부의 강력한 추진 의지
 - － 연구데이터의 공유 및 활용의 필요성이 널리 인식되면서 정부 차원에서 체계 구축을 위한 기반을 마련
- ICT 인프라 기술 선도 경험
 - － 한국은 전 세계 ICT 인프라 기술을 선도했던 경험을 보유하고 있음

□ 약점(Weaknesses)

- 법·제도 미비
 - － 연구데이터 공유·활용을 촉진하는 법·제도의 미비
- 연구데이터의 관리·공유에 대한 연구현장 인식 및 전문역량 부족
 - － 연구자 개인적으로 데이터를 관리중으로, 관리·공유에 대한 연구현장 인식이 부족(79%가 개인자산으로 인식)
 - － 데이터 집약형 연구 확산에 따라 특정 분야 차원에서 일부 데이터를 수집·관리 중이나, 활용은 저조

3.3.2. 대응전략

- 연구데이터의 공유·활용으로 새로운 과학적 발견과 새로운 가치 창출을 위한 추진방안 도출의 관점에서 SWOT 매트릭스를 활용하여 대응전략을 도출함

	강점(S)	약점(W)
기회요인(O)	SO 전략: 강점과 기회요인의 시너지 전략 [SO ①] 국가 연구데이터 플랫폼·인프라 구축·운영 [SO ②] 대·소분야 전문센터 플랫폼·인프라 지원	WO 전략: 약점 극복 전략 [WO ①] 연구데이터 공유·활용 위한 추진체계 구축 [WO ②] 연구데이터 공유·활용의 인식제고 및 활성화를 위한 시범사업 추진
위협요인(T)	ST 전략: 위협 극복 전략 [ST ①] 선진국과의 격차를 줄이기 위한 기술개발 [ST ②] 연구데이터 활용 기반 조성을 위한 법·제도 마련	WT 전략: 위협 회피 전략 [WT ①] 연구데이터 활용을 위한 전문인력 양성 [WT ②] 연구데이터 경쟁력 강화를 위한 대한 중장기·집약적 투자

<그림 3-17> SWOT 매트릭스

- [SO 전략] 연구데이터 공유·활용체계 전반의 연구데이터를 원스톱으로 검색하고 활용할 수 있는 국가연구데이터 플랫폼 및 대·소분야 전문센터 구축·운영
- 각종 연구데이터에 대한 색인·검색 기능과 연구자가 데이터 융합 및 분석에 활용할 수 있는 분석도구 및 가상작업환경을 제공하는 국가연구데이터 플랫폼 구축
 - 연구데이터 공유·활용 우선순위 도출을 통해 중점분야를 선정하고 분야별 대·소분야 전문센터 구축
- [WO 전략] 추진체계 구축 및 시범사업 추진
- 연구데이터 공유·활용 위한 추진체계 구축
 - － 다양한 분야에서 개별적으로 생성되는 연구데이터를 체계적으로 취합·관리·공유할 수 있도록 계층적 공유·활용 체계를 구축·운영할 추진체계 구축
 - 시범사업 추진
 - － 시범사업 실시를 통해 연구데이터 공유·활용의 인식제고 및 활성화

□ [ST 전략] 기술개발 및 법·제도 정비

- 기술개발을 통한 선진국과의 격차 해소
 - － 이미 연구데이터 공유·활용 인프라 및 관련 기술을 선점한 선진국과의 격차 해소를 위한 기술개발
- 법·제도 마련
 - － 연구데이터 활용 기반 조성을 위한 법·제도 정비 및 연구개발혁신을 위한 규제 완화

□ [WT 전략] 전문인력 양성 및 투자 강화

- 연구데이터 공유·활용을 위한 전문인력 양성
 - － 데이터 기반 집단연구를 통해 각 분야의 문제해결이 이루어지는 각 분야 전문센터에서 R&D와 인력양성 동시 추진으로 고급인재 육성
- 투자 강화
 - － 연구데이터 경쟁력 강화를 위한 대한 중장기·집약적 투자

사업의 추진전략

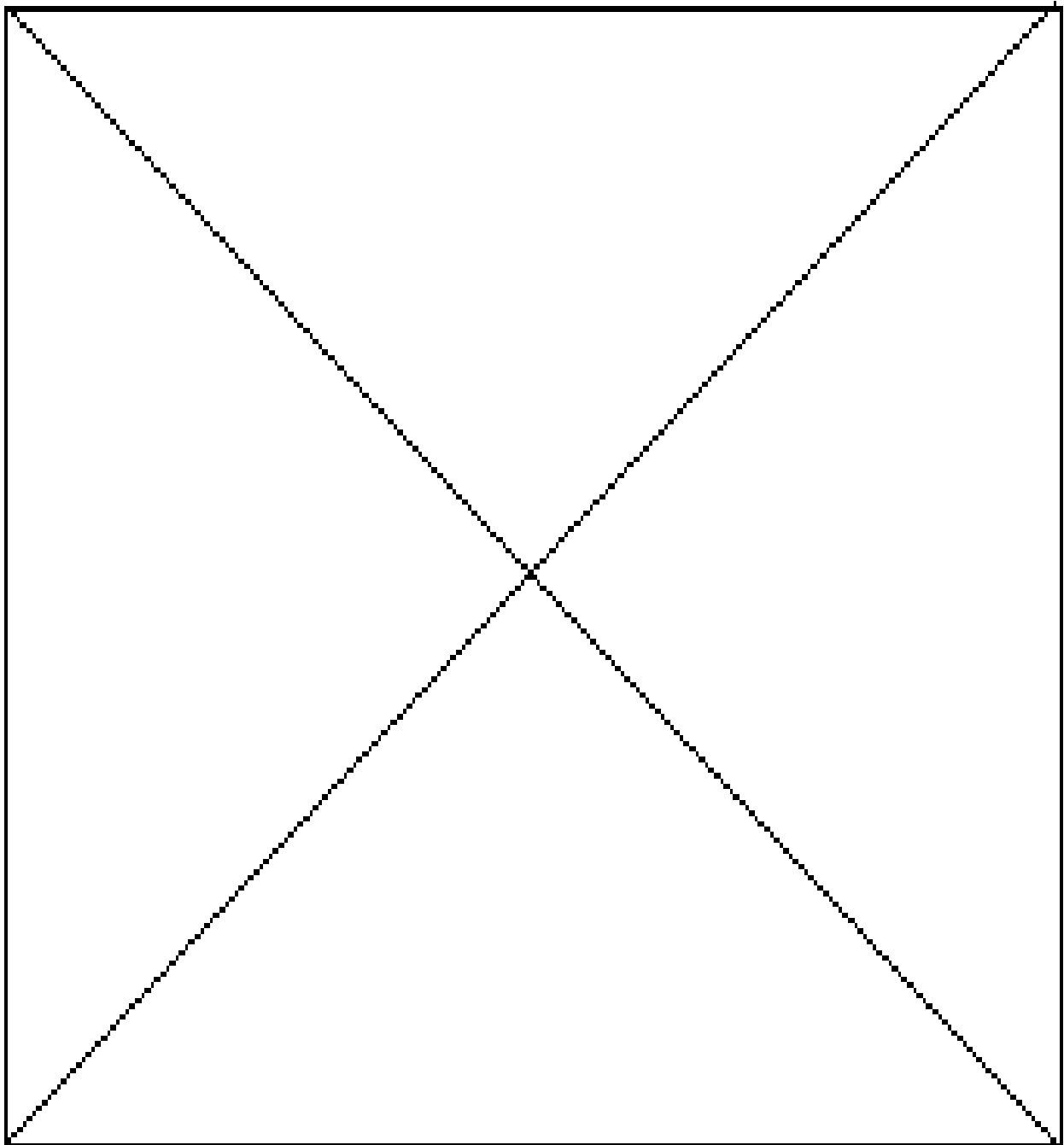
4.1. 목적과 범위

4.2. 사업의 추진방향

4.1. 목적과 범위



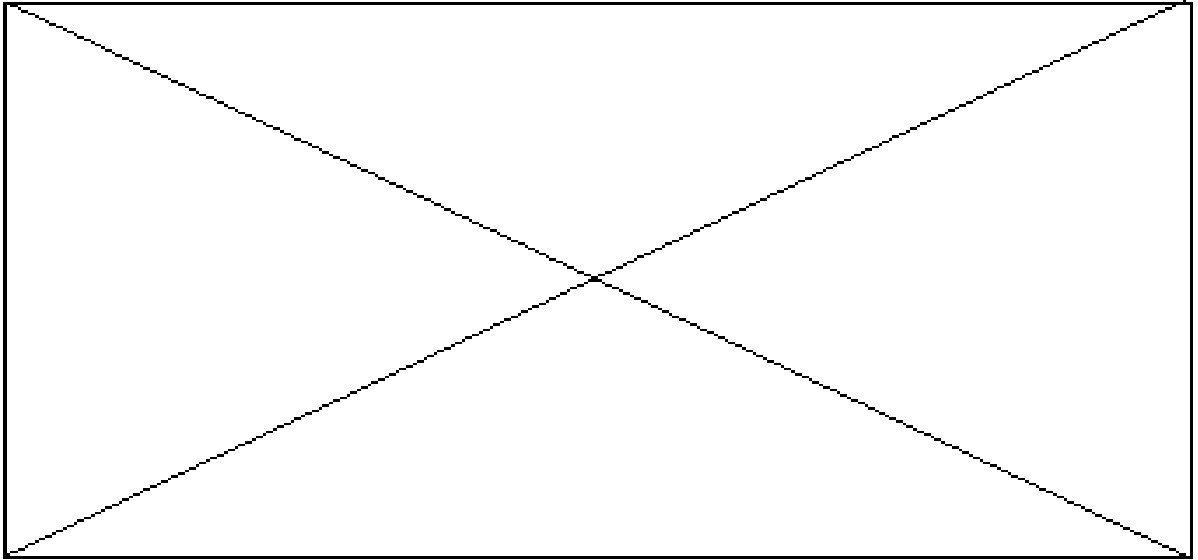
4.1.1. 사업의 비전



4.1.2. 사업 추진 목표

□ 사업의 목표

- 목표 1. 국가연구데이터 공유·활용 생태계 조성

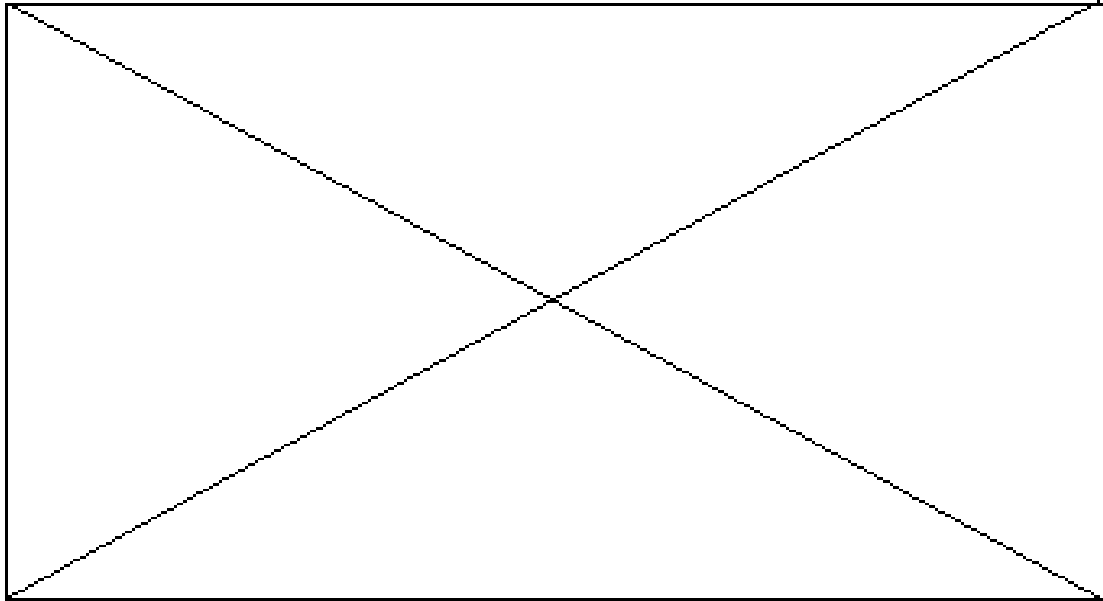


<그림 4-1> 국가연구데이터 공유·활용 생태계 개념도

※ 과학기술분야 R&D 전주기 연구 - 국내 생명 및 나노과학기술 연구자를 중심으로 - 참조

- 연구자가 연구를 기획하는 연구의 시작부터 연구결과를 도출하는 종료 시점까지 연구의 전 과정에서 생산되는 데이터를 연구 분야별로 취합·관리하고 종합적으로 공유·활용할 수 있는 생태계를 구축
- 각 연구자 및 연구기관은 소속된 분야의 대·소분야 전문센터를 통해 연구에 필요한 데이터를 검색 및 활용하여 연구를 수행하고 그 과정에서 생산된 데이터를 대·소분야 전문센터에 저장·관리
- 분야별 전문센터와 연계되는 국가연구데이터 플랫폼은 국가연구데이터를 총괄하는 센터로 지능적인 원스톱 검색을 통한 데이터 소재정보 제공과 연구데이터의 고속 분석환경을 제공
- 국가 R&D과제를 수행하고, 연구 과제를 관리하는 기관은 연구기획 단계에서부터 연구데이터에 대한 관리계획(DMP)을 제출받아 연구데이터가 계획대로 저장·관리 될 수 있도록 제도적 근거를 확보
- 연구를 통해 발생하는 연구과제정보, 연구결과물, 연구데이터가 서로 연결되는 연구생태계는 연구자를 중심으로 연구데이터가 제공되고 제공받는 연구데이터의 공유·활용 체계가 목표

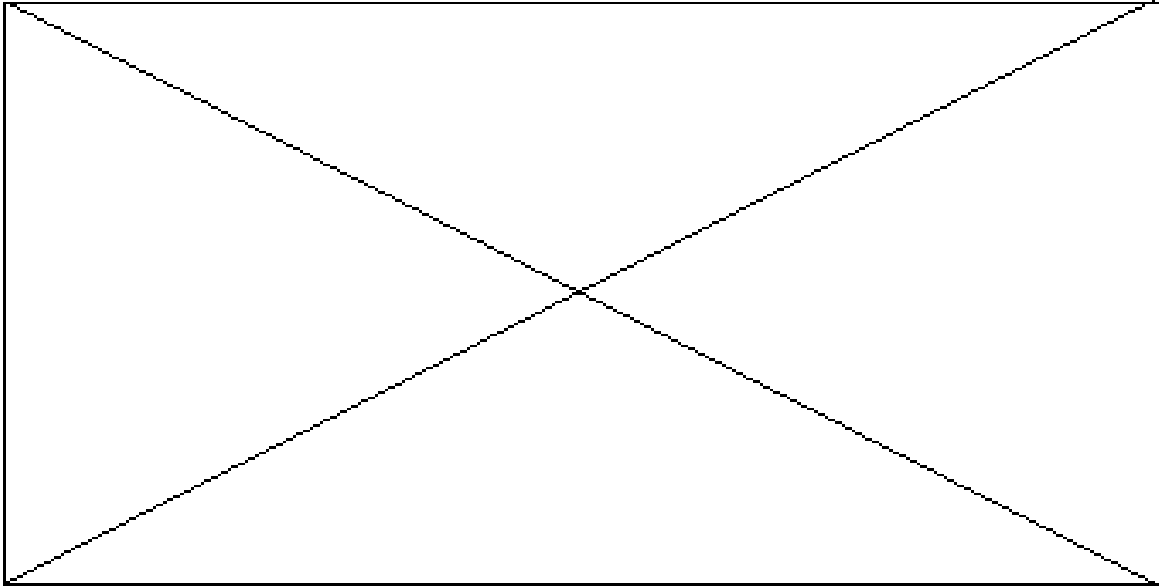
○ 목표 2. 전문분야별 연구데이터의 집약과 공동연구 기반확보



<그림 4-2> 전문분야별 데이터 집약과 공동연구 기반 조성

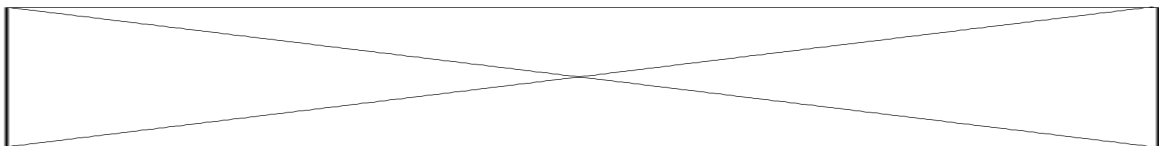
- 연구자간 협업연구의 증가에 따라 연구데이터가 실질적으로 공유되어 활용됨으로써 새로운 부가가치를 창출할 수 있는 전문분야별 연구 공동체를 위한 전문센터 구축
 - 과학기술의 복잡화, 거대화에 따른 다양한 분야, 여러 지역 연구자간 협력 증가
 - ☞ 새로운 아이디어 및 방법론 출현, 새로운 지역을 중심으로 신흥 연구 그룹 등장과 세계 과학 연구 중심의 변화 가능
- 문제해결을 원하는 커뮤니티, 소규모의 연구기관 및 새로운 과학기술이 태동하는 신규 커뮤니티에게 고성능 컴퓨팅 자원, 다양한 소프트웨어, 분석 도구 등의 통합 환경 제공
- 분야별 전문센터는 연구데이터 활용 방향을 설정하고, 데이터 공유를 위한 표준화, 큐레이팅 등 데이터 관리에 필요한 기준안을 마련하여 데이터 검색기능, 분석 SW 개발 등 연구데이터 공유에 필요한 기능을 확보
- 데이터 기반의 시뮬레이션, AI 등을 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
- 개별 연구자 등이 보유한 데이터(저장장치 미보유 기관 등)를 저장·관리·검색할 수 있도록 저장공간·데이터분석 도구 등 제공

○ 목표 3. 빅데이터, AI기술 등 데이터를 활용하는 R&D 확산



<그림 4-3> 데이터를 활용하는 R&D로의 확산

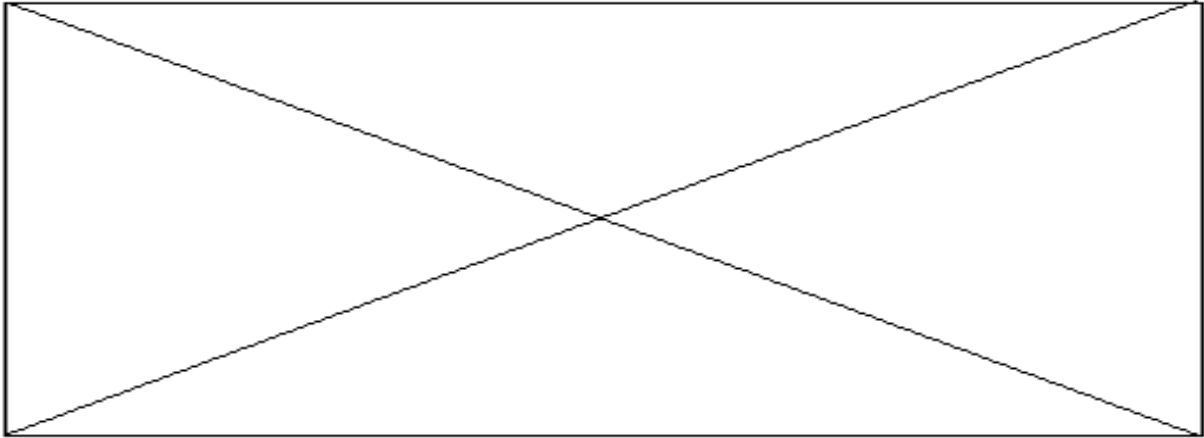
- ‘4차 산업혁명’과 ‘Open Science’는 데이터와 ICT를 근간으로, R&D 방식과 생산, 소비 방식을 혁신하는 것으로 빅데이터의 확보, 처리, 응용이 핵심
- R&D 분야에서도 시뮬레이션과 마이닝 기법 등을 통해 데이터 융합 및 분석만으로도 연구수행이 가능해짐에 따라 인공지능, 빅데이터 등의 디지털 기술을 통해 고부가가치의 연구수행 결과 확보가 가능
- 디지털 기술의 발전 단계에 따라 연구개발 활동도 데이터 중심의 융합 및 공동연구가 지속적으로 확대되고 있음



- 데이터분석을 위한 기본 지식(데이터 수집, 가공, 분석, 시각화 등)과 스킬을 위해 전문 교육기관의 교육과정을 이용할 수 있도록 교육환경 구축
- 연구데이터의 고부가가치의 창출은 연구자의 데이터 및 컴퓨팅 활용 역량이 확보되어하고 교육용 데이터 뱅크 구축을 통해 데이터 분석기법과 분야별 문제 해결에 대한 교육·실습환경 제공
- 데이터 분석의 기본지식을 보유하고, 해당 연구 분야 별 Domain에 대한 업무지식을 보유한 전문 고급인력 양성을 통해 대용량 연구데이터에 대한 활용 역량 강화

4.1.3. 사업의 범위

□ 대응전략에 따른 사업범위 도출



<그림 4-4> 대응전략과 사업범위

□ 연구데이터 관리 및 공유·활용 제도화

- 주요 선진국의 경우처럼 공공자금이 투입한 연구 성과물(논문, 보고서, 데이터 등)을 공유·활용 촉진 정책 추진. 특히, 연구 성과물 중 연구데이터의 쉬운 접근과 재사용을 촉진하기 위해 데이터관리계획(DMP) 의무화 추진
- 「국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률」, 「공동관리규정」등에 연구데이터를 ‘연구성과물’에 추가 규정, 연구자에게 소유권 부여 등 국가차원의 체계적인 수집·관리 및 공유·활용 체계 기반 구축
- 시범사업 등 분야별 추진경험을 바탕으로 중장기적으로 국가 연구데이터 법제 정비 및 개선
- 「국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률」에 정책 추진을 위한 근거 조항 마련
 - － 연구성과의 정의(제2조)에 연구데이터 추가
 - － 연구데이터의 관리 및 공유·활용, 지원을 위한 기본원칙(제3조) 정비 및 관련 규정 마련
 - － 연구성과 관리·활용계획(제12조)에 연구데이터를 포함시켜 종합적인 관리·활용 시책 추진 근거 마련
- 「공동관리규정」 개정을 통한 연구데이터의 세부적인 수집·관리 및 공유·활용 방법 규정
 - － 연구데이터의 보고 및 제출(제15조), 공개(제18조), 소유(제20조) 등 관련 규정에 연구데이터의 공유 및 활용을 위한 세부 절차 마련

- \ - 연구데이터를 ‘연구 성과물’에 포함시켜 거점기관을 통한 효율적 유통·관리 추진(제25조제13항)

□ 분야별 데이터 공유·활용 연구공동체 형성

- 분야별 전문센터는 수집된 데이터를 관리·검증 및 공유를 위한 큐레이팅, 활용 역할을 담당하며, 1차 아카이빙(Archiving)도 담당
- 인프라 보유여부와 무관하게 분야별 전문센터의 데이터 관리에 대한 권리·책임을 보유하는 연구 커뮤니티가 담당
- 데이터 집약적인 분야인 소재, 바이오, 인공지능 분야 등에서 수요가 있는 커뮤니티를 중심으로 전문센터를 지정 또는 공모
- 분야별 전문 센터의 데이터 및 컴퓨팅 자원은 우선적으로 센터 보유기관의 인프라를 활용하되, 인프라 미보유 센터에는 국가에서 클라우드 방식으로 인프라를 지원
- 분야지정 및 공모 방식으로 선정하되 일정기간의 예산, 컴퓨팅데이터 및 네트워크 자원을 함께 지원
- 소재, 바이오 분야 등의 우선 추진 분야에 대해 데이터 बैं크를 구축하여 서비스함으로써 연구자들의 활용을 통한 Data 기여를 유도

□ 데이터 및 컴퓨팅 활용 R&D 인력양성

- 연구자에 대한 데이터 및 컴퓨팅 활용 초급·중급 교육 제공
- 대학생·대학원생 및 현직 연구자들을 대상으로 전문분야에서의 데이터 컴퓨팅 활용을 위한 초급 및 중급 교육 프로그램 제공
- 인력육성 인프라 및 원격연구 환경으로서 분야별 데이터·컴퓨팅 활용 플랫폼과 교육용 데이터 बैं크를 구축하고, 수준별 교재 개발·보급
- 기계학습 도구, 계산과학 SW 및 광범위한 교육용 데이터 탑재 → 온라인 가상 실습 환경 제공
- 데이터 기반 연구참여를 통한 고급인력 양성 추진을 위한 고급교육
- R&D와 인력양성 동시 추진으로 고급인재 육성
 - 대학원생 인턴쉽, 산·학·연 방문연구자 프로그램 등
- 전문센터 기반 강화를 위한 전문인력 양성
 - 데이터 기반 연구 및 전문센터의 증가와 함께 수요가 크게 증가할 것으로 예상되는 ICT 기반 관리·운영·서비스 인력 양성

□ 국가 연구데이터플랫폼 구축 및 운영 서비스

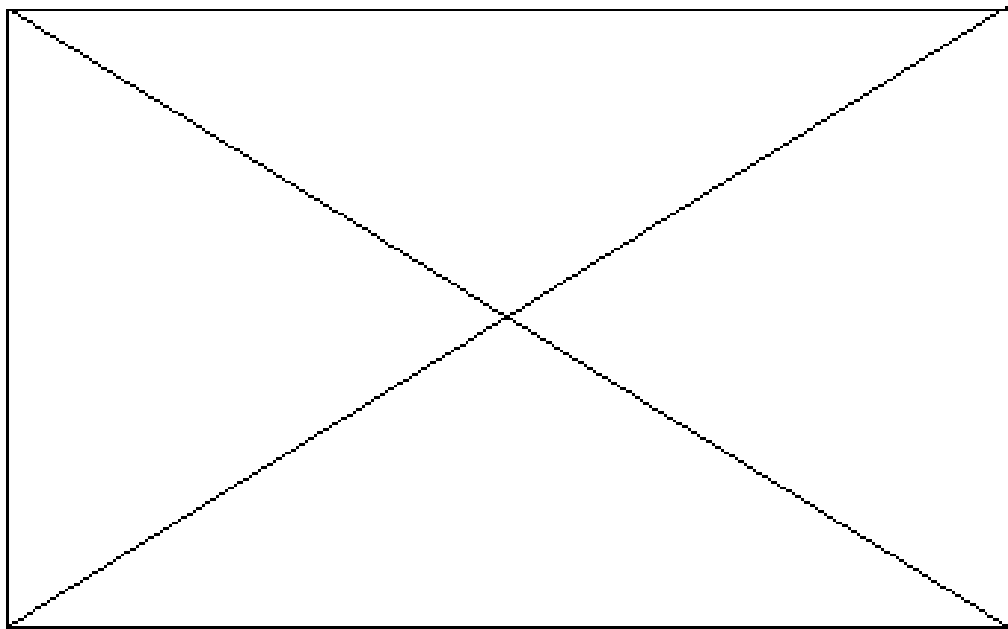
- 분야별 전문센터와 연계해서 연구데이터의 소재정보 등 연구데이터의 메타데이터 관리를 통해 여러 종류의 연구데이터를 동일한 방법으로 접근할 수 있는 연구 데이터 플랫폼 구축 및 운영
- 지능형 검색기술을 활용하여 국가연구데이터에 용이하게 접근할 수 있는 포털 기능과 고속의 컴퓨팅 환경을 제공하여 연구데이터를 활용한 제반 분석환경(인프라 및 분석S/W 등)을 제공
- 빅데이터 분석 SW 제공 및 SW의 자유로운 등록·활용 환경
- 인프라 미보유 전문센터에 클라우드 방식의 인프라 지원
- 개별 연구자에게 클라우드 방식의 데이터 관리용 플랫폼 제공(비공개 전제)
- 과거 과제 및 기존에 수행 중인 과제에서 생산되어 개별적으로 보유한 데이터의 유실 방지
- 공개 여부·수준 연구자 결정 : 비공개(기간제한), 메타데이터 공개, 연구용 공개, 완전 공개
- 성실실패 과제에 대해 과제 수행 중에 생산된 데이터 공개를 통해 추후 유관 과제 수행에 기여토록 유도

4.2. 사업의 추진방향

4.2.1. 사업전략

□ 사업 전략 도출

- 사업 비전 및 목표와 연계한 사업추진 전략 도출



<그림 4-5> 사업추진전략 도출

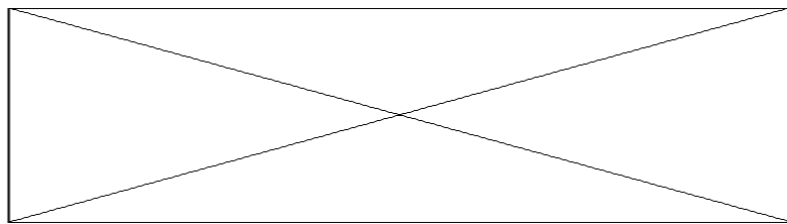
□ 연구데이터 공유·활용 원칙 및 제도 등을 반영한 플랫폼 설계

- 「단계별 법제 정비 및 개선」으로 연구데이터 관리·공유·활용에 관한 법·제도 마련
 - 단기적으로는 연구데이터 수집·관리 및 공유·활용을 위한 법·제도적 기반을 마련하고 중장기적으로 분야별 연구데이터 관련 정책의 추진 경험을 바탕으로 국가 연구데이터 법제 정비 및 개선
 - 연구데이터의 법적 정의 마련, 연구데이터 관리제도 도입, 데이터에 대한 연구자의 권리 보장 등을 위한 법령정비 추진
 - 국가연구개발 사업등의 성과평가 및 성과관리에 관한법률, 과학기술 기본법, 연구개발 사업 처리규정 등 관련법령 등
- 「국가연구개발사업 공동관리 규정」 개정을 통한 법적 정의 및 관리제도 마련

<표 4-1> DMP도입을 위한 관계법령

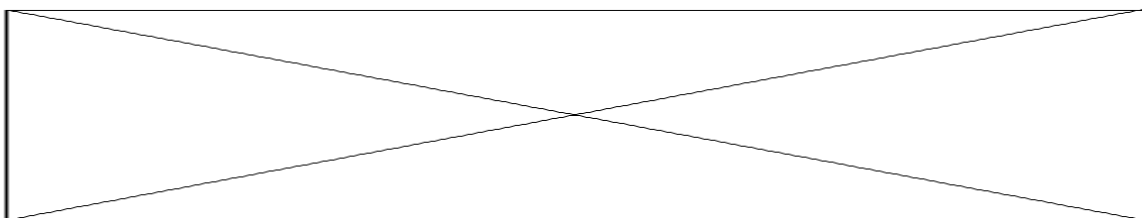
관련법령	① 국가연구개발사업 등의 성과평가 및 성과관리에 관한 법률 ② 과학기술기본법, 동법 시행령, 시행규칙 ③ 국가연구개발사업의 관리 등에 관한 규정/규칙 ④ 과학기술정보통신부 소관 연구개발사업 처리규정

- (연구데이터 정의) 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
- (데이터관리 계획의 도입) 주관연구기관이 비영리기관이고, 중앙행정기관의 장이 필요하다고 인정하는 과제에 적용하여, 협약체결 단계에서 연구 제안자로부터 데이터관리계획(DMP) 접수
- DMP : Data Management Plan
데이터의 양, 종류, 품질, 포맷, 표준 및 메타데이터, 데이터 수집 방법, 품질확인 및 백업절차 데이터 관리, 보존계획, 윤리적 고려사항, 데이터의 지적재산권, 데이터 관리와 보존의 책임사항 등(출처 : 한국문헌정보학회지, 47(3): 251-274.)



<그림 4-6> DMP의 내용구성

- 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
- 「연구관리 표준매뉴얼」 개선을 통한 연구자 권리의 근거 확보
 - (연구데이터의 인용) 논문, 특허와 마찬가지로 타인의 연구데이터를 인용(citation)한 경우 출처를 반드시 명시 할 수 있도록 매뉴얼 개선
- 단계별 법제 정비 및 개선의 내용이 관리 되도록 플랫폼 설계
 - 플랫폼에 DMP를 작성 또는 작성하기 위한 템플릿과 작성 가이드 등을 제공 하여 연구자가 DMP를 용이하게 작성하도록 서비스
 - DMP 작성 및 다운로드, DMP 템플릿 등록 및 관리, DMP 관리 이력 등 국가연구데이터 플랫폼의 포털을 통해 DMP 작성에 관련 된 기능이 서비스 될 수 있도록 설계



<그림 4-7> DMP 정보의 연계

－ DMP 정보 공유를 위해 플랫폼과 관련시스템 연계기능 설계

<표 4-2> DMP 관리기능

기능 구분		기능 설명
DMP 관리	• DMP 작성 및 다운로드	• DMP 작성 및 작성 DMP 다운로드
	• DMP 템플릿 등록 및 관리	• DMP 작성을 지원하는 템플릿 관리
	• DMP관리(이력, 공유, 공개)	• DMP 작성 이력 관리 등
	• DMP 작성가이드	• DMP 작성 가이드

○ 분야별 전문센터 및 국가연구데이터 플랫폼 구축 및 운영

- － (플랫폼지원) 대·소분야 전문센터 및 국가연구데이터 플랫폼을 통해 연구데이터를 공유·활용할 수 있도록 플랫폼화 하고 각종 데이터 분석 SW 개발과 운영을 통해 공동연구를 할 수 있는 협업 환경을 구축
- － (프로그램 지원) 연구자 가이드 교육 등 데이터를 통한 연구역량을 제고할 수 있는 프로그램을 제공하고 연구자간 협력 등의 컨설팅 교육 프로그램도 제공
- － (인프라 지원) 인프라를 보유하지 않은 기관 및 개별 연구자가 연구 협업을 할 수 있는 데이터 관리 시스템을 운영 지원

○ 데이터·컴퓨팅 기반 연구공동체 육성 해외사례를 적극 반영

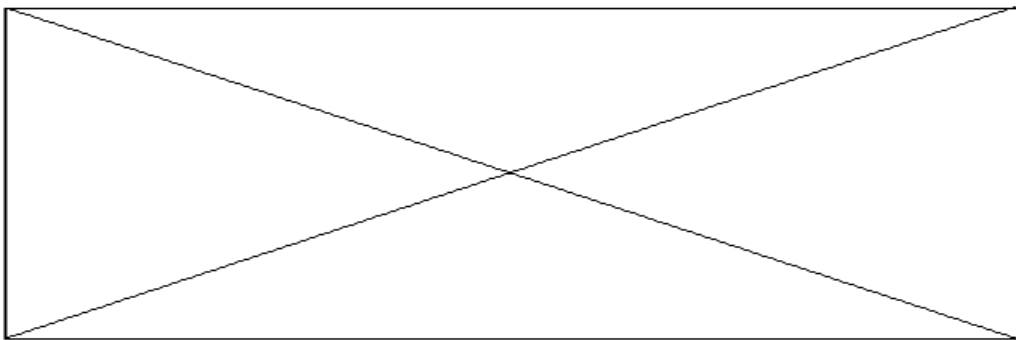
- － (미국) '95년부터 컴퓨팅 기반 교육·융합 연구 환경을 동적으로 지원하는 웹기반 플랫폼 허브제로 (HubZero) 구축·운영
 - * 60개 이상의 Hub('15년 기준)로 구성, 연간 60만 여명이 사용
- － (유럽) Horizon2020 사업에서 '14년부터 e-인프라에 기반한 분야별 전문센터(CoE, Center of Excellence) 운영('16년 9개)
- － (호주) '09년부터 분야별 가상실험실 플랫폼(Nectar) 운영('17년 12개)
 - * 미국 46%, 영국 43%, 호주 41% 연구데이터 공유 경험(Wiley 2014)

○ 데이터 기반으로 R&D 협업을 할 수 있는 공동체 형성 촉진 및 확산

- 빅데이터 기반의 시뮬레이션, 인공지능을 접목한 분석 등으로 신소재·신약 개발, 과학실험 등을 집단 연구를 통해서 성과를 달성할 수 있는 집단연구 지원 환경과 다양한 방안을 신설하여 제공
- 데이터 기반 연구 수요 수용 및 확산하기 위해 활용 가능한 데이터를 대상으로 소규모 연구 과제를 지속적으로 신설

□ 연구데이터의 검색·분석 Tool 등의 환경제공으로 연구자 부담경감

- 원스탑으로 검색하고 활용할 수 있는 국가연구데이터 통합 검색도구 제공
 - 해외의 데이터 공유·활용 프로젝트와 연계하여 해외 데이터 리포지터리와 상호운용성 확보
 - 분산파일시스템, NoSQL, 비디스크 기반 DBMS 등 원활한 검색을 위한 저장 기술 활용
- 연구자가 종합적인 분석을 처리할 수 있는 분석도구 제공
 - 데이터의 특징, 연관관계, 인과관계 등을 도출함으로서 종합적인 분석을 처리할 수 있는 분석도구·방법론·각종 S/W를 제공
 - 다양한 데이터 분석도구의 자유로운 공유·유통을 위해 데이터 분석도구 마켓 운영
 - 전통적 통계 분석, 데이터 마이닝, 텍스트 마이닝, 연관분석, SNS 분석 등 주요 방법론 활용



<그림 4-8> 연구데이터 분석 환경구축

- 거대용량의 연구 데이터를 실시간으로 시각화하는 도구 제공
 - 데이터 입출력, 시각적 변환, 상호작용적인 렌더링 등의 시각화 도구를 공개 소프트웨어 형태로 연구자에게 지원
 - 시각화의 주요 분야로서 전통적인 통계 결과의 그래픽 처리 뿐만 아니라, Web log 시각화, SNS 시각화, 지도 시각화 등 제공
 - Timeline, Modestmaps, Leafletjs, Wolframalpha, Visualizefree, d3js 등 시각화의 주요 도구 활용

□ 데이터 집약형 연구 분야부터 우선 추진, 단계적 확장

○ 전문센터 구축 및 연구데이터 플랫폼 연계를 3 단계로 나누어 추진

－ '18년 추진된 데이터 집약형 시범사업 분야를 1단계 사업으로 추진하여 성공사례를 조기에 창출하고 이후 파급성 등을 고려하여 전 분야로 확산

▶ (1단계, '20~'22년) 바이오, 소재, 인공지능 등 연구집약형 분야에서 우선 추진

▶ (2단계, '23~'26년) 국가전략성, 연구데이터 활용성, 파급성 등 우선분야 순위에 따라 우선분야 확산

▶ (3단계, '27~'29년) 우선분야 추가확산 및 연구데이터플랫폼을 기반 분야간 연계

• '18년 시범사업 추진현황

<바이오> 유전체 빅데이터 경진대회

인공지능 활용 신약개발 플랫폼 구축 및 운영

<소재> 열전소재 분야 연구데이터 플랫폼 구축 및 운영

축매분야 데이터 기반 R&D 환경 구축(KIST)

<대형장비> 이기종 대형연구장비 데이터 융합 분석사업

<인공지능> 인공지능 오픈이노베이션 허브 구축

－ 과학기술 분야에서 중심이 되는 실험, 관측, 참조 데이터를 중심으로 컴퓨팅을 비롯한 IT 인프라 활용이 많은 분야, 국가중점 연구개발 분야 등을 고려하여 1단계 3개 분야 전문센터 및 연구데이터 플랫폼 구축 및 운영 추진

○ 바이오 분야 전문센터 구축 추진방향

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

<표 4-3> <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

－ 중장기적으로 집단연구 및 기존 연구커뮤니티와의 연계를 추진하고 연구데이터 관리 고도화 및 성공적 활용모델 창출

○ 미래소재 분야 전문센터 구축 추진방향

－ 다양한 구조의 소량 데이터로 구성된 소재 연구데이터 특징을 고려하여 응용분야별 연

구데이터 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 산·학·연 활용도 제고
- 인공지능 분야 전문센터 구축 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

□ 플랫폼의 운영과 빅데이터 활용에 필요한 기술력의 체계적 확보

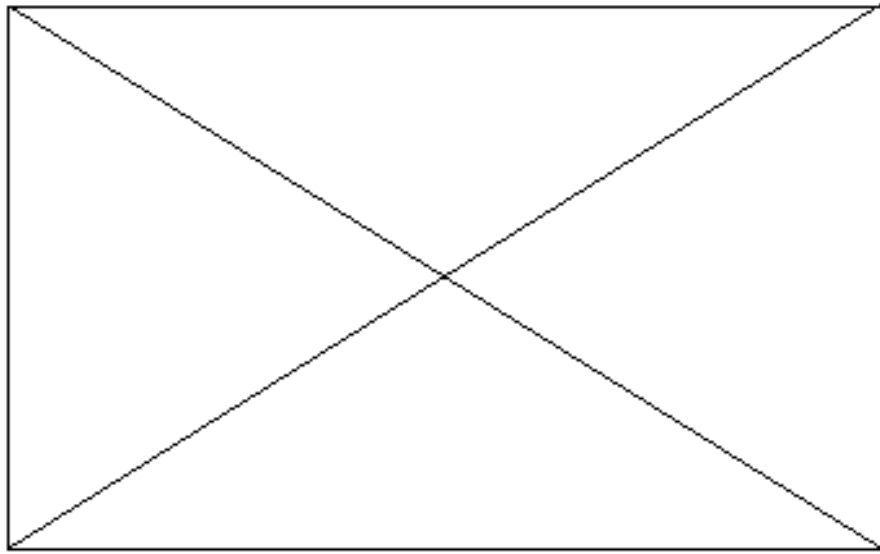
- 연구자에 대한 데이터 및 컴퓨팅 활용 교육 제공
 - 대학생·대학원생 및 현직 연구자들을 대상으로 전문분야에서의 데이터·컴퓨팅 활용을 위한 초급 및 중급 교육 프로그램 제공
 - 한국데이터진흥원 재직자 대상 빅데이터 전문인력 교육 프로그램 연계
 - 전 분야에서 공통적으로 활용되는 데이터 분석기법과 분야별 문제해결에 특화된 데이터 분석 및 활용 사례 교육 등 제공
 - 온라인 데이터·컴퓨팅 활용 교육·실습 환경을 제공하고 분야별 교육용 데이터 बैं크를 구축
 - 수준별 교재 개발하고 교육과정에 기계학습 도구, 계산과학 SW, 광범위한 교육용 데이터 등 교보재 탑재
 - 첨단사이언스교육허브(EDISON) 보유의 소재, 화학, 열유체 등 6개 분야 400여종 SW 활용
- 데이터 기반 연구 참여를 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

4.2.2. 사업의 구성

□ 사업 추진과제 도출

- 사업 추진전략을 이행할 방법을 통해 사업 추진과제 도출



<그림 4-9> 사업 추진과제 도출

□ 「국가연구데이터 공유·활용체제 구축」 사업은 3대 추진과제로 구성됨

- 사업기획 방향에 따른 수요조사
 - － 산·학·연 전문가 및 출연연 대상의 수요조사 등 총 5차례의 기술수요 조사 실시, 전문가 심층 회의 및 정성적 평가
 - － 연구데이터 공유·활용 체제 마련에 따르는 위협요인과 사전에 대응방안 마련, 기술선정 등 효과적 의사결정을 지원하는 3대 추진과제별 산·학·연 전문가 분과위원회 운영
- 사업추진 방향, 목표, 전략에 따른 이행 방법 도출
 - － 연구데이터 공유·활용 원칙 및 제도 등을 반영한 플랫폼 설계 등 5개의 전략에 따라 18개의 추진전략별 이행 방안을 수립
 - － 추진전략별 18개의 이행방안을 수립하고 이를 기반으로 국가연구데이터 플랫폼 구축 및 운영 등 3대 추진 과제 분야와 20대 세부과제를 도출·선정
 - 국가연구데이터 플랫폼 구축 및 운영
 - 대·소분야 전문센터 구축 및 운영지원
 - 데이터 및 컴퓨팅 활용 기반 R&D 인재양성

표 4-4 <<사업 추진전략별 18개 이행방안>>

사업 추진전략	이행 방안
연구데이터 공유·활용 원칙 및 제도 등을 반영한 플랫폼 설계	• 「단계별 법제 정비 및 개선」으로 연구데이터 관리·공유·활용에 관한 법·제도 마련 • 「국가연구개발사업 공동관리 규정」 개정을 통한 법적 정의 및 관리제도 마련 • 「연구관리 표준매뉴얼」 개선을 통한 연구자 권리의 근거 확보 • 단계별 법제 정비 및 개선의 내용이 관리 되도록 플랫폼 설계
융합·공동연구 역량 강화를 위해 분야별 연구공동체 형성	• 분야별 전문센터 및 국가연구데이터 플랫폼 구축 및 운영 • 데이터·컴퓨팅 기반 연구공동체 육성 해외사례 도입 적용 • 데이터 기반으로 R&D 협업을 할 수 있는 공동체 형성 촉진 및 확산
연구데이터의 검색·분석 Tool 등의 환경제공으로 연구자 부담경감	• 원스탑으로 검색하고 활용할 수 있는 국가연구데이터 통합 검색도구 제공 • 연구자가 종합적인 분석을 처리할 수 있는 분석도구 제공 • 거대용량의 연구 데이터를 실시간으로 가시화하는 도구 제공
데이터 집약형 연구 분야부터 우선 추진, 단계적 확장	• 전문센터 구축 및 연구데이터 플랫폼 연계를 3 단계로 나누어 추진 • 바이오 분야 전문센터 구축 추진방향 • 미래소재 분야 전문센터 구축 추진방향 • 인공지능 분야 전문센터 구축 추진방향 • 기타 9개 분야 전문센터 구축 추진방향
플랫폼의 운영과 빅데이터 활용에 필요한 기술력의 체계적 확보	• 연구데이터 관리를 위한 ICT기반 기술교육 • 연구데이터의 R&D 응용을 위한 분야별 전문활용 교육 • 연구데이터·활용 인재육성을 위한 교육용DataBank 및 서비스 구축

□ 3대 추진과제

○ 3대 추진과제의 구성

- 3대 추진과제는 국가 연구데이터 플랫폼 구축 및 운영, 대·소분야 전문센터 구축 및 운영 지원, 데이터 및 컴퓨팅 활용기반의 R&D 인재육성으로 구성됨
- 국가 연구데이터 플랫폼 구축 및 운영은 연구데이터 공유·활용체제 전반의 연구데이터를 원스탑으로 검색하고 활용할 수 있는 통합 플랫폼 구축 및 운영
- 대·소분야 전문센터 구축 및 운영지원은 클라우드 형태로 분야별 컴퓨팅·데이터 인프라를 서비스하고, 개별 연구자에도 데이터관리 클라우드 제공
- 데이터 및 컴퓨팅 활용기반의 R&D 인재육성은 대학·연구소 등 연구자의 데이터 분석·활용능력 제고를 위해 빅데이터 교육과정 운영

○ 3대 추진과제의 목표 및 기대효과

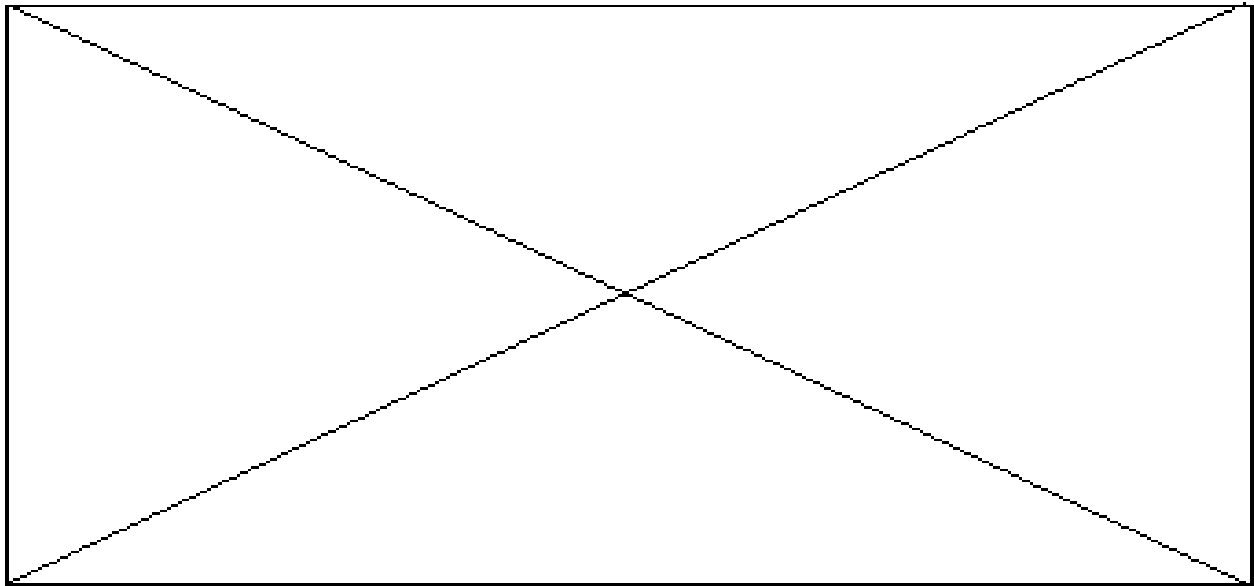
<표 4-5> 3대 추진과제의 목표와 기대효과

3대 추진과제	목표	기대 효과
국가 연구데이터 플랫폼 구축 및 운영	연구데이터를 효과적으로 공유·활용할 수 있는 환경과 기술 지원을 통해 국가 과학기술의 글로벌 경쟁력 제고	- 연구에 필요한 시간과 비용을 절감하여 국가 R&D 효율성을 제고 - 다양한 주제 분야들을 결합한 융합 연구 활성화 - 연구의 효율성 및 투명성과 연구결과물의 수준 제고
대·소분야 전문센터 구축 및 운영지원	데이터기반 연구역량 제고를 위한 분야별 연구데이터 공유·활용 전문데이터센터 구축	- 분야별 정보의 효율적인 활용을 통한 각 분야별 산업발전에 기여 - 빅데이터 활용으로 인해 데이터 엔지니어링, 빅데이터 분석기술제고 - 분야별 연구데이터 현황분석 및 관리를 위한 다양한 활용 인프라 구축
데이터 및 컴퓨팅 활용 기반의 R&D 인재육성	연구데이터 활용 R&D 인재의 지속적 육성을 위한 실무중심 교육체계 구축	- 데이터 전문과학자 양성으로 국가 데이터 산업육성과 발전에 기여 - 다양한 데이터 분석기술 습득 및 연구에 활용 - 실무 적용 및 활용 중심의 빅데이터 교육을 실시함으로써 현장 전문 인력 양성 및 배출

4.2.3. 세부과제별 개요

□ 추진과제별 세부과제 도출

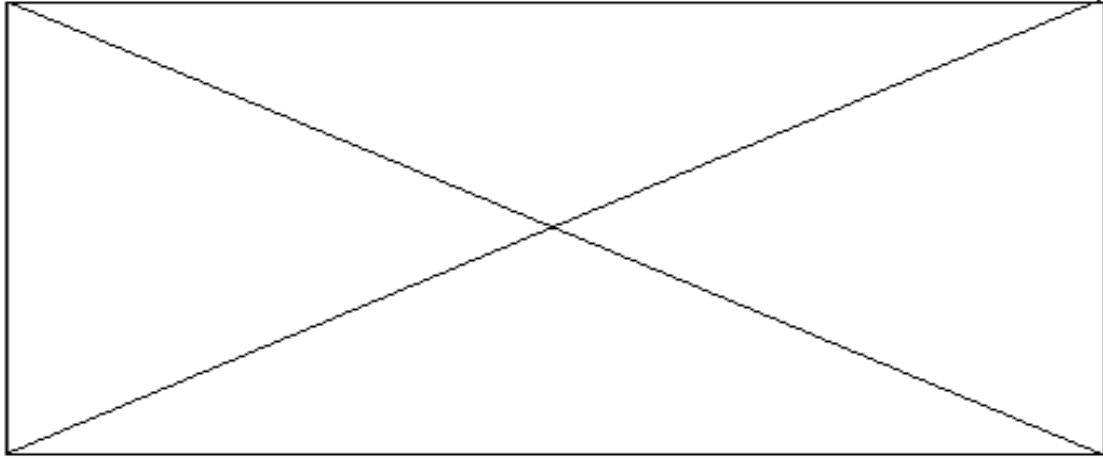
- 국가연구데이터 공유·활용 플랫폼 구축 및 운영 세부과제 도출



<그림 4-10> 국가연구데이터 플랫폼 구축 및 운영 사업 세부과제 도출

- 사업의 배경 및 필요성, 사업 추진방향 및 사업 추진전략 등을 통해 국가연구데이터 플랫폼의 위상과 역할을 정의
- 국가연구데이터 플랫폼의 역할과 구축방향에 따라 데이터 수집관리, 데이터 저장관리, 데이터 분석 등 85개 목표기능 및 서비스를 정의
- 목표기능을 구현하기 위한 국가연구데이터 플랫폼의 97개 요소기술과 요소기술 간의 결합 및 연계 구조를 반영한 14개의 세부기술을 도출
- 요소기술과 세부기술이 Operation될 인프라의 전체 아키텍처, 서버 및 네트워크 아키텍처, 상용 및 시스템 S/W아키텍처, 연계 아키텍처, 응용 아키텍처를 설계하여 전체 기술 구조의 논리적, 물리적 적합성과 타당성을 검증
- 국가연구데이터 플랫폼 구축 및 운영을 위한 기술요소들을 핵심 기술군으로 정립하고 연구개발의 효율성과 효과성을 고려하여
- 연구데이터 수집·연계 및 공유기술, 연구데이터 저장·보안기술, 연구데이터 품질·표준화기술, 지능형 연구데이터 고속검색 기술, 연구데이터 융합·분석 환경 구축기술, 연구데이터 활용·플랫폼 기술 등 6대 세부과제로 도출

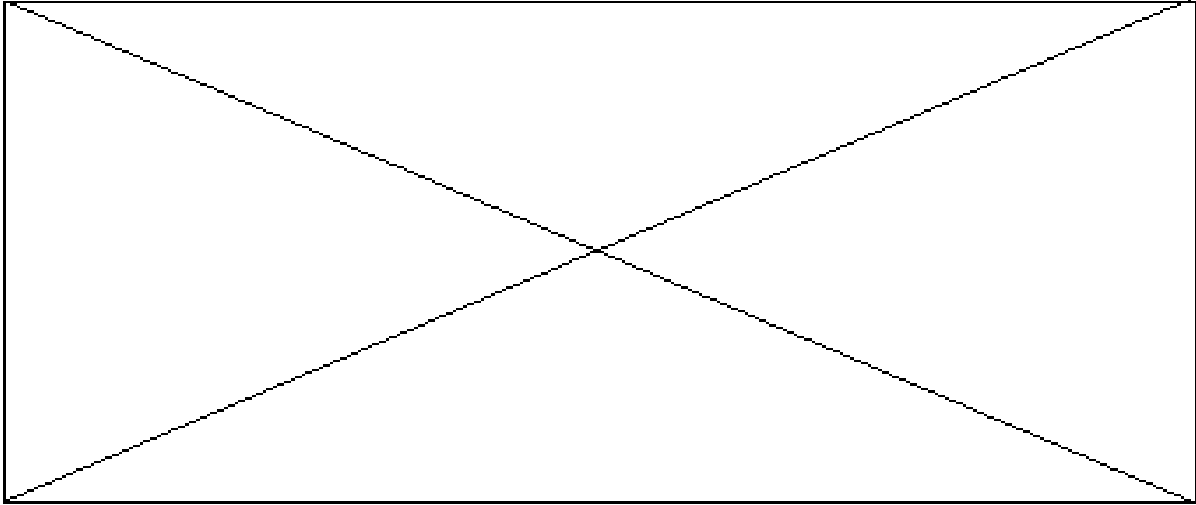
○ 대·소분야 전문센터 구축 및 운영 지원 세부과제 도출



<그림 4-11> 대·소분야 전문센터 구축 및 운영 사업 세부과제 도출

- 정부의 연구개발 중점분야, 국가의 중점과학기술 분야, 해외 사례 등을 통해 연구개발의 특정 범위나 부문을 도출하고 각 부문들의 공통성 및 유사성 분석을 통해 연구데이터 공유 및 활용의 대상이 되는 전문분야를 선정
- 우선 전문분야는 생명, 소재, ICT, 기계, 에너지 등 국가R&D 중점과제 및 연구개발 투자 영역 등을 반영하여 12개 영역으로 도출
- 사회적 요구에 부합하여 대상 분야의 국가전략적 중요성과 연구데이터 공유·활용의 필요성, 파급효과 측면에서 평가지표를 선정
 - 국가전략적 중요성 : 사회적 요구에 부합하여 대상 분야 활용의 중요성을 반영
 - 연구데이터 활용의 필요성 : 대상 분야가 생산하는 데이터양과 연구데이터 공유·활용의 필요성을 반영
 - 파급효과 : 과학기술분야, 시장, 사회적 가치 등의 파급효과의 정도를 반영
- 우선분야 선정의 객관성과 효율성 증진을 위해 평가지표별 중요도 반영(AHP 분석기법 이용)
- 선정된 대상분야의 AHP 계층 분석을 위한 평가지표를 계층적으로 정의하고 우선분야 평가지표별로 AHP 분석을 통해 각 분야별로 평가
- 정량적 평가결과에 항목별 중요도를 반영하여 분야별 최종점수 도출 및 우선순위를 선정하여 1단계로 3개 대소분야를 선정
 - 3개 분야에 대한 전문센터는 바이오, 소재, 인공지능으로 구체화하여 1단계에 우선 구축하고 나머지 9개 분야에 대해서는 1단계 구축과정에 대한 변화관리를 통해 추후 선정

\ ○ 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 세부과제 도출



<그림 4-12> 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 사업 세부과제 도출

- R&D의 수행절차 분석을 통해 실험 및 분석단계에서 연구데이터를 생성 및 수집/저장/전처리/분석/시각화 하여 연구에 활용되는 것을 도출
- 빅데이터 기반 데이터 분석의 기본지식을 보유하고, 해당 연구분야 별 Domain에 대한 업무지식을 보유한 전문인력 육성을 위해 Domain별 데이터분석이 가능한 전문 고급인력 양성을 목표로 수립
- ICT 기반 연구데이터를 관리 및 운영하는 활용 기반 강화형과 R&D에 응용 활용하는 성과창출형의 활용유형을 도출하고 유형에 맞는 인재양성 전략 수립
- 활용기반을 강화하거나 연구데이터에 대한 기본지식과 역량을 제고 하기위해서 데이터 전문 교육기관의 교육과정을 활용하고, 전문 Domain 지식을 활용하여 성과를 창출하는 고급인력에 대해서는 고급 분석가 양성을 전문적으로 담당하는 전문양성과정을 활용
- 국가 연구기관 및 대학 등 학계의 초·중급 연구원, 기업 등 산업계의 산업기술 개발인력 공공분야 연구인력을 대상으로 초·중급 데이터 관리 및 운영 인재육성을 목표로 연구데이터 관리를 위한 ICT 기반기술 교육 과제 도출
- 국가 연구기관 및 대학 등 학계의 석·박사급 고급인력을 대상으로 분야별로 고급 인재육성 데이터 관리 및 운영. 기업 등 산업계에 대해서는 필요 시 고급인력 교육 수행을 목표로 연구데이터의 R&D 응용을 위한 분야별 전문 활용교육 과제 도출
- 연구데이터 활용교육을 위한 교육 인프라 및 서비스 구축을 목표로 연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축 과제 도출

□ 20대 세부과제의 목표

<표 4-6> 추진과제별 20대 세부과제

추진과제	세부과제	세부과제 목표
국가 연구데이터 플랫폼 구축 및 운영	연구데이터 수집·연계 및 공유 기술	국내외 연구데이터/메타데이터 통합, 공유를 위한 수집·전송 시스템 개발
	연구데이터 저장·보안 기술	메타데이터, DMP정보 통합 관리, 대용량 연구데이터의 원활한 분석을 위한 연구데이터 저장, 보안 시스템 개발
	연구데이터 품질·표준화 기술	연구데이터/메타데이터의 품질제고를 위한 연구데이터 품질관리 시스템 개발
	지능형 연구데이터 고속 검색 기술	통합 검색 서비스제공 및 연구플랫폼 서비스 활성화를 위한 검색 기술 개발
	연구데이터 융합·분석 환경 구축 기술	손쉬운 연구데이터 분석 환경 제공을 위한 융합·분석 환경 개발
	연구데이터 활용·플랫폼기술	24·365 무중단 연구데이터 활용을 위한 클라우드 서비스 기술 개발
대·소분야 전문센터 구축 및 운영지원	바이오 분야 전문센터 구축 및 운영지원	데이터기반 연구역량 제고를 위한 분야별 연구데이터 공유·활용 전문데이터센터 구축
	소재 분야 전문센터 구축 및 운영지원	
	인공지능 분야 전문센터 구축 및 운영지원	
	농림수산·식품 분야 전문센터 구축 및 운영지원	
	환경·기상 분야 전문센터 구축 및 운영지원	
	기계·제조 분야 전문센터 구축 및 운영지원	
	우주·항공·해양 분야 전문센터 구축 및 운영지원	
	원자력·핵융합 분야 전문센터 구축 및 운영지원	
	에너지·자원 분야 전문센터 구축 및 운영지원	
	건설·교통 분야 전문센터 구축 및 운영지원	
	재난안전 분야 전문센터 구축 및 운영지원	
데이터 및 컴퓨팅 활용기반의 R&D 인재육성	연구데이터 관리를 위한 ICT 기반기술 교육	연구데이터 수집/저장/운영 등 ICT기반 활용역량 강화
	연구데이터의 R&D 응용을 위한 분야별 전문활용 교육	연구데이터 전처리/분석/시각화 등 분야별 R&D 응용역량 강화
	연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축	연구데이터 활용교육을 위한 교육 인프라 및 서비스 구축

□ 20대 세부과제 개요

○ 연구데이터 수집·연계 및 공유 기술

- 대소분야 센터, 국내외 연구기관의 연구데이터/메타데이터를 수집/전송을 위한 기술개발의 세부과제를 메타데이터 연계기술개발, 연구데이터 연계기술 개발
- 수집·연계 기술로 활용되고 있는 REST, XML-RPC, SOAP, FTP, OAI-PMH 등의 기술 중 적합한 기술을 이용하여 연구데이터/메타데이터를 효과적으로 수집하고, 대용량 데이터 전송을 위한 표준 기술 개발

○ 연구데이터 저장·보안 기술 개발

- 연구데이터/메타데이터를 저장하기 위한 기술로는 현재 클라우드 환경에서 표준으로 사용하고 있는 분산 데이터저장 기술을, 메타데이터의 개인정보 및 비공개 정보의 보안 확보
- 저장 프레임워크 기술 개발 분야에서는 빅데이터 저장 및 분석에 활용되고 있는 분산/병렬 파일시스템과 NoSQL 중심으로 구현 기술 개발
- 보안 프레임워크는 위변조 방지, 암호·복호화, 전송 및 접근제어, 비식별화·익명화 기술을 활용하여 데이터의 보안 및 개인정보 관리 기술 개발

○ 연구데이터 품질·표준화 기술 개발

- 수집,연계되는 메타데이터의 품질 확보를 위한 표준정의와 유통, 저장되는 메타데이터의 품질확보를 위한 관리 기술 개발
- 연구데이터의 출처 및 소유권 정보를 관리하는 저작권 관리 기술 개발
- 연구과제정보와 DMP정보를 기반으로 메타데이터 관리를 위한 스키마, 항목에 대한 표준 수립과 메타데이터 수집 시 수립된 메타데이터 표준(안)을 물기반의 자동 품질검증 기술개발

○ 지능형 연구데이터 검색 기술 개발

- 수집되는 메타데이터, 연구플랫폼의 클라우드 서비스 항목(소프트웨어, 분석도구)에 대한 활용성 확대를 위한 통합검색 기술 개발
- 연구데이터 검색 서비스는 사용자의 등록 연구데이터들에 편리하고 쉽게 조회할 수 있도록 하기 위해 등록된 메타데이터 기반으로 효율적인 검색기능 개발

○ 연구데이터 융합·분석 환경 구축 기술

- 연구플랫폼의 클라우드 서비스의 사용성을 높이기 위하여 클라우드기반의 빅데이터 분석 환경 제공을 위한 workflow, BI, AI, ML, DL, 가시화 기반 기술 개발
- 클라우드 환경에서 연구데이터를 실시간분석, 데이터 분석, 지능형 분석과 결과를 시각화 기술을 기반으로 분석 프레임워크 구성

- 연구데이터 활용·플랫폼관리 기술
 - － 연구플랫폼의 클라우드의 안정적인 서비스를 위하여 클라우드 어드민, 고가용성, 아키텍처, 보안, 가상머신 관리 기술개발
 - － 인프라 시스템(서버, 스토리지, 네트워크 등)에 대한 물리 자원 및 가상 자원 관리를 위한 기능 기술을 개발
 - － 물리적 자원을 개발, 실행 및 운용을 위해 가상 자원 관리 기능들을 기술 개발
- 바이오 분야 전문센터 구축 및 운영지원
 - － 바이오 연구데이터 전문센터와 국가연구데이터 플랫폼 간 메타정보 연계를 위한 기반 구축 및 연계
 - － (과기정통부 생명연구자원 정보 확보) 바이오 분야의 연구에서 산출된 바이오 연구데이터를 분야별 연구내용, 연구성과 발표 타입별, 생산된 연구데이터 타입별로 나누어 등록·관리하고, 제공하는 정보의 다변화
 - － (생명정보 연구성과물 확보) 연구성과물 등록시스템 기반의 연계 체계를 활용하여, 바이오 분야의 연구 방법론, 결과물 등을 수집하고 정제 강화
 - － 생명연구자원 정보기반 멀티 오믹스* 분석시스템 구축(개발 중) 및 정보연계 및 분석 지원을 위한 인프라 확보
- 소재 분야 전문센터 구축 및 운영지원
 - － 소재 전분야 연구데이터 수집·관리·공유·활용 전문센터 구축 및 계산과학, 에너지·환경, 스마트·IT, 안전(구조) 소재 4개 소분야 전문센터 구축
 - 에너지·환경, 스마트·IT, 안전(구조)소재 소분야 전문센터는 응용분야별 과제 수행
 - － 소재 R&D 및 소재산업을 위한 소재·정보·데이터 융복합 혁신기업 양성
 - 계산과학머신러닝, 실험데이터베이스 혁신기업 양성
 - － 데이터 이(利)활용을 통한 신소재 탐색·설계·개발·적용의 효율화, 인력양성, 논문/특허/소프트웨어, 신물질 개발
- 인공지능 분야 전문센터 구축 및 운영지원
 - － 정부 R&D 과제를 수행한 연구소, 학교, 기업이 연구를 위해 구축한 데이터를 쉽게 등록, 처리, 검색 활용 할 수 있도록 데이터 오픈 플랫폼 조성
 - － 연구데이터 관리시스템을 클라우드 구축을 기반으로 실시간 필요 연구 데이터 지원 시스템 구축 및 연구데이터의 효율적·안정적 관리체계 마련, 협력 네트워크 기반 조성
 - － 중장기적 AI연구생태계 활성화를 위한 제도 마련, 플랫폼 구축 및 인력 양성, AI 연구

- \ 데이터 거버넌스 정립, 개방형 AI연구 플랫폼 구축 AI데이터 전문인력 양성
 - AI연구데이터 지원 및 운영/유통센터(가칭)설립
 - 연구데이터 관리를 위한 ICT 기반기술 교육
 - 국가 연구기관 및 대학등 학계의 초·중급 연구원, 기업등 산업계의 사립기술 개발인력 공공분야 연구인력을 대상으로 초·중급 데이터 관리 및 운영 인재육성
 - 연구데이터 및 컴퓨팅 활용 R&D Task 중 데이터 관리영역인 연구데이터 수집과 연구데이터 저장, 데이터 운영영역인 연구데이터 운영 Task에 대한 교육
 - 데이터분석을 위한 기본 지식(데이터 수집, 저장, 운영 등)과 스킬에 대해서는 전문교육기관의 교육과정을 이용할 수 있도록 전문기관을 활용한 위탁교육을 수행
 - 연구데이터의 R&D 응용을 위한 분야별 전문활용 교육
 - 국가 연구기관 및 대학 등 학계의 석·박사급 고급인력을 대상으로 분야별로 고급 인재육성데이터 관리 및 운영, 기업 등 산업계에 대해서는 필요 시 고급인력 교육 수행
 - 연구데이터 및 컴퓨팅 활용 R&D Task 중 데이터 관리영역인 연구데이터 전처리, 연구데이터 분석, 분석결과 시각화 Task에 대한 교육
 - KISTI KSDS의 연구데이터 활용 분석 실무능력 향상 교육 중심, 데이터 전문과학자 고급인력 양성 체계 활용
 - ICT기반 기술교육 및 분야별 전문활용 교육을 운영 및 진행하는 가칭 데이터 교육지원센터를 운영하고 이를 지원하고 피교육생 및 바우처 등을 관리하는 지원시스템 구축
 - 연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축
 - EDISON 플랫폼에서 활용가능한 교육용 데이터 리포지토리 구축 및 서비스
 - 다양한 과학기술·사회과학 분야의 데이터를 전 처리를 통해 교육실습에 적합한 형태의 데이터를 제공할 수 있는 데이터 리포지토리 구축 및 서비스
 - 동영상 강의와 연계된 시뮬레이션 SW 및 데이터 분석툴, 인공지능 예측 모델링의 실행 가능한 반응형 콘텐츠를 개발하여 기존 계산과학공학 플랫폼인 EDISON 플랫폼에 탑재
 - 연구데이터를 수집, 저장 처리하고 분석 및 시각화의 데이터 관리 기술 S/W의 기술교재 개발 및 플랫폼인 EDISON 플랫폼에 탑재 운영

4.2.4. 세부과제별 성과목표(지표)

□ 성과목표 및 지표 도출

○ 국가연구데이터 플랫폼 구축 및 운영 세부과제별 성과목표(지표)

<표 4-7> 국가연구데이터 플랫폼 구축 성과목표(지표)

세부과제	성과 목표	성과지표
연구데이터 수집·연계 및 공유 기술	- 연구데이터/메타데이터 수집 및 연계 기술 개발 - 대용량 연구데이터를 분석환경에 전송하기 위한 연계 기술 개발	- 논문(SCI급) - 논문(KCI) - 특허 - 기술혁신 - 콘텐츠S/W - 기술활용효과 - 중소기업지원 - 연구개발서비스 - 인력양성 - 일자리창출 - 정책효과 - 홍보, 확산 - 국제협력 - 공동활용시설장비 - 전산시스템
연구데이터 저장·보안 기술	- 대용량 데이터 저장을 지원하는 클라우드 기반의 분산, 병렬 저장 기술 개발 - 데이터 연계 및 저장 시 보안 기술 개발	
연구데이터 품질·표준화 기술	- 연구데이터의 품질 관리를 위한 표준 정의, 관리기술 개발 - 연구데이터의 출처, 저작권 관리	
지능형 연구데이터 고속 검색 기술	- 메타데이터 기반의 효율적인 검색 방안을 개발 - 클라우드 서비스 항목(소프트웨어, 분석도구)에 대한 검색 기술 개발	
연구데이터 융합·분석 환경 구축 기술	- 다양한 BI, AI, ML, DL솔루션을 연구플랫폼에 탑재 - 분석 솔루션을 workflow기반으로 활용할 수 있도록 구축	
연구데이터 활용·플랫폼기술	- 연구자간 협업 및 커뮤니티 기능 개발 - 안정적인 클라우드 서비스 운영을 위한 기반 기술 개발	

○ 대·소분야 전문센터 구축 및 운영지원 세부과제별 성과목표(지표)

<표 4-8> 대·소분야 전문센터 구축 성과목표(지표)

세부과제	성과 목표	성과지표
바이오 분야 전문센터 구축 및 운영지원	• 바이오 연구데이터 연계체계 구축 • 바이오 연구데이터 수집 • 바이오 연구데이터 활용시스템 구축 및 지원을 위한 인프라 구축	• 논문(SCI급) • 논문(KCI) • 특허 • 기술혁신 • 콘텐츠S/W • 기술활용효과 • 중소기업지원 • 연구개발서비스 • 인력양성 • 일자리창출 • 정책효과 • 홍보, 확산 • 국제협력 • 공동활용시설장비 • 전산시스템
소재 분야 전문센터 구축 및 운영지원	• 소재분야 연구데이터 대·소전문센터 구축 • 소재·정보·데이터 융복합 혁신기업 양성 • 소재 데이터 활용 연구개발에 의한 경제적 효과 창출 • 정량적 연구개발 성과 목표	
인공지능 분야 전문센터 구축 및 운영지원	• 인공지능 학습용 데이터셋 6,000개 제공 • 인공지능 관련 연구데이터의 클라우드 허브 플랫폼 수용률 80% 달성 • 타분야 연구데이터가 참고가능한 데이터 표준모델 10종류 수립 • AI기술 이론과 실데이터 경험을 보유한 AI데이터 전문가 5,000명 양성	

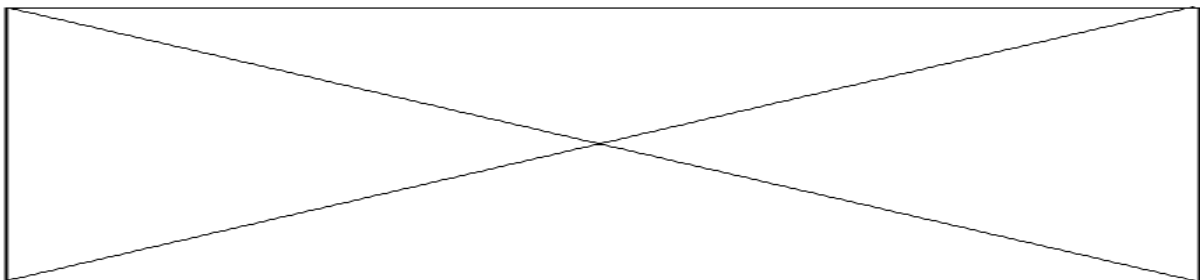
○ 데이터 및 컴퓨팅 활용기반 R&D 인재육성 세부과제별 성과목표(지표)

<표 4-9> R&D 인재육성 성과목표(지표)

세부과제	성과 목표	성과지표
연구데이터 관리를 위한 ICT 기반기술 교육	• 연구데이터 수집/저장/운영 등 ICT기반 활용역량 강화	• 중소기업지원 • 인력양성 • 일자리창출 • 정책효과 • 홍보, 확산 • 공동활용시설장비 • 전산시스템
연구데이터의 R&D 응용을 위한 분야별 전문활용 교육	• 연구데이터 전처리/분석/시각화 등 분야별 R&D 응용역량 강화	
연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축	• 연구데이터 활용교육을 위한 교육 인프라 및 서비스 구축	

□ 성과지표 설정근거

○ 성과지표 설정 프레임워크



* 10대 사업유형, 5대 성과지표 : 과학기술정보통신부 “국가연구개발사업 표준 성과지표 (성과목표·지표 설정 안내서)”참조

○ 본 사업의 사업유형

<표 4-10> 10대 사업유형 중 본사업의 사업유형

성격	유형	개념 및 분류 기준
연구개발	1. 기초연구	- 자연현상의 원리규명, 새로운 현상의 분석 등을 통해 창조적 지식 획득 연구(순수 기초형) - 현재 또는 미래에 광범위한 응용을 목적으로 문제해결의 근본원리 및 창의적 지식창출 연구(목적기초형)
	2. 단기산업 기술개발	단기간 내(3년 이내) 상용화를 목표로 한 신기술 및 신제품 개발을 위한 응용개발 연구사업
	3. 중장기산업 기술개발	중·장기적(3년 이상) 상용화를 목표로 추진 중인 응용개발 연구사업
	4. 공공기술개발	응용·개발단계 연구개발사업 중 최종적인 성과가 국민 건강증진, 재난방지 등 국민 삶의 질에 기여하는 형태로 나타나는 사업
	5. 지역연구개발	지역 대학과 연계한 산학협력 사업, 지역클러스터 육성사업, 특정 지역에 특정기술 개발 기반구축 사업
	6. 국방기술개발	응용개발단계 연구개발사업 중 국방력 강화 및 방위산업 발전을 목적으로 하는 사업
연구기반 조성	7. 인력양성	대학 및 전문대학 지원사업, 산업인력양성을 위한 전문인력양성사업, 초·중·고 과정의 과학기술 교육사업 등
	8. 시설장비구축	- 대형 연구시설 및 장비 구축 사업 * 사업 예산에 단순 시설 증축 및 장비 구입 등이 일부 포함된 경우는 제외
	9. 성과확산	사업목적이 각각 기술사업화, 표준화, 인증, 성과물 관리/확산, 정책지원 등인 사업
	10. 국제협력	- 해외기관유치, 다자 및 양자기관 협력 사업 등 * 연구방식이 해외와의 공동연구인 경우는 연구개발에 포함

- 연구개발 측면에서 본 사업은 연구데이터를 수집하고 저장하고 분석 활용하는 목적 기초형인 기초연구 사업이면서 중장기적으로(10년) 연구데이터를 기반으로 빅데이터 및 AI 산업 측면의 응용연구 개발사업임
- 연구기반 조성 측면에서 본사업은 데이터 및 컴퓨팅을 활용하여 R&D를 수행할 수 있는 인재 양성과 플랫폼 센터 및 대·소분야 전문센터를 위한 인프라를 구축하고 운영하는 시설장비 구축사업 임

○ 본 사업의 성과분야

<표 4-11> 본사업의 성과분야

사업유형/기간	초기	중기	장기
1. 기초연구	기술적 성과 과학적 성과	과학적 성과 기술적 성과	과학적 성과 기술적 성과 경제적 성과 사회적 성과
2. 단기산업 기술개발	기술적 성과 경제적 성과	기술적 성과 경제적 성과	기술적 성과 경제적 성과
3. 중장기산업 기술개발	과학적 성과 기술적 성과	과학적 성과 기술적 성과 경제적 성과	경제적 성과
4. 공공기술개발	과학적 성과 기술적 성과	과학적 성과 기술적 성과 사회적 성과	경제적 성과 사회적 성과
5. 지역연구개발	과학적 성과 기술적 성과 경제적 성과	기술적 성과 경제적 성과 사회적 성과	경제적 성과 사회적 성과
6. 국방기술개발	기술적 성과 경제적 성과 인프라 성과	기술적 성과 경제적 성과 인프라 성과	기술적 성과 경제적 성과 인프라 성과
7. 인력양성	과학적 성과 기술적 성과	과학적 성과 기술적 성과	경제적 성과 사회적 성과
8. 시설장비구축	인프라 성과	인프라 성과	경제적 성과 사회적 성과 인프라 성과
9. 성과확산	기술적 성과 경제적 성과 인프라 성과	기술적 성과 경제적 성과 인프라 성과	기술적 성과 경제적 성과 인프라 성과
10. 국제협력 과학적	과학적 성과 기술적 성과	과학적 성과 기술적 성과	경제적 성과 사회적 성과

- 본 사업은 장기간(10년) 동안 진행되는, 사업유형이 기초연구, 중장기산업 기술개발, 인력양성, 시설장비구축 성격의 혼합사업임
- 성과지표 분류체계상 5대 성과지표인 과학적성과, 기술적성과, 경제적성과, 사회적성과, 인프라성과 지표를 모두 적용해야 하는 사업임

○ 본 사업의 성과지표

<표 4-12> 5대 성과분야 중 본사업의 성과지표

성과유형		성과지표	비고
과학적 성과	논문	논문(SCI급)	논문건수
		논문(KCI)	논문건수
	신자원·물질	생명자원	
		화합물	
기술적 성과	사회적 평가	포상	
	지식재산	특허	특허건수
		비특허	
	비지식재산	기술혁신	공정혁신, 표준획득
	성장동력창출	콘텐츠S/W	S/W등록
		서비스개발	
		제품개발	
		플랫폼개발	
경제적 성과		신약/의료기기개발	
	사회적평가	포상	
	직접성과	기술료	
		경제효과	
	간접성과	기술활용효과	분석환경, 클라우드
		중소기업지원	교육, 분석환경
	기술사업화	기술사업화	
	연구개발서비스	연구개발서비스	표준화, 연구개발지원
사회적 성과	인적자원·고용	일자리창출	
	인적자원·고용	인력양성	인력수
		일자리창출	일자리 창출 수
	지역사회	지역발전	
	공공복지	정책효과	
		공공서비스	
	과학대중화	홍보, 확산	
	국제협력	국제협력	
인프라 성과	연구인프라	공동활용시설장비	DB, 계산(분석)시스템
		단독활용시설장비	
		전산시스템	응용시스템
	우주	우주개발	
		우주활용	
	국방	무기체계	

- 본 사업이 국가연구개발 사업임에 따라 성과물의 과학적 수준을 나타내는 논문의 게재 학술지 수준, 영향력 지수 등을 통해 과학적 성과를 측정가능
- 연구데이터를 집약하는 본사업의 특성상 빅데이터화, 클라우드기반 서비스, 고속의 컴퓨팅 환경 제공 등 기술적 혁신이 필요하고 이를 지원하는 특화된 S/W 개발등 기술적 성과를 측정 가능
- 플랫폼, 전문센터의 분석인프라 활용, 중소기업의 인재양성 교육과정지원, 연구데이터를 활용한 연구개발 지원 등 다양한 경제적 성과를 측정가능

\ - 인력양성 및 일자리창출 등 다양한 사회적성과 측정 가능

□ 과제별 성과목표

○ 사업전체 성과목표

<표 4-13> 성과목표 총괄

구분		1단계			2단계				3단계			계
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	
과학적 성과	논문 (SCI)	14	26	24	27	22	19	20	20	20	20	212건
	논문 (KCI)	34	57	54	60	47	39	43	44	43	43	464건
기술적 성과	특허	출원 (19) 등록 (7)	출원 (26) 등록 (11)	출원 (25) 등록 (10)	출원 (28) 등록 (11)	출원 (22) 등록 (9)	출원 (19) 등록 (7)	출원 (21) 등록 (8)	출원 (21) 등록 (8)	출원 (21) 등록 (8)	출원 (21) 등록 (8)	출원 (223) 등록 (87)
	기술혁신	표준 모델(6)	표준 모델(6)	표준 모델(8)	표준 모델(7)	표준 모델(7)	표준 모델(8)	표준 모델(4)	표준 모델(4)	표준 모델(4)	표준 모델(4)	58개
	콘텐츠 S/W		4	6		1		2				13건
경제적 성과	기술활용 효과					5조					5조	10조
	중소기업 지원					기업 양성 (1)					기업 양성 (1)	2개
	연구개발 서비스	분야 과제 (1)	분야 과제 (1)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	분야 과제 (2)	18개
사회적 성과	인력양성	2,605	2,747	2,843	2,943	3,049	3,208	3,323	3,443	3,569	3,700	31,430명
	일자리 창출	421	439	459	479	500	522	545	569	594	620	5,148 명
	정책효과	10%	10%	20%	20%	30%	30%	40%	40%	50%	50%	10건
	홍보확산	1	1	1	1	1	1	1	1	1	1	10건
	국제협력			1				1			1	3건
인프라 성과	공동활용 시설장비	데이터 set (250)	데이터 set (250)	데이터 set (500)	데이터 set (500)	데이터 set (500)	데이터 set (1000)	데이터 set (500)	데이터 set (500)	데이터 set (1000)	데이터 set (1000)	6,000 개
		교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (1)	교육용 DB구축 (2)	교육용 DB구축 (1)	교육용 DB구축 (1)	11개
			서버장 비(6)	서버장 비(9)				서버장 비(3)				18개
	전산 시스템	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	3Core 102TB	30Core 1PB
		지원 시스템 (1)	세부별 (2)	세부별 (3)				세부별 (1)				7개

○ 국가연구데이터 플랫폼 구축 및 운영

<표 4-14> <<국가연구데이터 플랫폼 구축 및 운영 성과목표>>

구분		1단계			2단계				3단계			계
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	
과학적 성과	논문 (SCI)	14	21	19	22	17	14	15	15	15	15	167건
	논문 (KCI)	34	52	49	55	42	34	38	39	38	38	418
기술적 성과	특허	출원 (14) 등록 (7)	출원 (21) 등록 (11)	출원 (20) 등록 (10)	출원 (23) 등록 (11)	출원 (17) 등록 (9)	출원 (14) 등록 (7)	출원 (16) 등록 (8)	출원 (16) 등록 (8)	출원 (16) 등록 (8)	출원 (16) 등록 (8)	출원 (173) 등록 (87)
	기술혁신											
	콘텐츠 S/W		4	6				2				12건
경제적 성과	기술활용 효과											
	중소기업 지원											
	연구개발 서비스											
사회적 성과	인력양성											
	일자리 창출											
	정책효과	10%	10%	20%	20%	30%	30%	40%	40%	50%	50%	10건
	홍보확산	1	1	1	1	1	1	1	1	1	1	10건
	국제협력			1				1			1	3건
인프라 성과	공동활용 시설장비		6	9				3				18개
	전산 시스템		2	3				1				6개

* 논문/특허 : 본 사업의 성과지표의 설정기준을 마련하기 위해 기 추진한 유사사업의 최근 3년간 논문 및 국제특허의 출원/등록 수를 분석한 결과 SCI급 논문은 10억 원당 1.15건, 국제 특허는 10억 원당 1.11건으로 조사(SW컴퓨팅원천기술개발 및 ICT융합원천기술개발사업 등 2개 사업)

* 콘텐츠S/W : 세부과제별 성과목표별로 개발 기술에 대한 S/W 출원

* 정책효과 : DMP 정책에 따른 연구데이터 활용 R&D 집행을 10%씩 증감

* 홍보확산 : 매년 1회 성과에 대한 홍보활동

* 국제협력 : 단계별 해외 세미나 참석 및 홍보

* 공동활용 시설장비 : 세부과제별로 Web/Was 서버, DB서버, Computing 서버 책정

* 전산시스템 : 세부과제별 1개로 책정

Ⅱ ○ 대·소분야 전문센터 구축 및 운영지원

<표 4-15> 대·소분야 전문센터 구축 및 운영지원 성과목표

구분		1단계			2단계				3단계			계
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	
과학적 성과	논문 (SCI)		5	5	5	5	5	5	5	5	5	45 건
	논문 (KCI)		5	5	5	5	5	5	5	5	5	45 건
기술적 성과	특허	5	5	5	5	5	5	5	5	5	5	50 건
	기술혁 신	표준 모델(6)	표준 모델(6)	표준 모델(8)	표준 모델(7)	표준 모델(7)	표준 모델(8)	표준 모델(4)	표준 모델(4)	표준 모델(4)	표준 모델(4)	58 건
	콘텐츠 S/W											
경제적 성과	기술활 용효과					5조					5조	10조
	중소기 업지원					기업 양성(1)					기업 양성(1)	2 개
	연구개 발서비 스	분야 과제(1)	분야 과제(1)	분야 과제(2)	분야 과제(2)	분야 과제(2)	분야 과제(2)	분야 과제(2)	분야 과제(2)	분야 과제(2)	분야 과제(2)	18 개
사회적 성과	인력양 성	500	550	550	550	550	600	600	600	600	600	5,700 명
	일자리 창출											
	정책효 과											
	홍보확 산											
	국제협 력											
인프라 성과	공동활 용시설 장비	데이 터셋 (250)	데이 터셋 (250)	데이 터셋 (500)	데이 터셋 (500)	데이 터셋 (500)	데이터 셋 (1000)	데이 터셋 (500)	데이 터셋 (500)	데이터 셋 (1000)	데이터 셋 (1000)	6,000 개
	전산 시스템	3Core 102T B	3Core 102T B	3Core 102T B	3Core 102T B	3Core 102T B	3Core 102T B	3Core 102T B	3Core 102T B	3Core 102T B	3Core 102T B	30Cor e 1PB

○ 데이터 및 컴퓨팅 활용기반 R&D 인재육성

<표 4-16> <<데이터 및 컴퓨팅 활용기반 R&D인재육성 성과목표>>

구분		1단계			2단계				3단계			계
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	
과학적 성과	논문(S CI)											
	논문(K CI)											
기술적 성과	특허											
	기술혁 신											
	콘텐츠 S/W					1						1개
경제적 성과	기술활 용효과											
	중소기 업지원											
	연구개 발서비 스											
사회적 성과	인력양 성	2,105	2,197	2,293	2,393	2,499	2,608	2,723	2,843	2,969	3,100	25,730 명
	일자리 창출	421	439	459	479	500	522	545	569	594	620	5,148 명
	정책효 과											
	홍보확 산											
	국제협 력											
인프라 성과	공동활 용시설 장비	교육 용 DB구 축 (1)	교육 용 DB구 축 (1)	교육 용 DB구 축 (1)	교육 용 DB구 축 (1)	교육 용 DB구 축 (1)	교육 용 DB구 축 (1)	교육 용 DB구 축 (1)	교육 용 DB구 축 (2)	교육 용 DB구 축 (1)	교육 용 DB구 축 (1)	11개
	전산 시스템	지원 시스 템(1)										1개

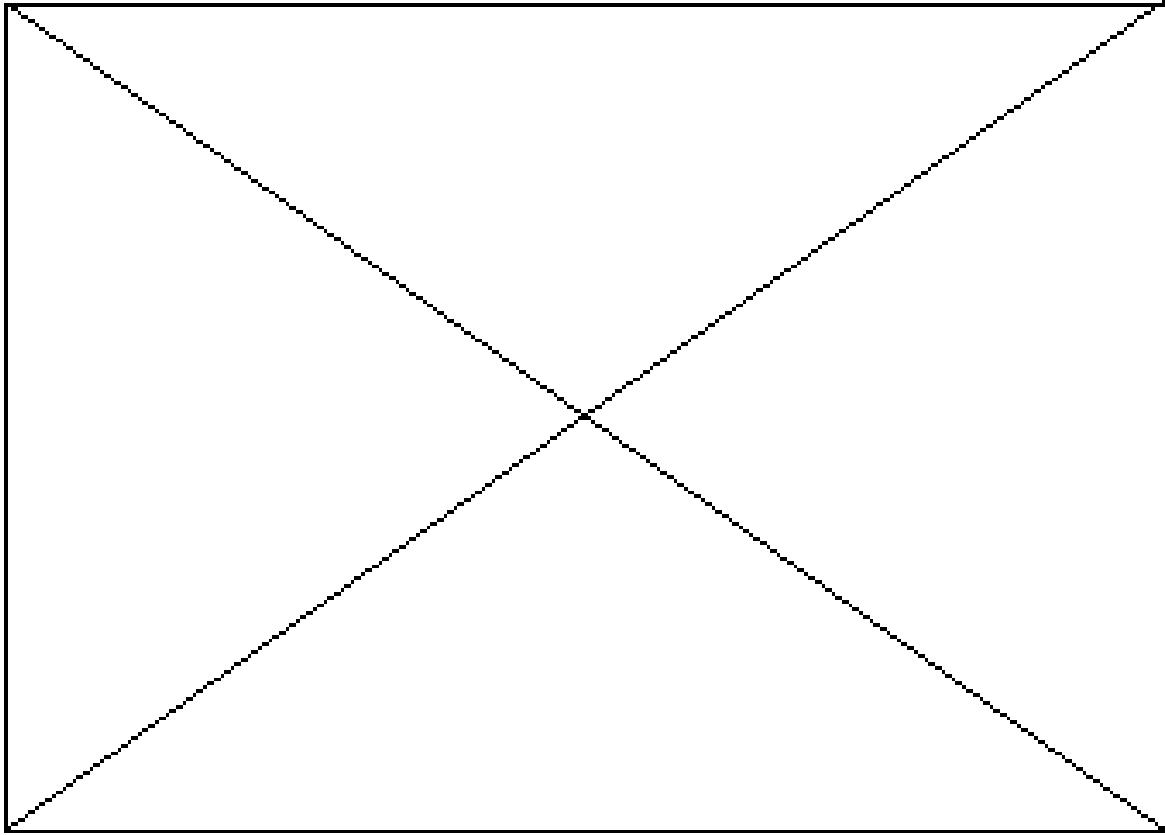
* 콘텐츠 S/W : 교육용 콘텐츠 저작도구

* 인력양성 : 국가기관, 대학 등 연구원 증가분을 반영한 대상인원 도출(데이터 및 컴퓨팅 활용기반 R&D 인재양성 과제 참조)

* 일자리창출 : 인력양성 대상 인력의 20%를 취업목표

* 공동활용시설장비 : 대소분야 전문센터 11개 분야에 대한 교육용 DB 구축

4.2.5. 과제별 연계성 강화방안

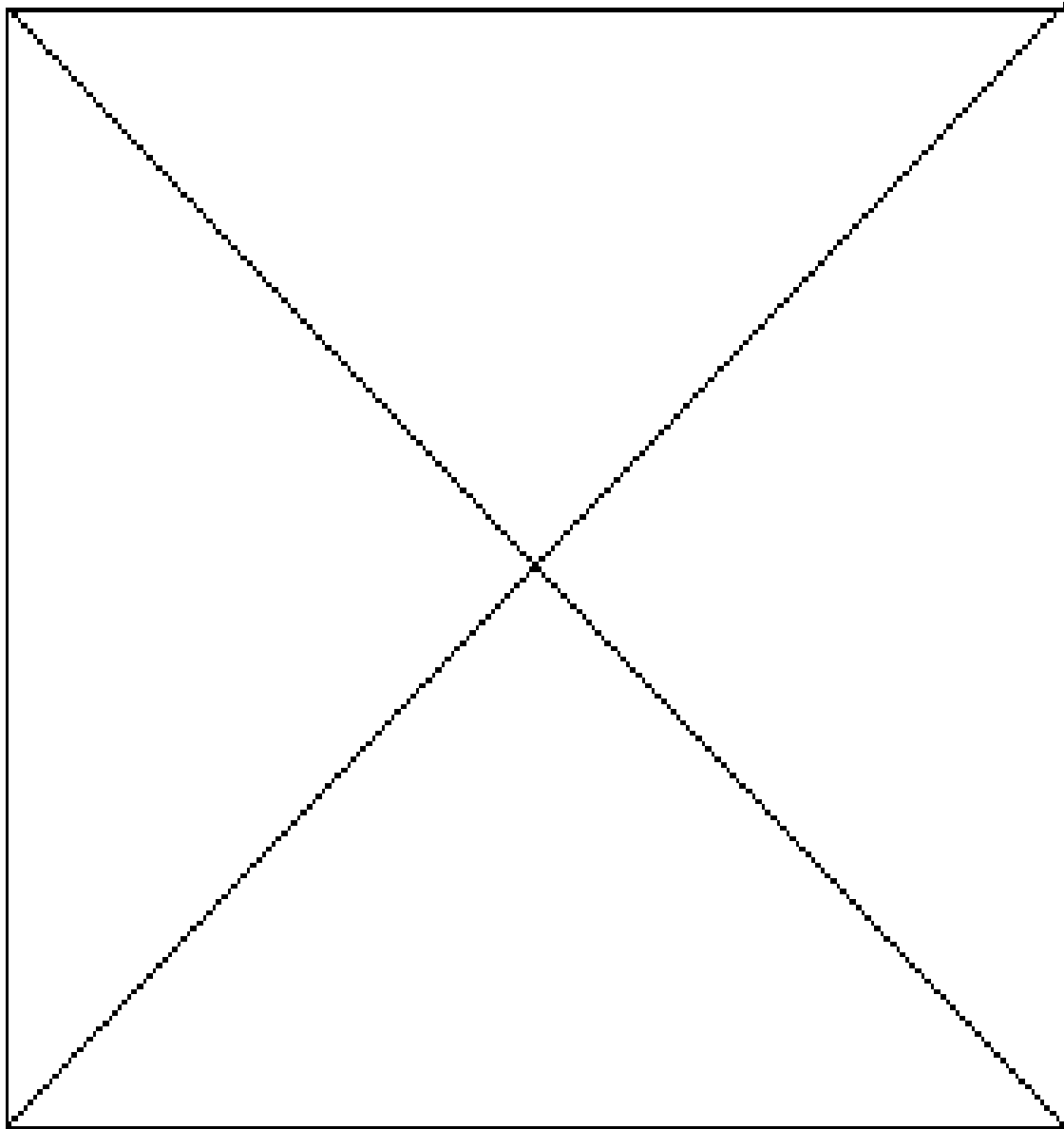


<그림 4-13> 추진 과제간 연계도

- 3개의 추진과제 간 연관 범위를 설정하고, 연계되는 데이터 및 정보의 활용 목적을 설정
 - 국가연구데이터 플랫폼 구축 및 운영 과제와 대소분야 전문센터 구축 및 운영지원 과제의 연계
 - －(연계의 목적)국가연구데이터 플랫폼은 국가의 연구데이터 공유·활용 체계의 총괄센터로서 대·소분야 전문센터를 지원하는 역할을 하고
 - －대·소분야 전문센터는 센터에서 저장 관리되는 연구데이터에 대한 메타정보를 국가연구데이터 플랫폼에 연계함으로써 국가연구데이터의 검색 정보를 제공
 - －(연계의 내용)국가연구데이터 플랫폼은 대·소분야 전문센터 또는 산하 연구기관, 연구자가 요청하는 연구데이터에 대한 검색요청에 대해 해당 연구데이터의 Location 정보 등을 DMP 및 메타정보를 활용해서 고속 검색하여 제공
 - －국가 연구데이터 플랫폼에 관리되는 분석도구, 분석을 위한 방법론, 분석 S/W 등을 전문센터에 지원하고 제공

- 국가연구데이터의 연계기술, Hosting 기술 등 플랫폼 운영을 위한 각종 플랫폼 S/W를 전문센터와 연계하여 지원하고 제공
- 대·소분야 전문센터는 분야별 연구데이터에 대한 메타 정보를 국가연구데이터 플랫폼에 제공하여 국가연구데이터를 검색하는 기반정보를 제공
- 대·소분야 전문센터는 필요 시 분야별 연구데이터를 국가연구데이터 플랫폼에 저장하거나 Hosting*하는 방식을 활용함으로써 전문센터별 추가적으로 필요한 클라우드 저장 및 분석을 국가연구데이터 플랫폼을 활용
 - * 클라우드 컴퓨팅 인프라 제공
- 국가연구데이터 플랫폼 구축 및 운영과제와 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 과제의 연계
 - (연계의 목적)연구데이터 공유·활용 체계의 총괄 기능을 하게 될 국가 연구데이터플랫폼을 운영하고 활용할 고급인재의 성장지원
 - (연계의 내용)박사과정 인력 및 석사과정 인력을 선발하여 연구데이터 분석전문인력 육성
 - 국가적 빅데이터 기반 산업의 중추를 담당하고 빅데이터 처리 이론을 다양한 산업·연구 분야에 접목할 능력을 갖춘 최고급 인재로 양성
 - 데이터 통합, 분석 방법론, 알고리즘, 보안, 시각화 등 다양한 학제간 융합에 초점을 맞춘 교과목 운영 등 석박사 중심 고급인력 양성과정 개설
- 대·소분야 전문센터 구축 및 운영지원 과제와 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 과제의 연계
 - (연계의 목적)분야별 연구데이터 공유·활용 기능을 하게 될 대·소분야 전문 센터를 활용하여 데이터 기반 연구 참여와 전문센터 기반 강화를 위한 전문인력 양성지원
 - (연계의 내용)데이터 연구분야에서 유용하게 활용될 수 있는 전문 연구관에서 운영 중인 KISTI의 EDISON 등을 적극 활용할 수 있도록 개방정책 확산
 - 우수한 R&D 인력, 교수진 등 교육 인프라 기반이 갖추어진 거점 대학들 간 교육과정을 유기적으로 운영 하는 등 기 운영 중인 전문 연구기관의 교육과정 심화
- 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 과제를 통한 연구자에 대한 교육연계
 - (연계의 목적)학교, 기업체 등의 연구자가 데이터 기반의 컴퓨팅을 통해 연구를 과학적으로 수행하도록 관련 교육기회를 제공하여 연구자의 R&D 역량을 제고
 - (연계의 내용)삼성SDS멀티미디어캠퍼스, SAS 등 빅데이터 스킬관련 전문기관에 교육을 효과적으로 수행 할 수 있는 교육바우처 제도도입 등 전문 교육기관을 활용한 위탁교육

□ 기술 개발과 과제 간 연계·활용 방안



<그림 4-14> 3대 추진과제의 기술 개발과 과제간 연계·활용 로드맵

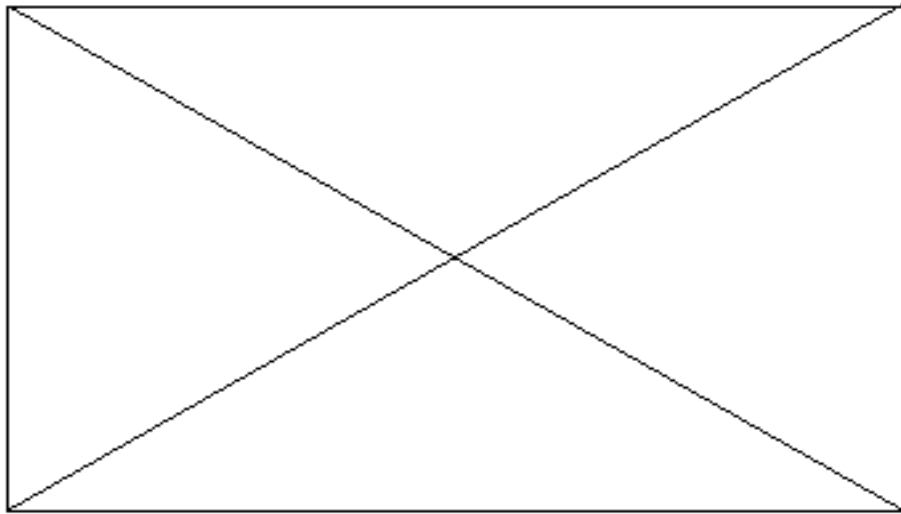
- 「국가 연구데이터 플랫폼 구축」 개발기술과 「대·소분야 전문센터 구축」 개발기술의 연계 활용
 - －수집, 연계, 품질, 표준화 등은 전문센터와 밀접한 연계기술 및 관계기술로 1단계 도입기에 동일한 기간에 기술 개발 및 기능 구축을 수행
 - －2단계 국가연구데이터 플랫폼 구축 최적화 및 안정화시기에 전문센터와의 연계 및 관계기술의 문제점 및 개선점을 도출하여 2단계 전문센터 구축 시 반영될 수 있도록 국가연구데이터 플랫폼 최적화 안정화 기간과 전문센터 기본 및 상세설계 일정을 병행

- 3단계 안정화 기간에 8개 이상의 전문센터가 국가연구데이터 플랫폼과 연계 연동이 최적화 될 수 있도록 최적화 및 안정화 기간을 수행
- 기본설계, 상세설계, 프로토 타입과 개발, 최적화 및 안정화 등 동일한 구축 절차를 활용하여 개발 기술을 서로 활용할 수 있도록 함
- 최적화 운영과 기술개발이 교차될 수 있도록 절차를 수립하여 최적화 운영의 결과가 기술개발에 지속적으로 반영될 수 있도록 함
- 「데이터 및 컴퓨팅 기반 R&D 인재육성」의 교육과정의 전략적 운영
 - 빅데이터 및 대량의 연구데이터를 수집하고 처리하는 기본과정을 사업이 시작되는 초기부터 지속적으로 운영하여 데이터를 활용하는 연구방법이 확산되고 분위기가 조성될 수 있도록 연계
 - ICT기반 기술 교육 프로그램 및 교육플랫폼을 활용한 R&D 활용 교육과정은 바이오, 소재, 인공지능 분야의 전문센터의 최적화 및 안정화 운영단계와 연계하여 교육과정을 운영함으로써 효율성을 제고
 - 고급인력 양성교육과정은 국가연구데이터 플랫폼 구축의 융합·분석 환경구축 기술이 개발 완료되는 시점과 연구데이터 활용 플랫폼기술 개발이 완료되는 시점에 맞춰 교육과정을 진행
 - 교육과정과 함께 교육용 DataBank 및 서비스 구축은 사업초기 핵심 서비스 및 콘텐츠로 구축하여 제공하고 플랫폼 및 전문센터 구축 시 데이터셋을 구성할 수 있도록 연계

4.2.6. 결과물 공유 및 부처 간 협력방안

□ 국가연구데이터 공유·활용체제 구축사업 결과물 공유

○ 국가연구데이터 공유·활용체제 구축사업결과물의 특징



<그림 4-15> 사업 결과물의 특징

- 국가연구데이터 공유·활용체제 구축 사업을 통해 정책적측면, 운영측면, 기술측면, 데이터측면에서 다양한 정책, 조직기능, IT기술, 연구데이터 등의 결과물을 생산
 - 생산된 결과물은 구체적인 기술 및 데이터뿐만 아니라 정책, 데이터센터 운영 등의 조직관리, 데이터 기반의 연구방식으로의 변화 등을 선도
 - 국가연구데이터 공유·활용체제 구축사업 결과물 활용 방안
 - (정책적 측면) 데이터 관리계획(DMP)에 따라 데이터에 대한 공개범위와 시기가 의무적으로 제공되어 연구데이터 공유·활용의 문화 정착과 연구데이터를 활용한 R&D 역량을 제고 할 수 있는 기회제공
 - 활용대상 : 연구자, 연구기관, 공공기관, 대학(연구기관), 중소기업(연구소)
- ▶ 데이터 관리계획(DMP)에 따라 공개되는 R&D에 필요한 관련 연구데이터를 검색하여 자유롭게 활용
 - ▶ 데이터기반 연구, 오픈사이언스 방식의 연구로 연구결과에 대한 객관성 확보
 - ▶ 데이터 기반 R&D 역량 제고를 위한 교육과정 참여(바우처제도)
- (운영 측면) 계층적으로 운영되는 국가연구데이터 플랫폼, 분야별 전문센터, 기관별 데이터 센터의 구조 하에 데이터 뿐만 아니라 플랫폼, 표준기술, 상용기술 등을 계층구조에 맞게 운영함으로써 활용대상 기관들이 최적합한 연구데이터를 관리할 수 있는 운영기회 제공

- 활용대상 : 연구기관, 공공기관, 대학(연구기관)

- ▶ 계층별 데이터, 플랫폼, 표준기술 등의 가이드라인 활용
- ▶ 기관에서 제공하지 않는 연구자측면의 다양한 서비스 활용(PaaS, 분석기술 등)
- ▶ 데이터기반 R&D 역량제고를 위한 교육과정 참여(전문기관의 연구데이터 교육 플랫폼 활용)

－ (기술 측면) 빅데이터 기술, 인공지능기술, 머신러닝 기술 등 다양한 4차 산업혁명의 기반이 되는 핵심기술 활용하고 플랫폼 및 클라우드 기술을 활용할 수 있는 기회 제공

- 활용대상 : 연구자, 연구기관, 공공기관, 대학(연구기관), 중소기업(연구소)

- ▶ 빅데이터 처리기술, 플랫폼 기술, 클라우드 기술 등의 활용 및 기술이전
- ▶ 상용 연구데이터 분석 SW R&D를 활용 및 기술이전
- ▶ 우수SW는 국가연구데이터 플랫폼 센터가 라이선스 구입하여 보급·활용 및 기술이전
- ▶ 데이터기반 R&D 활용 제고를 위한 교육과정 참여(고급 연구역량 교육활용)

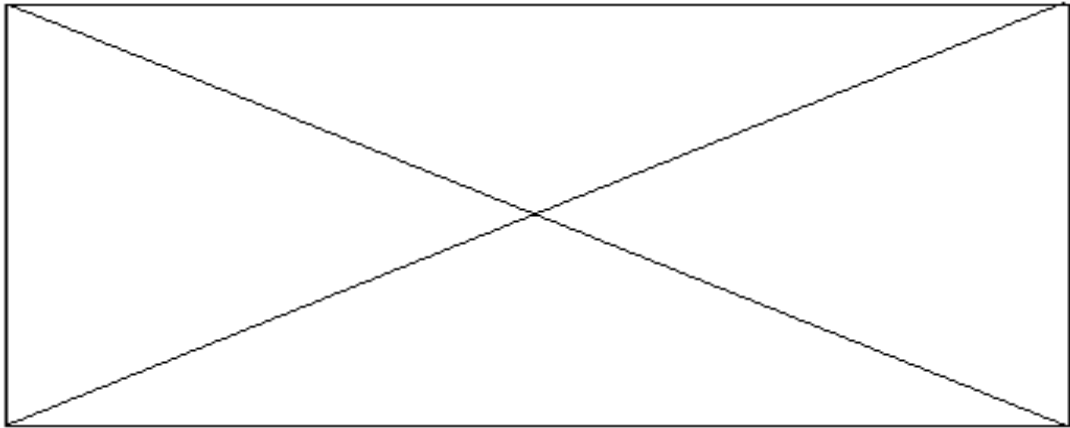
－ (데이터 측면) 관찰, 센싱 등 대용량 연구데이터, 비정형, 반정형의 다양한 연구데이터, 연구결과를 검증하기 위한 가공 연구데이터 등을 DMP 범위 내에서 활용

- 활용대상 : 연구자, 연구기관, 공공기관, 대학(연구기관), 중소기업(연구소), 사기업

- ▶ 활용대상별 R&D 성과제고를 위해 연구데이터의 활용
- ▶ 기업이 사업화 R&D 및 시제품 개발에 연구데이터를 손쉽게 활용
- ▶ 전문센터의 기업대상 정기적 활용교육 프로그램, 데이터 활용 컨설팅 활용

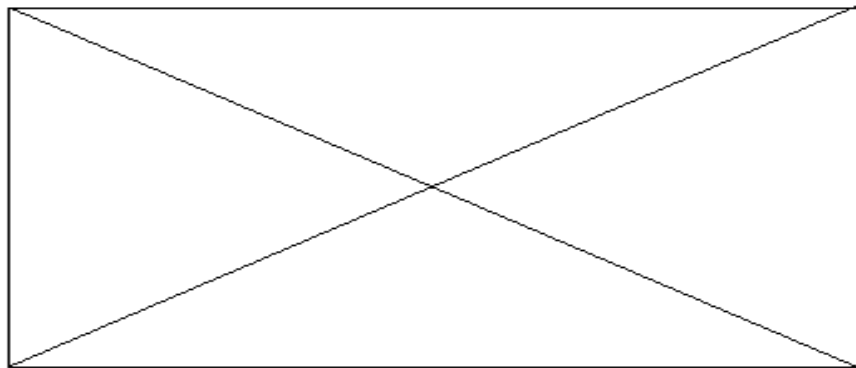
□ 부처 간 협력방안

- 과기 정통부는 산·학·연·관이 필요로 하는 4차 산업혁명 관련 원천기술을 개발·공유
 - － 빅데이터, 인공지능, 클라우드 등에 대한 기술을 농업·물류·제조 등 타 분야로 공급, 새로운 비즈니스 모델 및 혁신기업 육성을 위한 토대 마련
 - － 아울러, 혁신기술 개발에 관련 S/W, 기초기술 등이 규제로 인해 장애가 되지 않도록 관계부처와 협력하여 국내 기업의 성장 기반 구축



<그림 4-16> 사업관련 부처의 협력

- 국가연구데이터 플랫폼, 전문센터 구축을 위해 관련 정부부처의 적극적인 참여 및 지원이 유지되도록 협력체계를 운영
- － 과기부는 사업을 총괄할 전담기관 및 사업단과 범부처 협의회*를 운영하여 부처간 이견 및 요구사항 등을 사업에 원활하게 반영



<그림 4-17> 범부처 협의회 운영

- 범부처 협의체는 국가연구데이터 공유·활용 체제 구축 사업의 종료시점까지 운영하고 정기적으로 범부처 협의체 회의를 개최
- 연간 사업계획에 따른 사업 추진경과 점검 및 성과 공유·확산 방안, 사업추진에 중요하다고 판단하는 사항에 대한 범부처 의사결정
- － 국가연구데이터 공유·활용체제 구축 사업과 관련된 정부의 협의기구*를 통해 국가 R&D 사업의 혁신적인 변화를 위한 주요 의사결정의 창구로 활용
- 4차산업혁명 위원회, 정보통신전략위원회, 국가과학기술심의회(운영위원회, 특별위원회), 생명공학 종합정책심의회 등

국가연구데이터플랫폼 구축 및 운영

- 5.1. 과제개요
- 5.2. 추진전략
- 5.3. 세부과제별 추진내용
- 5.4. 기대효과
- 5.5. 연구개발대상 기술 선정과정

5.1. 과제개요

5.1.1. 추진배경

- 주요 선진국들 중심으로, 4세대 R&D의 핵심인 융합연구 및 오픈 사이언스의 주요 기반이 되는 연구데이터 공유 플랫폼 구축을 경쟁적으로 추진하고 있음
- 바이오·소재 등 일부 분야의 연구데이터 수집이 진행 중이나, 국가 차원의 연구데이터 공유·활용을 위한 통합 인프라가 부재
- 국내·외 연구데이터를 수집·연계하여 다양한 연구자가 공유하고 융합·분석에 활용할 수 있는 국가 차원의 연구데이터플랫폼 구축이 시급

5.1.2. 필요성

- (R&D 메가트렌드 측면) 연구데이터 공유·융합을 통한 가치 창출 필요성 증대
 - 최근 R&D는 데이터의 축적 및 연관성 도출 활동과 데이터를 공유·융합하고 분석하여 문제를 해결하는 추세로 패러다임이 변화함
 - 공공연구의 성과를 개방하는 오픈사이언스 대두로 중복 연구 감소, 연구결과의 검증, 연구성과의 파급성 등을 증대시키기 위해 연구데이터가 개방과 공유의 대상으로 정의됨
 - 해외 주요국은 연구데이터 공유·활용 목적의 플랫폼을 마련하여 연구에 활용하고, 연구자들 간 협업 연구가 증가하며 가치를 발굴하는 성과를 창출하고 있음
 - 점차 복잡해지고 어려워지는 과학연구에서 새로운 발견과 효율적 연구를 위해 연구데이터의 공동 활용 및 협력이 중요해지고 있음
 - ※ 대형연구장비 데이터 공유 및 공동연구 : 힉스입자 발견('13), 중력과 관측('16) 등 성공
 - ※ 미(美) 소재게놈전략(Material Genome Initiative) : 소재 개발기간비용의 1/2 감축 목표
- (정책 및 제도적 측면) 연구데이터 중요성 증대 및 국가 경쟁력의 원천으로 대두
 - 데이터의 중요성과 활용 가치는 점점 증대되고 있어 단순히 경제적 가치를 넘어 국가 경쟁력의 원천이며, 미래사회의 가치를 창출할 엔진으로 인식됨

- 주요 선진국은 연구데이터의 중요성을 인지하여 국가적인 투자와 제도적인 정책을 마련하고 시행 중이며, 미국, 영국 등 주요국들은 논문, 특허 등 최종 성과물의 공유를 넘어 국가적 차원에서 연구데이터를 공유·활용하는 체계 구축을 위해 노력하고 있음

<표 5-1> 주요 국가별 연구데이터 정책추진 현황

국가	정책
미국	- 과학재단(NSF)를 중심으로 2010년부터 연구데이터 정책 추진 - 백악관 과학기술정책실 ‘연구데이터 관리 및 공유에 대한 지침’(‘13년)에서 원칙 제시 및 빅데이터 연구 개발 계획 발표(‘16년)
영국	- 데이터정책공동원칙(RCUK Common Principles on Data Policy)으로 2011년부터 연구데이터 관리 및 공유 원칙 제시 - 정보자유법에서 공개 정보의 범위에 데이터 셋을 포함시키고, 영국연구위원회의 공동 원칙을 기반으로 공유 데이터 센터 및 데이터 큐레이션 센터 운영
호주	- 호주책임연구수행법(Australian Code for the Responsible Conduct of Research)으로 ‘연구진실성’ 확보 측면에서 접근 ‘책임 있는 연구 수행을 위한 국가 지침’에 연구 산출물 관리 규정을 마련하고, 호주국가 데이터서비스에서 연구 기관들의 메타 데이터 수집·제공

□ (연구데이터의 공유·활용 측면) 연구데이터 공유·활용을 위한 국가적인 환경 구축 필요

- 국가 R&D 예산의 지속적인 증가와 데이터의 중요성이 증대됨에도 불구하고, 국내 연구데이터의 관리와 활용이 미흡함
- 특히, 보유처 정보부족 및 소유주체 간 칸막이로 개인의 인적 네트워크에 의존·활용하고 있는 상황으로 데이터 공동 활용이 제한적임

* 출처: KISTI 실태 조사 결과(‘11)	* 출처: KISTI연구데이터 관리실태 조사결과(‘18)

<그림 5-1> 국내 연구데이터 관리 및 공유 현황

- (인프라 구축 및 운영 측면) 국가적 차원의 연구데이터 활용과 공유·연계를 위한 플랫폼 구축 및 서비스 운영 필요
- 미국, 영국 등 주요국들은 연구데이터 수집·저장·연계·유통을 위한 국가연구데이터센터 구축 및 컨설팅·교육을 제공하는 지원 조직 운영하고 있음
 - (미국) 주요 연방기관 및 대학에서 연구데이터 관리 및 공유를 위한 데이터 센터와 리파지토리를 운영하고 있음
 - Dataverse Network Project, SciDAC(Scientific Discovery through Advanced Computing)-2 등을 통해 연구데이터 공유 및 다학제 기반의 협업연구를 지원함
 - (유럽) FP7의 e-Infrastructure 구축사업을 추진하여 과학데이터와 컴퓨터 자원의 실시간 공유 및 가상 연구 환경을 제공
 - Horizon2020 사업에서 '14년부터 e-인프라에 기반한 분야별 전문센터(CoE, Center of Excellence) 운영('16년 9개)
 - (호주) ANDS를 중심으로 국가 차원의 과학데이터 수집, 관리 및 서비스 구축·운영
 - 호주국가데이터서비스(ANDS)에서 연구데이터 관리·공유 인프라 구축, 연구자 가이드·교육 제공 및 연구자간 협력 지원 담당함
 - '09년부터 분야별 가상실험실 플랫폼(Nectar) 운영('17년 12개)
 - (영국) 연구데이터 공유센터 및 연구데이터의 공유·활용을 지원하는 디지털 큐레이션 센터²⁴⁾(DCC) 구축·운영함
 - 국내의 경우 NTIS를 중심으로 국가연구개발정보와 8대 성과물은 공유연계가 이루어지고 있으나, R&D 성과물 중 하나인 연구데이터의 공유는 미흡한 실정임
 - 데이터 집약형 연구 확산에 따라 특정 분야 차원에서 일부 데이터를 수집·관리 중이나, 활용은 저조함
 - ※ 유전체(국가생명자원정보센터), 소재 물성(소재정보은행) 등

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

5.2. 추진전략

5.2.1. 추진방향 및 목표

가. 추진방향

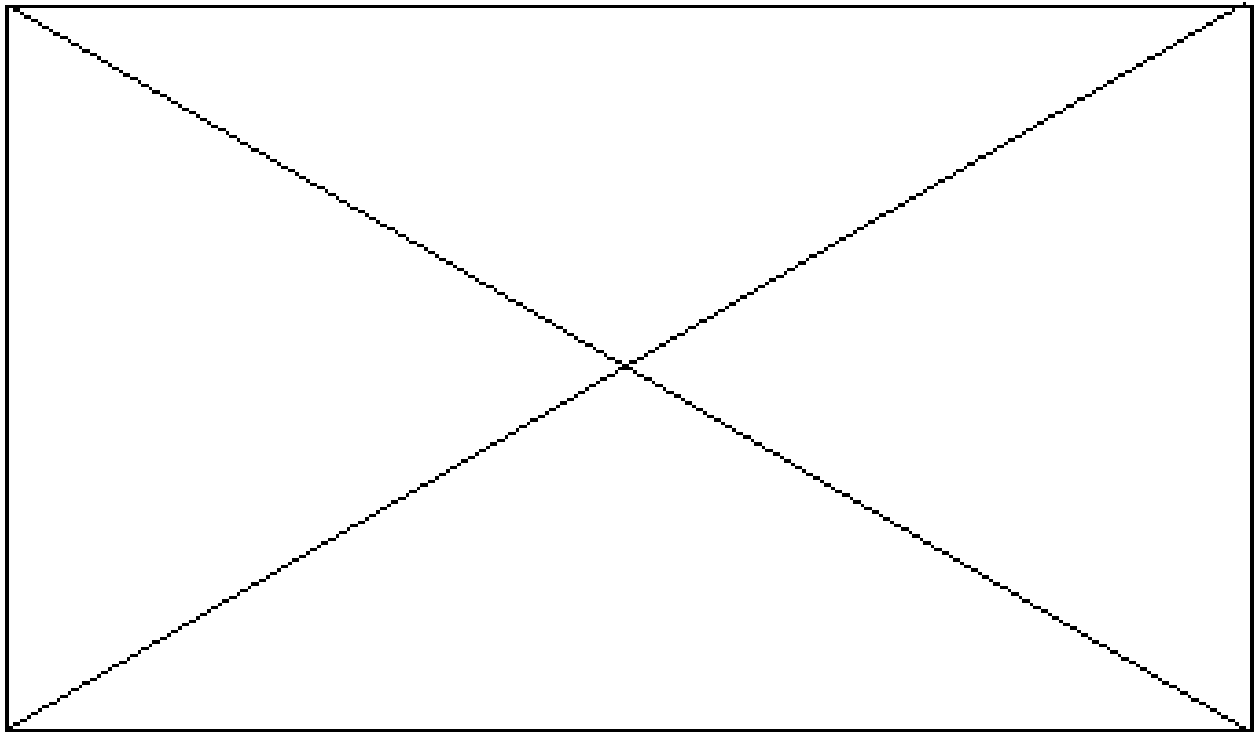
□ 국가연구데이터플랫폼·인프라 구축·운영

- 연구데이터 공유·활용 관련 법/제도, 표준화 등의 반영
 - － 연구데이터 관리를 위한 공통, 전문분야별 표준 가이드라인 관리 및 배포 지원
 - － 연구데이터 관리계획(DMP) 의무화에 따른 관리 및 사용자 편의성 지원
 - － 전문분야별 특화된 메타데이터 및 표준화를 통한 상호운용성 확보 지원
 - － 워크플로우 기반의 데이터 품질, 분석모델 등 관리업무 표준화 지원
- 연구자의 권리보호 및 데이터 소유권 보장
 - － DOI, 엠바고 및 공개범위 등의 설정 기능 제공
 - － 연구데이터 등록, 사용, 만족도 등과 같은 정량적 활용실적의 추적과 이력 관리
 - － 데이터 유통이력의 투명한 관리와 위변조 방지 등을 위한 블록체 기술 접목
- 연구데이터 공유·활용 활성화를 위한 사용자 중심의 서비스 제공
 - － 연구데이터 등록 시 DMP 작성 및 메타데이터 자동 생성 지원
 - － 사용자 패턴 분석 등을 통한 맞춤형, 지능형 검색 서비스
 - － 연구자가 데이터 융합 및 분석에 활용할 수 있는 분석도구 및 가상작업 환경 구축
 - － 연구자간 협업연구를 위한 커뮤니티 활동 지원
- 안전한 시스템 및 데이터를 보장하는 인프라 구축
 - － 시스템의 보안사고(비인가자 접속 및 접근, 개인정보유출 등) 방지 및 대처를 위한 기술 적용
 - － 데이터의 보호 및 안정적 관리를 위한 저장, 보존, 보안, 백업, 복구 기술 적용
 - － 무중단, 고가용성, 확장성 보장을 위한 검증된 최신 기술(클라우드, 분산/병렬 데이터 처리 등) 적용
 - － 국가연구데이터 플랫폼의 운영과 대·소분야 전문센터 임무수행 지원을 위한 컴퓨팅·스토리지 등 인프라 구축

□ 대·소분야 전문센터 플랫폼·인프라 지원

- (플랫폼 지원) 전문센터의 연구데이터 공유·활용 플랫폼 및 데이터 분석 SW 개발·운영을 지원하고 컨설팅 및 수집·연계 프로그램 제공
- (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·검색할 수 있도록 저장 공간(예: 개인당 100TB)·데이터분석 도구 등 제공

나. 추진목표



<그림 5-2> 국가연구데이터플랫폼 추진목표 및 개념도

5.2.2. 추진전략 및 세부과제

가. 추진전략

□ 전략1. 국가연구데이터플랫폼의 기능적 역할 정의 및 요소기술 선정

- 국가연구데이터플랫폼은 다양한 기관과 시스템에 산재되어 있는 연구데이터 및 메타데이터의 수집·연계, 분석 및 활용 지원 등의 기능적 역할 정의
 - － 국가연구데이터플랫폼 구축을 위한 기능 구성 및 기능요건, 상세 기능의 정의
 - － 국가연구데이터플랫폼 구축 시 필요한 요소기술과 세부기술 선정에 따른 세부과제 구성

□ 전략2. 데이터의 상호운용성 확보를 위한 표준화 및 품질 관리

- 다양한 연구환경에서 생산 관리되는 연구데이터의 공유·활용 시 상호운용성 확보를 위한 표준화 관리기능 제공
 - － 연구데이터 관련 분류체계 및 코딩표준, 데이터베이스 구축 및 공유를 위한 표준, 서비스(정보공개, 표현, 교환 및 공유 관련) 표준 등 관리
 - － 메타데이터 등록과 인증을 통해 표준화된 메타데이터 관리 및 공유할 수 있는 레지스트리 구축
 - － 분야별로 연구데이터의 생산·관리·활용 등과 관련한 데이터 처리 절차 및 방법 등 워크플로우에 따른 관리
 - － 바이오, 소재, 인공지능 등 영역별로 데이터 공유 및 재사용을 위한 메타데이터 표준 관리
 - － 데이터의 신뢰성 및 유효성 확보를 위한 품질검증 및 표준 준수여부 관리

□ 전략3. 연구데이터 수집·연계를 위한 국내·외 연구기관 및 협력 네트워크와 협업

- 국내·외 R&D 사업으로부터 발생하는 방대한 연구데이터를 수집·연계하기 위해 국내·외 연구기관들과의 협업 추진
 - － 국가연구데이터플랫폼 구축을 위한 수집·연계 요소기술 개발 시 분야별 전문센터, 개별 연구데이터센터와 협업 및 연계 API 보급
 - － 해외 데이터 공유·활용 프로젝트와 연계하여 데이터 리포지토리와 상호운용성 확보
 - － 국내·외 Domain Scientist 네트워크 구축
- 연구데이터 수집·연계 대상과 역할 수립
 - － 연구데이터 공유·활용을 위해 국가연구데이터플랫폼과 수집·연계 대상은 분야별 연구데이터플랫폼, 연구과제관리시스템, 해외 연구데이터플랫폼 등으로 구분됨

<표 5-2> 국가연구데이터플랫폼과 수집·연계 대상

수집·연계대상	역할 및 설명	관리·운영주체
분야별 연구데이터플랫폼	- 소분야(Tier-2) 전문센터의 연구데이터 수집·연계 등 관리 - 데이터 관리·분석SW 개발, 교류·협력 프로그램 운영 등	분야별 전문센터 (Tier-1)
연구과제관리시스템	- 국가 과학기술 연구과제 정보 및 연구원 정보 관리(NTIS) - 국가 과학기술 연구과제, 성과 및 DMP 정보 관리(성과마루)	- KISTI - 한국연구재단
해외 연구데이터플랫폼	분야별 표준/비표준 연구데이터 및 메타데이터 관리	해외 개별 운영기관
Open Data 사이트	개방형 데이터 웹 사이트를 구축 및 운영	정부, 공공기관, 연구단체 등
분석환경지원시스템	고성능 컴퓨터(슈퍼컴퓨터), 외부 클라우드와 같이 연구데이터에 대한 분석환경을 지원	- KISTI - 민간 클라우드
개별 연구관리시스템	자체 연구데이터 관리 및 분석	개별 연구기관

※ (분야별 전문센터) 분야별 연구데이터의 공유·활용을 위해 소분야 전문센터의 연구데이터를 취합·관리하기 위해 분야별 연구데이터플랫폼을 관리·운영

※ (개별 연구기관) 분야별 전문센터에 속하지 않는 연구기관으로 자체적인 연구관리시스템을 보유하며 관리·운영

□ 전략4. 연구자가 데이터 융합 및 분석에 활용할 수 있는 다양한 환경 지원

- 대용량 정보 분석기술 및 분산/병렬 인프라 기반 확장을 통한 융합·분석 환경 지원
 - － 데이터 분석지원을 위한 분석도구 및 가상작업 환경 구축, 대용량 데이터 분석환경지원을 위해 가상슈퍼컴퓨터(KISTI)와 연계 및 활용
 - － 연구현장 수요 변화에 따른 탄력적 운용성 확보 등을 위해 민간 클라우드 적극 활용
 - － 다양한 데이터 분석도구의 자유로운 공유·유통을 위해 데이터 분석도구 마켓 운영

□ 전략5. 안전하고 투명한 데이터 유통체계 확보

- 데이터 유통이력의 투명한 관리와 위변조 방지 등을 위한 데이터 저장·유통 시스템에 보안기술 적용
- 연구데이터 생산자의 권리보호를 위한 엠바고, 인용표기, 이용조건 및 라이선스 관리
- 연구데이터의 질적 수준 확보를 위한 표준화와 품질검증 관리

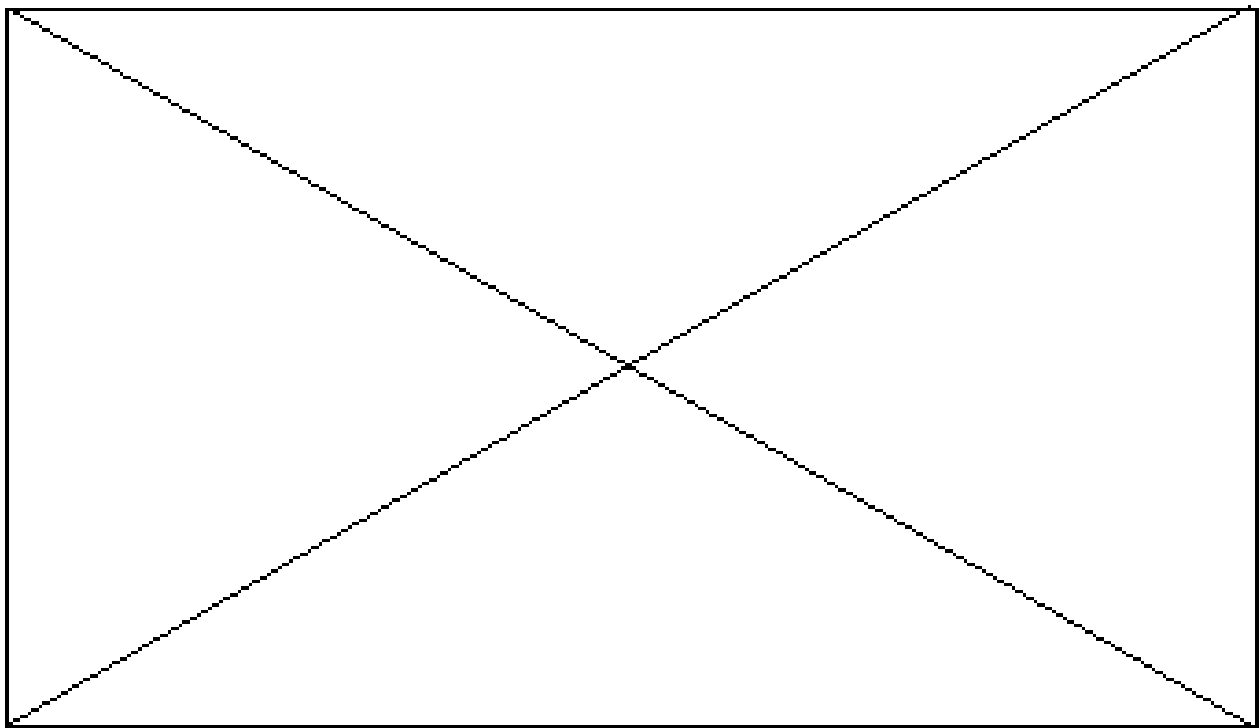
□ 전략6. 범 국가 차원의 연구데이터 공유·활용 체계 활성화 지원

- 자체적으로 인프라 구축이 어려운 대·소분야 전문센터, 연구기관이나 연구자에게 인프라

- \ 를 지원하고, 연구데이터 공유·활용을 지원할 수 있도록 분석도구 및 프로그램 개발 등 인프라 측면의 자원 제공
 - 인프라 지원을 필요로 하는 기관에게 국가연구데이터플랫폼의 클라우드서비스(IaaS, PasS) 제공
 - o 대·소분야 전문센터, 연구기관 등을 위한 연구자용 가이드·교육자료 제공 및 연구자간 협력 등 컨설팅·교육 프로그램 제공

나. 국가연구데이터플랫폼 기능 구성 및 요소기술

- 연구데이터 공유·활용체제 지원을 위한 국가연구데이터플랫폼의 기능은 4개(수집·연계 프레임워크, 서비스 프레임워크, 보안 프레임워크, 시스템 관리 프레임워크)의 영역으로 구성함



<그림 5-3> 국가연구데이터플랫폼 기능 구성도

- o 수집·연계 프레임워크
 - 연구데이터 및 메타데이터 연계 그룹, 과제정보 연계 그룹과의 다양한 연구데이터 및 메타데이터 수집과 연계 인터페이스 제공

- 데이터 분석, 데이터 기반 R&D 수행을 지원할 수 있도록 슈퍼컴퓨터(KISTI) 등 분석 환경 연계 인터페이스 제공
- 서비스 프레임워크
 - 연구데이터에 대한 메타데이터의 사용자/인증관리, 데이터 등록 관리, 데이터 큐레이션, 분석 및 활용, 클라우드서비스(IaaS, PaaS)를 위한 기능 제공
- 보안 프레임워크
 - 시스템 보안 관리, 위험 관리, 네트워크 보안 등의 관리 기능 제공
- 시스템 관리 프레임워크
 - 국가연구데이터플랫폼의 물리적, 논리적 인프라스트럭처와 클라우드서비스 및 SW 관리 기능을 제공함. 클라우드 및 분산처리 관점에서의 관리 뷰를 제공하며, 이 프레임워크 의해 제공되는 기능들은 클라우드서비스 운용자 및 클라우드 관리자에게 각각의 역할에 맞는 정보를 제공하고 관리활동을 지원

1) 국가연구데이터플랫폼의 기능 요건

□ 수집·연계 프레임워크

- 연구데이터 및 메타데이터 연계
 - 국가연구데이터플랫폼 기반의 국가연구데이터센터, 분야별 전문센터 및 연구센터들은 등록된 연구데이터들의 공유 활용 및 센터들 간의 자원 공유를 위한 상호연계가 필요함
 - 국내 분야별 전문센터와 연구데이터 및 메타데이터를 연계하는 기능
 - 자체 연구관리시스템을 보유하고 있는 국내 정부 출연 연구기관들과 연구데이터 및 메타데이터를 연계하는 기능
 - 해외 주요 연구데이터플랫폼에서 공유·활용 중인 연구데이터들을 연계하는 기능
 - 주요 Open Data 사이트의 데이터를 수집하는 기능
- 연구과제정보 및 DMP 연계
 - NTIS, 성과마루 등 과제관리시스템 연계를 통해 국가 과학기술인 정보 및 국가 과학기술 프로젝트들에 대한 참조 데이터와 DMP를 수집함. 국가 과학기술인 정보는 국가연구데이터 센터의 사용자 등록을 위한 참조 정보로 활용하며, 과학기술 프로젝트 정보는 연구데이터를 등록하기 위한 참조 데이터로 공식적인 국가 과학기술 프로젝트 정보를 제공해야 함
 - 국가연구데이터플랫폼과 NTIS와 성과마루 간 연계 기능

- \
 - 국가연구데이터플랫폼은 NTIS에 등록된 과학기술인 정보를 수집하여 국가센터의 사용자 참조 데이터로 저장
 - 국가연구데이터플랫폼은 NTIS에 등록된 국가 과학기술 프로젝트 정보를 수집하여 연구데이터의 프로젝트 참조 데이터로 제공
 - 국가연구데이터플랫폼은 NTIS에 등록된 DMP정보를 수집하여 연구데이터 참조 데이터로 제공

○ 분석환경 연계

- 국가연구데이터플랫폼은 등록된 연구데이터 기반의 분석환경을 제공하고, 등록된 연구데이터의 분석 기능을 제공하는 시스템(KISTI 슈퍼컴퓨터, 외부 클라우드)들과의 일련의 분석 요청, 분석 실행 및 분석 결과 수집을 지원해야 함
- 사용자의 분석 서비스 활용을 위한 요청, 승인 기능
- 승인된 분석서비스에 대해 사용자는 등록된 연구데이터와 메타데이터를 슈퍼컴퓨터에 전송하고, 슈퍼컴퓨터는 이를 기반으로 분석 과정을 수행하여 그 결과를 제공하는 기능

□ 서비스 프레임워크

○ 사용자/인증관리

- (회원관리) 국가연구데이터플랫폼은 클라우드 및 연구데이터 서비스에 대한 사용자 등록 및 인증을 위한 사용자 정보 관리 기능을 제공하고, 사용자 정보 관리를 위한 사용자 유형을 분류하여 관리함
- 국가연구데이터플랫폼 사용자를 위한 회원가입과 로그인 지원 기능
- 회원의 로그인 아이디와 비밀번호를 관리하는 기능
- 회원의 개인 정보를 관리하는 기능
- 국가연구데이터플랫폼의 사용자 그룹 및 커뮤니티 등을 관리하는 기능
- (인증/권한관리) 국가연구데이터플랫폼은 사용자 인증 메커니즘을 통하여 연계된 웹사이트의 정보 공유·활용 시 통합 ID 및 인증관리 기능을 제공하여야 함
- ※ 연구데이터플랫폼과 연계된 웹사이트들 간 인증, 개인정보, 본인확인 정보를 공유함으로써 하나의 ID를 이용하여 서비스 사용자들에 대해 Single Sign-On 기능
- 연계된 웹사이트에서 사용되는 계정 정보를 관리하는 기능
- 연구데이터플랫폼 내 사용자들의 웹사이트 별 또는 서비스 별 접근 권한에 대한 관리 기능
- 국가연구데이터플랫폼은 사용자에게 대한 권한 관리를 통한 연구데이터 접근제어 기능
 - 서비스 사용자에게 대한 권한레벨 등록 관리

- 서비스 사용자별 접근 가능 연구데이터 권한 할당 기능
- 서비스 운용자에 대해 접근 가능 커뮤니티 서비스에 대한 권한 할당 기능
- 센터 관리자에 대한 접근 가능 시스템들에 대한 권한 할당 기능
- 사용자 그룹별로 시스템 또는 서비스 사용 가능 프로토콜 권한 할당 기능

○ 데이터 등록 관리

- (DMP관리) 연구자 또는 연구기관이 연구 개발 프로젝트를 진행하며 연구데이터 생산 계획을 관리할 수 있도록 DMP(연구데이터 관리계획서) 작성 시 지원 및 활용 가능한 기능을 제공하여야 함
 - 연구자가 개별 연구관리시스템에 등록 시 활용할 수 있도록 DMP 작성용 템플릿 관리, DMP 등록 및 다운로드 기능
 - 등록된 DMP를 열람하거나 DMP의 수정, 변경이력, 공유이력 등 관리를 위한 기능
 - DMP 작성 시 참조할 수 있는 가이드라인 관리 기능
- (등록 및 퍼블리싱) 연구자 또는 연구기관이 보유한 연구데이터를 시스템의 관리 영역에 등록하는 서비스 기능을 제공하여야 함
 - 국가연구데이터플랫폼에 등록된 연구데이터에 대한 등록자 및 사용자들의 의견을 관리하는 기능
 - 연구데이터를 등록하거나 이미 등록된 연구데이터를 관리하는 기능
 - 등록된 연구데이터를 공유·활용할 수 있도록 출판하거나 관리하는 기능
 - 연구데이터를 등록할 경우 메타데이터 등록이 가능한 지원 기능
 - 연구데이터 등록 시 연구데이터의 유형에 따라 시스템 저장환경 설정 기능
 - 연구데이터 등록 시 등록된 연구데이터에 대한 공유·활용 데이터 정책을 설정할 수 있는 기능
 - 연구데이터 등록을 위한 영구 식별자(DOI, Digital Object Identifier)를 부여하는 기능
 - 등록하는 연구데이터와 관련된 메타데이터와 데이터 정책에도 식별자를 부여하는 기능

○ 데이터 큐레이션

- (데이터 저장관리) 연구자 또는 연구기관이 보유한 연구데이터를 시스템의 관리 영역에 저장하는 기능으로 연구데이터의 물리적 파일들을 국가연구데이터플랫폼에서 관리할 수 있도록 기능을 지원해야함
- 연구데이터플랫폼에 등록, 연계되어 있는 메타데이터에 대한 관리 (업데이트, 반출 등) 기능

- \
 - 과학기술표준분류와 같은 연구데이터의 관리를 위해 필요한 각종 표준 정보들을 수집하여 관리하는 기능
 - 대용량 연구데이터에 대한 저장과 갱신, 색인 등의 제반 관리 기능
 - 국내·외 유관기관에서 수집·연계된 연구데이터들을 통합 저장할 수 있는 표준 데이터 스키마들을 설계하고 실제로 연구데이터를 구축하고 관리하는 기능
 - 연구자가 생산, 보유하고 있는 각종 데이터(연구데이터 포함)를 저장하기 위한 클라우드 데이터 저장소를 관리하는 기능
 - 연구자의 (로컬)연구장비와 (원격)클라우드 간 연구데이터를 자동으로 동기화하는 기능
 - 연구데이터플랫폼과 연계되어있는 다양한 종류의 여러 리포지토리들 간 연구데이터 자동 동기화와, 연구데이터플랫폼과 리포지토리 간의 연구데이터 자동 동기화를 지원하는 기능
 - 연구데이터플랫폼에서 등록, 관리되고 있는 각종 데이터들에 대한 식별자 형식을 정의하고 신규 입수(등록) 데이터에 대해 식별자를 등록하고 관리하는 기능
- (데이터 품질 관리) 연구데이터 품질 관리는 국가연구데이터플랫폼에 등록된 연구데이터와 메타데이터의 중복, 무결성, 표준적용 등을 통해 데이터의 품질을 확보할 수 있도록 기능을 제공하여야 함
 - 국가연구데이터플랫폼에 등록되어 있는 연구데이터의 신뢰성과 유효성을 분석한 정보를 수치화하여 제공하는 기능
 - 국가연구데이터플랫폼에 데이터(연구데이터, 소프트웨어 등)들이 신규 등록, 업데이트 되는 경우 이에 대한 이력 관리 기능
 - 연구데이터에 포함된 레코드들 간 중복과 동일 리포지토리 또는 타 리포지토리에 같은 연구데이터가 존재하는지 여부를 검증하는 기능
 - 연구데이터가 조작되었거나 실효성이 없는 가짜 데이터인지의 여부를 검증하는 기능
 - 스키마의 각 항목별 제약조건에 기반을 두어 연구데이터의 각 항목 값들의 오류 존재 여부를 검증하는 기능
 - 국가연구데이터플랫폼에서 저장되어 관리·공유되고 있는 모든 데이터들에 대하여 시스템 장애 및 재해로부터 데이터 손실을 방지하기 위해 주기적으로 백업하고, 이벤트가 발생했을 때 복구하는 기능
 - 국가연구데이터플랫폼에서 저장되어 관
 - 리·공유되고 있는 데이터들 중 오래된 데이터로 사용성이 낮거나 대량 외부 공유가 필요한 경우 아카이빙하는 기능(재해 발생 시 복구 기능 포함)
- (데이터 출처/저작권 관리) 국가연구데이터플랫폼에 등록 및 출판된 연구데이터에 대한 접근권한을 관리하고, 저작권에 관련된 사항을 관리하는 기능을 제공하여야 함

- 등록·출판된 연구데이터에 대한 타 연구자들의 접근권한(사용 기간, 사용 범위, 사용 목적 등)을 등록하고 관리하는 기능
- 연구데이터의 저작권 관련 사항들을 등록 및 관리하는 기능
- (데이터 보안 관리) 국가연구데이터플랫폼에 등록된 연구데이터의 이력관리와 데이터 처리 시 요구되어지는 보안을 통해 데이터를 보호하기 위한 관리 기능을 제공하여야 함
- 연구데이터의 메타데이터 변경 시 블록체인 기술을 적용한 이력 관리 기능
- 연구데이터 전송, 수신, 저장 등의 상황에서 요구되는 제반 보안 기능
- 분석 및 활용
 - (연구데이터 검색) 연구데이터 검색 서비스는 사용자가 등록된 연구데이터를 편리하고 쉽게 조회할 수 있도록 하기 위해 메타데이터 기반으로 효율적인 검색 방안을 제공하여야 함
 - 등록된 연구데이터의 메타데이터에 대하여 키워드 검색 기능
 - 연구데이터의 원시데이터에 포함된 값 또는 문자열을 기준으로 검색 기능
 - 사용자가 입력한 키워드와 관련된 의미 기반의 검색 기능
 - 각종 필터를 사용한 검색 기능
 - 사용자의 이력에 기반을 둔 맞춤형 검색 기능
 - 연구자가 사용하고자 하는 소프트웨어를 검색할 수 있는 기능
 - 연구데이터 스토어에 등록된 연구데이터 검색 기능
 - 연구데이터를 사용하여 생산된 과학기술정보, 과제 등 참고정보 조회 기능
 - 국가연구데이터플랫폼에 등록된 커뮤니티 검색 기능
 - 연구자가 분석 작업에 사용할 수 있는 분석도구(소프트웨어, 프로그램 등) 검색 기능
 - (데이터 분석 관리) 국가연구데이터플랫폼에 등록된 연구데이터 또는 개별 기관에서 관리하는 연구데이터를 활용하여 분석업무를 지원하기 위한 환경을 제공하여야 함
 - 연구데이터 분석을 위해 데이터 포맷변환, 조건 값 추출 등 전처리 기능
 - 대용량 연구데이터 처리를 위해 슈퍼컴퓨팅 환경 지원
 - 분석을 위한 워크플로우 기반의 분석 환경 지원
 - 분석용 도구(소프트웨어, 프로그램 등)를 등록 및 관리할 수 있는 기능
 - 분석결과를 시각적으로 표현할 수 있는 기능
 - 사용자가 필요로 하는 분석도구를 추천하는 기능
 - 사용자가 직접 분석작업이 가능하도록 환경 지원
 - 데이터 분석을 위한 클라우드 제공 기능
 - 연구자에게 최적화된 분석모델을 추천해주는 기능
 - (연구 및 협업지원) 국가연구데이터플랫폼은 과학기술 커뮤니티 또는 기관들을 위한

\ 서비스를 제공할 수 있어야 함

- 연구자의 개인정보 및 관심분야 기반의 연구데이터 및 전문가를 추천하는 기능
- 연구자가 검색한 연구데이터 또는 과학기술정보와 관련된 다양한 데이터들을 추천하는 기능
- 연구데이터에 대한 다양한 통계정보를 활용하여 영향도를 분석하여 정보를 제공하는 기능
- 연구자들이 활용한 연구데이터의 통계정보를 제공하는 기능
- 국가연구데이터플랫폼의 활용 방법 등에 관한 다양한 콘텐츠 관리 기능
- 연구데이터 스토어를 관리할 수 있는 기능

□ 보안 프레임워크

- 컴퓨터가 사용하는 물리적인 자원에 대한 보안, 가상화를 이루는 운영체제와 게스트 운영체제, 즉 가상화 계층에 대한 보안과 관리 계층에 대한 보안, 그리고 서비스 계층에 대한 보안을 모두 포함하여 관리하여야 함
- 국가연구데이터플랫폼의 운영을 위해 사용되는 각종 서버에 대한 보안관리 기능
- 국가연구데이터플랫폼 운영 시 발생할 수 있는 각종 위험에 대한 대비·대응 관리 기능
- 국가연구데이터플랫폼 운영 시 사용되는 네트워크 장비 관련 보안 관리 기능

□ 시스템 관리 프레임워크

○ 호스팅 관리

- 국가연구데이터플랫폼의 클라우드 컴퓨팅 서비스 제공 모델(Service delivery Model)은 IaaS, PaaS 및 SaaS의 형태로 나뉘어 질 수 있으며, 각각의 형태로의 서비스 제공이 가능하여야 함
- 연구자에게 저장공간, 분석환경 등 시스템 자원을 동적으로 할당하는 기능
- 연구자에게 가상 분석환경을 제공하기 위한 관리 기능
- 연구자의 클라우드서비스(IaaS, SaaS) 요청사항을 관리하기 위한 기능
- 연구자가 분석환경을 제공받기 위한 클라우드서비스(PaaS) 요청사항을 관리하는 기능

○ 사용환경 관리

- 국가연구데이터플랫폼은 외부 유관기관 및 사용자에게 공유·활용에 필요한 데이터를 연계할 수 있도록 지원하고, 사용자가 디바이스나 브라우저에 독립적인 환경에서 이용할 수 있도록 지원해야함
- 연구데이터를 외부 유관기관, 연구자 등에게 연계할 수 있는 OpenAPI 관리 기능

- 국가연구데이터플랫폼이 디바이스와 브라우저에 독립적으로 동작할 수 있도록 사용환경 지원 기능
- 백업/복구 관리
 - 국가연구데이터플랫폼은 등록된 연구데이터와 메타데이터들의 상태에 대해 이상 현상 또는 오류가 발생되었을 경우, 이를 원래의 상태로 복구할 수 있는 방법들을 제공하여야 함. 이를 위해서는 현재 또는 과거의 어느 시점에서 국가연구데이터플랫폼에 유지되는 데이터들에 대한 주기적인 백업 및 복구 방안을 제공하여야 함
 - 연구데이터플랫폼에서 저장되어 관리·공유되고 있는 모든 데이터들에 대하여 주기적으로 백업하고, 이벤트가 발생했을 때 복구하는 기능
 - 연구데이터플랫폼에서 저장되어 관리·공유되고 있는 데이터들 중 재해 발생 시 복구 기능, 오래된 데이터로 사용성이 낮거나 대량 외부 공유가 필요한 경우 아카이빙하는 기능
- 통계관리 및 모니터링
 - 국가연구데이터플랫폼은 효율적이고 안정적인 운영·관리를 위해 시스템 관리자에게 다양한 통계정보 및 모니터링 정보를 제공하여야 함
 - 국가연구데이터플랫폼에 등록된 데이터의 통계 및 모니터링 정보 제공 기능
 - 국가연구데이터플랫폼에서 제공중인 서비스의 통계 및 모니터링 정보 제공 기능
 - 국가연구데이터플랫폼의 기능개선 및 서비스 추가 등을 파악할 수 있는 분석자료 및 모니터링 정보 시각화 기능
 - 국가연구데이터플랫폼에서 사용되는 어플리케이션의 이력을 관리하는 기능

2) 기능 상세

<표 5-3> 국가연구데이터플랫폼 세부 기능별 정의

기능구분				설명
연계 프레임워크	연구데이터 및 메타데이터 연계	분야별 전문센터 연계		국내 분야 별 전문센터에서 공유·활용 중인 연구데이터들을 연계하는 기능
		개별기관 연계		자체 연구관리시스템을 보유하고 있는 국내 정부 출연 연구기관들에서 공유·활용 중인 연구데이터들을 연계하는 기능
		해외 연구데이터플랫폼 연계		해외 주요 연구데이터플랫폼(OpenAIRE, ANDS 등)에서 공유·활용 중인 연구데이터들을 연계하는 기능
		Open Data		주요 Open Data 사이트로부터 Data를 수집하는 기능
	과제 정보 및 과학기술정보 연계	NTIS 연계		연구데이터플랫폼에서 공유·활용 중인 연구데이터와 과제정보 및 과학기술정보(논문, 특허, 보고서 등)를 수집하기 위해 NTIS와 연계하는 기능
		성과마루 연계		연구데이터플랫폼에서 공유·활용 중인 연구데이터와 과제정보 및 과학기술정보(논문, 특허, 보고서 등), DMP 정보를 수집하기 위해 성과마루와 연계하는 기능
	분석환경 연계			연구데이터 분석 기능을 제공하는 고성능 컴퓨터 및 외부 클라우드와 연계하는 기능
서비스 프레임워크	사용자/인증 관리	회원관리	회원가입/ 로그인	연구데이터플랫폼을 사용하고자 하는 연구자들의 회원가입과 로그인을 지원하는 기능
			아이디/ 비밀번호 관리	회원들의 로그인 아이디와 비밀번호를 관리하는 기능
			회원 정보 관리	회원들의 개인 정보(주소, 전화번호, 관심 분야, 경력 등)를 관리하는 기능
			연구자 및 기관/커뮤니티 관리	연구데이터플랫폼에 사용자(연구자, 시스템 또는 서비스의 관리자 등), 사용자 그룹(연구자 커뮤니티 포함), 프로젝트 등의 등록을 위해 필요한 기능

기능구분				설명
	인증/권한 관리	인증/권한 관리	통합 ID 인증, ID 연계 인증	연구데이터플랫폼과 연계된 웹사이트들 간 인증, 개인정보, 본인확인 정보를 공유 함으로써 하나의 ID를 이용하여 모든 연 계 웹사이트의 로그인 및 서비스를 이용 하는 기능(특히, 연구데이터플랫폼)
			통합 계정 관리	연계된 웹사이트에서 사용되는 계정 정보 를 관리하는 기능
			사용자 권한 관리	연구데이터플랫폼 내 사용자들의 웹사이 트 별 또는 서비스 별 접근 권한에 대한 관리 기능
	데이터 등록관리	DMP관리	DMP 작성 및 등록	수행하는 연구에서 활용·생산할 데이터에 대한 관리계획(DMP)을 작성(DMP 템플 릿* 활용)하여 등록하는 기능과 DMP 작 성의 편의를 위한 DMP 템플릿을 작성하 여 등록하는 기능 *DMP템플릿:DMP작성시 자주사용되는항목 을선정하고값들을미리작성해서저장해놓음 으로써DMP작성편의성향상
			DMP 열람 및 관리	등록된 DMP를 열람하거나 다운로드하는 기능과 DMP 관리를 위한 제반 기능(수 정, 삭제, 변경 이력 관리, 공유 이력 관 리, 공개 여부 관리, DMP 템플릿 관리 등)
			DMP 작성 가이드	DMP 작성 가이드라인을 제공하고 연구자 가 DMP 작성 중 쉽게 가이드라인을 참고 할 수 있도록 항목별·단계별로 화면에 보 여주는 기능
		등록 및 퍼블리싱	WSC(Wiki Style Commentary)	연구데이터플랫폼에 등록된 데이터(연구 데이터, 소프트웨어 등)에 대한 등록자 또 는 사용자들이 작성한 의견에 대하여 Wiki와 같이 협업 편집을 가능하게 하는 기능
			연구데이터 등록·관리	연구자들이 생산하거나 보유하고 있는 연 구데이터들을 연구데이터플랫폼에 등록하 거나 이미 등록된 연구데이터를 관리하는 기능
			연구데이터 출판·관리	등록되어 있는 연구데이터를 타 연구자들 과 공유·활용하기 위해 출판(DOI 발급/부 여)하거나 이미 출판된 연구데이터를 관리 하는 기능
			메타데이터 자동 추출	연구데이터를 등록할 때 작성해야 하는 메타데이터 항목 값들을 자동으로 추출하 여 연구자 편의를 증대하는 기능

기능구분				설명
데이터 큐레이션	데이터 저장관리	메타데이터 관리		연구데이터플랫폼에 등록, 연계되어 있는 메타데이터에 대한 관리(업데이트, 반출 등) 기능
		연구데이터 표준 정보 관리		과학기술표준분류와 같은 연구데이터의 관리를 위해 필요한 각종 표준 정보들을 수집하여 관리하는 기능
		대용량 데이터 관리(색인 효율화 포함)		대용량 연구데이터에 대한 저장과 갱신, 색인 등의 제반 관리 기능
		연구데이터 통합 DB 스키마 변경 설계 구축·관리		국내·외 유관기관에서 수집·연계된 연구데이터들을 통합 저장할 수 있는 표준 데이터 스키마들을 설계하고 실제로 연구데이터를 구축하고 관리하는 기능
		클라우드 데이터 저장소 관리		연구자가 생산, 보유하고 있는 각종 데이터(연구데이터 포함)를 저장하기 위한 클라우드 데이터 저장소를 관리하는 기능
		개인 연구데이터 보관·관리 위한 클라우드		연구자가 생산하거나 보유하고 있는 연구데이터를 보관·관리하기 위한 클라우드 저장 공간을 제공하는 기능
		퍼스널 클라우드 기반의 연구데이터 동기화		연구자의 (로컬)연구장비와 (원격)클라우드 간 연구데이터를 자동으로 동기화하는 기능
		복수 리포지토리 자동 동기화 관리		연구데이터플랫폼과 연계되어있는 다양한 종류의 여러 리포지토리들 간 연구데이터 자동 동기화와, 연구데이터플랫폼과 리포지토리 간의 연구데이터 자동 동기화를 지원하는 기능
		식별자 등록·관리(연구자, 기관, 연구데이터, 소프트웨어 등)		연구데이터플랫폼에서 등록, 관리되고 있는 각종 데이터들에 대한 식별자 형식을 정의하고 신규 입수(등록) 데이터에 대해 식별자를 등록하고 관리하는 기능
	데이터 품질관리	연구데이터 신뢰성, 유효성 분석		연구데이터플랫폼에 등록되어 있는 연구데이터의 공유·활용을 촉진하여 연구 성과 제고를 위해 연구데이터의 신뢰성과 유효성은 분석한 정보를 수치화하여 연구자들에게 제공하는 기능
		연구데이터 및 소프트웨어 버전 관리		연구데이터플랫폼에 데이터(연구데이터, 소프트웨어 등)들이 신규 등록, 업데이트 되는 경우 이에 대한 이력 관리를 지원하는 기능
		연구데이터 중복		연구데이터에 포함된 레코드들 간 중복과

기능구분				설명
			검증	동일 리포지토리 또는 타 리포지토리에 같은 연구데이터가 존재하는지 여부를 검증하는 기능
			Fake 데이터 검증	연구데이터가 조작되었거나 실효성이 없는 가짜 데이터인지의 여부를 검증하는 기능
			스키마 기반 품질 관리	스키마의 각 항목별 제약조건에 기반을 두어 연구데이터의 각 항목 값들의 오류 존재 여부를 검증하는 기능
		데이터 출처/저작권 관리	연구데이터 접근권한 등록·관리	등록·출판된 연구데이터에 대한 타 연구자들의 접근권한(사용 기간, 사용 범위, 사용 목적 등)을 등록하고 관리하는 기능
			라이선스 관리	연구데이터의 저작권 관련 사항들을 등록 및 관리하는 기능
		데이터 보안관리	블록체인 기반 데이터 이력 관리	연구데이터의 메타데이터 변경 시 블록체인 기술을 적용한 이력 관리 기능
			데이터 보안 관리	데이터 전송, 수신, 저장 등의 상황에서 요구되는 제반 보안 기능
	분석 및 활용	데이터 검색	키워드 검색	연구데이터의 메타데이터에 포함되어 있는 항목 값들을 색인하고 연구자가 입력한 키워드에 기반하여 검색하는 기능
			데이터 내용 검색	연구데이터의 원시데이터에 포함되어 있는 값 또는 문자열을 검색하는 기능
			의미 기반 검색	사용자가 입력한 키워드와 관련된 시맨틱 네트워크 상의 정보를 이용하여 언어의 의미적 정보를 사용하여 검색함으로써 검색 결과의 질을 향상시키는 기능
			분야별 상세 검색	분야, 년도, 데이터 종류, 데이터 생산자 등의 각종 필터를 사용하여 자세한 정보를 검색할 수 있는 기능
			패턴 기반 지능형 검색	연구자의 다양한 이력(검색, 데이터 열람, 분석, 커뮤니티 활동 등)에 기반을 두어 연구자 맞춤형 검색 결과를 제시하는 기능
			소프트웨어 검색	연구자가 사용하고자 하는 소프트웨어를 검색할 때 필터 등을 적용하여 검색 편의를 제공하는 기능

기능구분				설명
			연구데이터 스토어 검색	연구데이터플랫폼에서 운영되는 연구데이터 스토어에 등록된 데이터들을 검색하는 기능
			연구데이터 인용·출처 브라우징	연구데이터를 사용해서 생산된 과학기술 정보(논문, 특허, 보고서 등)나 수행된 과제, 연구데이터 사용 시 참고할 수 있는 자료들에 대한 브라우징 기능
			분석도구 스토어 검색	연구자가 분석 작업을 수행할 때 사용할 수 있는 소프트웨어 또는 프로그램 등을 검색할 수 있는 기능
	데이터 분석관리		데이터 내용 시각화	연구데이터에 포함된 실제 데이터들을 원본, 차트, 그래프 등 다양한 시각화 방법을 통해 보여줌으로써 연구자들의 연구데이터에 대한 이해를 높이는 기술
			데이터 전처리	연구데이터 분석을 위해 데이터 포맷 변환, 특정 컬럼 삭제, 조건에 맞는 값 추출 등 다양한 전처리 기능들을 제공
			데이터 처리용 클라우드	연구데이터 처리를 위해 사용되는 클라우드를 제공하는 기능
			거대 계산용 슈퍼컴퓨터 활용	(대용량) 연구데이터를 처리하거나 계산량이 많아 계산 시간이 오래 걸리는 데이터를 처리하기 위해 한국과학기술정보연구원에서 보유하고 있는 슈퍼컴퓨터 활용 환경을 지원하는 기능
			워크플로우 기반 분석	분석용 툴을 위젯 또는 아이콘 형태로 화면에 나타내어 분석 워크플로우를 작성(예: 아이콘 배열)함으로써 편의성을 제공하는 기능
			분석 도구 등록	분석을 위한 소프트웨어, 프로그램 등을 연구데이터플랫폼에 등록하는 기능
			분석 결과 가시화 도구 제공	분석의 결과를 연구자가 직관적으로 이해할 수 있도록 가시화 도구를 제공하는 기능
			데이터 맞춤형 분석도구 추천	연구자의 분야, 사용하는 연구데이터, 기존의 분석 패턴 등을 고려하여 연구자가 요구하는 분석도구를 자동으로 추천하는 기능
			CLI 기반 분석 환경 지원	연구자들이 연구데이터플랫폼에서 제공하는 분석환경을 사용하여 직접 분석 작업을 하는 경우 CLI(Command Line Interface)를 활용할 수 있도록 환경을 제공하는 기능

기능구분				설명
			IDE 기반 분석 환경 지원	연구자들이 연구데이터플랫폼에서 제공하는 분석환경을 사용하여 직접 분석 작업을 하는 경우 IDE(Integrated Development Environment)를 활용할 수 있도록 환경을 제공하는 기능
			데이터 분석용 클라우드	연구데이터 분석을 위해 사용되는 클라우드를 제공하는 기능
			인공지능 기반 분석모델 추천	연구자들이 분석을 위해 사용할 모델을 자동으로 추천해주는 기능
		협업 및 커뮤니티 지원	관심분야 연구데이터 추천	연구자의 개인정보에 입력된 관심분야에 해당하는 연구데이터를 추천하는 기능
			전문가 추천	연구자의 개인정보에 입력된 관심분야에 해당하는 전문가를 추천하는 기능
			검색 결과 연관 데이터 추천	연구자가 검색한 연구데이터 또는 과학기술정보(논문, 특허, 보고서 등)과 관련된 다양한 데이터들을 추천하는 기능
			연구데이터 영향도 분석	연구데이터의 조회 수, 다운로드 수, 인용 수 등의 다양한 통계정보를 활용하여 영향도를 분석하여 수치 정보를 보여주는 기능
			연구데이터 활용도 통계 정보 제공	연구데이터를 사용한 연구자들이 입력한 코멘트, 유용성 지수 등 다양한 통계정보를 활용하여 활용도를 분석하여 수치 정보를 보여주는 기능
			온라인 교육 Contents(Webin ar 포함) 관리	연구자의 연구데이터플랫폼 활용 방법 습득 및 연구 역량 강화를 위한 다양한 콘텐츠 등록·관리 기능
			연구데이터 스토어 관리	연구데이터플랫폼에서 제공되는 연구데이터 스토어에 등록된 연구데이터, 소프트웨어, 분석 도구 등의 다양한 데이터의 관리 및 판매/구매 활동을 지원하기 위한 기능
			커뮤니티 개설관리	사용자가 연구데이터플랫폼에 커뮤니티 개설을 요청하고, 요청에 대한 승인/거절, 커뮤니티에 대한 클라우드서비스 내용을 관리하는 기능
			커뮤니티 검색	연구데이터플랫폼에 등록되어 있는 커뮤니티들을 열람할 수 있으며 연구자에게 맞는 커뮤니티들을 다양한 조건을 부여하여 검색할 수 있는 기능

기능구분			설명
보안 프레임워크	연구데이터플랫폼 시스템 보안 관리		연구데이터플랫폼의 운영을 위해 사용되는 웹서버, 데이터베이스 서버, 색인 서버 등의 보안을 위한 제반 기능
	발생 위험 관리		연구데이터플랫폼 운영 시 발생할 수 있는 각종 위험 요소들에 대한 항목들을 정의하고 대응·대비하기 위해 체계적으로 관리하는 기능
	네트워크 보안 관리		연구데이터플랫폼 운영을 위해 사용되는 네트워크 장비 및 관련 보안 이슈들에 대해 대응·대비하기 위한 제반 기능
시스템관리 프레임워크	호스팅 관리	동적 자원 할당	연구자들에 제공되는 저장공간, 분석환경 등을 연구자의 패턴과 시스템 자원 운용 상황을 고려하여 동적으로 할당하는 기능
		VM Workspace 관리	연구자들에게 제공되는 가상 분석환경의 설정 전반을 관리하는 기능
		개인연구자 클라우드서비스(IaaS, SaaS)	연구자들에게 제공되는 저장공간의 환경 설정, 요청 사항들을 관리하는 기능
		분석용 클라우드 PaaS 서비스	연구자들에게 제공되는 분석환경의 환경 설정, 요청 사항들을 관리하는 기능
	사용환경 관리	Open API 관리	연구데이터플랫폼의 연구데이터를 외부 유관기관, 개인 연구자 등에게 연계하기 위해 필요한 OpenAPI를 등록, 관리하는 기능
		이종 디바이스(PC·모바일) 지원	연구데이터플랫폼이 디바이스와 브라우저에 독립적으로 동작할 수 있도록 지원하는 기능
	데이터 백업/복구 관리	연구데이터 백업, 복구	연구데이터플랫폼에서 저장되어 관리·공유되고 있는 모든 데이터들에 대하여 시스템 장애 및 재해로부터 데이터 손실을 방지하기 위해 주기적으로 백업하고, 이벤트가 발생했을 때 복구하는 기능
		연구데이터 재해복구 및 아카이빙	연구데이터플랫폼에서 저장되어 관리·공유되고 있는 데이터들 중 오래된 데이터로 사용성이 낮거나 대량 외부 공유가 필요한 경우 아카이빙하는 기능(재해 발생 시 복구 기능 포함)

기능구분			설명
	통계관리 및 모니터링	연구데이터 현황 모니터링	연구데이터플랫폼의 데이터 관련 제반 관리 기능(데이터 등록, 갱신, 삭제 등에 대한 통계 정보와 모니터링)
		서비스 컴포넌트 모니터링	연구데이터플랫폼의 서비스 관련 제반 관리 기능(제공되고 있는 주요 서비스들의 사용도, 오류 등에 대한 통계 정보와 모니터링)
		플랫폼 서비스 분석 및 모니터링 시각화	연구데이터플랫폼의 데이터와 서비스의 각종 통계를 활용하여 향후 플랫폼의 방향설정, 추가 개발 기능 등을 수행할 때 활용할 분석 내용 제공과 모니터링 기능
		플랫폼 서비스 어플리케이션 버전 관리	연구데이터플랫폼의 서비스들을 위해 사용되는 어플리케이션의 신규 등록, 갱신, 삭제 등의 이력을 관리하는 기능

3) 요소기술

- 국가연구데이터플랫폼 구축을 위해 기술개발 시 필요한 요소기술을 기능에 따라 구성하고, 이를 기반으로 세부과제를 선정함

<표 5-4> 국가연구데이터플랫폼 구축을 위한 요소기술 목록

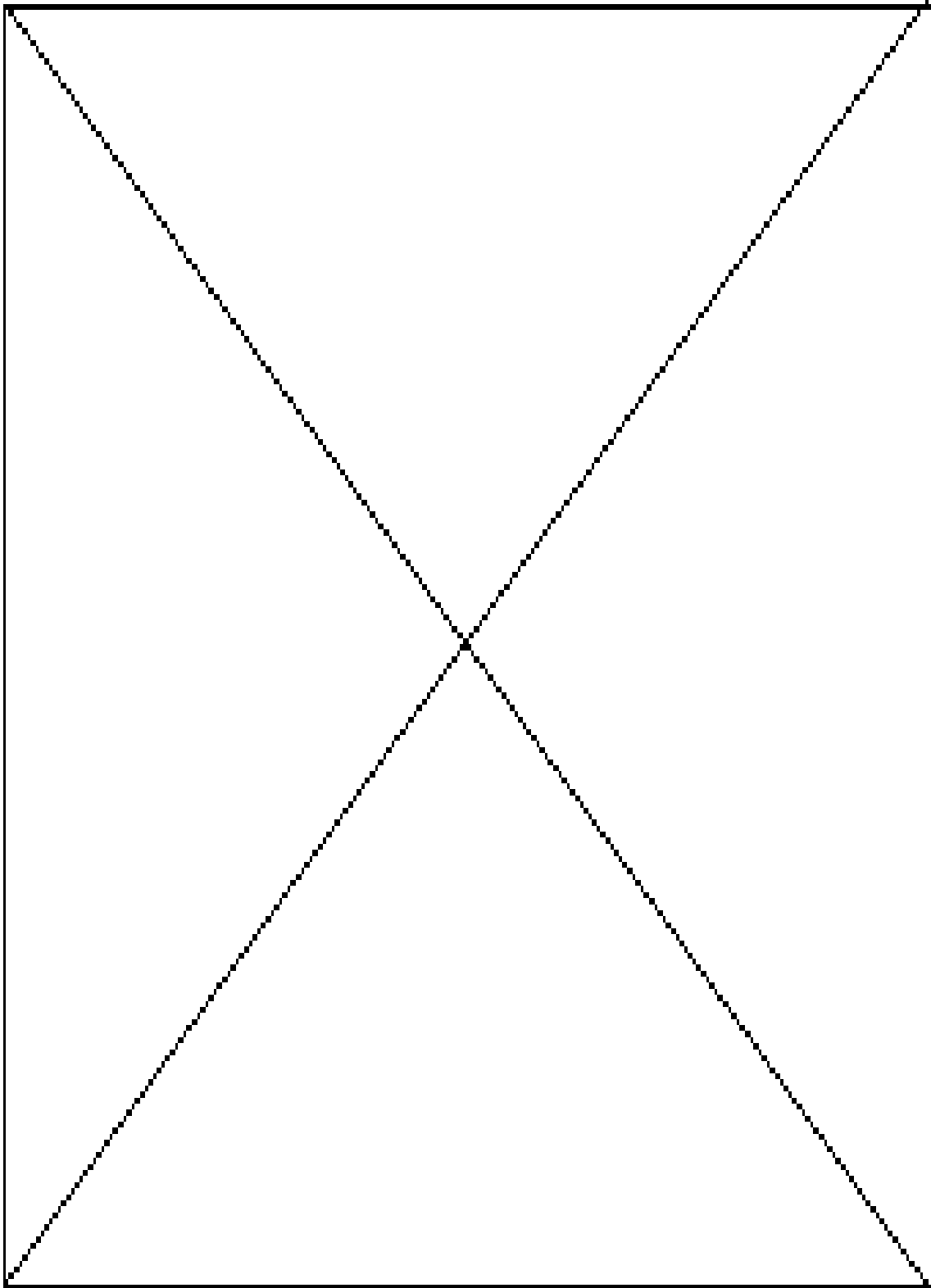
기능구분				요소기술	
연계 프레임워크	연구데이터 및 메타데이터 연계	분야별 전문센터 연계		• 데이터 수집 에이전트 개발 및 관리 기술 • 연구데이터 연계API 기술 • 연구데이터 개체식별 기술 • 대용량데이터 고속 송수신 기술 • 복수리포지토리 자동동기화 관리기술 • 데이터전송보안 기술 • 표준메타데이터정의/관리자동화기술 • 이종데이터스키마 매핑관리 기술	
		개별기관 연계			
		해외 연구데이터플랫폼 연계			
		Open Data			
	과제 정보 및 과학기술정보 연계	NTIS 연계			
		성과마루 연계			
	분석환경 연계				
서비스 프레임워크	사용자/인증관리	회원관리	회원가입/로그인	• 연구자, 기관, 커뮤니티 등록·관리 기술 • 개인정보보호비식별화·익명화기술	
			아이디/비밀번호 관리		
			회원 정보 관리		
			연구자 및 기관/커뮤니티 관리		
	인증/권한 관리	인증/권한 관리	통합 ID 인증, ID 연계 인증	• ID Federation/Open ID 인증 관리 기술 • 사용자접근제어기술	
			통합 계정 관리		
			사용자 권한 관리		
	데이터 등록관리	DMP관리	DMP 작성 및 등록	• DMP 등록 및 저장 기술 • DMP관리(이력,공유,공개등)기술 • maDMP기술(MachineActionableDMP)	
			DMP 열람 및 관리		
			DMP 작성 가이드		
		등록 및 퍼블리싱	등록 및 퍼블리싱	WSC(Wik Style Commentary)	• online rich-text editor 기술 • 메타데이터자동생성 및 추출기술 • DOI등록 및 관리기술 • 대용량데이터(데이터셋)저장 및 등록기술 • 클라우드기반데이터저장 및 조작기술
				연구데이터 등록·관리	
				연구데이터 출판·관리	
				메타데이터 자동 추출	
	데이터 큐레이션	데이터 저장관리	메타데이터 관리	• 메타데이터 관리 기술 • 대용량 데이터 분산 저장 기술 • 클라우드 기반 데이터 저장 및 조작 기술	
			연구데이터 표준 정보 관리		

기능구분			요소기술
		대용량 데이터 관리(색인 효율화 포함)	• 대용량 데이터 색인 관리 기술 • 다수의 이중 클라우드서비스 연계 및 중개서비스 기술 • 연구데이터 가상드라이브 저장 기술 • 인메모리 데이터베이스 기술 • 로컬·가상 드라이브 동기화 기술 • 복수 리포지토리 동기화 기술 • 식별자 등록 및 관리 기술
		연구데이터 통합 DB 스키마 변경 설계 구축·관리	
		클라우드 데이터 저장소 관리	
		개인 연구데이터 보관·관리 위한 클라우드	
		퍼스널 클라우드 기반의 연구데이터 동기화	
		복수 리포지토리 자동 동기화 관리	
		식별자 등록·관리(연구자, 기관, 연구데이터, 소프트웨어 등)	
	데이터 품질관리	연구데이터 신뢰성, 유효성 분석	• 중복데이터 판정 및 제거 기술 • 데이터 무결성 검증 기술 • 메타데이터 결측치 및 이상치 자동 보정 기술 • 기관 별 스키마·이종 메타데이터 변경이력 관리 및 동기화 기술 • 연구데이터셋 버전 관리 기술 • 연구 중간산출물 리니지 관리 기술 • 연구용 SW 버전 관리 기술 • Fake 데이터 판별 기술 • 연구분야별 DB 표준스키마 구축 및 관리 기술 • 데이터 표준화 적용 및 저장 기술
		연구데이터 및 소프트웨어 버전 관리	
		연구데이터 중복 검증	
		Fake 데이터 검증	
		스키마 기반 품질 관리	
	데이터 출처/저작권 관리	연구데이터 접근권한 등록·관리	• 연구데이터 인용·출처 등록·관리 기술 • 데이터 라이선스 관리 기술
		라이선스 관리	
	데이터 보안관리	블록체인 기반 데이터 이력 관리	• 블록체인 기반의 데이터 이력 관리 기술 • 데이터 전송보안·접근제어기술 • 데이터 암호·복호화기술 • 개인정보보호 위한 비식별화·익명화 기술 • 데이터베이스 접근제어 기술
		데이터 보안 관리	
	분석 및 활용	데이터 검색	• 키워드 고속 검색 기술 • LSM기반 고속분산색인 기술 • 다국어검색 기술
		키워드 검색	
		데이터 내용 검색	

기능구분			요소기술
			의미 기반 검색
			분야별 상세 검색
			패턴 기반 지능형 검색
			소프트웨어 검색
			연구데이터 스토어 검색
			연구데이터 인용·출처 브라우징
			분석도구 스토어 검색
		데이터 분석관리	데이터 내용 시각화
			데이터 전처리
			데이터 처리용 클라우드
			거대 계산용 슈퍼컴퓨터 활용
			워크플로우 기반 분석
			분석 도구 등록
			분석 결과 가시화 도구 제공
			데이터 맞춤형 분석도구 추천
			CLI 기반 분석 환경 지원
			IDE 기반 분석 환경 지원
			데이터 분석용 클라우드
			인공지능 기반 분석모델 추천
		협업 및 커뮤니티 지원	관심분야 연구데이터 추천
			전문가 추천
			검색 결과 연관 데이터 추천
			연구데이터 영향도 분석
			연구데이터 활용도 통계 정보 제공
			온라인 교육 Contents(Webinar 포함) 관리
			연구데이터 스토어 관리
			의미기반검색 기술
			대용량 데이터 분산검색 기술
			데이터 내용 검색 기술
			대용량 데이터 실시간 색인기술
			사용 패턴 기반 검색 기술
			데이터 내용 시각화 기술
			연구데이터 전처리 기술
			Automated AI 모델 추천 엔진 기술
			데이터 분석 모델 등록·관리 기술
			워크플로우 모델링 기반의 연구데이터분석 기술
			데이터 분석 클라우드 지원 기술
			대용량 데이터 분산·병렬 분석 기술
			원시데이터 포맷 전처리 및 변환 기술
			분석도구 버전 관리 기술
			분석자원 동적 할당 기술
			사용자 패턴 기반 VM 분석환경 최적화 기술
			분석 리니지 관리 기술
			분석결과 가시화 기술
			모델링 및 시뮬레이션 기술
			연구데이터 영향도 평가 기술
			워드클라우드 생성 기술
			데이터 간(보고서, 과제, SW 등) 연관성추론 기술
			지능형 연구데이터 자동 추천 기술
			온라인 교육 Contents(Webinar 포함) 관리 기술
			연구자 맞춤형 큐레이션 기술
			연구자 추천 기술
			연구데이터 활용도 평가 기술
			연구데이터 프로파일링 정보(데이터 분포, 샘플 등) 생성 기술
			사용자·커뮤니티 간 소통 환경

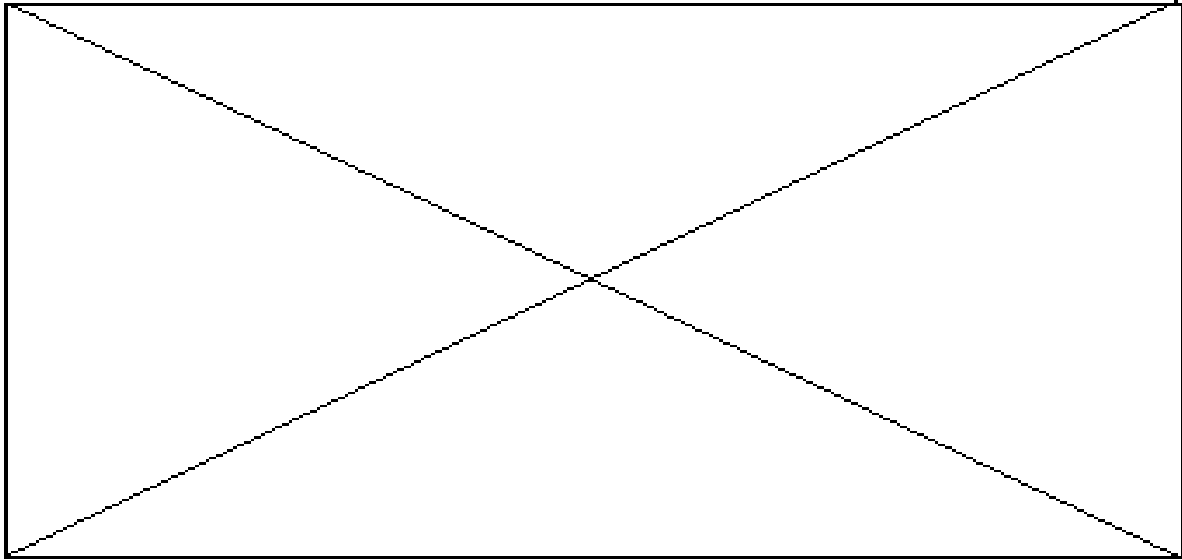
기능구분				요소기술
			커뮤니티 개설관리	기술
			커뮤니티 검색	
보안 프레임워크	연구데이터플랫폼 시스템 보안 관리			• 통합 보안관리 기술 • 종합 위험관리 기술 • 유해 트래픽탐지 기술 • 패킷분석 기술 • 침입차단 및 방지 기술
	발생 위험 관리			
	네트워크 보안 관리			
시스템관리 프레임워크	호스팅 관리	동적 자원 할당		• 클라우드 기반 데이터 저장 및 조작 기술 • 이종 대용량 파일시스템 통합 관리 기술
		VM Workspace 관리		
		개인연구자 클라우드서비스(IaaS, SaaS)		
		분석용 클라우드 PaaS 서비스		
	사용환경 관리	Open API 관리		• 연구데이터 연계 API 생성 및 관리 기술
		이기종 디바이스(PC·모바일) 지원		• 디바이스 독립적인 사용 환경 지원 기술
	데이터 백업/복구 관리	연구데이터 백업, 복구		• 연구데이터 백업 및 복구 기술 • 데이터 보존 및 아카이빙 기술
		연구데이터 재해복구 및 아카이빙		
	통계관리 및 모니터링	연구데이터 관리컴포넌트 모니터링		• 연구데이터 관리 모듈 모니터링 기술 • 분석 및 모니터링 시각화 기술 • 위젯 추가방식의 실시간 로그 분석 기술 • 연구데이터 현황 모니터링 기술 • 연구데이터 활용 현황 분석 기술 • 시스템, 리소스 사용 현황 모니터링 기술 • 사용자별 대시보드 지원 기술
		서비스 포트폴리오 컴포넌트 모니터링		
		플랫폼 서비스 분석 및 모니터링 시각화		
		플랫폼 서비스 어플리케이션 버전 관리		

4) 기술트리



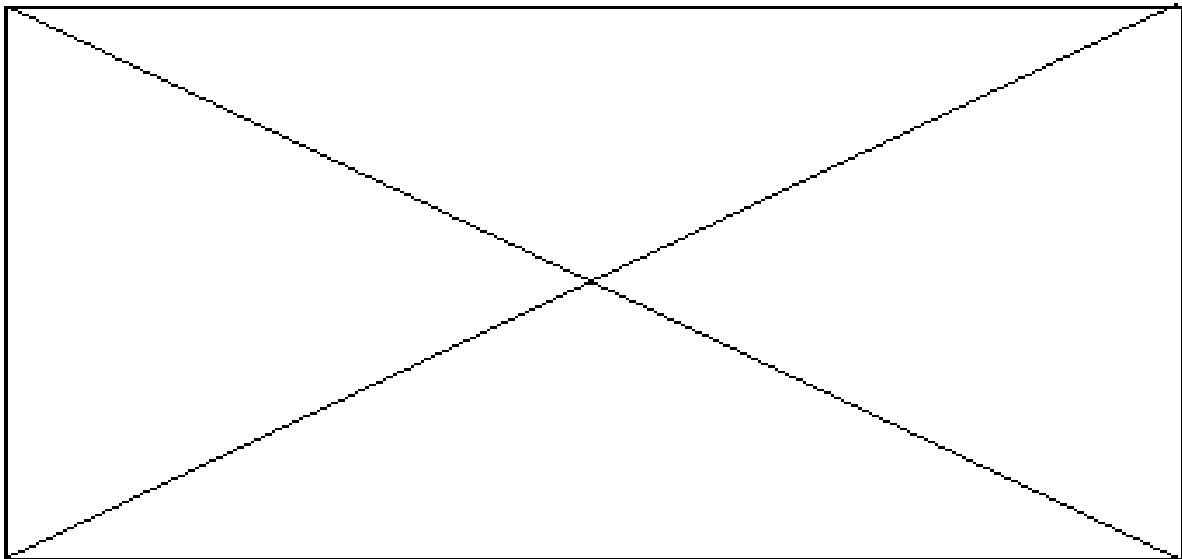
5) 국가연구데이터플랫폼 아키텍처 구성

- 국가연구데이터플랫폼 구축을 위한 아키텍처는 인프라, 하드웨어(H/W), 소프트웨어(S/W)로 구분하여 구성함
 - 국가연구데이터플랫폼 인프라 구성 방안



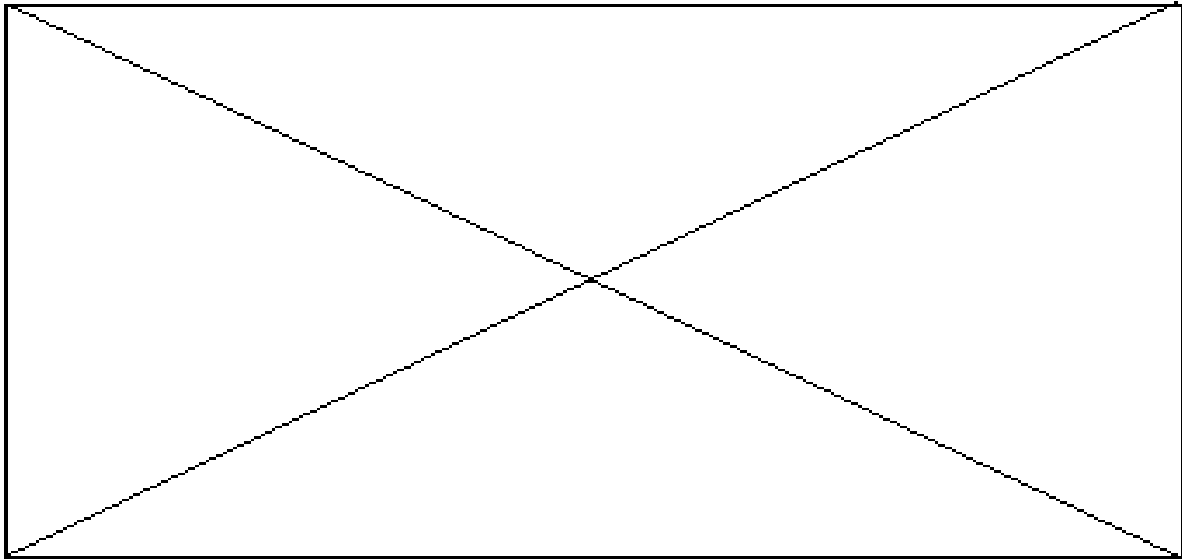
<그림 5-4> 국가연구데이터플랫폼 인프라 구성도

- 국가연구데이터플랫폼 H/W 구성 방안



<그림 5-5> 국가연구데이터플랫폼 H/W 구성도

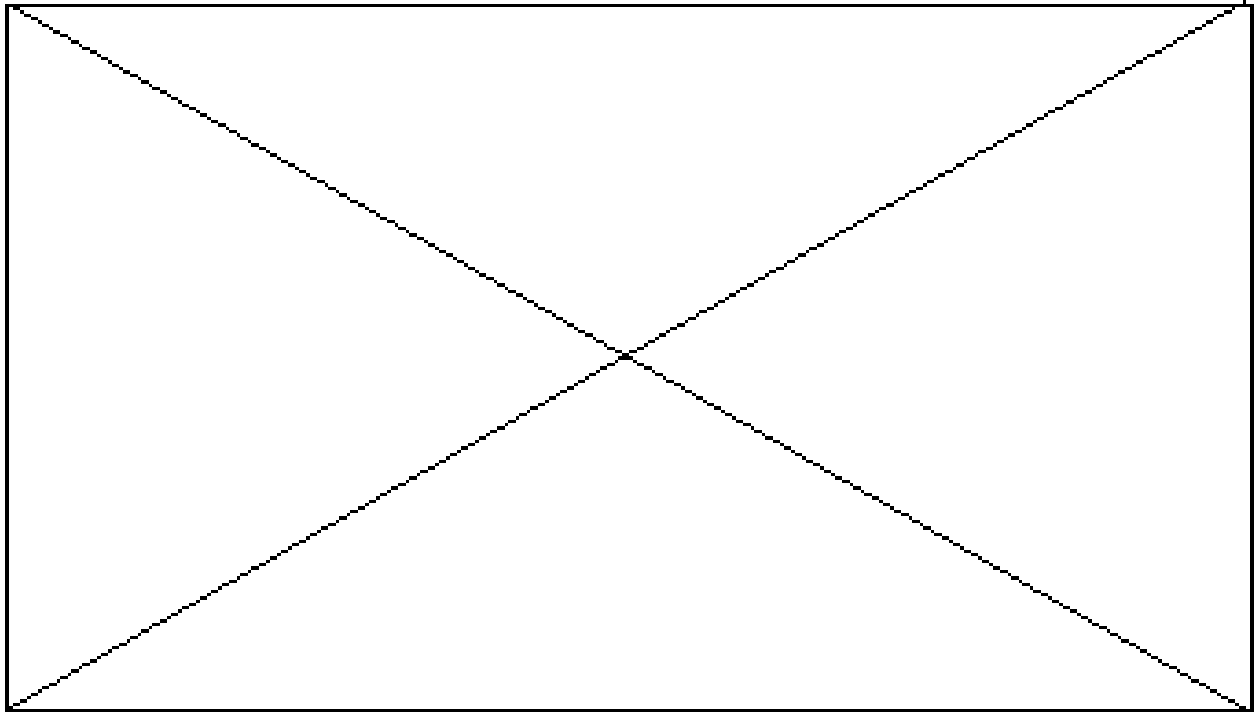
\ ○ 국가연구데이터플랫폼 S/W 구성 방안



<그림 5-6> 국가연구데이터플랫폼 S/W 구성도

다. 세부과제 구성

- 국가연구데이터플랫폼 구축에 필요한 요소기술 및 세부기술을 기반으로 전략 목표를 달성하기 위한 6개의 세부과제를 선정함



<그림 5-7> 국가연구데이터플랫폼 구축 및 운영을 위한 세부 연구개발 과제

5.3. 세부과제별 추진내용

5.3.1. 연구데이터 수집·연계 및 공유 기술 개발

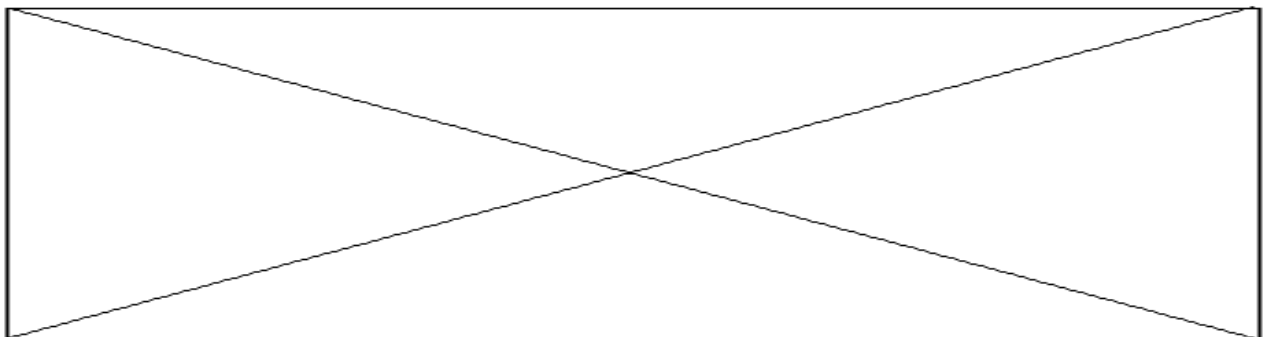
가. 개발 및 추진 목표

개발 목표	국내외 연구데이터/메타데이터 통합, 공유를 위한 수집·전송 시스템 개발
세부 추진 목표	① 연구데이터/메타데이터 수집 및 연계 기술 개발 ② 대용량 연구데이터를 분석환경에 전송하기 위한 연계 기술 개발

- 데이터 수집·연계 관리, 연구과제 정보 및 DMP 연계, 분석환경 연계의 기능을 개발하고 연계 데이터의 모니터링을 위한 연계관리 기능 개발

□ 세세부 과제 도출

- 대소분야 센터, 국내외 연구기관의 연구데이터/메타데이터를 수집/전송을 위한 기술개발의 세세부과제를 메타데이터 연계기술개발, 연구데이터 연계기술 개발로 도출



<그림 5-8> 연구데이터 수집·연계 및 공유 기술 개발 세세부 과제 도출

□ 기술 확보전략

기술 구분	현재 기술수준	목표 기술수준	세세부 과제	세부기술 확보 전략		
				자체개발	기술도입	글로벌 협력
연계프레임 워크	TRL 3	TRL 6	연구데이터/메타데이터 연계	●		
			연구데이터 연계기술 개발	●	○	

● : 주, ○ : 부

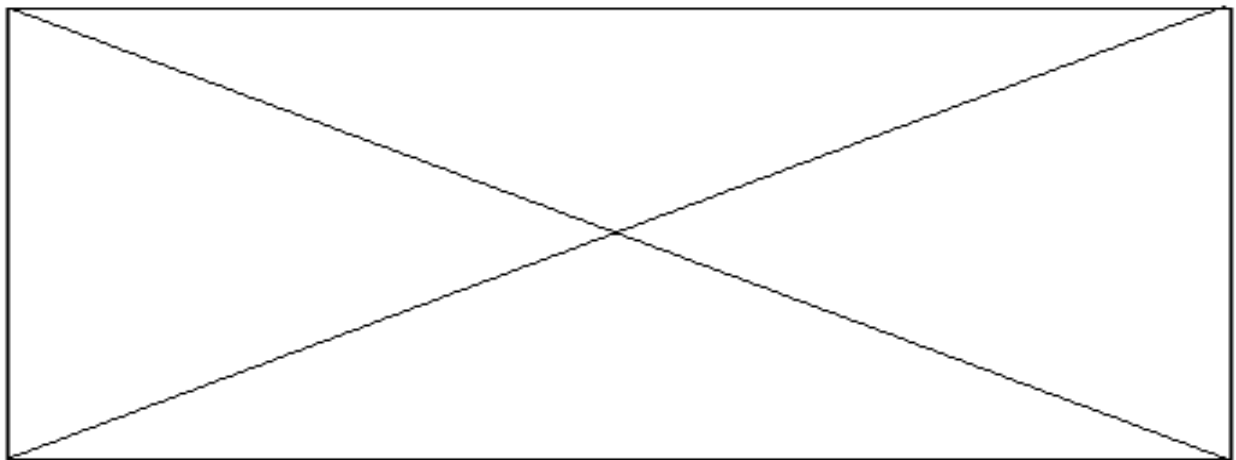
－ 연구데이터/메타데이터 연계 : REST API, OAI-PMH 표준을 준수하여 자체 개발

－ 연구데이터 연계기술 개발 : FTP, Cloud환경의 데이터 공유 기술을 활용하여 자체 개발

나. 핵심기술 구성 및 내용

□ 핵심기술 구성

- 수집·연계 기술로 활용되고 있는 REST, XML-RPC, SOAP, FTP, OAI-PMH 등의 기술 중 적합한 기술을 이용하여 연구데이터/메타데이터를 효과적으로 수집하고, 대용량 데이터 전송을 위한 표준 기술 개발

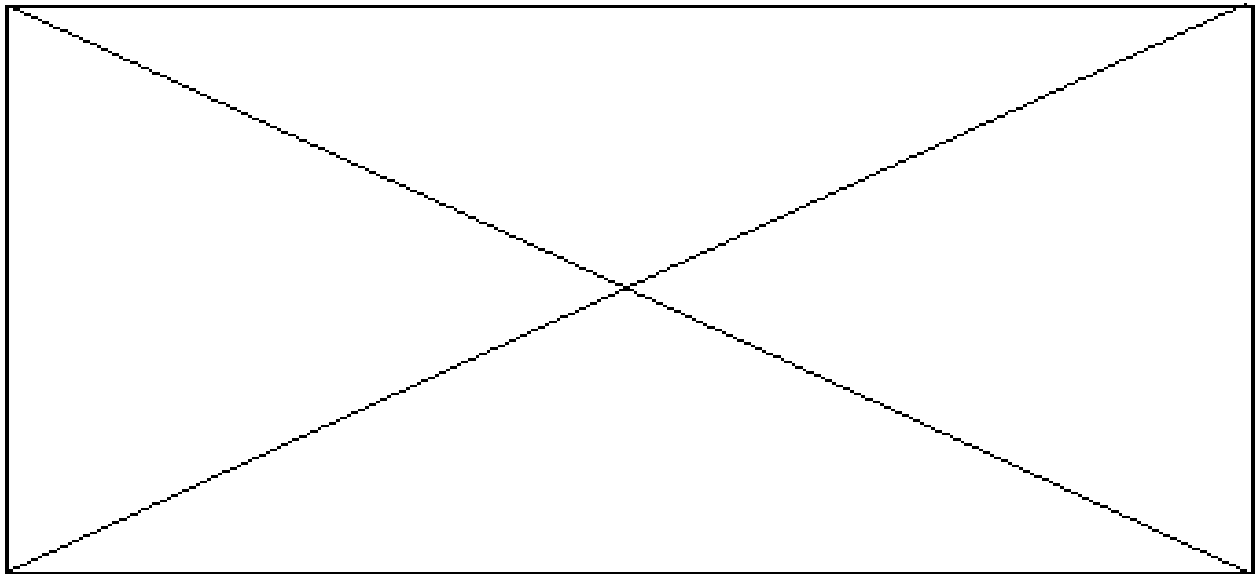


□ 핵심기술 내용

구분	기술	설명	비고
대외연계 (Open API)	REST (Representational State Transfer)	- REST는 월드 와이드 웹 같은 분산 하이퍼미디어 시스템을 위한 소프트웨어 아키텍처의 한 형식 - 도메인 지향 데이터를 http위에서 soap이나 쿠키를 통한 세션 트래킹 같은 부가적인 전송 레이어없이 전송하기 위한 아주 간단한 인터페이스를 의미 - REST는 네트워크 아키텍처 원리의 모음으로 리소스를 정의하고 리소스에 대한 주소를 지정하는 방법 - xml과 http 인터페이스를 이용하여 커다란 소프트웨어 시스템 설계가 가능	InHouse
	XML-RPC	- RPC프로토콜의 일종으로 인코딩 형식은 xml을 선택하고 전송방식은 http를 사용 - Xml-RPC의 경우는 클라이언트/서버 방식으로 동작하여 서버가 반드시 필요	InHouse
	SOAP (Simple Object Access Protocol)	- HTTP, HTTPS, SMTP 등을 사용하여 XML 기반의 메시지를 컴퓨터 네트워크 상에서 교환하는 형태의 프로토콜 - 웹 서비스에서 기본적인 메시지를 전달하는 기반 - 원격 프로시저 호출 패턴으로 네트워크 노드에서 다른 쪽 노드쪽으로 메시지를 요청 하고, 서버는 메시지를 즉시 응답 - SOAP는 XML을 근간으로 헤더와 바디를 조합하는 디자인 패턴으로 설계	open source / 상용
	FTP (File Transfer Protocol)	- TCP/IP 네트워크 상에서 컴퓨터들이 파일을 교환하기 위한 통신 규약 - 파생형으로 FTP에 TLS 프로토콜이 적용된 FTPS가 있으며, 비슷한 기능을 제공하는 SSH 프로토콜 기반의 SFTP - FTPS와 SFTP는 보안 전송이라는 특성 상 FTP보다 파일 전송 속도가 느린 편	InHouse
	OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)	- 저장소 상호 운용성을 위한 메커니즘 - 데이터 제공은 OAI-PMH를 통해 구조화 된 메타 데이터를 노출하는 리포지토리 제공 - 서비스 공급자는 해당 메타 데이터를 수집하기 위해 OAI-PMH 서비스 요청을 수행	open source
	crawling / scraping	크롤링(crawling) 혹은 스크레이핑(scraping)은 웹 페이지를 그대로 가져와서 거기서 데이터를 추출해 내는 행위	InHouse
분산환경 연계	Link	스토리지상의 연구데이터에 대한 위치 정보를 제공하며, 연구자가 참조 형태로만 활용	
	FTP	대외 분석 환경으로 연구원 자료를 이동시켜 원자료에 변형을 가공하여 사용할 경우	수퍼컴퓨터센터
	Copy(복제)	센터 내부의 분석 환경으로 연구원 자료를 이동시켜 원자료에 변형을 가공하여 사용할 경우	

□ 연구데이터/메타데이터 연계 기술 개발

- 바이오·소재·인공지능등의 전문 센터 연계 기술 개발
 - 대·소분야 전문센터와 국가 연구데이터 플랫폼간 연구데이터를 상호 검색하여 연구데이터를 DB로 자동 구축할 수 있는 API Server, API Client 개발
 - 검색되어 수집·공유되는 연구데이터의 효과적인 연동을 위하여 페이지 처리, Partial Response 처리, 검색 기능을 API에 탑재되도록 설계
 - API인증, API인가, 네트워크 레벨 암호화, 메시지 무결성 보장, 메시지 본문 암호화 등의 기술을 활용하여 수집·공유되는 데이터의 보안이 확보되도록 기술 설계
 - ※ API Key, API Token, 클라이언트 인증 적용
 - 대·소분야 전문센터의 신규 서비스 시 API를 개발하지 않고 관리자가 자동으로 생성하고 수정 등의 버전 관리를 할 수 있도록 설계
 - 대·소분야 전문센터와 연동되는 API의 상태, 오류 현황, 연동 현황을 실시간적으로 파악되도록 모니터링 기능을 설계하고 개발



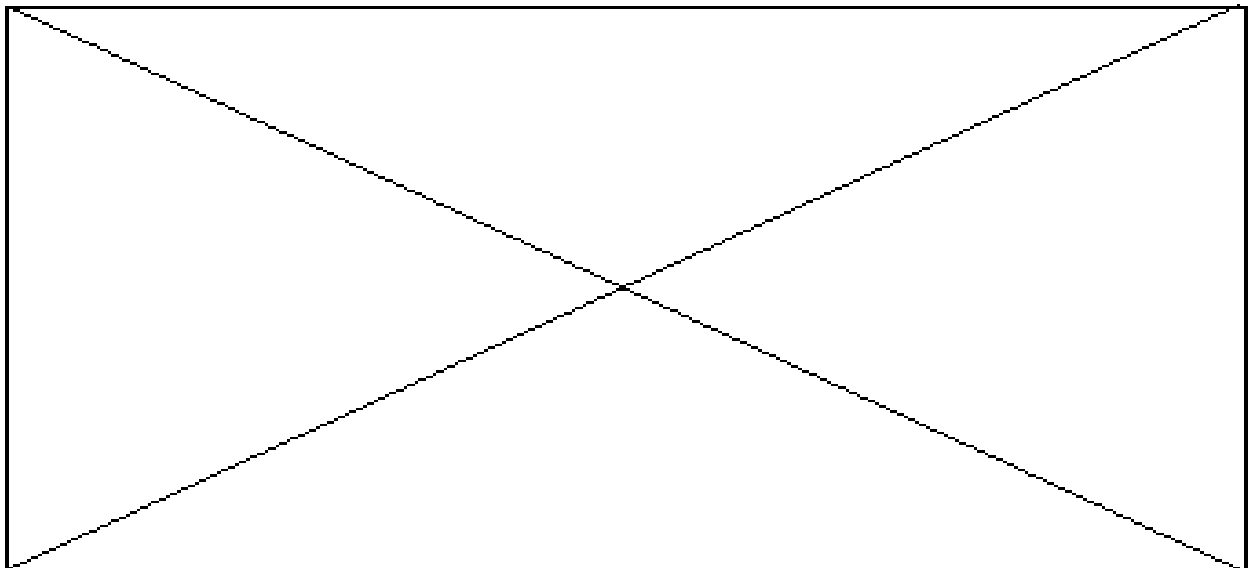
<그림 5-9> 대·소분야 전문센터 연계·공유 기술 개발 구성도

- 국내외 과학기술 기관 연계 및 공유 기술 개발
 - OAI-PMH 프로토콜 기반의 수집 프로그램을 개발하고 수집 스케줄러, Validation Check, DataExtractor 기능이 포함된 수집 서버 설계

표 5-5> OAI-PMH 요청 인터페이스

구성	설명
Identify	전문정보센터의 명칭, 관리자 이메일 주소, 전문정보센터 URL, DP 설치일, OAI-PMH 버전 등의 식별 정보를 수집
ListMetadataFormats	메타데이터 Schema 정보를 수집
ListSets	전문정보센터 자체 분류정보 목록을 수집
ListIdentifiers	메타데이터의 레코드 아이디 목록을 수집
Listrecords	메타DB에 적재되어 있는 메타데이터 중 전달 조건에 맞는 메타데이터들을 수집
GetRecord	메타DB에 적재되어 있는 메타데이터 중 해당하는 레코드 아이디의 메타데이터를 수집

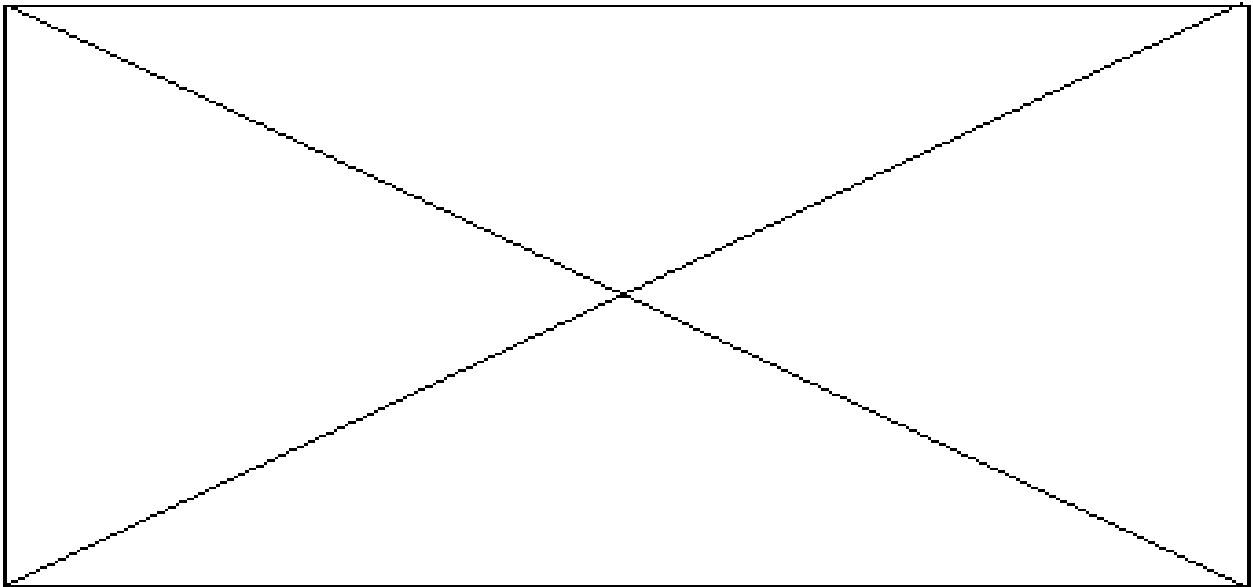
- 연구데이터의 메타데이터를 표준에 맞게 설계하고 메타데이터를 등록하고 관리할 수 있는 DOI 관리시스템 설계 개발
- 수집한 메타데이터를 기반으로 검색, 브라우징, 연계, 원문제공 등과 같은 서비스를 제공하는 SP(Service Provider)와 각종 디지털 컬렉션을 수집, 보유하고 있는 DP(Data Provider) 시스템을 구성하여 수집 요청에 대응하여 적합한 메타데이터를 제공하도록 설계
- 인증, 인가, 네트워크 레벨 암호화, 메시지 무결성 보장, 메시지 본문 암호화 등의 기술을 활용하여 수집·공유되는 데이터의 보안이 확보되도록 기술 설계



<그림 5-10> 국내외 과학기술 기관 연계 및 공유 기술 개발 구성도

○ NTIS·성과마루 연계 기술 개발

- Rest API, Crawling 프로토콜을 활용한 연구데이터 수집 및 공유 기술을 개발
- NTIS, 성과마루 시스템과의 연계를 위한 표준 프로토콜을 선정하여 표준안을 설계
- 인증, 인가, 네트워크 레벨 암호화, 메시지 무결성 보장, 메시지 본문 암호화 등의 기술을 활용하여 수집·공유되는 데이터의 보안이 확보되도록 기술 설계
- 대상 시스템(NTIS, 성과마루)에 설치한 수집 Agent를 개발
- 연동되는 연구데이터의 상태, 오류 현황, 연동 현황을 실시간적으로 파악되도록 모니터링 기능을 설계하고 개발
- 수집·공유된 연구데이터의 사용현황을 파악하기 위한 일/월/연별 통계 기능 설계



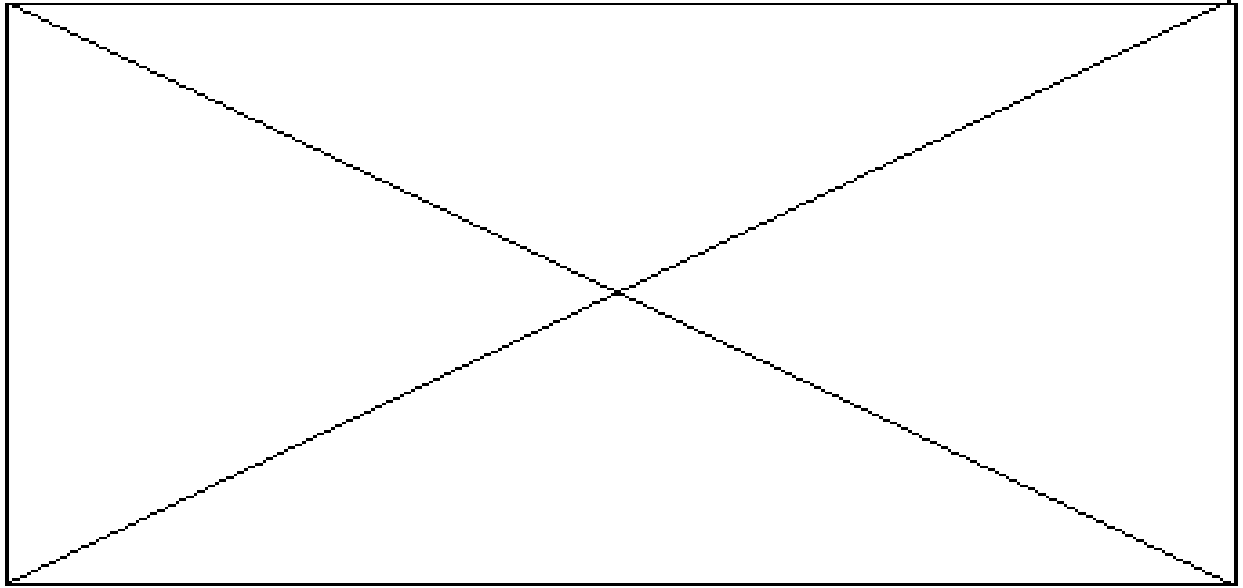
<그림 5-11> NTIS·성과마루 연계 및 공유 기술 개발 구성도

□ 대용량 연구데이터 전송기술 개발

○ 연구내용

- 대·소분야 전문센터, 연구자 보유 연구데이터, 타 기관의 연구데이터를 국가연구데이터 플랫폼의 분석 시스템에 전송 및 복제할 수 있는 기술을 도출
- 국가 연구데이터 플랫폼 이외의 시스템 및 연구자가 전송하는 기술과 국가연구데이터 플랫폼 내에서 연구데이터를 복제하는 제공하는 기술을 분리하여 설계
 - ※ 클라우드 내에서의 데이터 전송은 클라우드 매니지먼트 플랫폼을 활용
- 국가 연구데이터 플랫폼과 슈퍼컴퓨터센터로 연구데이터를 전송하는 기술 설계

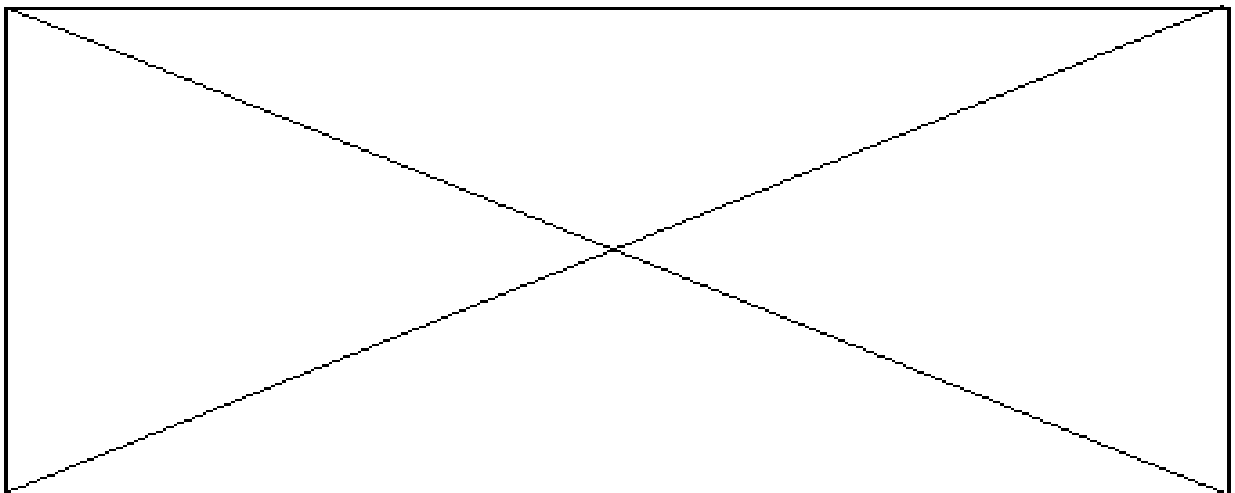
- \ - 연동되는 연구데이터의 상태, 오류 현황, 연동 현황을 실시간적으로 파악되도록 모니터링 기능을 설계하고 개발
- 수집·공유된 연구데이터의 사용현황을 파악하기 위한 일/월/연별 통계 기능 설계
 - ※ 모니터링 및 통계 시스템은 국가 연구데이터 플랫폼 내부와 외부의 연동이 통합되어 관리될 수 있도록 설계



<그림 5-12> 대용량 연구데이터 전송 기술 개발 구성도

□ 수집·연계 기능구성

- 수집·연계 기능은 연구데이터/메타데이터를 연계하고 모니터링하는 기능과, 클라우드 환경 내의 분석 환경으로 외부 자료를 업로드, 클라우드 환경 내에서의 전송 기능으로 구성



다. 연차별 연구개발 계획

번호	플랫폼	1단계			2단계				3단계		
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
1	연구데이터/메타데이터 연계	설계 및 기술개발	최적화 및 안정화		기능 고도화						
2	연구데이터 연계기술 개발										

라. 소요예산

○ 전체 예산 : 20,774 백만원

단위 : 백만원

번호	플랫폼	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
	계	1,804	2,522	2,405	2,739	2,066	1,671	1,861	1,909	1,898	1,899	20,774
1	연구데이터/메타데이터 연계	305	427	407	463	349	283	315	323	321	321	3,514
2	연구데이터 연계기술 개발	1,499	2,096	1,998	2,276	1,717	1,388	1,546	1,586	1,577	1,578	17,261

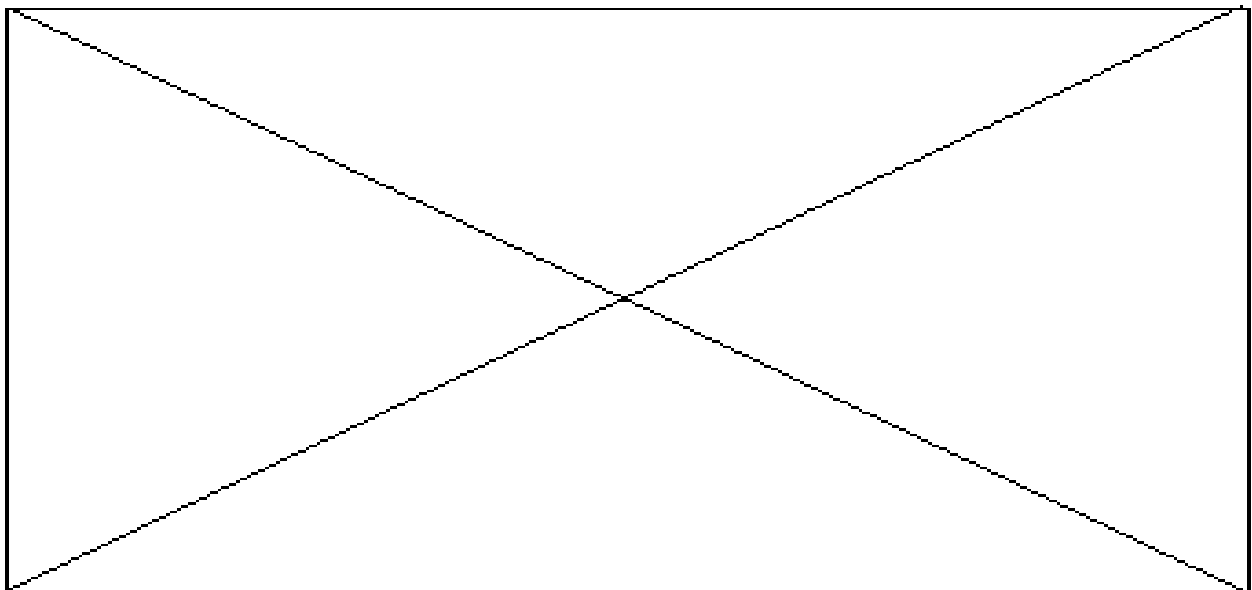
5.3.2. 연구데이터 저장·보안 기술 개발

가. 개발 및 추진 목표

개발 목표	메타데이터, DMP정보 통합 관리, 대용량 연구데이터의 원활한 분석을 위한 연구데이터 저장, 보안 시스템 개발
세부 추진 목표	① 대용량 데이터 저장을 지원하는 클라우드 기반의 분산, 병렬 저장 기술 개발 ② 데이터 연계 및 저장 시 보안 기술 개발

□ 세세부 과제 도출

- 연구데이터/메타데이터를 저장하기 위한 기술로는 현재 클라우드 환경에서 표준으로 사용하고 있는 분산 데이터저장 기술을, 메타데이터의 개인정보 및 비공개 정보의 보안 확보 방안을 세세부 과제로 도출



<그림 5-13> 저장·보안 기술 개발 세세부과제 도출

□ 기술 확보전략

기술 구분	현재 기술수준	목표 기술수준	세세부 과제	세부기술 확보 전략		
				자체개발	기술도입	글로벌 협력
저장·보안 프레임워크	TRL 3	TRL 6	클라우드 기반 분산,병렬 저장 기술 개발	●		
			데이터 저장 및 연계 보안기술 개발	○	●	

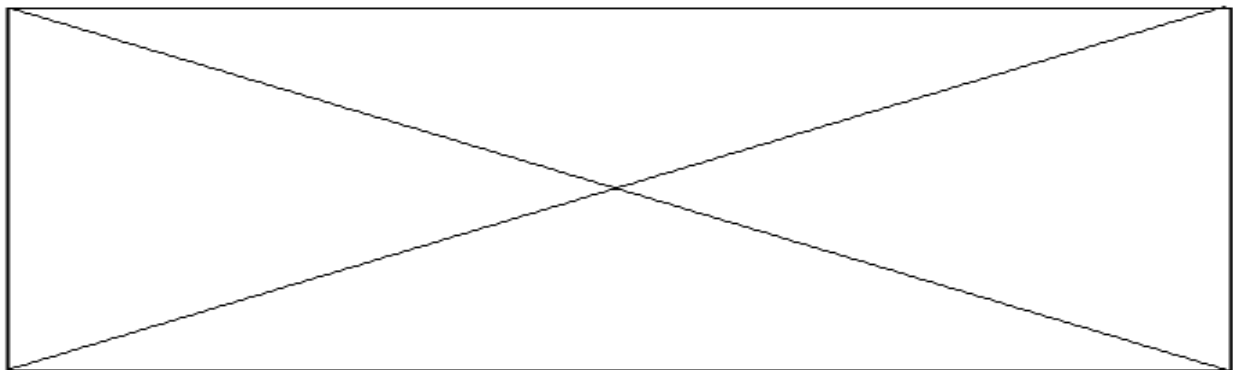
● : 주, ○ : 부

- 분산, 병렬 저장 기술 : Hadoop 기술을 활용하여 자체 개발
- 보안기술 : SSO, 암호화 솔루션을 도입, 통합계정관리는 자체 개발

나. 핵심기술 구성 및 내용

□ 핵심기술 구성

- 저장 프레임워크 기술 개발 분야에서는 빅데이터 저장 및 분석에 활용되고 있는 분산/병렬 파일시스템과 NoSQL 중심으로 구현 기술을 도출
- 보안 프레임워크는 위변조 방지, 암호·복호화, 전송 및 접근제어, 비식별화·익명화 기술을 활용하여 데이터의 보안 및 개인정보 관리 기술 개발



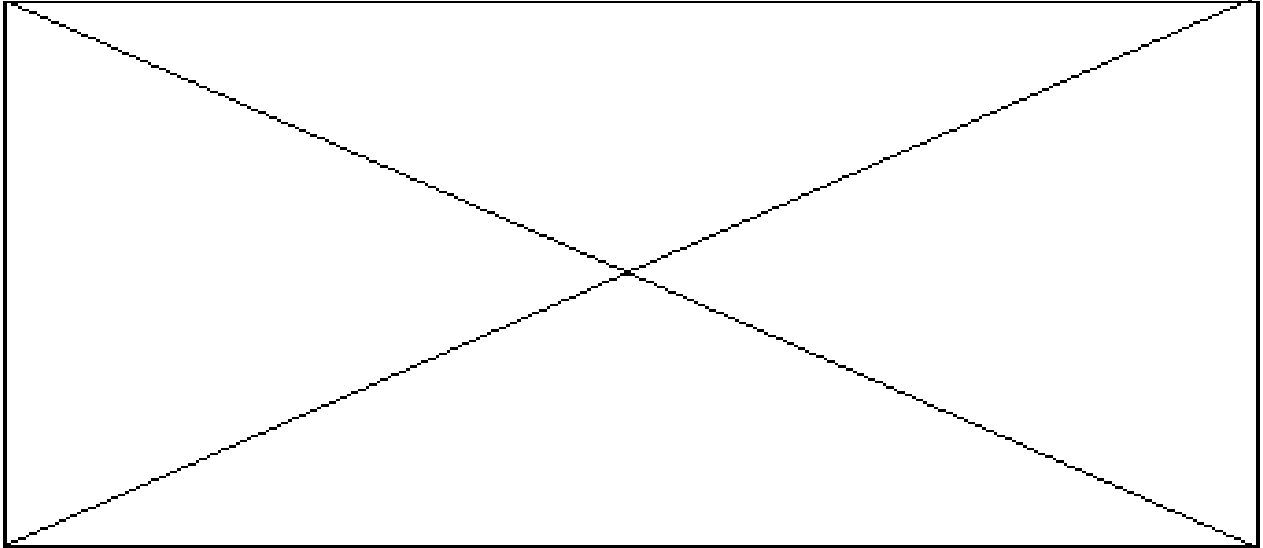
□ 핵심기술 내용

구분	기술	설명	비고
저장 프레임워크	분산/병렬 파일시스템	클라이언트 측에서 분산된 서버에 저장된 데이터에 접근하여, 마치 로컬 시스템에 저장되어 있는 데이터인 것처럼, 처리할 수 있는 클라이언트/서버 기반의 애플리케이션	
		HDFS (Hadoop Distributed File System) - Java로 작성된 Open Source로 확장성과 이식성을 제공 - 데이터를 저장하는 데이터노드와 메타데이터를 저장하는 네임노드로 구성되며 다수의 데이터노드에 중복하여 저장하므로 하나의 노드에서 고장이 발생하거나 느려지더라도 데이터에 접근할 수 있도록 신뢰성을 제공 - 파일은 일반적인 파일시스템보다 많은 양의 블록 단위로 저장 - HDFS를 이용할 때는 HDFS 고유의 API를 이용하며, 네임노드의 이중화 구성을 통해 가용성을 향상시킴	open source
		GFS (Google File System) - 대용량 분산 데이터전문 애플리케이션을 위한 유연한 분산파일 시스템 - GFS는 저가의 일반적인 하드웨어 상에서 동작하면서도 무 정지 기능과 많은 수의 클라이언트에 대한 높은 군집성능(high aggregate performance)을 제공함 - 앞으로 예측되는 어플리케이션 부하(workload)와 기술적 환경에 중점을 두고 설계 - GFS는 구글의 주 저장 플랫폼으로 채택되어 오랫동안 구글 서비스뿐만 아니라 대량의 데이터 셋을 요구하는 연구와 개발 분야에 사용 - 큰 클러스터는 수 천개의 디스크에 걸쳐 구성된 수 백 테라바이트의 사이즈를 가지며 동시에 수 백 명의 클라이언트에 의해 접속	open source
	NoSQL (Not-Only SQL)	- 전통적인 관계형 데이터베이스 보다 덜 제한적인 일관성 모델을 이용하는 데이터의 저장 및 검색을 위한 매커니즘을 제공 - SQL 계열 쿼리 언어를 사용할 수 있음 - 단순 검색 및 추가 작업을 위한 매우 최적화된 키 값 저장 공간으로, 레이턴시와 스루풋과 관련하여 상당한 성능 이익을 내는 것이 목적이며 NoSQL 데이터베이스는 빅데이터와 실시간 웹 애플리케이션의 상업적 이용에 널리 쓰임	
	BigTable	- 매우 큰 규모의 구조화 된 데이터를 관리하기 위한 분산 스토리지 시스템 - 구글 인덱싱, 구글 어스 그리고 구글 금융까지 구글의 많은 프로젝트가 빅테이블에 저장되어 있으며, 이들 애플리케이션들은 큰 규모 데이터, 실시간 처리를 요구 - BigTable은 모든 구글 제품들에 대한 유연하고 높은 수 행 해결책을 성공적으로 제공 - 빅테이블은 산재해있고(sparse), 분산되어있으며(distributed), 영구적인 다차원의 잘 정리된 맵(multi dimensional sorted map) - 맵은 row key, column key, time-stamp로 인덱싱 - 맵의 각 value는 해석되지 않은 바이트 어레이(bytes	open source

구분	기술	설명	비고
		array)	
	Hbase	• HBase는 일반에게 논문으로 공개된 구글의 BigTable을 모델로 만들어졌으며, 2007년 10월에 첫 번들을 시작으로, 2010년에는 아파치 최고 수준의 프로젝트가 됨 • HBase는 확장성 문제를, 기존 데이터베이스와는 다른 관점으로 접근해서, 단지 노드만 추가하면 선형적으로 확장될 수 있도록 개발 • 관계형 데이터베이스가 아니며 SQL도 지원하지 않지만, RDBMS가 할 수 없는, 웹테이블(수집된 웹페이지와 웹페이지 URL을 키로하는 속성 테이블)과 같은 기능을 제공 • 웹테이블은 로우(row) 개수만 10억개가 넘을 수도 있음	open source
	mongodb	• C++ 언어로 10gen에 의해 개발 • 문서형 데이터베이스로 높은 성능과 확장성을 가지고 있는 데이터베이스 • 스키마가 없고, Key값에 대한 복잡한 검색이 가능하며, BSON(JSON을 바이너리화 한 것) 형식으로 데이터 저장 • 웹 애플리케이션과 인터넷 기반을 위해 설계된 데이터베이스 관리 시스템으로 대부분의 상품 정보를 하나의 도큐먼트로 표현할 수 있음 • 내장된 맵리듀스 방식의 집계연산이나 공간 정보 색인과 같은 다양한 기능 제공 • [단점] 조인이나 트랜잭션성 처리가 안됨 • [단점] 데이터 입력 및 갱신 시점에 바로 디스크에 쓰지 않음	open source
	cassandra	• Java로 만듦 • 아파치 프로젝트로 진행되는 오픈소스 소프트웨어 • Google의 BigTable과 아마존에서 만든 Dynamo를 기반으로 만들었으며 컬럼 단위로 관리함 • 효과적인 분산된 대량 데이터 처리로 분산환경에 적합하며 확장성이 뛰어남	open source
	Redis	• Remote Dictionary Server의 약자로서, "키-값" 구조의 비정형 데이터를 저장하고 관리하기 위한 오픈 소스 기반의 비관계형 데이터베이스 관리 시스템 • 모든 데이터를 메모리로 불러와서 처리하는 메모리 기반 DBMS • 외부층에서 레디스의 데이터 모델은 키를 값에 맵핑하는 자료구조의 디렉토리 • 스트링뿐만 아니라 자료구조의 추상적 자료형(Abstract Data Type)도 지원	open source
	RDB	Oracle, mySQL, MS-SQL	

구분	기술		설명	비고
보안	보안 관리	사용자 인증	• 누가 어떤 데이터에 어떤 조치를 취할 수 있는가를 미리 정한 바에 따라, 데이터나 데이터 관리 시스템에 접근하는 사람의 접근 자격을 확인 • 세부 기술 : ID/Password / 일회용 패스워드(OTP) , 전자인증, 생체인증, 통합사용자 인증	inhouse
		접근제어	• 어떤 주체가 어떤 객체를 읽고자 하거나, 객체에 기록하고자 하거나, 객체를 실행(객체가 실행 과 일일 경우)시키하고자 할 때 마다 그 주체가 그 객체에 대한 권한을 가지고 있는지를 체크하고 통제하는 것 • 세부 기술 : 강제 접근제어, 임의 접근제어, 역할기반 접근제어	inhouse /상용
		암호화	• 평문을 해독 불가능한 형태로 변형하거나 또는 암호화된 암호문을 해독 가능한 형태로 변형하기 위한 원리, 수단, 방법 등을 취급하는 기술 • 암호화는 복호화 가능 여부에 따라 양방향 암호화와 단방향 암호화로 분류되며, 양방향 암호화는 암호·복호화 시 사용되는 키의 대칭여부에 따라 대칭키 암호화와 비대칭키 암호화로 분류 • 암호화 알고리즘 : DES, AES, RSA, MD5, SHA 등	상용
	개인정보 처리	비식별화	• 수집된 데이터에 포함된 개인정보의 일부 또는 전부를 삭제, 다른 정보로 대체 또는 다른 정보와 결합(Mash-up)하여도 특정 개인을 식별하기 어렵도록 하는 일련의 조치 • 가명기법 : 개인정보 중 주요식별 요소를 다른값으로 대체하여 개인식별을 어렵도록 함 • 데이터 값 삭제 : 데이터의 공유 , 개방 목적에 따라 데이터셋에 구성된 값 중에서 필요없는 값 또는 개인 식별에 중요한 값을 삭제 • 범주화 : 데이터의 값을 범주의 값으로 변환하여 명확한 값을 감춤 • 데이터마스킹 : 공개된 정보 등과 결합하여 식별하는데 기여할 확률이 높은 주요 개인식별 자가 보이지 않도록 처리하여 개인을 식별하지 못하도록 함	inhouse /상용
		암호화	• 데이터베이스 전체에 대한 보호가 아닌 개인정보가 포함된 특정 필드에 대한 보호기술 • 개인정보의 암호화 저장 후 데이터베이스에 저장된 개인정보의 정상적인 이용을 위해 데이터베이스를 안전하고 효율적으로 인덱싱하는 기술	inhouse /상용

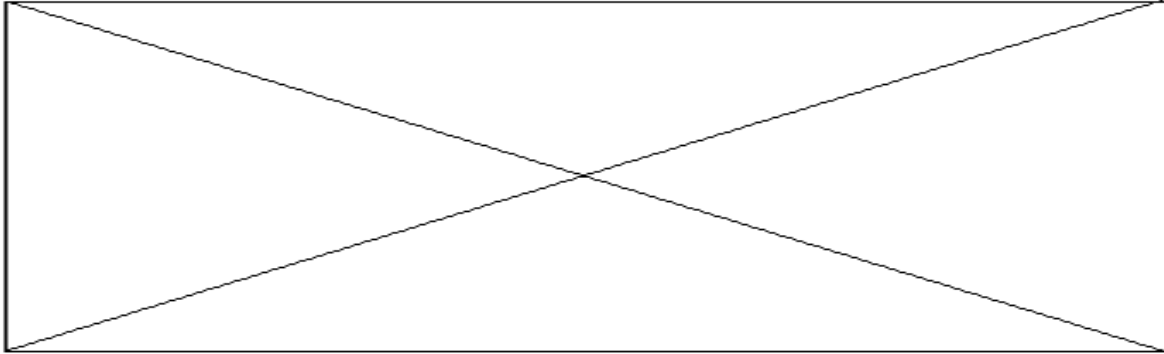
□ 분산, 병렬 처리 기술 개발 구성



<그림 5-14> 분산, 병렬처리 아키텍처 구성도

- 분산처리 프레임워크의 아키텍처는 클라우드 기반 하둡 분산처리 기술을 활용하여 과학데이터를 저장 및 관리. 일반적으로 하둡은 대용량 및 다양한 이질적인 데이터를 저장하고 이를 분석하기 위한 기술이지만 이 프레임워크는 연구데이터를 저장하고 이를 관리함
- 하둡의 장점인 대용량파일 및 데이터 보관의 안정성 및 향후 분석 기술의 확장성을 고려하여 하둡환경의 분산처리 기술을 활용하여 데이터 및 파일을 HDFS 형태로 저장하고, TB이상의 데이터는 스토리지에 저장하여 이에 대한 사용자의 접근 제어를 통한 공유 체계를 제공
- 연구데이터를 저장하는 형태는 하둡기반 기술을 활용하여 분산처리 HDFS에 저장하는 방법과 스토리지에 저장하는 형태로 구분 되어 질 수 있음

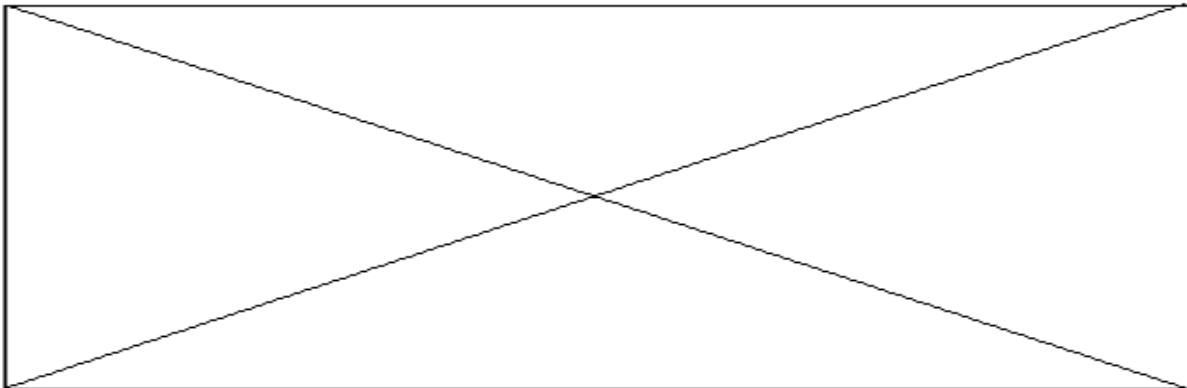
□ 통합계정관리, SSO/접근제어 기술 개발 구성



<그림 5-15> 통합계정관리, SSO, 접근제어 기능 구성도

- 인증기능 그룹 : 데이터에 대한 접근 통제를 위해 개인 인증 및 SSO 기능 등을 포함한 보안 프레임워크와 연동
- 데이터 보안기능 그룹 : 데이터를 공유 또는 제공하기 위해서 PKI를 통한 인증체제를 활용하고 SSL을 적용하여 데이터의 암호화

□ DB, 분산환경 보안 관리 구성



<그림 5-16> DB, 분산환경 보안 관리 구성도

- 분산저장 처리 환경을 제공하는 시스템(클러스터)에 대한 접근제어 기능을 제공
- 분산저장 처리 환경에 저장된 데이터들에 대한 보안 방안 수립
- 분산처리 프레임워크의 기반을 제공하는 Hadoop 및 Hadoop 에코시스템의 설정 및 관리 기능들은 플랫폼관리 프레임워크의 관리 기능 통합

다. 연차별 연구개발 계획

번호	플랫폼	1단계			2단계				3단계		
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
1	클라우드 기반 분산,병렬 저장 기술 개발		설계 및 기술 개발	최적화 및 안정화	기능 고도화	정상운영					
2	데이터 저장 및 연계 보안기술 개발										

라. 소요예산

□ 전체 예산 : 42,072 백만원

단위 : 백만원

번호	플랫폼	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
계		3,315	5,324	4,984	5,544	4,186	3,385	3,771	3,869	3,846	3,848	42,072
1	클라우드 기반 분산, 병렬 저장 기술 개발	2,354	3,780	3,539	3,936	2,972	2,404	2,677	2,747	2,731	2,732	29,874
2	데이터 저장 및 연계 보안기술 개발	961	1,544	1,445	1,607	1,214	981	1,093	1,122	1,115	1,116	12,199

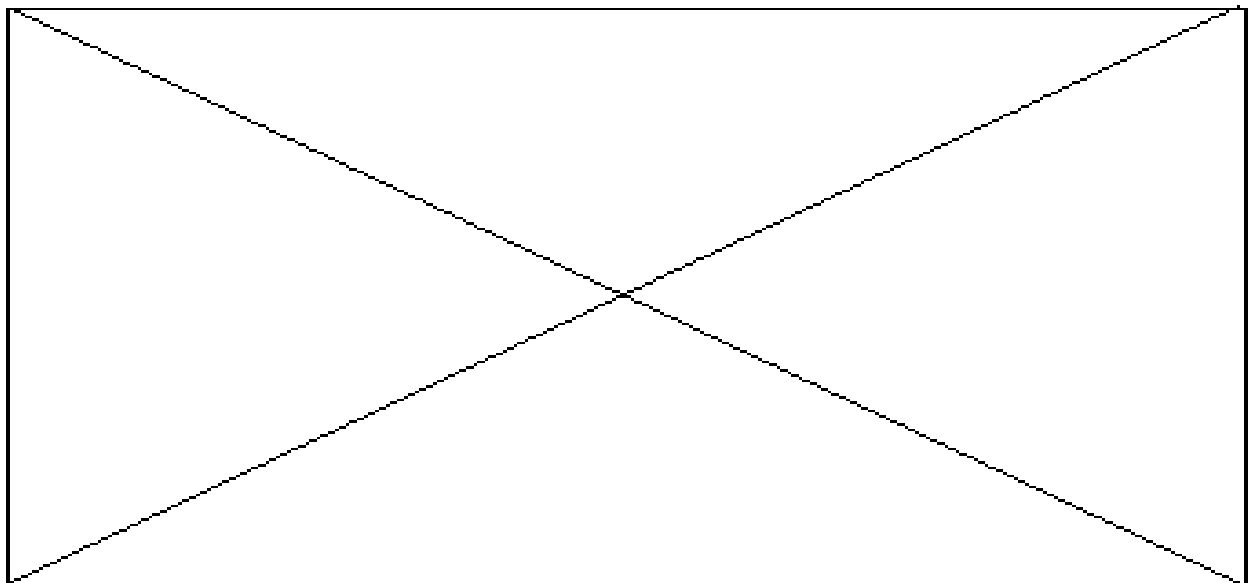
5.3.3. 연구데이터 품질·표준화 기술 개발

가. 개발 및 추진 목표

개발 목표	연구데이터/메타데이터의 품질제고를 위한 연구데이터 품질관리 시스템 개발
세부 추진 목표	① 연구데이터의 품질 관리를 위한 표준 정의, 관리기술 개발 ② 연구데이터의 출처, 저작권 관리

□ 세세부 과제 도출

- 수집·연계되는 메타데이터의 품질 확보를 위한 표준정의와 유통, 저장되는 메타데이터의 품질확보를 위한 관리 기술 개발 도출
- 연구데이터의 출처 및 소유권 정보를 관리하는 저작권 관리 기술 개발을 도출



<그림 5-17> NTIS·성과마루 연계 및 공유 기술 개발 구성도

□ 기술 확보전략

기술 구분	현재 기술수준	목표 기술수준	세세부 과제	세부기술 확보 전략		
				자체개발	기술도입	글로벌 협력
품질/표준 프레임워크	TRL 3	TRL 6	데이터 표준 수립 및 품질관리 기술 개발	●	○	
			연구데이터 저작권 관리 기술 개발	●		

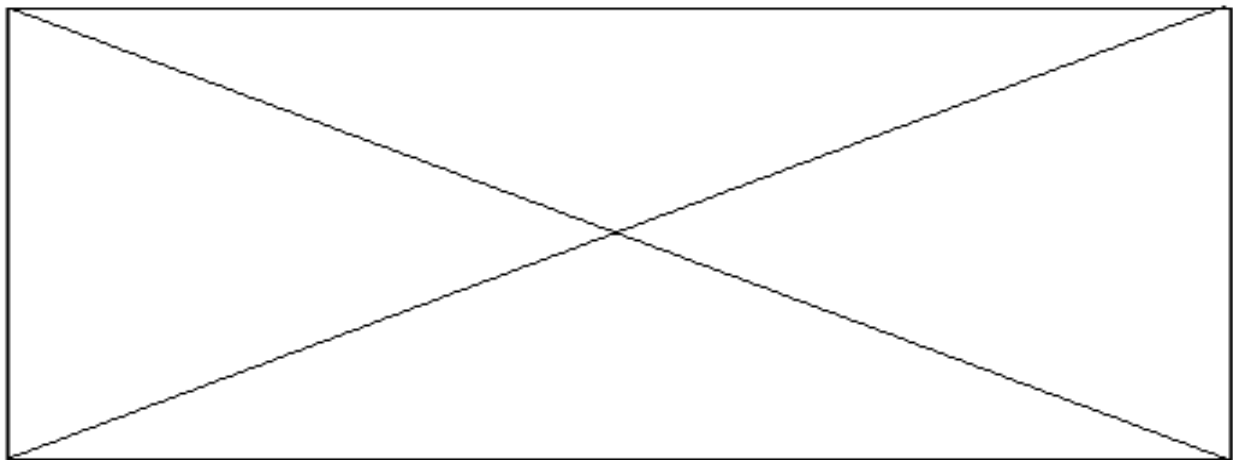
● : 주, ○ : 부

- 표준수립 : 연구과제정보, DMP표준을 기반으로 표준 정의
- 품질관리 : 품질관리 솔루션을 도입
- 저작권관리 : 오픈소스 기반 자체 개발

나. 핵심기술 구성 및 내용

□ 핵심기술 구성

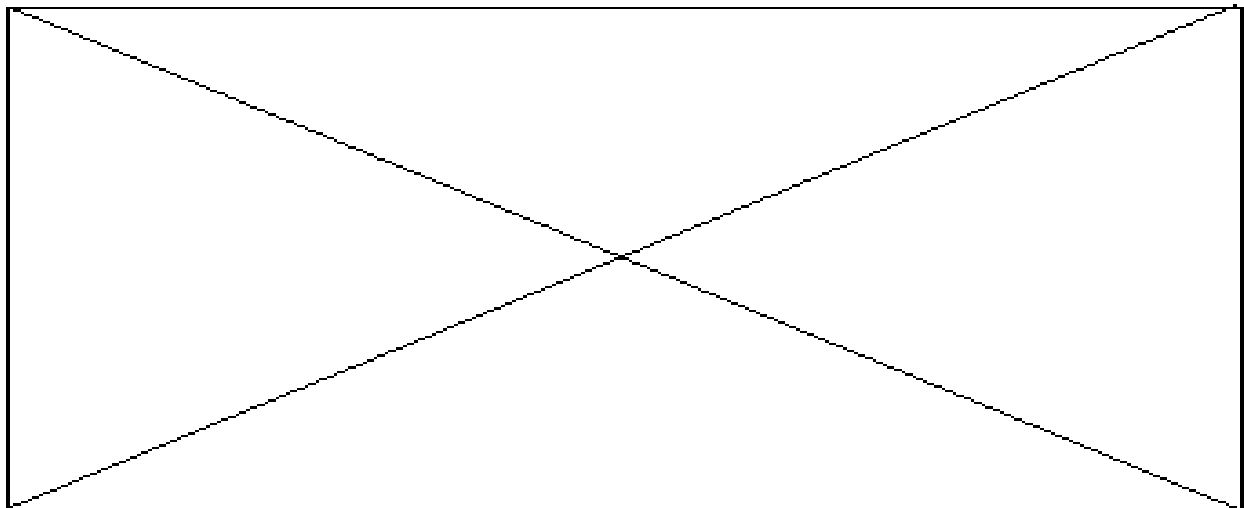
- 연구과제정보와 DMP정보를 기반으로 메타데이터 관리를 위한 스키마, 항목에 대한 표준 수립과 메타데이터 수집 시 수립된 메타데이터 표준(안)을 물기반의 자동 품질검증 기술개발을 세세부 과제로 도출



□ 핵심기술 내용

구분	기술		설명	비고
표준 정의	메타 형식	컬렉션	• 적별(연구사업관리, 과제관리, 주제분야관리 등)로 과학 데이터를 관리	InHouse
		아이템	• 컬렉션 하위객체로서 생성되는 아이템은 공유와 활용을 위한 과학데이터 파일의 묶음 • 인용을 위한 DOI(Digital Object Identifier) 부여	InHouse
		파일	• 공유 활용의 가치가 있는 개별 단위의 과학데이터로서 최종적으로 연구자들이 활용하는 대상 • 파일 그 자체로도 공유 활용이 가능하며, DOI 부여도 가능하다	InHouse
	항목 (스키마)	컬렉션	• CollectionID, CollectionName, CollectionDates, CollectionDescription, AccessUser, AccessUser, CollectionOrganization	InHouse
		아이템	• ItemID, ItemTitle, ItemDates, ItemSubject, ItemAccesspolicy, ItemCreator, ItemPublisher, ItemPublicationYear, ReferenceID, DOIIssue	InHouse
		파일	• Identifier, Creator, Title, Publisher, PublicationYear, Subject, Date, AccessPolicy,	InHouse
품질 점검	룰	정의	• 연계 또는 입력되는 메타데이터의 항목별 “형태”, “길이”, “범위” 등을 정의	InHouse / solution
		검증	• 연구데이터의 연계 또는 입력 시 실시간/배치 룰 검증을 통해 저장되는 메타 데이터의 품질 확보	InHouse / solution

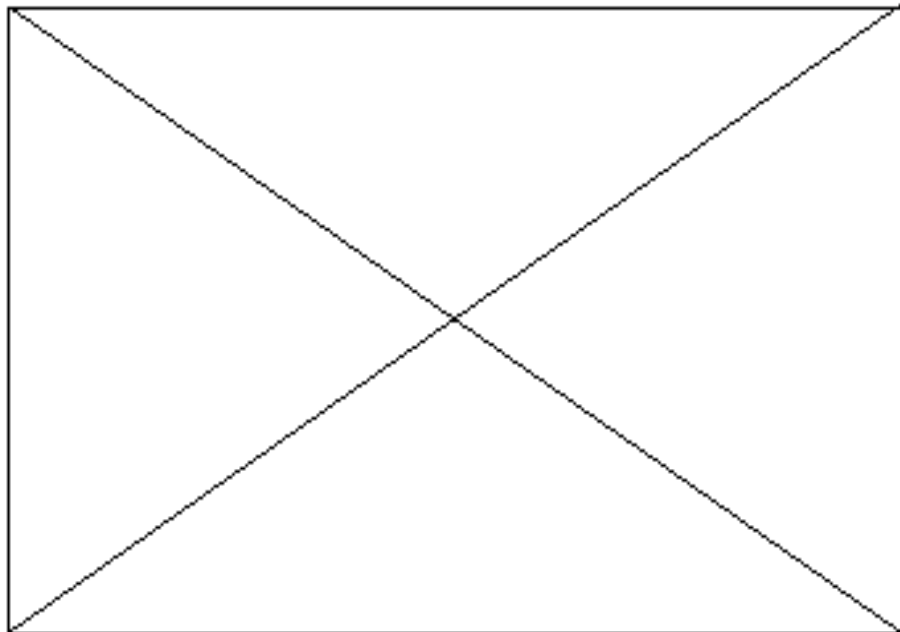
□ 메타데이터 표준정의



<그림 5-18> 메타데이터 스키마 표준 정의 구성도

- 메타데이터 요소 중 주제(subject)를 작성할 경우, 국가과학기술분류체계의 표준을 우선적으로 준수
- 메타데이터 요소 중 연구자를 식별하는 Identifier (nameIdentifier)를 작성할 경우, 과학기술인등록번호를 우선적으로 사용할 것을 권장하며, 그 외에 정해진 리스트 중에서 선택하여 기입
- 기관별 구축, 보유한 연구데이터의 메타데이터를 연계하고자 할 때는 XML 형태를 권장

□ 데이터 품질, 출처/저작권 관리 기능 구성



- 표준관리 : 단어, 용어(컬럼), 도메인, 유사어 및 공통코드에 대한 조회 및 표준 체크, 변경에 따른 변경 요청 및 승인 관리하는 기술 개발
- 도메인, 코드, 금치어 등과 같은 데이터 표준의 세부 사항을 관리하는 데이터 사전 구축
- 데이터 모델(스키마) 관리 : 개념, 논리, 물리 데이터 모델과 운용 데이터베이스를 관리
- 메타데이터 생명주기 관리를 위한 최적화된 워크플로우 개발
- 데이터 품질 관리 활동의 세부적인 시나리오 설계 및 주기적인 실행 관리하는 기능 개발
- 데이터 점검 규칙(비즈니스 룰)으로 업무 담당자의 지식을 관리하며 규칙에 위배되는 데이터를 추출하는 기능 개발
- 품질 측정 결과(오류 데이터 목록) 및 개선 결과에 대한 다양한 보고서 생성 기능 개발

다. 연차별 연구개발 계획

번호	플랫폼	1단계			2단계				3단계		
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
1	데이터 표준 수립 및 품질관리 기술 개발	설계 및 기술 개발	최적화 및 안정화		기능 고도화						정상운영
2	연구데이터 저작권 관리 기술 개발										

라. 소요예산

□ 전체 예산 : 32,556 백만원

단위 : 백만원

번호	플랫폼	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
	계	2,828	3,953	3,769	4,292	3,238	2,618	2,916	2,992	2,975	2,976	32,556
1	데이터 표준 수립 및 품질관리 기술 개발	2,217	3,099	2,955	3,366	2,539	2,053	2,287	2,346	2,332	2,334	25,529
2	연구데이터 저작권 관리 기술 개발	610	853	814	926	699	565	629	646	642	642	7,027

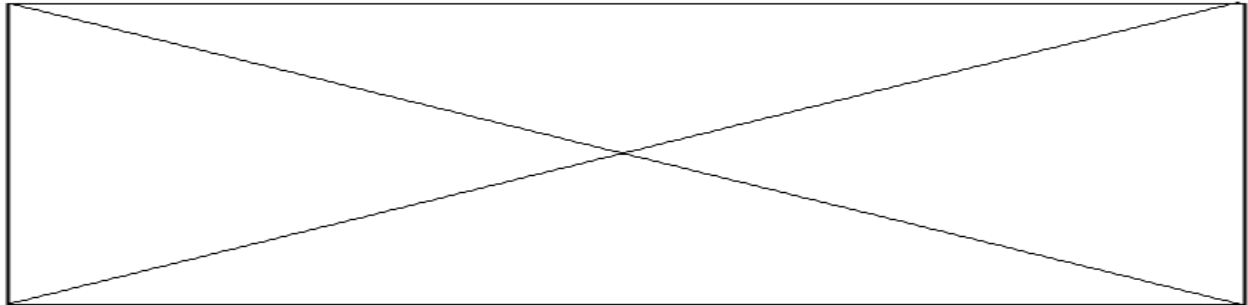
5.3.4. 지능형 연구데이터 검색 기술 개발

가. 개발 및 추진 목표

개발 목표	통합 검색 서비스제공 및 연구플랫폼 서비스 활성화를 위한 검색 기술 개발
세부 추진 목표	① 메타데이터 기반의 효율적인 검색 방안을 개발 ② 클라우드 서비스 항목(소프트웨어, 분석도구)에 대한 검색 기술 개발

□ 세세부 과제 도출

- 수집되는 메타데이터, 연구플랫폼의 클라우드 서비스 항목(소프트웨어, 분석도구)에 대한 활용성 확대를 위한 통합검색 기술 개발을 세세부 과제로 도출



<그림 5-19> 연구데이터 검색 기술 개발 세세부 과제 도출

□ 기술 확보전략

기술 구분	현재 기술수준	목표 기술수준	세세부 과제	세부기술 확보 전략		
				자체개발	기술도입	글로벌 협력
검색 프레임워크	TRL 3	TRL 6	연구데이터 검색 기술 개발	●	○	

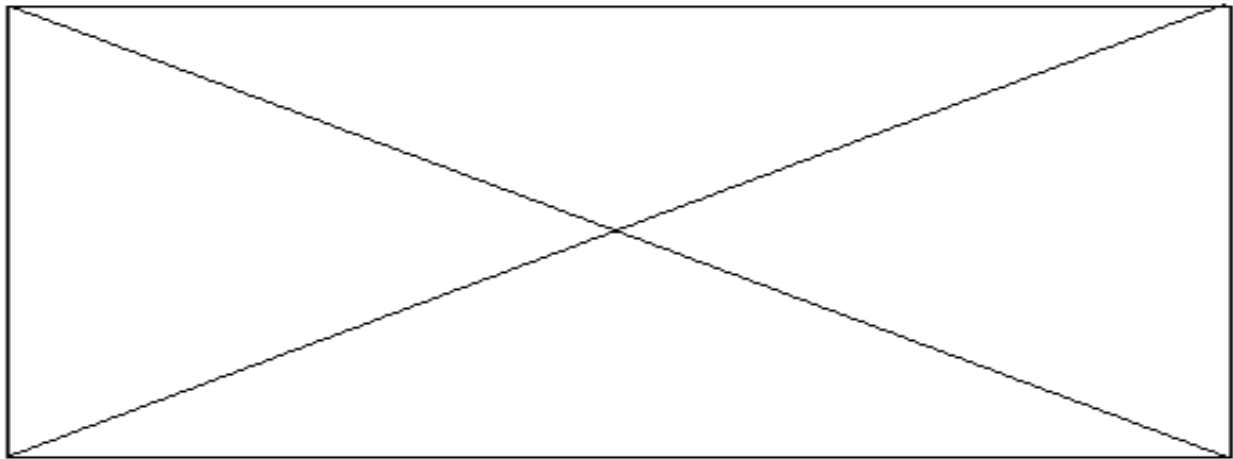
● : 주, ○ : 부

－ 오픈소스 분산 검색엔진 기반의 자체 개발

나. 핵심기술 구성 및 내용

□ 핵심기술 구성

- 연구데이터 검색 서비스는 사용자의 등록 연구데이터들에 편리하고 쉽게 조회할 수 있도록 하기 위해 등록된 메타데이터 기반으로 효율적인 검색 방안을 마련
- 연구데이터 플랫폼은 등록된 연구데이터 검색을 위한 검색 엔진 도입

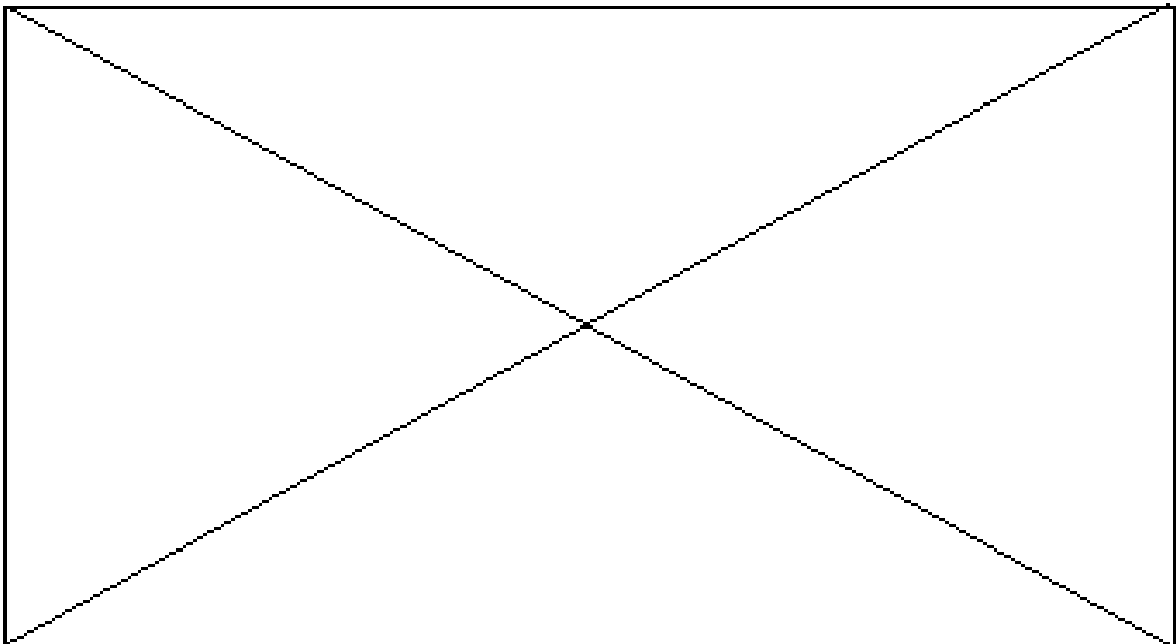


□ 핵심기술 내용

구분	기술	설명	비고
검색	메타데이터	• 일반 검색 키워드 검색 (기본 키워드 설정 검색) • DOI기반 검색 • 상세 키워드 및 필터 설정 검색	
	리포지토리	• 전체 리포지토리 조회 • 국가/기관/분야에 따른 리포지토리 조회 • API 및 조회 결과 XML Schema는 re3data 규격 준수	
	큐레이션	• 과학데이터 큐레이션 및 검색 기능의 Collection, Item, File에 대한 Link 클릭을 통한 다운로드 실행 • 리포지토리 검색은 re3data 웹사이트를 통한 개인 영역으로 파일 다운로드 진행	file download
솔루션	BigData 검색엔진 + YARN + Hadoop	• 하둡, YARN(Yet Another Resource Negotiator)과 검색 솔루션 연동 • 검색엔진과의 연동을 통해서 대량의 데이터를 인덱싱(Indexing)하여 필요한 때에 실시간으로 데이터를 분석하고 검증	solution

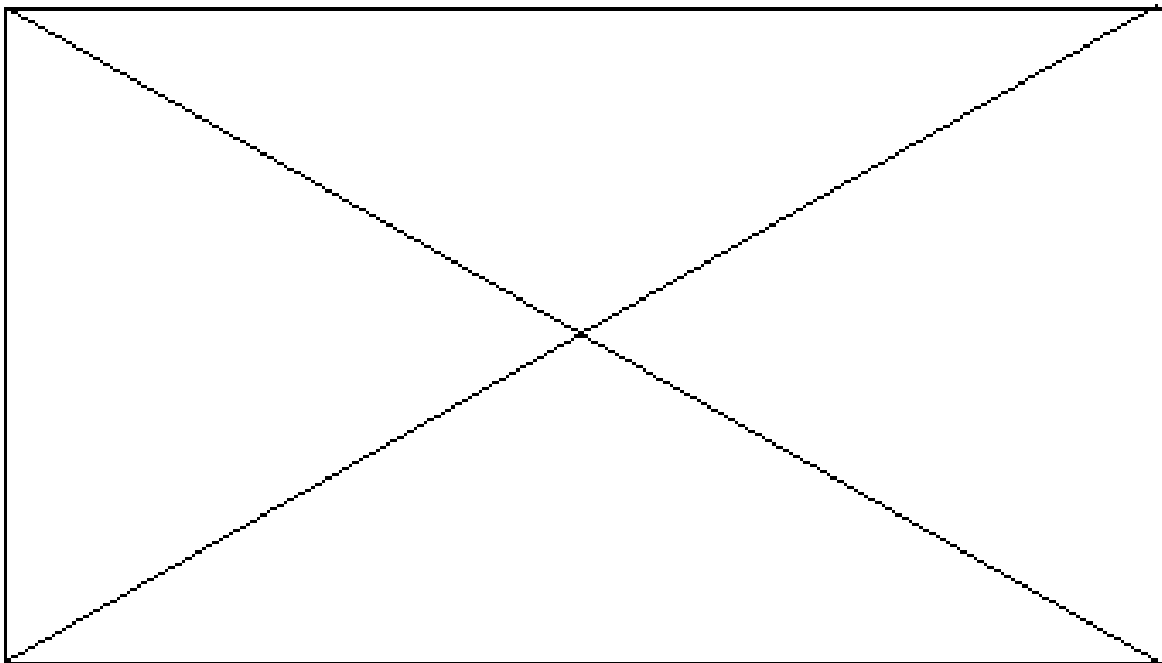
구분	기술	설명	비고
	Apache Lucene	• 자바로 개발된 오픈소스 정보검색 라이브러리 • 루씬은 확장 가능한 고성능 정보검색(IR, Information Retrieval) 라이브러리임. 루씬은 소프트웨어 프로그램에 색인과 검색 기능을 간단하게 추가할 수 있도록 지원 - 루씬을 사용할 때 검색에 대한 전문적인 지식을 반드시 알 필요는 없음. 꼭 필요한 몇 가지 기본 클래스들을 사용하는 방법만 익히면 색인과 검색 기능을 직접 추가할 수 있음 - 루씬은 독립 프로그램이 아닌 단순한 소프트웨어 라이브러리 • 프로그램에 텍스트 색인과 검색 기능을 추가할 수 있도록 지원	InHouse / solution
	Apache Solr	• 자바 기반 오픈소스 엔터프라이즈 검색 플랫폼 • 전체 텍스트 검색, 히트 하이라이트, 패킷 검색, 실시간 인덱싱, 동적 클러스터링, 데이터베이스 통합, NoSQL 기능 및 풍부한 문서 (예 : Word, PDF) 처리 • 분산 검색 및 인덱스 복제를 제공 • 엔터프라이즈 검색 및 분석 사용 사례에 널리 사용	InHouse / solution
	ElasticSearch	• HTTP에서 JSON을 사용하여 데이터를 인덱스할 수 있고 다중 타입을 가지는 다중 Tenant를 지원하는 분산 검색 엔진 • 모든 종류의 문서를 검색, 가변 검색 및 실시간에 가까운 검색을 제공	InHouse / solution

□ ElasticSearch 엔진 기반의 검색 시스템 기술개발



<그림 5-20> ElasticSearch 검색시스템 개념도

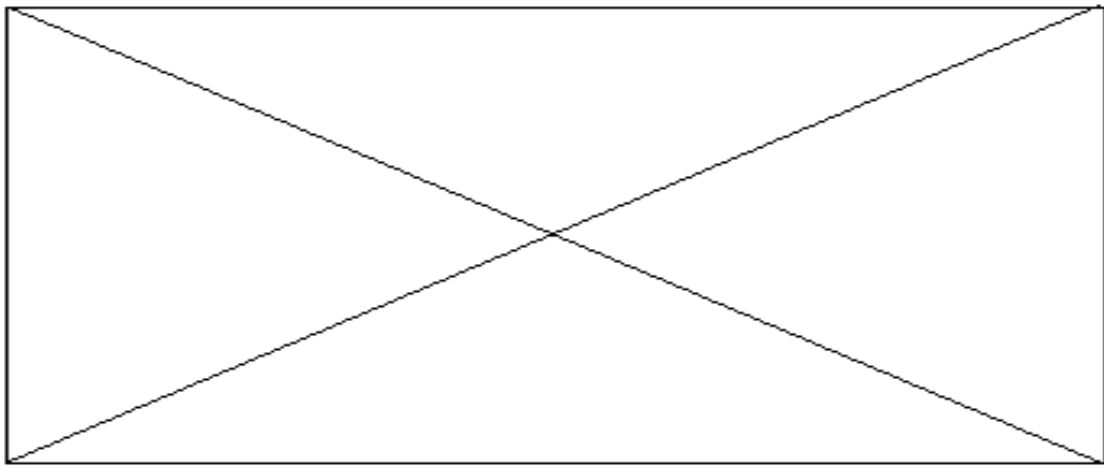
- \ ○ 클라우드 분석환경과 융합되는 아키텍처를 수립하여 연구데이터 분석 정보를 실시간 수집 검색에 활용
- Hadoop 데이터를 Elastic Stack으로 색인하여 고성능 Elasticsearch 엔진과 Kibana 시각화 기능의 이점을 활용하는 아키텍처 수립
- Hadoop을 활용하여 동적 임베디드 검색 응용 프로그램 구축하고, 전체 텍스트 검색과 geo 쿼리, 어그리게이션을 통해 보다 깊이 있고 지연 시간이 짧은 Hadoop 데이터 분석 아키텍처 수립
- Spark, Spark Streaming, SparkSQL과의 연동, Hive, Pig, Storm, Cascading, 표준 MapReduce 중 연구데이터 플랫폼의 분석환경 구성시 Elasticsearch 에서 색인을 하고 쿼리를 수행할 수 있는 기본 인터페이스를 제공하는 아키텍처 수립 (대용량 데이터 분산 검색, 대용량 데이터 실시간 색인, LSM기반 고속 분산 색인)



<그림 5-21> ElasticSearch + hadoop 개념도

□ 검색기술 기능 구성

- 연구데이터 검색 기술은 관리 영역의 색인관리, 검색관리와 상세검색 영역으로 구분하여 기능 구현



<그림 5-22> 검색 기능 구성도

다. 연차별 연구개발 계획

번호	플랫폼	1단계			2단계				3단계		
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
1	연구데이터 검색 기술 개발		설계 및 기술 개발	최적화 및 안정화	기능 고도화	정상운영					

라. 소요예산

□ 전체 예산 : 20,738백만원

단위 : 백만원

번호	플랫폼	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
	계	1,634	2,624	2,457	2,733	2,063	1,669	1,859	1,907	1,896	1,897	20,738
1	연구데이터 검색 기술 개발	1,634	2,624	2,457	2,733	2,063	1,669	1,859	1,907	1,896	1,897	20,738

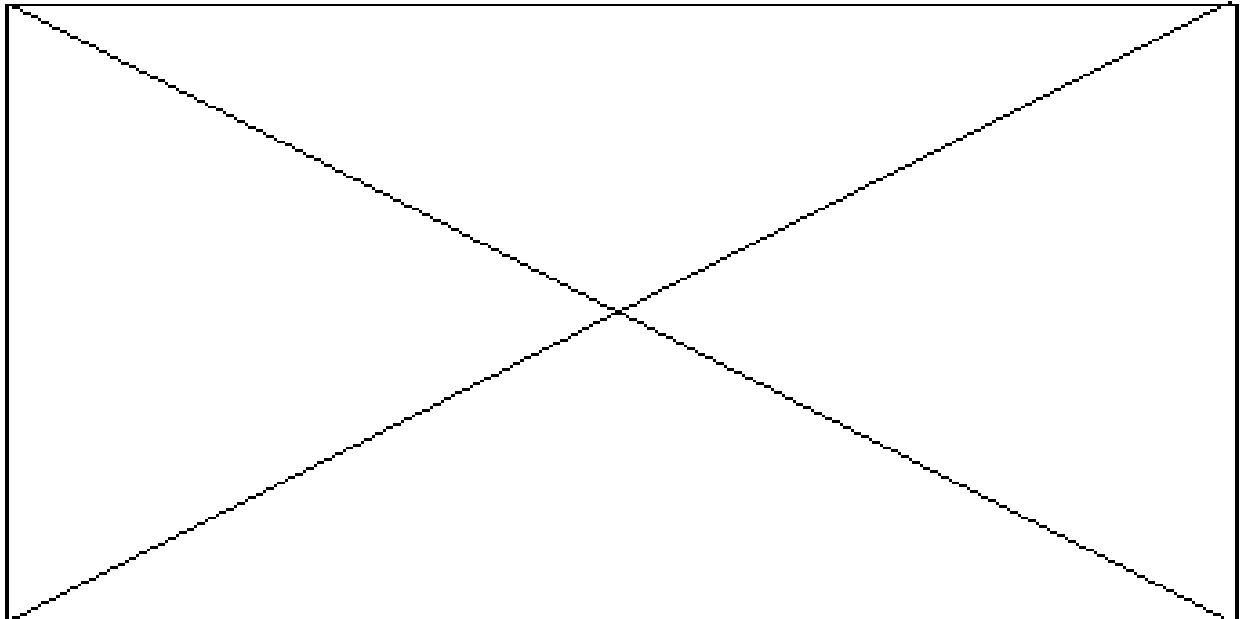
5.3.5. 연구데이터 융합·분석 환경 구축 기술

가. 개발 및 추진 목표

개발 목표	손쉬운 연구데이터 분석 환경 제공을 위한 융합·분석 환경 개발
	
세부 추진 목표	① 다양한 BI, AI, ML, DL솔루션을 연구플랫폼에 탑재 ② 분석 솔루션을 workflow기반으로 활용할 수 있도록 구축

□ 기술 도출

- 연구플랫폼의 클라우드 서비스의 사용성을 높이기 위하여 클라우드기반의 빅데이터 분석 환경 제공을 위한 workflow , BI, AI, ML, DL, 가시화 기반 기술 개발을 세세부 과제로 도출



<그림 5-23> 연구데이터 융합·분석 기술 개발 세세부 과제 도출

□ 기술 확보전략

기술 구분	현재 기술수준	목표 기술수준	세세부 과제	세부기술 확보 전략		
				자체개발	기술도입	글로벌 협력
융합·분석 프레임워크	TRL 3	TRL 6	Workflow 서비스 기술 개발	●	○	
			Batch & Realtime 분석 기술 개발	●	○	
			BI 서비스 기반기술 개발	●	○	
			AI 서비스 기반기술 개발	●	○	
			ML 서비스 기반기술 개발	●	○	
			DL 서비스 기반기술 개발	●	○	
			가시화 기반 기술 개발	●	○	

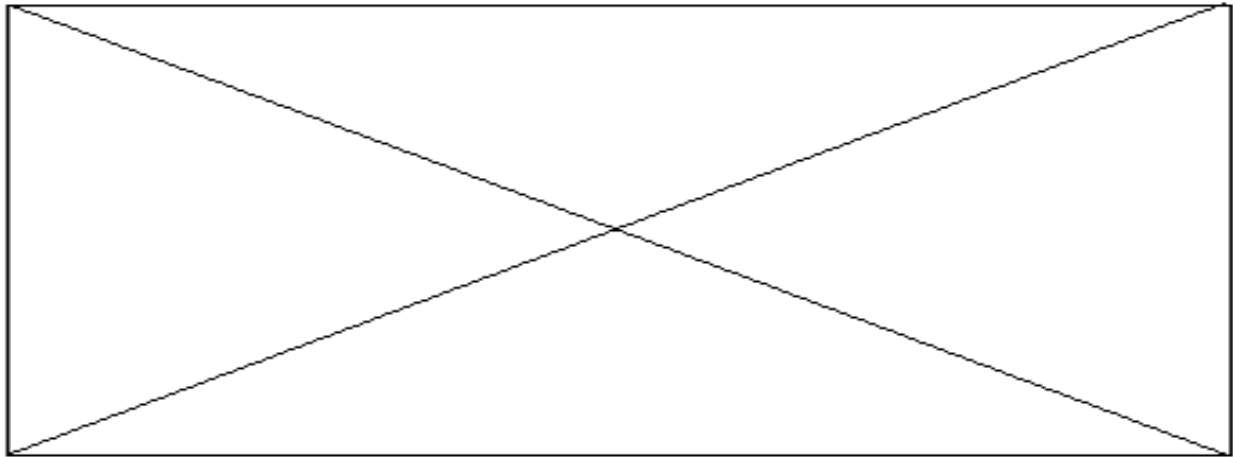
● : 주, ○ : 부

- 오픈소스 분석 기술 기반의 자체 개발

나. 핵심기술 구성 및 내용

□ 핵심기술 구성

- 클라우드 환경에서 연구데이터를 실시간분석, 데이터 분석, 지능형 분석과 결과를 시각화 기술을 기반으로 분석 프레임워크 구성



□ 핵심기술 내용

구분	기술	설명	비고
분석 처리	실시간분석	- 특정 사용자 메시지, 특정 키워드가 아닌 전체 메시지에 대한 분석과 데이터 수집과 동시에 제한된 분석 수행 - 분석 엔진이 수집기에 내장되어 어답터 역할과 동시에 분석을 수행 - 수집기에서 분석 툴 관리와 분석 결과 데이터에 대한 정합성을 제공	
	지능형분석	- 빅데이터를 다룰 때에 정보들 간의 관계 및 의미 정보를 기계가 처리할 수 있는 형태로 표현된 데이터로부터 정보들 간의 관계와 의미 정보의 표현을 메타데이터로 추출 - 추출된 데이터를 데이터 분석 및 검색 등의 처리과정에서 활용 - 메타데이터를 시맨틱 애노테이션으로써 온톨로지를 구축	
	통계처리	- 통계 기술로 수집 및 분석된 데이터를 통계적인 분석 결과로 추출 - 각종 로그 정보와 클릭스트림 등의 정보를 검색 및 수집 - 새로운 의미 정보를 추출하기 위해 텍스트 마이닝, 자연어 처리 기능 제공	
	시각화	- 평면 데이터와 입체 데이터로 표현하는 기능 제공 - 시각화된 데이터에 의해 새로운 분석 결과가 도출 가능하도록 데이터 조직화, 데이터 차별화, 인-메모리 기술 활용 - Spatial Information Flow, SClustergram, History Flow - Google Chart, Plot, Raphael, D3, Highcharts, Processing	
	머신러닝	- 컴퓨터 혹은 기계가, 입력된 데이터를 바탕으로 자동적으로 혹은 반자동적으로 학습할 수 있게 해주는 알고리즘 - 인터넷 정보검색(web mining), 텍스트마이닝(text mining), 자연어 처리, 음성인식, HCI(Human-Computer Interface) 등이 머신 러닝을 지원 - 애플의 Siri, IBM Watson, IBM SystemML, Hadoop의 Mahout	
TOOL ML 분석	R	- 통계 기술을 활용한 분석 언어 - 가시화(visualization) 기능을 제공하는 개발환경으로 고도의 기능을 제공하는 Open Source - 통계뿐만 아니라, 객체지향 언어 특성과 함수형 언어 특성을 모두 지원	open source
	Action Vector	- 빅데이터 고속분석용 DBMS 엔진 - 초대용량 분석용 데이터의 저장 - RDBMS로서 표준 SQL을 통한 분석결과 검출 및 제공	open source
	Elastic Search	- 분산검색엔진 - 분산구조로 데이터를 저장하고, 안정성과 성능을 최대화	open source

구분	기술	설명	비고
		하여 해당 검색 결과를 도출	
	BI, Visualization Too	- 데이터 분석결과 표현 및 시각화 - 서비스 UI를 위한 다양한 데이터 표현, 그래프 표현, 다양한 레포팅 기능을 제공	open source/상용
	Hive, Pig, SPARK	- 빅데이터 집계/분석/처리 - 대용량 데이터의 분석 지원 - 데이터 흐름 기반의 스크립트 프로그래밍 언어	open source
	Mahout	MapReduce 패러다임을 사용하여 하둡 상에서 실행되는 클러스터링, 협업 필터링 및 분류를 위한 알고리즘	open source
	MapReduce	분산 컴퓨팅을 지원하기 위한 소프트웨어프레임워크	open source
	Pig	데이터-흐름 기반의 스크립트 프로그래밍 언어	open source
	pandas	파이썬 데이터 분석 라이브러리	open source
	Oozie	하둡의 Workflow 스케줄러 시스템	open source
	Apache Avro	오픈소스 데이터 직렬화 프레임워크. 이기종간 데이터 타입을 교환할 수 있는 체계를 제공, 다양한 데이터 구조 지원	open source
	Apahce Giraph	빅데이터의 그래프 분석 처리 시스템	open source
	Apache Hama	아파치 Top 레벨 프로젝트로, 대규모 매트릭스와 그래프 연산을 제공하는 프레임워크.	open source
	Apache Impala	하둡기반 실시간 SQL 쿼리 엔진	open source
	Apache Sqoop	RDBMS와 아파치 하둡간의 대용량 데이터들을 변환하여 주는 Command-Line Interface 애플리케이션. 데이터 커넥터와 일치하지 않는 데이터 포맷을 사용할 수 있도록 지원	open source
	Apache Tajo	빅데이터 분석엔진. 하둡 기반의 대용량 데이터웨어하우스 시스템	open source
	Cascading	하둡에 데이터 어플리케이션을 만들기 위해 사용하는 개발 플랫폼	open source
	Gephi	1백만 노드 규모의 네트워크와 다양한 크래프 타입을 지원하는 시각적 분석 도구	open source
	HPCC(High-Performance Computing Cluster)	대용량 병렬처리 플랫폼	open source

□ PYTHON , SCALA, R 언어별 라이브러리 지원 현황

- 데이터 과학 및 기계 학습 솔루션을 구현을 위해 가장 많이 사용하고 있는 PYTHON, SCALA, R의 세 언어를 가장 많이 지원하는 분석 도구를 파악하여 연구데이터 플랫폼에 구축 시 활용
- Machine Learning : 기계 학습 알고리즘을 구현하고 워크플로우를 작성하며 일반적으로 기계 학습 문제를 해결, 서로 다른 분류, 회귀 및 기타 문제에 대한 기본 툴킷 제공

구분	PYTHON		SCALA		R	
	TOOL	libraries	TOOL	libraries	TOOL	libraries
1	TensorFlow	34,689	Deeplearnin g4j	22,938	H2O	22,947
2	Learn	22,811	Spark	22,073	mlr	4,052
3	PyTorch	11,763	PredictinIO	4,383	XGBoost	3,296
4	Keras	4,591	bigDL	2,358	caret	1,741

- Visualization : 데이터를 보다 잘 이해하고 해석하는 데 필요한 시각적으로 데이터를 표시, 다양한 시각화 차트와 표현을 위한 다양한 옵션을 포함

구분	PYTHON		SCALA		R	
	TOOL	libraries	TOOL	libraries	TOOL	libraries
1	matplotlib	25,984	VEGAS	215	ggplot2	4,070
2	Bokeh	17,066	Breeze-viz	29	plotly	3,185
3	plotly	2,910			ggvis	2,159
4	Seaborn	2,093			DataTbales	1,921

- Mathematics & Engineering : 수학 및 공학용 라이브러리는 수치 데이터를 편리한 형식으로 저장하고 복잡하고 진보 된 수학 연산 및 과학 계산을 수행 할 수 있는 기능을 제공. 또한 패키지는 텍스트 및 콘텐츠와 같이 더 복잡하게 해석되는 데이터를 처리하는 데 사용

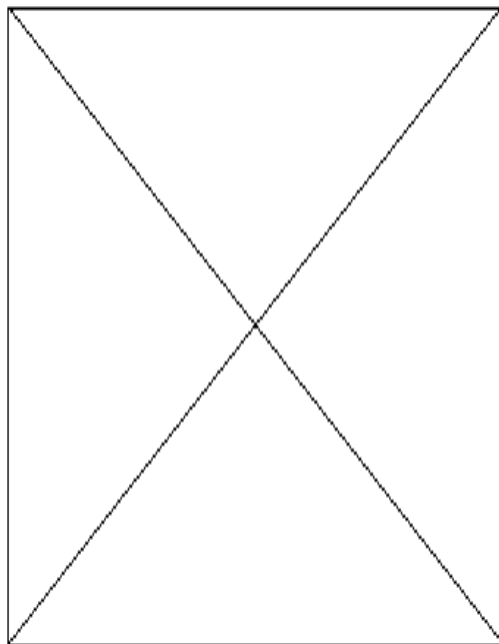
구분	PYTHON		SCALA		R	
	TOOL	libraries	TOOL	libraries	TOOL	libraries
1	SciPy	19,255	akka	22,298	—	—
2	NumPy	18,162	kafka	5,145	—	—
3	Natural Language ToolKit	13,053	Breeze	3,431	—	—
4	SM	10,244	Epic	1,790	—	—

- Data Manipulation & Analysis : 데이터 과학의 필수적인 부분 인 데이터 조작 및 분석 필드 는 데이터 스크래핑, 처리, 클리닝, 사전 처리 및 기타 작업을 수행하여 데이터를 재생할 수있는 라이브러리를 나타내며 결과적으로 분석 자체를 수행

구분	PYTHON		SCALA		R	
	TOOL	libraries	TOOL	libraries	TOOL	libraries
1	pandas	17,243	Slick	2,037	dplyr	4,990
2	Scrapy	6,643	ADAM	1,775	data.table	3,416
3					lubridate	1,501
4					jsonlite	914

- Reproducible Research : 재현 가능한 연구를 위한 패키지 는 코드, 데이터 및 내용을 결합하는 문서를 만드는 아이디어를 구현. 기본적으로 도움을 받아 프로젝트를 즉시 출판 할 수 있는 새로운 작업 수행

구분	PYTHON		SCALA		R	
	TOOL	libraries	TOOL	libraries	TOOL	libraries
1	—	—	—	—	knitr	5,528
2	—	—	—	—	R Markdown	2,373
3					slidify	302



<그림 5-24> top python, scala, r libraries in data science

다. 연차별 연구개발 계획

번호	플랫폼	1단계			2단계				3단계		
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
1	Workflow 서비스 기술 개발				설계	기술 개발	최적화 및 안정화	기능 고도화		정상운영	
2	Batch & Realtime 분석 기술 개발										
3	BI 서비스 기반기술 개발										
4	AI 서비스 기반기술 개발										
5	ML 서비스 기반기술 개발										
6	DL 서비스 기반기술 개발										
7	가시화 기반 기술 개발										

라. 소요예산

□ 전체 예산 : 33,923 백만원

단위 : 백만원

번호	플랫폼	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
계		2,612	4,029	3,854	4,500	3,566	2,851	3,104	3,190	3,106	3,111	33,923
1	Workflow 서비스 기술 개발	452	697	667	779	617	494	537	552	538	539	5,873
2	Batch & Realtime 분석 기술 개발	452	697	667	779	617	494	537	552	538	539	5,873
3	BI 서비스 기반기술 개발	452	697	667	779	617	494	537	552	538	539	5,873
4	AI 서비스 기반기술 개발	452	697	667	779	617	494	537	552	538	539	5,873
5	ML 서비스 기반기술 개발	328	506	484	565	448	358	390	401	390	391	4,261
6	DL 서비스 기반기술 개발	205	317	303	354	281	224	244	251	244	245	2,669
7	가시화 기반 기술 개발	270	416	398	464	368	294	320	329	321	321	3,502

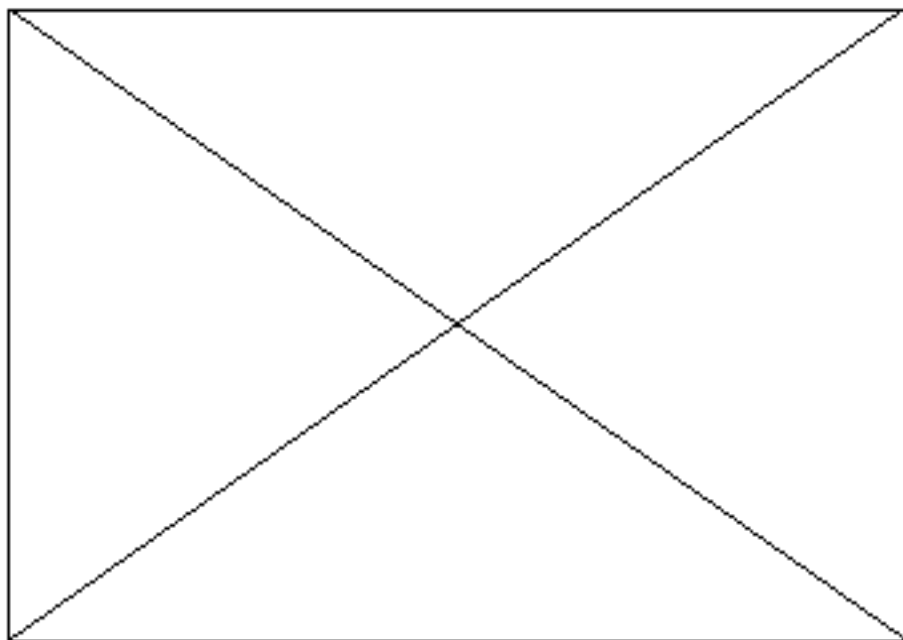
5.3.6. 연구데이터 활용·플랫폼관리 기술

가. 개발 및 추진 목표

개발 목표	24·365 무중단 연구데이터 활용을 위한 클라우드 서비스 기술 개발
세부 추진 목표	① 연구자간 협업 및 커뮤니티 기능 개발 ② 안정적인 클라우드 서비스 운영을 위한 기반 기술 개발

□ 기술 도출

- 연구플랫폼의 클라우드의 안정적인 서비스를 위하여 클라우드 어드민, 고가용성, 아키텍처, 보안, 가상머신 관리 기술개발을 세세부 과제로 도출



<그림 5-25> 연구데이터 활용·플랫폼 기술 개발 세세부 과제 도출

□ 기술 확보전략

기술 구분	현재 기술수준	목표 기술수준	세세부 과제	세부기술 확보 전략		
				자체개발	기술도입	글로벌 협력
품질/표준 프레임워크	TRL 3	TRL 6	협업 서비스 및 관리 기술 개발	●		
			커뮤니티 생성 및 관리 기술 개발	●		
			클라우드 고가용성 기술 개발	●	○	
			클라우드 어드민/아키텍처 관리 기술 개발	●	○	
			클라우드 보안 관리 기술 개발	●	○	
			클라우드 가상머신 관리 기술 개발	●	○	

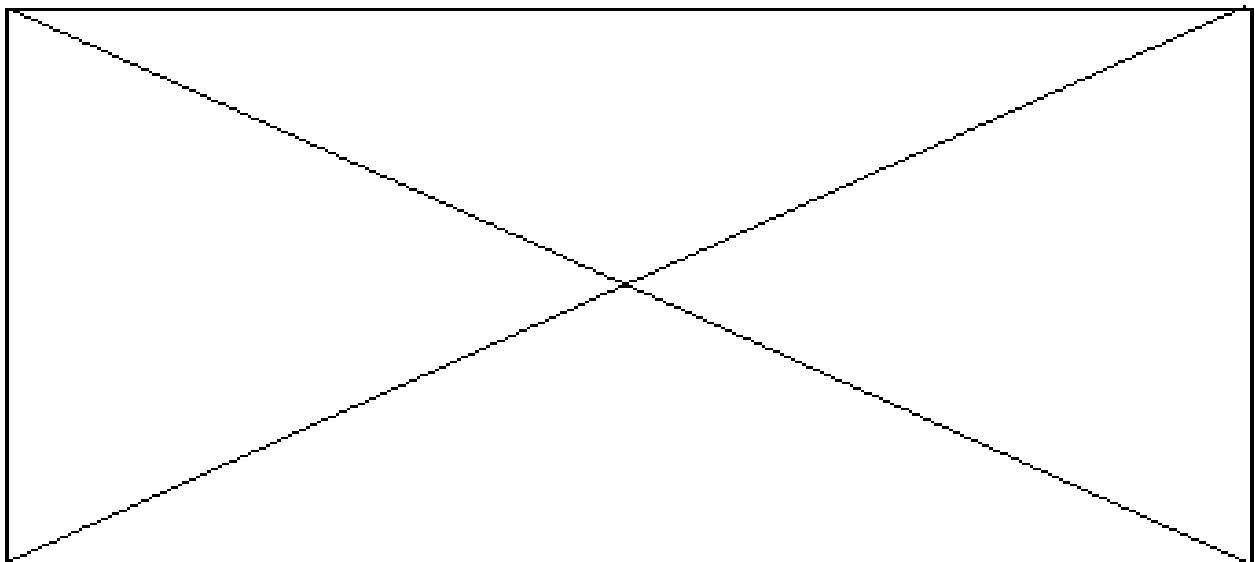
● : 주, ○ : 부

- 협업서비스, 커뮤니티 생성 기술은 자체 기술개발을 통해 구현
- 클라우드 서비스 관리 기술은 오픈소스 솔루션을 기반으로 기술 개발

나. 핵심기술 구성 및 내용

□ 핵심기술 구성

- 인프라 시스템(서버, 스토리지, 네트워크 등)에 대한 물리 자원 및 가상 자원 관리를 위한 기능 기술을 개발
- 물리적 자원을 개발, 실행 및 운용을 위해 가상 자원 관리 기능들을 기술 개발
- 관리환경은 플랫폼에서 제공하는 개발/서비스 프레임워크, 분산 프레임워크, 메타데이터 프레임워크 및 공통 기능을 각각의 제공 서비스에 배포 및 관리하기 위한 기능 모듈로 구성



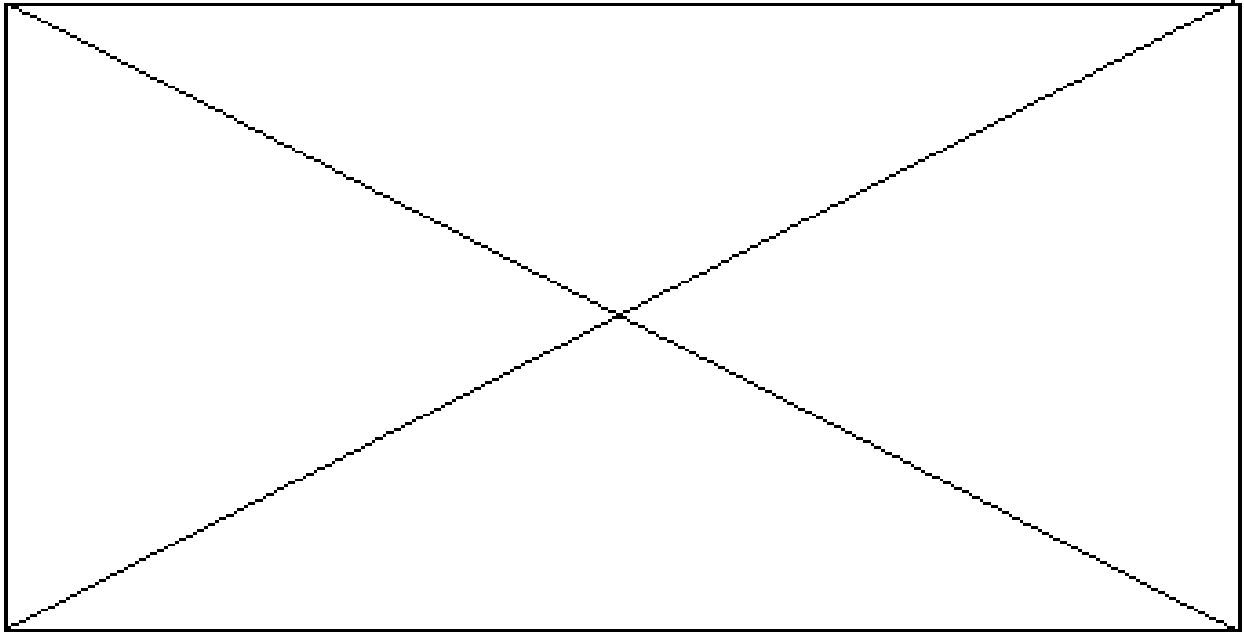
□ 핵심기술 내용

구분	기술	설명	비고
가상화 관리	Docker	• 리눅스컨테이너 가상화 기술. SW컨테이너 내에 애플리케이션 배포를 자동화하는 공개SW	open source
	KVM	• x86, x86_64, S/390, PowerPC, IA64 등 아키텍처와 리눅스, 솔라리스, BSD, 윈도우 등의 게스트 OS를 지원하는 가상화 공개SW 솔루션	open source
	VirtualBox	• Windows, Mac, Linux 등 다른 OS를 함께 실행시킬 수 있는 가상OS 구동 프로그램	open source
	Xen project	• 다양한 아키텍처 지원 하이퍼바이저	open source
	Libvirt	• 리눅스 가상화를 위한 API로 Xen, KVM 등 다양한 하이퍼바이저와 QEMU 및 일부 가상화 제품을 지원	open source
	Ganeti	• 소규모 가상화 네트워크 관리 툴	open source
	Open vSwitch	• 가상화 관리 솔루션	open source
모니터링	Nagios	• 네트워크 서비스 모니터링 (SMTP , POP3 , HTTP ,NNTP , ICMP , SNMP , FTP , SSH)	open source
	Hyperic	• 물리적, 가상 및 클라우드 환경 전체 모니터사용자 지정 웹 응용 프로그램	open source
데이터 테이블 및스토리지 관리	Hcatalog	• 하둡을 사용하여 생성된 데이터 테이블 및 스토리지관리 서비스	open source
분산 관리 시스템	Zookeeper	• 분산 환경에서 노드들 간의 분산 동기화, 그룹서비스, 공유, 락 등을 관리하는 시스템	open source
분산 로그관리 시스템	Flume	• 대용량 로그데이터를 중앙집중화된 데이터 저장소로 효율적으로 수집, 통합, 저장하기 위한 시스템	open source
	Scribe	• 여러 서버로부터 실시간으로 스트림되는 로그데이터를 종합 처리	open source
	Chukwa	• 분산 환경에서 서버의 로그 데이터를 수집하고 그 데이터를 저장하고 분석하는 분산 시스템	open source
분산 캐싱 시스템	Memcached	• 분산 메모리 캐싱 시스템, 데이터베이스 부하를 줄여 웹 애플리케이션 성능향상도움	open source
	couchbase	• 분산 메모리 캐싱 시스템, Membase Server 1.7 이후 버전은 Couchbase Server 1.8로 변경	open source
분산데이 터 관리시스 템	Cloudata	• 수집 테라바이트 이상의 데이터를 관리목적데이터를 여러 대의 서버에 분산 배치하여 관리	open source

구분	기술	설명	비고
빅데이터 관리	Apache Ambari	• Hadoop에 대한 웹기반의 설치, 관리, 모니터링 툴.	open source
	Flamingo	• 빅데이터 기반 애플리케이션을 쉽게 개발하기 위한 맵-리듀스 워크플로우, 하둡 클러스터와 HDFS 관리시스템을 웹 클라우드 IDE상에서 제공	open source
	Hue	• 하둡을 위한 Web 인터페이스. 웹 브라우저를 통한 시스템 접근 및 파일 시스템 열람, 사용자 계정 생성 및 관리, 클러스터 모니터링 등 다양한 기능 제공	open source
클라우드 관리	Juju	• 클라우드 인프라 관리 솔루션	open source
	Scalr	• 클라우드 관리 플랫폼	open source
	CloudForms	• 하이브리드 클라우드를 IaaS 클라우드로 구축 및 관리 기능을 제공하는 프레임워크	open source
클라우드 플랫폼	Apache jclouds	• 오픈소스 멀티 클라우드 Toolkit	open source
	Apache Stratos	• PaaS 프레임워크	open source
	Cloudify	• 파이썬으로 만들어진 TOSCA(Topology andOrchestration Specification for Cloud Applications)기반의 클라우드 PaaS 솔루션	open source
	Nimbus	• WSRF -based 또는 Amazon EC2 WSDL web serviceAPI를 통해 사용자에게 IaaS 클라우드를 제공하는 오픈소스 툴킷	open source
	OpenNebula	• IaaS 클라우드를 구성하기 위하여 데이터 센터의 가상 인프라스트럭처를 관리하는 가상화 솔루션	open source
	OpenQRM	• 데이터센터 관리시스템 -모니터링, 고가용성, 클라우드 컴퓨팅, 여러 가상화 기술을 지원	open source
	OpenStack	• Rackspace와 NASA 가 주축이 되어 시작한 프로젝트로, 서버, 스토리지, 네트워크, 가상화를 묶어서 제어하고 운영하기 위한 클라우드 플랫폼	open source
	OpenShift	• 표준 컨테이너 기반의 PaaS 및 클라우드 운영 플랫폼	open source
	Osv	• 클라우드 환경에 맞게 고안된 오픈소스 운영체제. Hypervisor 기반의 클러스터 배포 관리를 지원	open source
	Eucalyptus	• NASA에서 사용한 서버, 스토리지, 네트워크 인프라를 클라우드로 통합 관리하는 클라우드 플랫폼	open source
	CloudStack	• 세련된 GUI, 웹기반의 관리콘솔, 다양한 하이퍼바이저와의 연동, 소프트 방화벽, 로드밸런싱 기능이 있는 클라우드 플랫폼	open source
	Cloud Foundry	• 클라우드 플랫폼 서비스 솔루션	open source

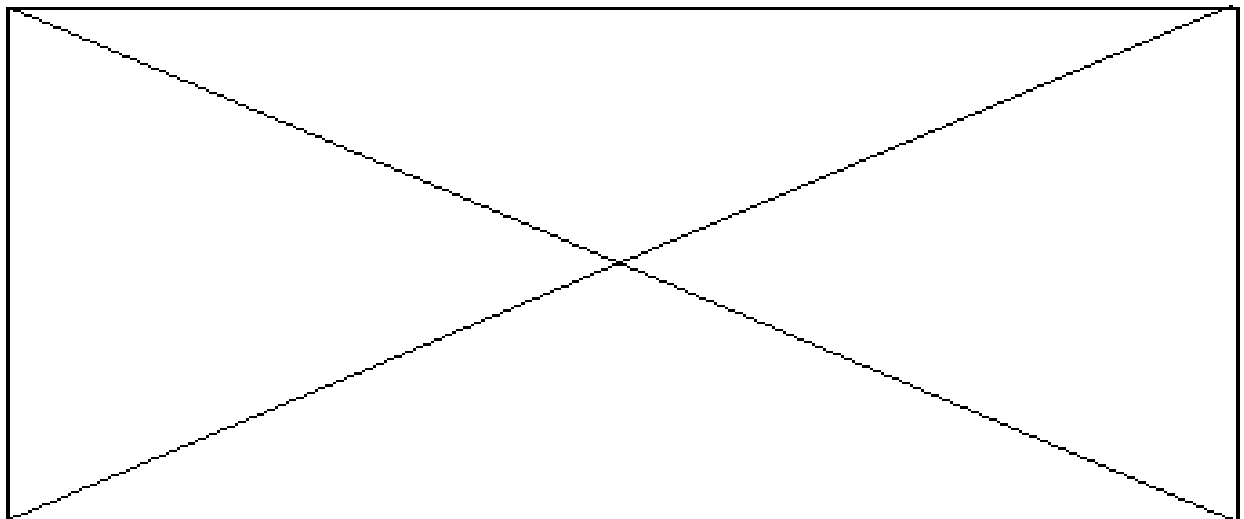
□ 관리시스템 기능구성

- 시스템관리 구성은 물리 시스템 자원관리, 가상 시스템 자원관리로 구성



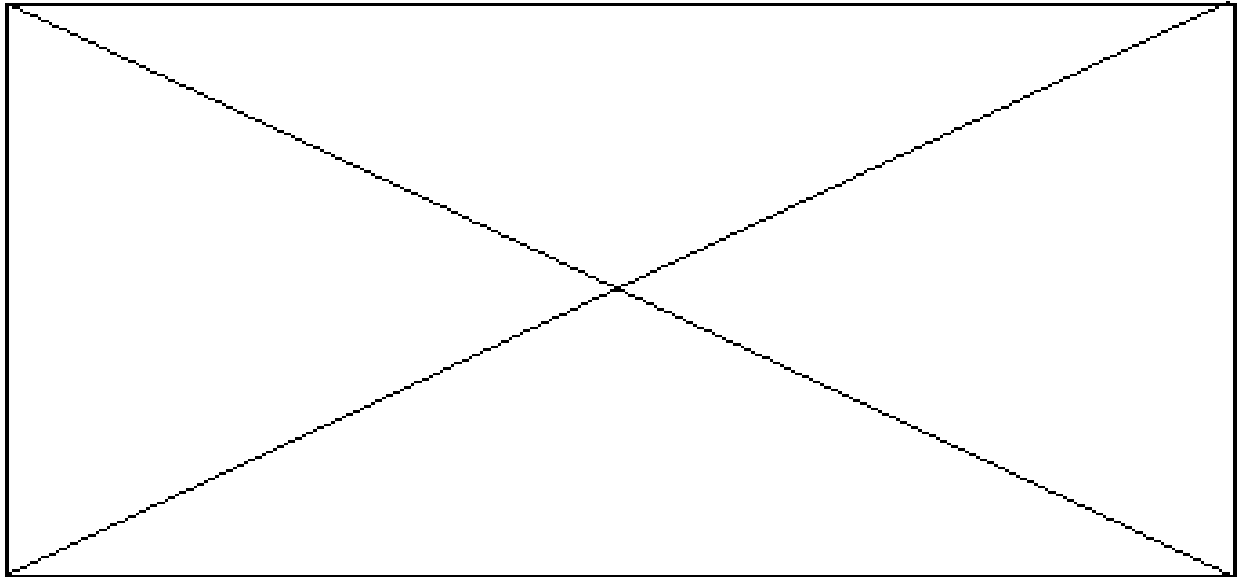
□ 모니터링 및 서비스관리 기능구성

- 서비스 성능 현황을 모니터링, 서비스 장애 현황을 모니터링, 운영 환경을 유연하게 확장 또는 축소할 수 있는 기능, 단위 서버 및 컴포넌트 단위 수준까지의 모니터링, 대쉬보드 기능, 이력정보 관리, 가시화 기능으로 구성



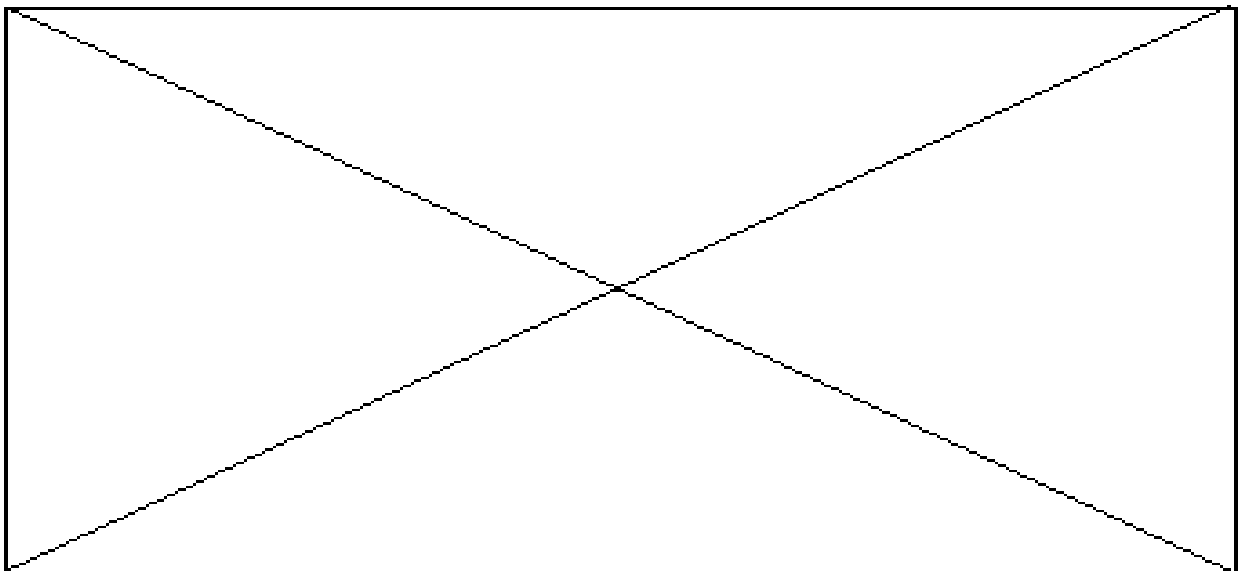
□ 클라우드 운영관리 기능구성

- 플랫폼관리 구성은 SW 현황관리, SW 변경관리, SW 형상관리, SW 배포관리, SW 라이선스 관리, 플랫폼 개발 환경 관리, 플랫폼 실행 환경 관리로 구성

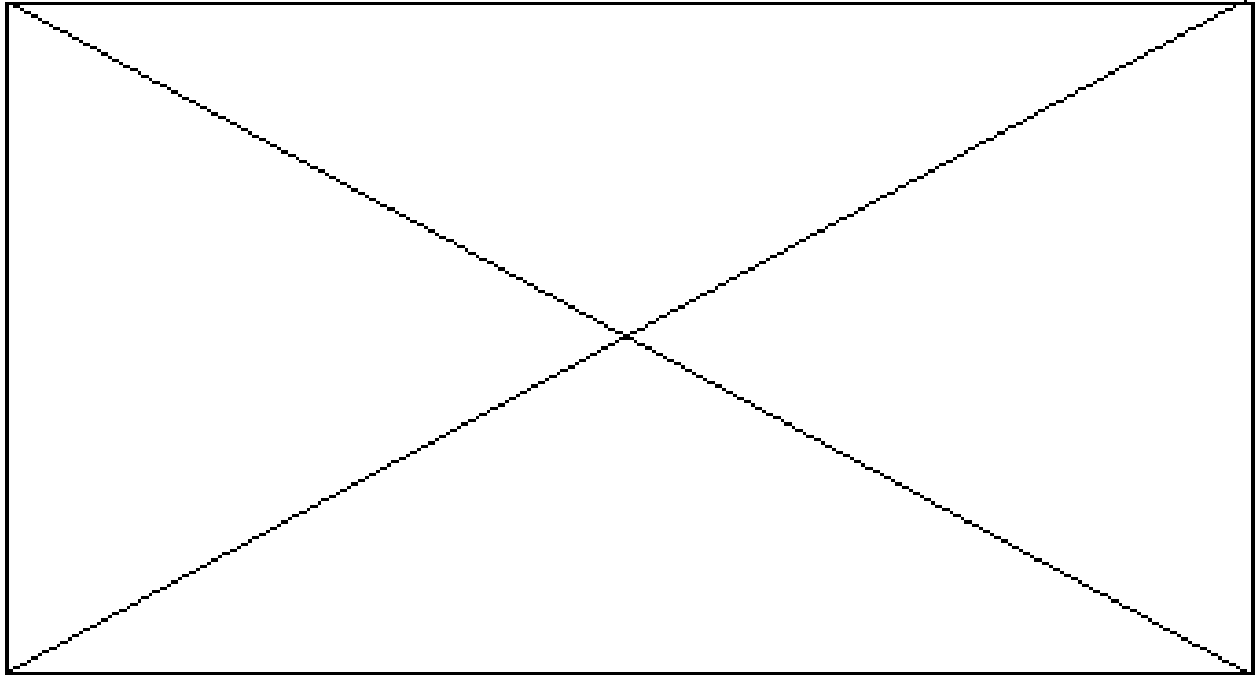


□ 클라우드 제어관리 기능구성

- 물리적 하드웨어와 소프트웨어의 운영을 위한 가상 환경 프로비저닝 기능

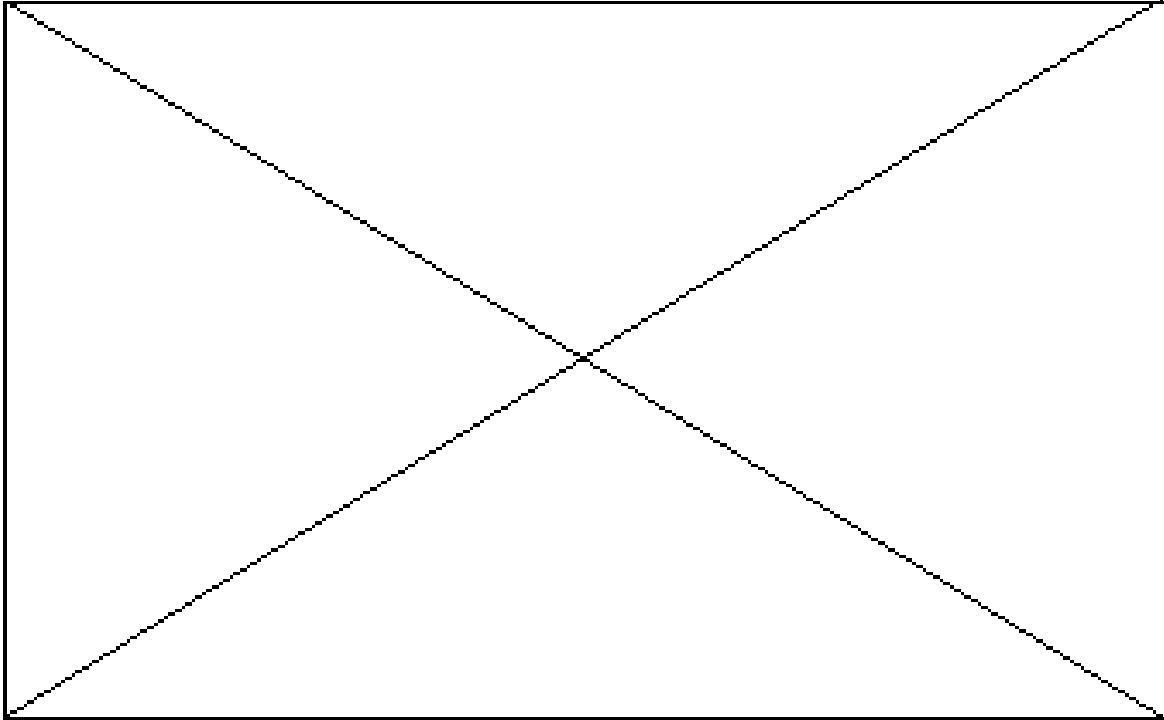


□ 클라우드 IaaS 아키텍처(Open Stack)



- 모든 클라우드 서비스들은 API 요청, 선처리 및 서비스 다른 파트로 전달하는 적어도 하나의 API프로세스를 가지고 있음
- 한 서비스의 프로세스 간 통신을 위해 AMQP 메시지 브로커(RabbitMQ, Qpid) 사용, 서비스의 상태는 데이터베이스(MySQL, MariaDB, 및 SQLite 등)에 저장
- 사용자는 클라우드 서비스 Dashboard나 명령어-라인 클라이언트로 구현된 웹 기반 사용자 인터페이스를 경유해서 클라우드 서비스에 접근

□ 클라우드 PaaS 아키텍처



- PaaS의 개발환경은 기존 IaaS에서 개발용 가상서버를 제공받아 PaaS 개발환경을 구성
- 개발환경과 운영환경에서 생성할 플랫폼 타입과 환경 구성 템플릿으로부터 플랫폼을 생성
- 생성된 플랫폼에 시험할 어플리케이션 바이너리를 등록
- 개발환경에서 검증한 어플리케이션 바이너리를 운영환경 구성 후 배포해서 운영

다. 연차별 연구개발 계획

번호	플랫폼	1단계			2단계				3단계		
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
1	협업 서비스 및 관리 기술 개발	설계	기술 개발	최적 화 및 안정 화	기능 고도화						정상운영
2	커뮤니티 생성 및 관리 기술 개발										
3	클라우드 고가용성 기술 개발										
4	클라우드 어드민/아키텍처 관리 기술 개발										
5	클라우드 보안 관리 기술 개발										
6	클라우드 가상머신 관리 기술 개발										

라. 소요예산

□ 전체 예산 : 42,101백만원

단위 : 백만원

번호	플랫폼	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
	계	3,402	5,326	4,948	5,508	4,190	3,387	3,772	3,871	3,848	3,850	42,101
1	협업 서비스 및 관리 기술 개발	905	1,416	1,316	1,464	1,114	900	1,003	1,029	1,023	1,024	11,195
2	커뮤니티 생성 및 관리 기술 개발	939	1,471	1,366	1,521	1,157	935	1,042	1,069	1,062	1,063	11,625
3	클라우드 고가용성 기술 개발	467	732	680	757	576	465	518	532	529	529	5,785
4	클라우드 어드민/ 아키텍처 관리 기술 개발	467	732	680	757	576	465	518	532	529	529	5,785
5	클라우드 보안 관리 기술 개발	234	366	340	378	288	233	259	266	264	264	2,892
6	클라우드 가상머신 관리 기술 개발	390	610	567	631	480	388	432	443	441	441	4,821

5.4. 기대효과

□ [과학기술적 측면]

- 데이터 중심의 지능형 연구환경 구축
 - 데이터에 근거한 분석-예측-최적화를 기반으로 과학기술 분야에 대한 통찰력(Insight)을 제공함으로써, 경쟁력 높은 과학기술을 확보
 - 연구 빅데이터 경쟁력 함양을 위한 국가적, 사회적 기반을 정비하고, 산·학·연 협력을 통한 원천기술 개발로 빅데이터 핵심 기반 기술 확보
 - 연구데이터 고성능 분석 애플리케이션 및 분석환경을 제공함으로써, 과학기술 내 방대한 양의 데이터로부터 가치 있는 정보를 확보
- 연구에 필요한 시간과 비용을 절감하여 국가 R&D 효율성을 제고
 - 연구자가 필요로 하는 데이터를 원스톱 검색, 데이터 생산, 전송, 재사용에서의 중복성 및 비용 감소
 - 국가연구데이터플랫폼의 고성능 분석 인프라 및 다양한 분석 S/W 등 자원의 공동활용으로 투자 비용 절감
- 연구의 효율성 및 투명성과 연구결과물의 수준 제고
 - 다양한 연구자들이 자유롭게 참여해 연구결과의 폭넓은 재연(replication)과 입증을 통해 연구의 투명성과 연구의 질 제고
 - 국내 연구데이터 및 메타데이터 표준화를 통한 관리·분석의 효율성 증대

□ [사회·경제적 측면]

- 다양한 주제 분야들을 결합한 융합 연구 활성화
 - 연구 빅데이터를 활용함으로써, 연구자간 데이터 생산과 유통을 통한 공유, 고부가가치 지식의 범용적 활용이 가능
 - 타 연구자의 데이터를 활용함으로써 다양한 연구의 시도가 가능하고, 다양한 데이터와 커뮤니티를 중심으로 협업연구 기회 증가
 - 온라인 커뮤니케이션과 협력 데이터 분석을 촉진하는 소프트웨어 툴을 활용하여 다양하고 폭넓은 협업 연구의 가능성이 증가

- 국가 차원의 이슈 대처 및 글로벌 협업을 통한 국가의 위상 제고
 - － 연구데이터플랫폼 활성화를 통해 기후, 환경, 건강한 삶 유지와 수명 연장 등의 글로벌 연구 이슈에 신속한 대처 가능
 - － 플랫폼 세계 유명 연구기관들과의 협업 및 공동 연구를 통해 세계 시장에서 글로벌 기업/연구소보다 경쟁력 높은 기술을 제공

5.5. 연구개발 대상 기술 선정과정

□ 연구데이터플랫폼 분과위원회 운영계획 수립(2018. 06. 01)

- ‘연구데이터 공유 활용체제 구축 운용 사업’의 추진을 위한 분과별 기술위원회 구성
- 산·학·연 연구데이터 및 플랫폼 기술 관련 전문가 10명 내·외로 위원회를 구성하고, 연구데이터플랫폼에 대한 수요 및 요소기술에 대한 의견 수렴
- 분과위원회 역할
 - － 플랫폼 개발 및 운영에 대한 방향성 제시
 - － 플랫폼 수요 및 요소기술에 대한 토론 및 의견 수렴
 - － 연구데이터 공유·활용체제 구축사업 기획 협력

□ 1차 분과위원회 전문가 검토(2018. 06. 08)

- 분과위원회 전문가를 통해 연구데이터플랫폼의 개발 및 운영, 수요 및 요소기술 관련 의견 수렴

□ 2차 분과위원회 전문가 검토(2018. 08. 16)

- 분과위원회 전문가를 통해 ‘①국가연구데이터플랫폼의 방향과 역할, ②국가연구데이터플랫폼의 기능 구성(안)’에 대한 의견 수렴
 - － 국가연구데이터플랫폼 구축 방향과 역할
 - 사업의 비전 및 목표
 - 연구데이터 정의 및 관리 주체
 - 국가연구데이터 관리를 위한 연계 개체
 - － 국가연구데이터플랫폼의 기능구성
 - 해외 주요 연구데이터 관리시스템 기능 현황
 - 국가연구데이터플랫폼 구축 목적
 - 국가연구데이터플랫폼 기능(안)

□ 3차 분과위원회 전문가 검토(2018. 09. 20)

- 분과위원회 전문가를 통해 ‘국가연구데이터플랫폼의 세부설계 방향’에 대한 의견 수렴
 - － 국가연구데이터플랫폼 세부설계
 - 플랫폼의 기능별 요소기술
 - 세부기술 및 세부과제 선정
 - 세부과제별 아키텍처 구성

대소분야 전문센터 구축·운영 지원

- 6.1. 과제개요
- 6.2. 추진전략
- 6.3. 바이오분야 전문센터 구축
- 6.4. 소재분야 전문센터 구축
- 6.5. 인공지능분야 전문센터 구축
- 6.6. 전문센터 우선분야 선정과정

6.1. 과제개요

6.1.1. 추진배경

□ 디지털 기술의 발전으로 데이터 기반의 연구 및 융·복합 연구 증가

- 첨단연구장비, 센서 및 데이터 처리 기술의 발달로 대용량 과학데이터의 지속적 생산
 - 데이터를 분석하여 새로운 이론이나 현상을 발견하는 데이터기반 4세대 연구패러다임* 등장
 - * 4세대 연구패러다임은 데이터 분석을 통해 새로운 이론이나 현상을 발견하는 데이터 기반 연구(Data-intensive Research)²⁵⁾
 - 실험 중심 연구(관측 및 관찰), 이론 중심 연구(모델링), 컴퓨팅 자원 중심 연구(컴퓨팅 시뮬레이션)에서 데이터 기반 연구(Data-intensive Research)로의 연구 패러다임 전환
- 분야 간 다양한 융합연구 및 국제공동연구 지속적 확대
 - 대용량 데이터를 개인 혹은 단일기관이 분석하고 연구하는데 한계 존재
 - 혁신적인 연구개발을 위한 융·복합 수요로 인해 데이터의 공유 및 이를 통한 협력 필요성 대두
 - 1990년 이후 과학기술 분야 단일 연구자의 논문은 지속적으로 감소하고 있는 추세이며, 대형 협업 연구는 지속적으로 증가²⁶⁾

□ 해외의 경우 데이터·컴퓨팅 기반 연구공동체(전문센터)를 지속적으로 육성

- 주요 선진국은 연구개발 과정에서 다수의 연구자가 연구데이터 공유·활동 수행 경험*
 - * 미국 46%, 영국 43%, 호주 41% 연구데이터 공유 경험(Wiley 2014)
- (미국) '95년부터 컴퓨팅 기반 교육·융합 연구 환경을 동적으로 지원하는 웹기반 전문센터로서 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 전문센터로서 가상실험실 플랫폼(Nectar) 운영('17년 12개)

25) T. Hey et al, The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, 2009

26) Adams, Jonathan. Collaborations: The rise of research networks. Nature 490.7420 (2012): 335-336. 2012

6.1.2. 필요성

□ 분야별 연구 데이터 공유·활용 생태계 미흡

- 데이터 분석을 통한 R&D에는 관심이 많으나 연구에 요구되는 샘플 수의 부족과 데이터 생성 비용 부담으로 연구 수행에 어려움
- 개인 연구자 및 소규모 연구소에서는 연구 데이터 확보에 어려움을 겪고 있을 뿐만 아니라 해당분야에 필요한 연구자의 니즈 파악이 어려움
- 특정 분야* 차원에서 일부 데이터를 수집·관리 중이나, 복잡한 절차 및 부족한 홍보로 인해 연구자들의 데이터 서비스 활용은 저조
 - * 유전체(국가생명자원정보센터), 소재 물성(소재정보은행) 등
- 데이터 획득과 활용을 위한 분석환경 및 인프라 제공 등 국가 차원의 분야별 연구데이터 공유·활용 지원체제 미흡

□ 공유·활용 기반의 연구로 투자효율성과 연구 투명성 확보 필요

- 연구데이터의 공유 및 재사용 방식의 연구수행으로 연구투자의 효율성 제고
 - 공개된 연구결과물(논문, 데이터, 방법론 등)을 재활용하여 새로운 과학적 발견을 위한 입증을 시도할 수 있기 때문에 불필요한 중복투자 없이 연구수행 필요
 - ※ 연구데이터 공개로 인한 데이터 가용성 서비스 투자 대비 2-10배의 수익을 창출할 것으로 추정²⁷⁾
- 공개적이고 자유로운 논의와 다양한 관점의 재현 및 입증을 통해 연구에 대한 정확한 검증 및 연구의 투명성이 요구됨
- 연구자 뿐 아니라 기업과 사회 구성원에게도 쉽게 연구데이터를 이용할 수 있는 공유 인프라를 기반구축으로 빠른 지식 확산과 연구결과물에 동등한 접근 보장이 필요

□ 연구데이터 공유·활용을 위한 분야별 전문 데이터센터 필요

- 연구 분야는 범위가 방대하기 때문에, 데이터 공유가치와 수요가 높은 세부분야를 선정하여 현실적인 데이터 공유·활용 체계를 수립하고 분야별 전문 데이터센터 구축이 필요
 - 주로 개인적인 네트워크를 통해서 공유, 연구이후 관리되지 않음
- 국내에서 운영 중인 연구데이터 수집·관리·활용 서비스는 추진 기관 및 목적에 따라 산재되어 있어 연구자들이 필요한 데이터가 어디에 있는지 찾기 어려우며 데이터 기탁 및 활용도 저조함

27) Beagrie, N. and Houghton J.W.. The Value and Impact of Data Sharing and Curation A synthesis of three recent studies of UK research data centres. 2014

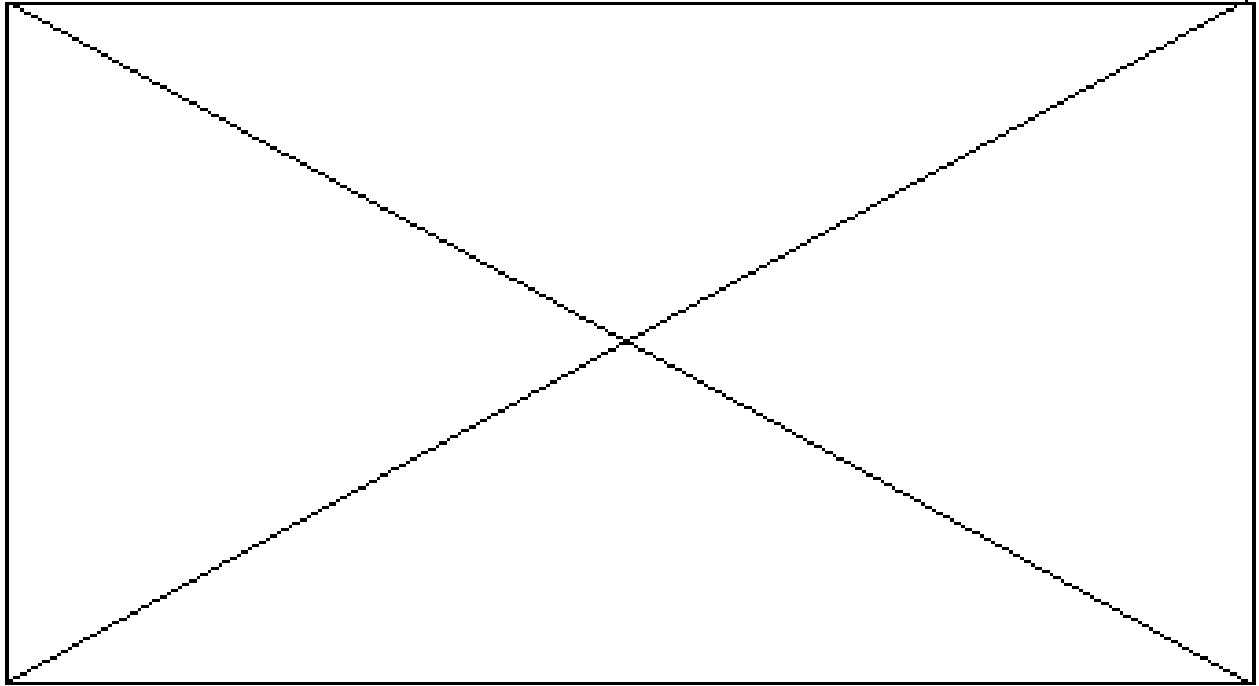
6.2. 추진전략

6.2.1. 추진방향 및 목표

가. 추진방향

- 연구개발 중점 투자분야 및 데이터 집약형 연구 분야에 우선하여 추진
 - 분야 선정을 위해 국가 전략적 중요성, 연구데이터 활용성, 파급효과 등으로 구분하여 우선분야 선정의 평가요소로 고려
 - －(국가전략적 중요성) 국가 정책 또는 상위계획과의 부합성, 사업추진 시급성 등
 - －(연구데이터 활용성) 생산하는 데이터량과 연구데이터 공유의 필요성, 활용 가능성 등
 - －(파급효과) 과학기술분야 파급성, 시장(산업) 파급성, 사회적(고용유발) 파급성 등
- 분야별 전문센터 평가체계에 따라 평가 및 선정하고 단계적으로 구축
 - 전문센터의 해당분야에서의 역할과 범위, 전문센터가 속한 해당분야와 분야 내에서의 전문센터의 역할과 연구데이터 확보 범위 등에 대한 평가
 - 전문센터의 연구데이터 활용도, 활용 시 연구성과에 대한 파급성 등의 평가*
 - * 과학기술정보통신부 「국가연구개발 과제평가 표준지침」 준수, 동지침의 취지 및 내용을 고려하여 평가지침을 마련하고 과제평가를 수행(연구성과평가법 시행령 제9조)
 - 전문센터의 선정평가는 사업계획서에 대한 전문기관 사전검토, 전문가 평가, 전문기관 검토 등의 절차를 통해 해당분야의 전문센터를 최종 선정하여 확정하고 단계적으로 구축
- 데이터 센터로서의 역할 및 기능 등의 도출과 분야별 전문센터 구축 시 공통 적용
 - (연구데이터 관리체계측면) 분야별 전문센터와 국가연구데이터 플랫폼과의 메타데이터 통합 및 연계, 연구기관 및 연구자에게 표준정보, 인프라 및 관리시스템 제공
 - (연구데이터 수집연계측면) 데이터 및 메타데이터의 표준화 등을 공개하여 데이터를 원활하게 수집 연계하여 공유할 수 있는 체계구축
 - (연구데이터 활용측면) 분야별 데이터 취합관리 및 활용 촉진 등의 서비스 제공

나. 추진목표

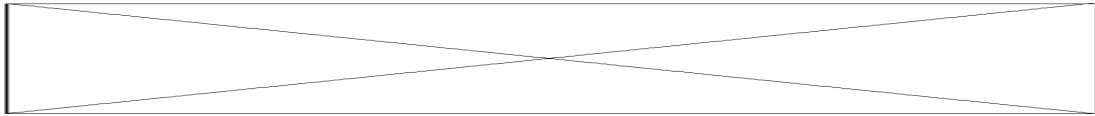


- (전문센터의 역할)분야별 연구자에게 연구데이터 제공과 연구협력을 유도하는 과학 지원 플랫폼
 - 분야별 데이터의 유형과 성격에 따라 다양한 방법의 표준화, 품질관리를 통해 생성된 연구데이터를 연계하고 수집하여 저장 관리하는 서비스 제공
 - 연구자들의 분야별 데이터 관리, 분석 방법 등을 지원하고 데이터 제공자와 연구자간 공동 연구 창출을 위한 커뮤니티 활동 및 협업이 가능한 서비스를 제공
 - 분야별 특성에 맞는 개별 연구자를 위한 데이터 관리 등 인프라적인 지원 서비스 제공
 - 글로벌 및 기존 데이터 서비스와 협력 및 연계를 통해 데이터 상호 교류 및 등록 서비스를 제공
- (전문센터의 기능)실험, 관찰 등으로부터 데이터를 생성, 저장, 유통하는 데이터 저장소
 - 국가연구데이터 플랫폼 및 개별 연구기관과의 메타데이터 통합 및 연계
 - 전문센터별 연구데이터관리 정책에 따라 연계 범위 및 방식 고려하되, 소분야 및 개별 연구기관과 연구데이터에 대한 최소기준, 표준화 등에 대한 협의 후 연계하여 수집
 - 분야별 원(raw) 데이터 및 관련 메타데이터 수집, 관리 및 표준화
 - 전자 연구노트 등 원(raw) 데이터의 품질관리

6.2.2. 추진전략 및 세부과제

가. 추진전략

□ 전략1. 현황분석 및 설문을 통한 전문센터 우선분야 선정



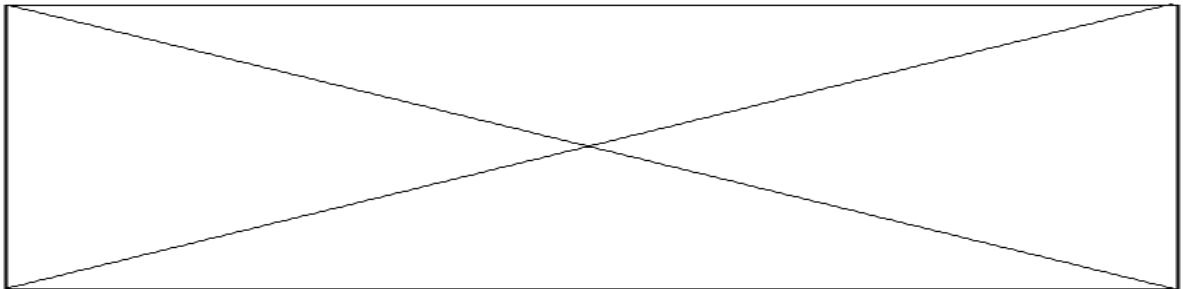
- 정부의 연구개발 중점분야*, 국가의 중점과학기술 분야** 등을 통해 연구개발 분야도출
 - － 연구개발의 특정범위나 부문들의 공통성 및 유사성 분석을 통해 연구데이터 공유 및 활용의 대상이 되는 연구분야를 선정
 - － 국내 R&D는 국가가 투자하는 중점 과학기술 분야 및 연구개발 투자 계획의 기술분야에 따라 사업이 기획되고 진행과정이 관리되어짐으로, 생산되는 연구 데이터도 동일 범위로 공유·활용 될 수 있도록 연구분야를 선정
- * 2019년도 정부연구개발투자 방향 및 기준: ICT·S/W, 생명·보건의료, 기초연구 등 11개 분야
- ** 제4차 과학기술기본계획(2018~2022): ICT·S/W, 생명·보건의료, 재난안전 등 11개 분야, 국가융합기술 시행계획(2017): 빅데이터,차세대반도체,융합형 콘텐츠,스마트 자동차 등 15개 분야
- 평가지표, 평가기준 등의 개발 및 정량적 평가 등을 통한 우선분야 도출 및 순위선정
 - － 국가의 전략적 중요성은 상위계획에 포함된 전략적 분야여부, 정부정책 동향의 부합성, 해외 사례여부 등을 평가지표로 도출
 - － 연구데이터 활용성은 연구데이터 생산량, 공유에 대한 수요, 연구데이터 활용의사, 타분야 연계성, 전략기술별 연관구조 등을 평가지표로 도출
 - － 파급효과 측면에서는 과학기술적 파급효과, 정부의 연구개발사업 투자효율성 포토폴리오 영역, 타산업에 미치는 영향, 산업연관 생산유발효과 등을 평가지표로 도출
 - － ICT·SW, 생명·보건의료, 농림수산·식품, 환경·기상, 소재·나노 등이 상위 5위로 도출됨

<표 6-1> 분야별 우선순위 도출 결과

순위	1	2	3	4	5	6	7	8	9	10	11
분야명	ICT·SW	생명·보건의료	농림수산·식품	환경·기상	소재·나노	기계·제조	우주·항공·해양	원자력·핵융합	에너지·자원	건설·교통	재난안전

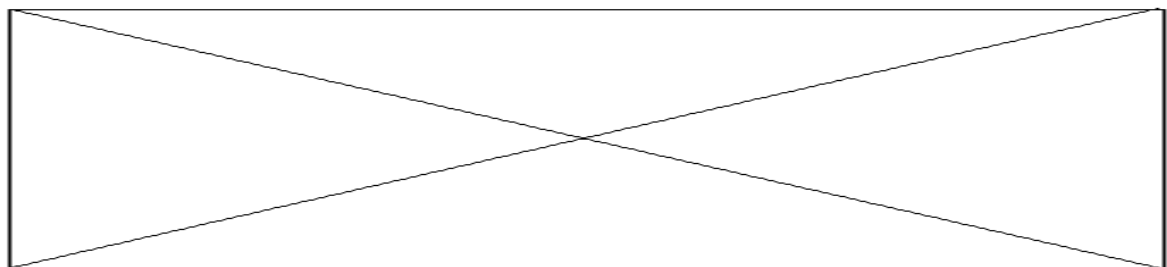
□ 전략2. 전문센터의 체계적 선정평가를 통해 데이터 집약형 분야부터 단계적 구축

- 우선분야 선정 결과에 따라 단계별로 평가절차에 따라 대·소분야 전문센터를 선정하여 구축
 - － 평가단계를 선정평가 → 중간평가 → 최종평가로 구분 및 정의하여 순차적으로 평가
 - － 평가위원 Pool 구성*은 NTIS에서 보유·관리하고 있는 평가위원 후보단 중에서 세부기술별로 적정 규모의 전문가를 확보하여 평가의 전문성을 유지
 - * 한국 연구재단이 보유하고 있는 한국학술진흥재단의 대학 연구자 그룹 DB, 한국과학재단의 출연 연구기관 DB를 활용
 - － 전문센터의 평가는 선정평가, 중간평가, 최종평가의 순으로 절차를 정의
 - － 선정평가는 신청 받은 국가연구데이터 전문센터 사업계획서에 대한 검토와 심의 등을 거쳐 신규로 지원할 대상 전문센터를 선정



<그림 6-1> 선정 평가단계 절차

- 데이터 집약형 전문분야를 우선 구축하고, 전분야 확산, 분야간 연계의 3 단계로 나누어 구축



- － 1단계는 사전 전문센터 평가를 통해 선정된 데이터 집약분야인 바이오, 소재, 인공지능 분야를 '20년 ~ '22년에 구축
- － 2단계는 전 분야 확산 단계로 환경·기상, 농림수산·식품분야, 기계·제조분야 등을 구축
- － 3단계는 분야간 연계성을 강화하는 단계로 원자력·핵융합, 에너지·자원 분야 등을 구축

□ 전략3. 데이터 기반 연구공동체 지원환경 구축

- 분석도구, 연구 및 협업 지원, 연구자료 공유, 커뮤니티 등 분야별로 연구자들에게 가장 필요한 기본적인 연구지원 서비스 제공
- 분석 도구 및 연구자료(소프트웨어, 학술자료 등) 등 대상이 광범위하며 수요가 많고 개별 연구자들이 접근하기 어려운 것을 우선적으로 구축

<표 6-2> 전문센터 제공콘텐츠 예시

분류	제공 콘텐츠		
분석도구	BLAST, FASTA, Multi-alignment(ClustalW, MAFFT, MUSCLE) Moleye 2D-gel ViPS Primer3 VectorScreen GeneChaser_P	MOTIF(MEME) Pathway(PathChaser) Clustr Search transeq pepwindow MSDfold DALI Expression Profiler	BWA & BWAmem Picard Mutec Pathway Studio IPA (Ingenuity Pathway Analysis)
연구지원 자료	<소프트웨어> Aspera: 대용량 데이터를 빠르게 전송하기 위해 채택한 솔루션으로 NCBI 및 EMBL-EBI 등에서 채택. <학술자료> F1000: F1000 추천위원들이 주기적으로 생명의료분야 출판 논문 추천하는 그룹에서 출발해 비상업적, 완전공개 Open Access 온라인 저널운영 및 개인 문헌 관리 도구 제공 (조건부 유료) Authorea: 온라인 협력 문서 작성 도구로 각종 저널 서식 변환 및 공동 편집, 커멘트 작성 기능 지원 및 citation 기능 제공. 무료 <교육자료(e-Learning)> Coursera, EdX 등에서 생명의료 관련 강의가 무료로 진행되고 있음.		

- 검색된 데이터를 분석하고, 협업연구 할 수 있는 연구자 지원 서비스
 - 데이터 관리 인프라(하드웨어 및 네트워크 등)를 기반으로 분석과 가시화 할 수 있도록 지원하고, 다수의 연구자들이 연구자원 및 데이터를 공유하면서 협업연구를 수행할 수 있는 공동연구 및 협업 지원 서비스
 - 연구자들 간 교류를 통해 정보를 수집하거나 의견을 교환하는 온라인 포럼 역할을 수행 할 수 있는 커뮤니티 활동 지원 서비스
- 데이터 지원 및 활용 관점에서 연구자 수요를 반영한 서비스 개발
 - 국내 과학기술 연구데이터의 전반적 현황을 파악하고 필요한 정보를 쉽게 찾아서 공유 할 수 있는 데이터맵 기능

- \ - 데이터와 콘텐츠에 대한 다양한 방식의 검색 기능과 데이터 미리보기 지원서비스, 검색 결과를 외부서비스와 연계하여 저장하거나 스크랩할 수 있는 기능
- 연구자가 연구목적에 따라 데이터를 통해 해결하고자 하는 문제를 등록하고 다른 사용자들이 참여하여 솔루션을 제시하는 오픈 포럼 서비스 제공

나. 세부과제 구성

추진과제	우선분야	세부과제
대·소분야 전문센터 구축 및 운영지원	ICT·SW	인공지능 분야 전문센터 구축 및 운영지원
	생명·보건의료	바이오 분야 전문센터 구축 및 운영지원
	에너지·자원	에너지 분야 전문센터 구축 및 운영지원
	소재·나노	소재 분야 전문센터 구축 및 운영지원
	기계·제조	기계분야 전문센터 구축 및 운영지원
	농림수산·식품	농림수산 식품분야 전문센터 구축 및 운영지원
	우주·항공·해양	우주항공분야 전문센터 구축 및 운영지원
	건설·교통	건설교통분야 전문센터 구축 및 운영지원
	환경·기상	환경기상분야 전문센터 구축 및 운영지원
	원자력·핵융합	원자력핵융합분야 전문센터 구축 및 운영지원
	재난안전	재난안전분야 전문센터 구축 및 운영지원

6.3. 바이오 전문센터 구축

6.3.1. 사업개요

가. 추진배경 및 필요성

□ 바이오기술 개발 및 융합으로 2030년경 바이오경제 시대로 진입이 예상되면서, 생명연구 자원*은 바이오경제의 핵심소재이며, 국제 이슈 해결의 잠재력 보유

* 생명공학연구의 기반이 되는 자원으로서 산업적으로 유용한 동물, 식물, 미생물, 인체유래 연구 자원 등 생물체의 실물(實物)과 정보를 의미(근거: 생명연구자원법 제2조제1항)

<표 6-3> 생명연구자원의 분류 및 정의

분류	정의
생물자원 (Biological resource)	• 인간에게 실제 혹은 잠재적 유용한 가치가 있는 생물체 및 그 유래 자원에 관한 정보 • 배양 가능한 생물체(미생물, 식물, 동물, 인간세포 등), 인체유래생물자원, 복제 가능한 부분, 배양 불가능한 생물체 및 이와 관련된 분자, 생리, 구조적 정보 • 범위: 생물가공품, 생물제재, 추출물, 분양, 유전체소재, 전통소재, 정보서비스
생물다양성 (Biological diversity)	• 생태계내의 생물종 및 생물체 다양성을 말하며, 종(種)내·종(種)간 생물 서식지와 생태계 다양성 정보 • 생태계에 존재하는 종(species) 수준의 야생동식물과 생물표본 개념에 한정하여 정의 • 범위: 동물계(Animals), 식물계(Plants), 균계(Fungi), 원생생물계(Protists), 원핵생물계(Bacteria), 고세균계(Archaea), 비세포생물계(Non-cell)
생명정보 (Bio-information)	• 생물자원, 생물다양성 그리고 생명정보 정보의 전산화 및 데이터베이스화 • 정보연계 및 통합시스템 개발생명정보학 기법을 이용한 대량 생명정보의 처리 및 분석 • 범위: 데이터베이스 구축을 위한 생명연구자원 데이터 표준안 마련/ 네트워크 기반의 정보연계 시스템 개발/ 정보통합을 위한 시스템 개발/ 대량데이터 관리 및 가공처리 기술개발

<출처: 제1차 생명연구자원관리기본계획('11~'20)>

- 에너지 고갈, 신종 전염병, 환경오염, 식량부족, 질병치료 등 인류 난제를 해결하는 바이오연구의 필수 자원으로서 중요성 증대

※ 1980년대 중반 이후 개발된 신규 의약품 중 60%가 천연물(동물, 식물, 미생물)에서 유래

□ 세계 주요국들은 바이오경제 시대의 자국 이익보호를 위해 생명연구자원 확보 및 활용을 위한 정책과 관리체계 마련

- 주요 선진국은 생명연구자원 확보를 위한 대규모 투자, 국가 차원의 통합적 관리시스템 구축 등 대응 정책 추진
 - － (미국) 국가생명공학정보센터(NCBI)를 통한 생명연구자원 핵심 DB 구축, 국립인간게놈연구소(NHGRI), 공동유전체연구소, 농업연구청 등 주요 국가기관을 통해 자원 확보 및 통합 관리
 - － (EU) 유럽 생물정보센터(EBI), 유럽 바이오정보기반정비계획(ELIXIR), 유럽 미생물자원연구기반(MIRRI) 등 유럽 내 생명연구자원 관리 기관 운영을 통해 유럽 생명연구자원(유전체, 미생물자원 등)의 정보화 및 통합 관리기능 강화
 - － (일본) 국가생물자원프로젝트(NBRP)를 통한 생명연구자원의 종합 관리 체제 정비, 국가빅데이터센터(NDBC)를 설치하고, 생명연구자원 DB를 통합·관리
- 우리나라는 ‘생명연구자원의 확보·관리 및 활용에 관한 법률’에 근거하여 관계부처에서 소관 분야 및 정보에 관한 사항을 관리, 국가생명연구자원정보센터를 중심으로 범부처 통합 연계
 - － (과기정통부) 생명연구자원 정보와 통합적 관리·유통을 위하여 국가생명연구자원정보센터(이하, KOBIC)를 지정·운영
 - － (관계부처) 소관분야별로 기탁등록보존기관, 책임기관, 생물자원정보센터를 지정 및 운영하고, 부처별 책임기관을 중심으로 소관분야 생명연구자원의 확보·관리·활용 체계를 구축, 효율적 운영을 위해 책임기관협의회 구성
 - － (관리시스템) 부처별 구축·운영되는 생명연구자원 관리시스템은 KOBIC의 국가생명연구자원통합정보시스템(이하, KOBIS) 중심으로 범부처 정보 연계

□ 디지털 기술의 확산과 함께 국가별, 지역별 특성의 차이로 인한 폐쇄적 연구 문화를 벗어나, 연구 성과와 과정을 개방하는 추세 확대

- 연구자 사회에서 일상적으로 연구 결과뿐만 아니라 생명연구자원을 포함한 연구자원을 공개·공유
 - － (연구데이터) 데이터기반·디지털 과학의 성장으로 연구데이터의 가치가 제고되고, 연구 성과 출판 이후 및 연구 진행 단계에서의 연구데이터 공개·공유 사례 증대
 - － (개방형 연구협력) 연구데이터의 공개·공유가 활발해지면서 이를 지원하는 디지털 인프라와 서비스, 커뮤니티도 성장

- (온라인 연구협력 플랫폼) 온라인에서 연구관심이 동일한 연구자들 간 연구데이터를 공개·공유하는 협업 플랫폼도 다양하게 성장
- 데이터 중심 R&D 활성화로 연구데이터 활용 수요가 점증함에 따라, 국가 R&D를 통해 축적되는 연구데이터를 다양한 연구자가 공유·활용할 수 있는 체계를 구축하기 위한 ‘연구데이터 공유·활용 전략’ 마련(국과심, '18.1)
- (4대 전략) 연구데이터 관리체계 구축 및 공유커뮤니티 형성 촉진, 국가연구데이터플랫폼 구축 및 서비스 제공, 데이터 및 컴퓨팅 활용 R&D 인재 성장 지원, 연구데이터 관리·공유·활용에 관한 법·제도 마련
- 최근 생물자원뿐만 아니라, 자원에 담긴 생명정보의 중요성 증대로 빅데이터 이니셔티브, 국제 연구협력 등을 통한 다양한 형태의 개방형 연구 모델 등장

◆ 세계적인 연구데이터 공개·공유 정책의 확대와 ‘연구데이터 공유·활용 전략(안)’의 이행을 위해 생명연구자원을 포함한 바이오분야 전문센터 구축·운영 필요

6.3.2. 환경분석

가. 생명연구자원 관련 국외 정책

□ 장치를 마련하고 생물자원 확보 경쟁하는 국가 관리주의 등장

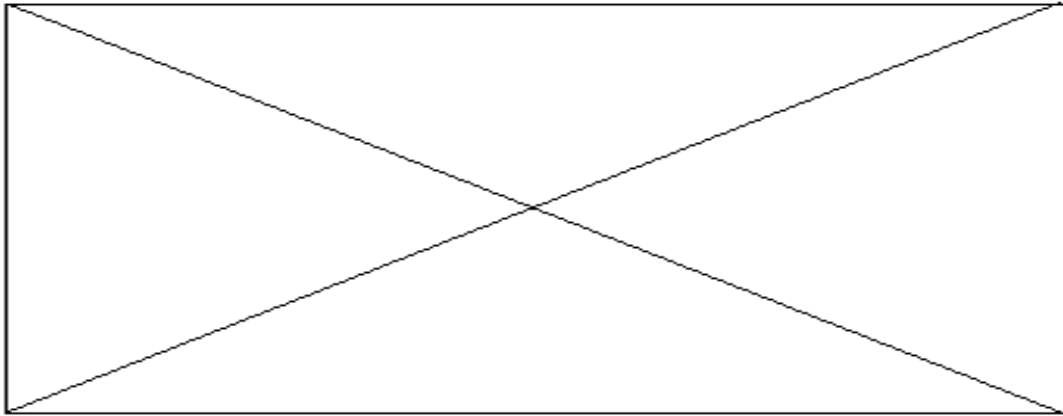
- (미국) 생명연구자원을 국가 자산으로 인식하여, '60년대 초반부터 전세계 생명연구자원 확보를 위한 국가 차원의 체계 마련 적극적인 투자 지속
 - 국가기관을 중심으로 생명정보 수집, 연구 및 유용 생명연구자원 확보 등 집중
 - ATCC(American Type Culture Collection)를 주축으로 하여 세계 최대 생물자원 보유, 국가-민간 조화 통한 전세계 생물자원 확보와 관리
 - 생물학적 연구자원 관리 효율화 및 시설 확충 등 생명연구자원을 핵심 인프라로 인식하여 지속적으로 지원 확대
 - 보건 분야에 가장 많은 비국방 R&D 예산 투자(22%, 327억 달러), 이 중 가장 많은 예산 할당 받은 NIH(National Institutes of Health)는 대부분의 보건의료 R&D 담당
 - 생물연구자원 연구와 인프라 시설 확충 등을 위한 지속적인 사업 수행 및 생물다양성이 풍부한 국가와 협력네트워크 진행
 - 간질, ALS, 뇌종양 등의 치료법 개발 목적으로 뇌신경조직은행과 빅데이터 센터(University of Illinois NeuroRepository) 설립('17.01)
- (EU) 범 유럽차원에서 공동연구 수행 및 정보 네트워크 구축
 - Horizon 2020에서는 식량자원 확보를 위한 농수산업 및 식품산업 지원 연구 관련 프로그램 진행('16년 141.5백만 유로)
 - 생물기반 연구 인프라 구축 및 지원 등을 위해 회원국 간(영국, 프랑스 등 19개국) 인프라 사업(BBMRI) 수행('16년 5.4백만 유로)
 - 바이오뱅크 연합(German Biobank Alliance)를 통해 바이오뱅크 노드와 소재은행을 활용하여 생명정보 분석 서비스를 직접 제공·통합하는 시스템 구축 추진('17.05)
 - (영국) 성인들의 신체활동 모니터링 정보를 제공('17.02)하거나 바이오뱅크를 통해 50만명의 유전자 데이터를 온라인에서 공공개방('17.11)하는 등 정보 공유
- (독일) 세계 최고 수준의 미생물자원은행인 DSMZ(Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH) 보유, 유럽 최대의 생물자원은행을 기반으로 세균자원 확보·관리 주도
 - DSMZ는 미생물, 인간·동물·식물 세포 배양, 연구자 및 과학 공동체 지원 서비스 제공하며, 특히 미생물자원 분야에서 세계 최대 보유 기관

- (일본) 선택과 집중을 위한 정부차원에서의 이니셔티브(initiative) 추진
 - 이화학연구소(RIKEN) 산하 생물자원센터인 JCM(Japan Collection of Microorganisms)를 통해 생물자원 확보, 개발 및 공동연구 지원
 - 질환 특이적 iPS세포은행 운영하며 이화학연구소 바이오리소스센터에서 일본 내 11개 기관이 제작한 300종 질환의 iPS 세포를 연구에 분양·활용('17.11)
 - (국가생물자원프로젝트, NBRP) 국가의 전략적인 생물자원을 체계적으로 확보, 관리, 제공을 위해 제4기 NBRP('17~'21)를 운영하여 시대 변화 반영한 정비
- (중국) 신흥 자원 부국으로 주도권 확보를 위한 중장기 프로젝트 착수
 - 풍부한 생물다양성 바탕(세계 8위, 북반구 1위)으로 생명연구자원의 미래 가능성을 예측하여 일관성 있는 생명연구자원 확보에 집중
- (정보) 생물자원 자체의 확보보다 생명정보 연계 중요성 높아지면서 통합된 빅데이터 구축을 위한 개방형 모델 등장
 - (생명정보 플랫폼) 20세기 초 개방형 연구 규범의 확산이 현대의 정보관리 표준화 요구로 이어져 선진국 중심의 통합관리 시도
 - 주요 생명정보 관련 인프라로 꼽히는 미국 생명공학정보센터(NCBI), 유럽 생물정보학연구소(EMBL-EBI), 일본 DDBJ(DNA database of Japan)의 경우 국제핵산서열데이터컨소시엄(INSDC) 참여를 통해 공동성과 창출
 - 미국 NIH 산하기관 중 하나인 국가암연구소 주도의 caBIG(cancer Biomedical Informatics Grid) 프로그램이 개방형 플랫폼을 통해 맞춤형 의료 제공 시도('04~'11)
 - 빅데이터 기반의 정밀의학 실현을 위한 유전체 분야에서의 투자 지속
 - (미국) NCBI의 경우, DNA, 단백질서열, 화합물 정보 등 생명정보 관련 데이터 기탁 및 공유를 통해 과학 및 보건 발전 목적
 - Entrez라는 검색시스템으로 사용자에게 논문, 유전자 염기서열, 생물체 분류, 구조 데이터 등에 대한 정보, 통합검색, 분석서비스 등을 제공하고 있으며, 외부 과학자와의 협력 촉진을 위한 과학자 방문 프로그램 지원
 - (유럽) 생물정보학연구소(EMBL-EBI)에서는 자유롭게 이용 가능한 최신 분자 데이터 베이스를 세계에서 가장 포괄적인 범위로 제공
 - 전세계 협력기관들에 의해 개발된 서비스를 통해 데이터 공유, 다양한 방식의 분석과 통합 액세스 제공에 집중하고 있으며, 생물학자들이 광범위한 데이터 자원을 활용할 수 있도록 실제적 생물정보학 교육을 제공

- \ ○ (일본) 10개 정부 부처에서 유전체 분야를 지원, 생명공학 전체에서 유전체 차지 비중이 약 5.8%로 타국에 비해 높은 편
 - DDBJ는 유전체서열 데이터를 수집하고, 자유롭게 이용 가능한 염기서열 데이터 및 슈퍼컴퓨터 시스템을 제공하여 생명과학 분야의 연구 활동 지원
- (중국) 베이징유전체연구소(BGI)는 중국 정부의 확고한 유전자원 지식재산권화 의지에 따라 세계에서 가장 큰 유전체 연구기관으로 발전
 - 대규모 바이오 저장소와 오픈 DB를 통합한 최초의 중국유전자은행(China National GeneBank)를 통해 저장시스템을 제공하고, 유전체 메커니즘 이해를 위한 새로운 플랫폼을 개발
- (자원·정보 융합) 바이오산업 내에서도 생명연구자원 및 생명정보 관련 사업의 성장 가능성이 유망함에 따라, 국가별 또는 국제적인 관리·활용 시스템 마련
 - (국가별 대응) 글로벌 유전체 시장규모는 46.2억 달러로 평가('15년), 새로운 플랫폼 활용에 따라 성장 예상되며 이에 따라 각국에서는 다양한 대응방법 추진
 - (미국) 간질, ALS, 뇌종양 등의 치료법 개발 목적으로 뇌신경조직은행과 빅데이터 센터(University of Illinois NeuroRepository) 설립('17.01)
 - (독일) 바이오뱅크 연합(German Biobank Alliance)를 통해 바이오뱅크 노드와 소재은행을 활용하여 생명정보 분석 서비스를 직접 제공·통합하는 시스템 구축 추진('17.05)
 - (영국) 영국 성인들의 신체활동 모니터링 정보를 제공('17.02)하거나 바이오뱅크(British Biobank)를 통해 50만 명의 유전자 데이터를 온라인에서 공공개방('17.11)하는 등 정보 공유
 - (일본) 질환 특이적 iPS세포은행 운영하며 이화학연구소 바이오리소스센터에서 일본 내 11개 기관이 제작한 300종 질환의 iPS 세포를 연구에 분양·활용('17.11)
 - (국제적 대응) 국제미생물연합정보센터(WDCM), 인간 단백질 Atlas, 암유전자 Atlas(TCGA) 등 생명연구자원 관련 국제컨소시엄을 통해 국제적 대응 강화

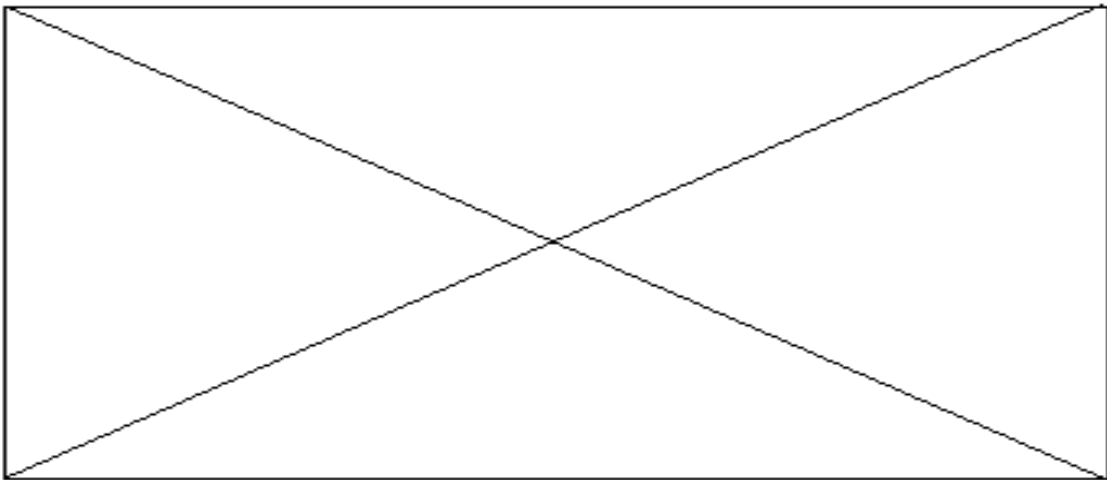
나. 생명연구자원 관련 국내 정책

- '07년 수립된 '국가생명자원 확보·관리 및 활용을 위한 마스터플랜'을 바탕으로, '09년 생명연구자원법 제정, '16년 '제2차 생명연구자원관리 기본계획('16~'20)' 수립 등을 통해 국가 차원의 체계적인 생명연구자원을 관리 중



<그림 6-2> 생명연구자원 관련 주요경과

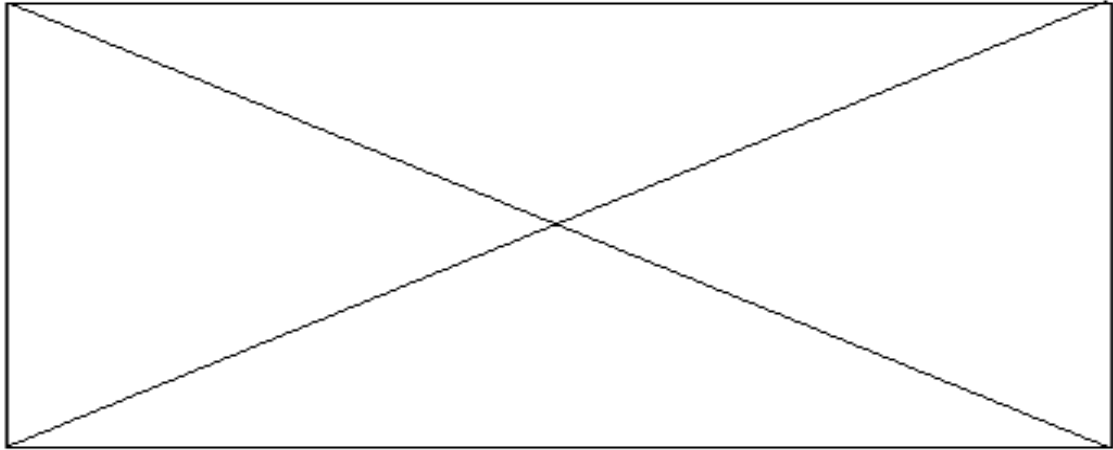
- ‘생명연구자원관리 기본계획’에 연계하여 연차별 시행계획 수립을 통해 체계적인 지원 시작
 - ※ (3대 추진전략) 국가전략생명연구자원의 안정적 확보, 생명연구자원의 이용가치 제고, 생명연구자원 관리 체계 고도화
- 생명연구자원법에 근거하여, 각 부처는 소관분야를 관리하고 있으며, 국가생명연구자원정보센터를 중심으로 범부처 통합 연계
 - (자원 관리) 부처별 책임기관을 중심으로 소관분야 생명연구자원의 확보·관리·활용 체계를 구축하고, 효율적 운영을 위해 책임기관협의회 구성·운영



<그림 6-3> 생명연구자원 관리체계

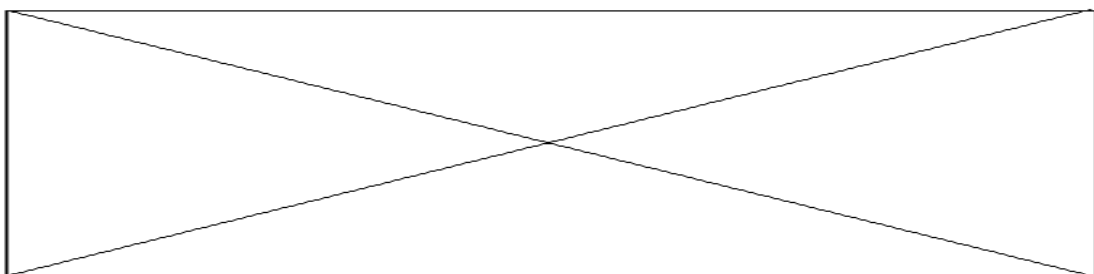
- (정보 관리) 과기정통부는 생명연구자원 정보와 통합적 관리·유통을 위하여 국가생명연구자원정보센터(KOBIC)를 지정 및 운영
 - － 부처별로 구축·운영되는 생명연구자원 관리시스템은 국가생명연구자원정보센터의 국가

\ 생명연구자원통합정보시스템(KOBIS) 중심으로 범부처 정보를 연계



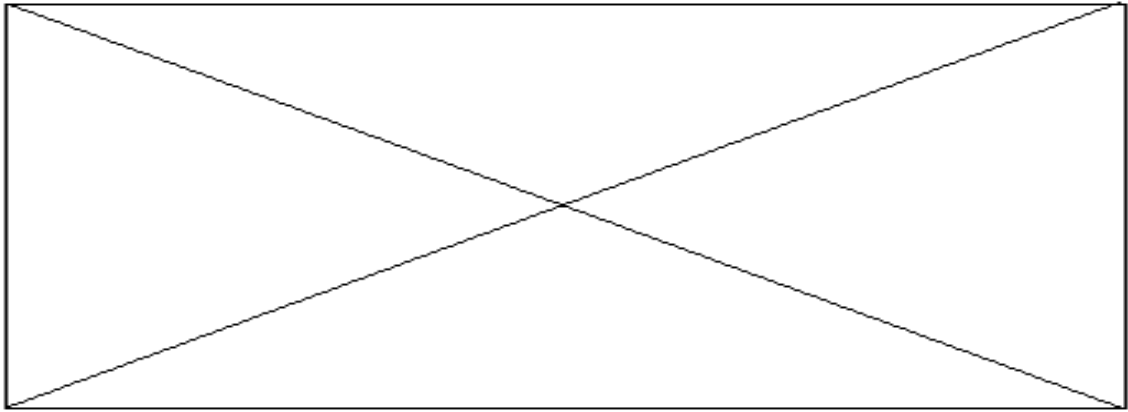
<그림 6-4> KOBIC 연계 부처

- KOBIC은 생명연구자원법 제11조에 따라 '10.3월 지정되어, 범부처 생명연구자원 정보의 통합·연계하는 범부처 센터로 역할 중이며, 생명정보 분야 전문연구를 위한 인프라 기관 역할 수행
- (설립 배경) 바이오경제의 핵심 소재인 생명연구자원의 중요성 증대에 선제적으로 대비하여, 과기부는 '10.3월 국가생명연구자원정보센터를 지정
- (역할) KOBIC은 국가 차원에서 생명연구자원의 효율적인 활용이 이루어질 수 있도록 부처별로 분산된 정보를 통합·연계*하고, 사용자가 활용하도록 정보공개 노력
 - * 과기정통부, 농식품부, 환경부, 해수부는 KOBIS를 통해 범부처 생명연구자원 정보가 연계되어 있으며, 산업부와 보건복지부는 유전체 정보만 연계 중
- 우리나라는 「국가연구개발사업의 관리 등에 관한 규정」을 통해 국가 R&D에서 도출된 연구성과를 등록(정보) 또는 기탁(실물)하도록 의무화



<그림 6-5> 연구성과물 제도

- KOBIC은 국가연구개발사업에서 산출되는 9대 연구성과물 중 생명정보에 대한 연구성과 관리·유통 전담기관으로 지정되었으며, 국가과학기술종합정보시스템(NTIS)과 연계하여 생명정보의 관리·유통 체계를 구축·운영



<그림 6-6> 9대 연구성과물에 대한 연구성과 관리·유통 전담기관 현황

다. 생명연구자원 관련 기술 동향

- (자원) 생명연구자원은 원천기술개발의 핵심 역할을 수행하고, 다양한 응용분야에서 활용 가치 증대 및 새로운 유형의 자원에 대한 산업 수요 증가
 - (기초·원천 분야·자원) 생명연구자원은 바이오 연구개발 수행을 위한 핵심 자원으로 나고야의정서 발효에 따라 해외 생명연구자원 활용이 제약되면서 국내 생산되는 고부가가치 자원의 보존·활용에 대한 국가 차원의 관리 중요
 - － 신약개발 등 바이오분야 원천기술 개발의 핵심도구로 활용되는 고부가가치 생명연구자원의 개발·확보 필요성 증대
 - － 국가별 원천기술 선점을 위한 경쟁이 진행되고 있는 만큼, 국가적 역량을 결집하여 우위를 점하는 전략 필요하나 부처별로 분산된 상태
 - (치료·응용 분야) 진단·치료를 위한 기초·표적 기술은 순수연구 이외에도 생물자원의 상품화처럼 제약회사 등의 치료제 개발 등을 통해 특허로 보호되는 개발성과 존재
 - － (표적 치료) 기존 약물 개발 기간에 비해 맞춤형 치료제의 개발이 시간을 단축시킬 수 있다는 점에서 기술 개발이 시급하며, 유전체 활용을 통한 약물 개발은 개인정보 문제와 함께 특허 문제가 발생할 수 있는 여지 존재
 - － (바이오마커) 단백질, DNA, RNA, 대사물질 등의 물질 또는 지표를 의미하는 바이오마커는 암, 심혈관질환, 뇌신경질환 등 난치 질병의 맞춤치료 위한 연구

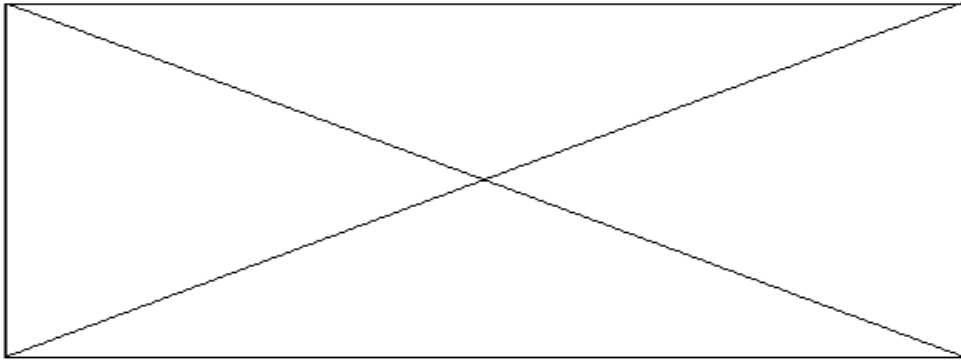
- \ - (동반진단) 특정 질병의 진단-치료제를 함께 구성하는 동반진단은 처방 전 특정 유전자를 확인하여 진단하고 맞춤형치료제 처방하는 형태
- (약물유전체학) 환자의 유전적 특성을 바탕으로 약물 종류와 복용량을 제시하고 효능이나 부작용을 예측하는 형태로 약물 개발
- (자원·정보 융합) 유전체, 바이오융합 등 기초 및 응용연구를 통해 보건·의료, 농업, 식품, 화학, 환경·에너지 산업 등 다양한 산업분야로 활용 확대
- 예방·관리, 개인맞춤형 의료 등 바이오 헬스케어 트렌드 변화에 따라 유전체정보 등 생명정보 빅데이터 구축·분석 수요 증가
- IoT, 3D 프린팅 등 타 분야의 혁신기술이 바이오(의료, 건강)에 접목되어 IoB, 3D 바이오 프린팅 등으로 진화
 - ※ 3D 프린팅 등 ICT 기술을 활용하여 가상자원 제작, 연구개발에 활용
- (정보) 기술 및 산업 발전과 함께 차세대 염기서열분석기술(NGS)의 급속한 발전에 따라 생명정보 분석 수요 증대
- (분석 수요) NGS의 발전에 따라 맞춤 의학의 본격화와 개인유전체해독 비용의 절감은 유전체 정보의 대중화, 보편화와 함께 수요자 맞춤형 분석 서비스, 대용량 클라우드 서비스에 대한 수요 증대로 이어짐
- 유전체 데이터 용량이 증가하면서 정보관리·활용 관련 요구 대응과 바이오 빅데이터 구축을 위해 IT 기술과 연계 필수적, 빅데이터 기반의 정밀의학 실현을 위해 미국에서는 유전체 분야에 투자 지속
- 인간유전자클론 특성정보·오믹스·표준화 정보를 확보·관리하는 통합 DB 등 수요에 따른 시스템 마련
- (미래 수요) 바이오산업 발전 및 기술 발전에 따라 맞춤의료 등 새로운 분야 관련 수요가 증대되면서, 이에 대비한 유전체 분야의 기초·원천기술의 확보와 인프라 구축
- 우리나라의 경우, 각 부처별로 유전체 연구 지원 및 기반구축 사업을 진행하고 있으며, 부처 공동으로 중첩영역 공동연구*를 통해 성과 극대화 노력
- (한국인 표준 게놈지도) 한국인 특성에 맞춘 대표 표준 게놈지도를 완성하면서 한국인의 유전적 특징에 따른 질병에 대한 치료제 개발 등 참고자료로 활용
- (맞춤의료 기술) 맞춤의료 기술은 생명정보기술 분야와 바이오 및 의료기술 분야로 구분되며, 병원 내 의료시스템 변화 및 제도 개선 등 많은 분야와 연계
 - ※ 개인정보 활용에 따른 악용 우려, 유전체/유전자 수집 및 상업화의 문제, 질병관련 유전자검사 및 특허 허용 범위 등 기술 발전에 따라 변화해야 할 내용 확대

- (기술 발전) 유전체 분야의 소프트웨어는 유전정보 처리기술, 유전정보 보안기술이 성장 전망되고, 미국이 압도적인 장비·기술을 가지고 있으나, 아직 타국의 경우 기반 구축이 제대로 이루어지지 않아 추격 가능성 높음
- － (유전정보 처리기술) 유전정보의 생산 속도 및 양의 폭발적 증가로 인해 정보의 처리와 분석 기술 분야가 성장하면서, 클라우드 서비스의 수요 증가하고 분석 장비 확보
- － (유전정보 보안기술) 유전정보 특성상 개인의 민감한 사항을 포함하므로, 정보를 보호하면서도 안전하게 활용하기 위한 보안기술 요구
- ※ 유전정보의 경우, 특히 범죄 수사 및 보험 등 민감한 사안에서 중요하게 활용되는 경우가 많아 정보보안기술 발전 필요

라. 생명연구자원 관련 시장 동향

□ 생명연구자원은 4차 산업혁명 성장 동력의 필수소재

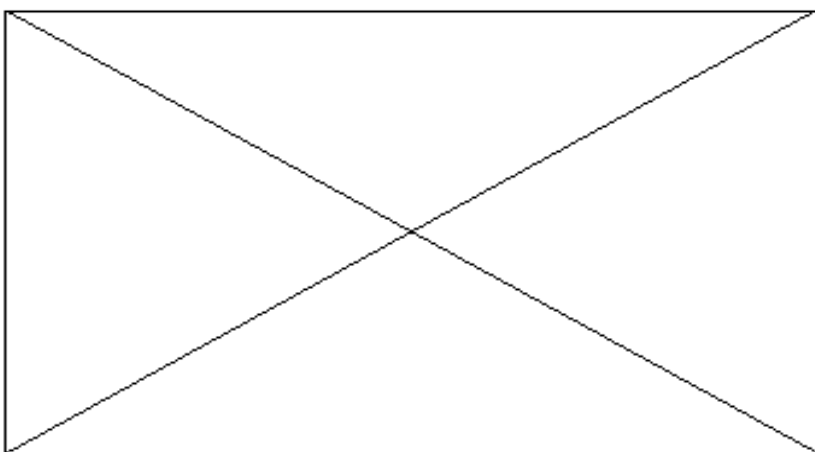
- 생명공학 시장 규모는 2024년까지 775억 달러를 넘어설 것으로 예상(근거: Global Market Insight, 2016)
- － 최근에는 맞춤형의료, 유전체의학 등 패러다임의 변화를 맞아 첨단·융합기술 바탕의 바이오시장의 급속한 성장이 예상됨
- － 사료첨가제와 항암제, 발효식품, 진단키트 등 4개 분야 시장을 선도할 것으로 전망되고, 백신과 기타 바이오의약품, 혈액제제, 식품첨가물, 호르몬제, 바이오 화장품 및 생활 화학제품, 바이오공종 및 분석기기, 동물약품은 상대적으로 성장세는 둔화되었으나 시장의 CashCow 역할 기대
- － 신개념치료제와 기타 바이오화학, 연구개발 서비스, 기타 바이오공정 및 기기, 생체의료 기기 및 진단기 등도 향후 시장 확대 전망
- 세계 의약품 시장은 2005년 이후 연평균 6%대의 안정적 성장세를 유지, '16년 1,260조원 규모를 형성, 앞으로도 연평균 4~7%의 성장, '20년 최대 1,650조원 규모의 시장으로 대폭 확대될 전망(근거: 한국 제약산업 길라잡이, 2017)
- － 최근 5년간 미국을 비롯한 EU 5개국, 일본 등 선진 제약 시장은 연평균 3~6%의 성장률을 기록하였고, 특히 지난 18개월 동안 중국의 상위 20대 제약 회사의 가치는 4,500억 달러~5,340억 달러로 19% 증가, 향후 50년 동안 중국 제약 산업의 규모는 4배로 확대될 것이며 미국은 2배, 유럽 시장은 2배 수준으로 성장 전망



<그림 6-7> 주요 6개국 제약 지출 규모

※ 출처: <https://endpts.com/top-20-pharma-companies-worldwide-by-valuation/>

- 글로벌 유전체 시장규모는 '15년 46억 2천 달러로 평가되었으며 새로운 플랫폼의 활용은 NGS에 대한 수요를 증가시킬 것으로 기대
- Metabolomics, Proteomics, Transcriptomics 분야의 R&D 프로젝트의 증가로 기술에 대한 수요가 증가
 - NGS기술에 대한 과학자와 약물 개발자의 연구가 커짐에 따라 많은 수의 생물체의 유전적 구성에 대한 통찰력을 얻으려는 플랫폼의 수요 증가로 인해 NGS 시장은 2025년까지 성장할 것으로 예상
 - 북아메리카 NGS 시장은 '15년 최대 점유율을 차지, 일루미나는 '15.6월 영국에서 실시된 10K 게놈 프로젝트를 통해 수익을 더 많이 창출하고 시장 입지를 강화하기 위해 케임브리지에 유럽 본사 설립
 - 아시아 태평양 지역은 건강관리 발전, 지역 인구의 의료 인식 증가, 1인당 소득 및 잔여 소득 수준 증가로 인해 예측기간 내내 연평균 18%이상 성장할 것으로 예상



<그림 6-8> 대륙별 Genomics 시장 규모주요 6개국 제약 지출 규모

* 출처: <https://www.marketsandmarkets.com/Market-Reports/genomics-market-613.html>

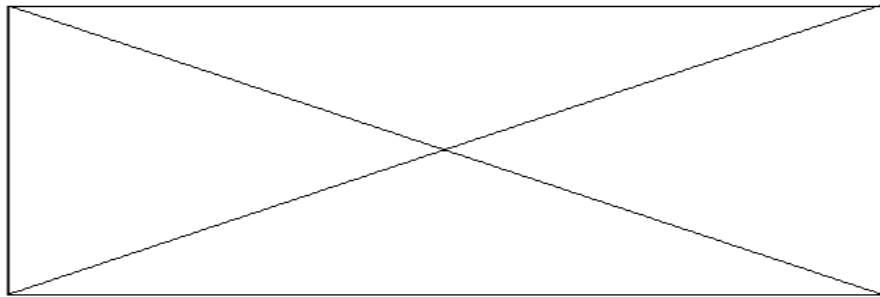
6.3.3. 추진전략

가. 사업추진 방향 및 목표

1) 국가생명연구자원정보센터(KOBIC) 운영 현황

□ 생명연구자원의 통합적 관리와 유통을 위하여 국가생명연구자원정보센터 지정('10.3)

- 2030년 경 바이오경제로의 진입 전망과 글로벌 이슈 해결 키워드로서 생명연구자원 부각에 따라 2010년 3월 교육과학기술부는 한국생명공학연구원 생명정보센터를 국가생명연구자원정보센터(KOBIC)로 지정
 - － (법적 근거) 「생명연구자원의 확보·관리 및 활용에 관한 법률(국가생명연구자원정보센터의 지정 등)」('09.05) 제11조에 따라 국가생명연구자원정보센터 설립
 - － KOBIC은 '10.3월 지정 이후 범부처 생물자원정보의 총괄관리 역할과 세계적으로 급속한 발전 및 수요 증가 중인 유전체 인프라 사업 등을 중점 수행



<그림 6-9> KOBIC 주요 연혁

- (주요 기능) KOBIC은 국가생명연구자원 통합정보시스템(KOBIS), 연구성과물 등록시스템(Bio-Data), 유전체 통합관리시스템(GenomeInfraNet), 유전체 분석 클라우드 서비스(Bio-Express)을 운영

<표 6-4> KOBIC 소개

비전	국가 바이오 빅데이터 연구를 선도하는 인프라 기관
기능	국내 생물자원정보의 총괄관리 및 생명정보분야 전문연구를 위한 범부처 국가센터
목표('21)	생명연구자원 확보·관리·활용 선순환체계 구축 유전체 정보 통합관리 및 활용 서비스체계 구축
전략 및 과제	생물자원정보 인프라 활용 기반 확산 KOBIS 및 생명정보 연구성과물 등록시스템 고도화 생명연구자원법에 따른 범부처 중추 역할과 제반 업무 수행

	바이오 빅데이터 활용 기반 구축 유전체 정보의 체계적 확보, 고도화 및 활용 기반 조성 국내 최고의 바이오빅데이터 기반 조성
	생명정보 인프라 연계 R&D 역량 제고 선도적 R&D에 기여할 수 있는 고급 분석기반 구축 산학연 연구지원 및 공동연구를 통한 대형성과 창출
	인력 양성 및 홍보 강화 생명정보 분석 전문인력 양성 총괄 역할 수행 대국민 홍보 강화를 통한 인지도 제고

－ (생명연구자원 정보) 범부처 센터로서 국가생명연구자원 통합정보시스템(KOBIS)을 구축 및 운영하고 있으며, 부처 연계를 통해 확보한 정보를 정제하고 품질을 향상시켜 수요자가 원하는 형태의 정보로 제공하기 위한 노력

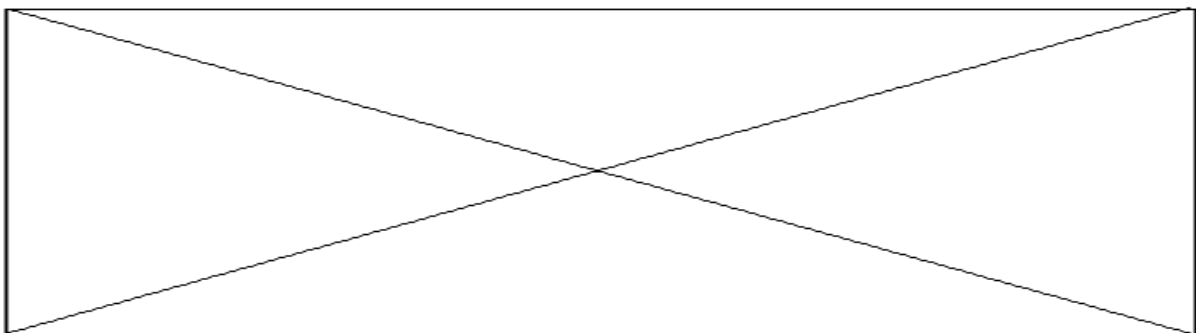
－ (유전체 및 연구성과물) 유전체 분석기술인 NGS의 발전으로 개인유전체 해독의 비용 및 대용량 데이터의 저장·관리에 대한 수요가 높아지면서, 대용량 클라우드 서비스 및 연구성과물* 등록시스템 구축

* 「국가연구개발사업의 관리 등에 관한 규정」에 따라 생명정보 연구성과물의 등록 전담기관으로 지정

□ 국가생명연구자원정보센터 조직 및 인력

○ (조직·인력) 생명연 내 산하 조직으로 1개 위원회, 3개의 실, 6개의 팀으로 조직 구성, 센터장 1명 포함 총 29명

－ 정규직 9명(센터장 1인 포함), 계약직연구원 17명, Post-Doc 2명, 계약직사무원 1명



<그림 6-10> KOBIC 조직도

－ (자원정보실) 생명연구자원법 및 동법 시행령에 따라 KOBIC에 부여되거나 위탁된 업무 수행, 자원관리팀과 자원활용팀으로 구성

- (생명정보실) 주로 생명정보 대외지원, 외부 대응 업무 담당하며 연구성과물 등록시스템 관련 업무 수행, 생명정보분석팀과 생명정보서비스팀으로 구성
- (전산개발실) 운영 서비스 중 KOBIS, Biodata, Bio-Express, GenomeInfraNet 등 설계·개발 및 하드웨어 인프라 구축 및 운영 업무, 개발팀과 전산팀으로 구성

□ 국가생명연구자원센터에서 구축 및 운영 중인 시스템

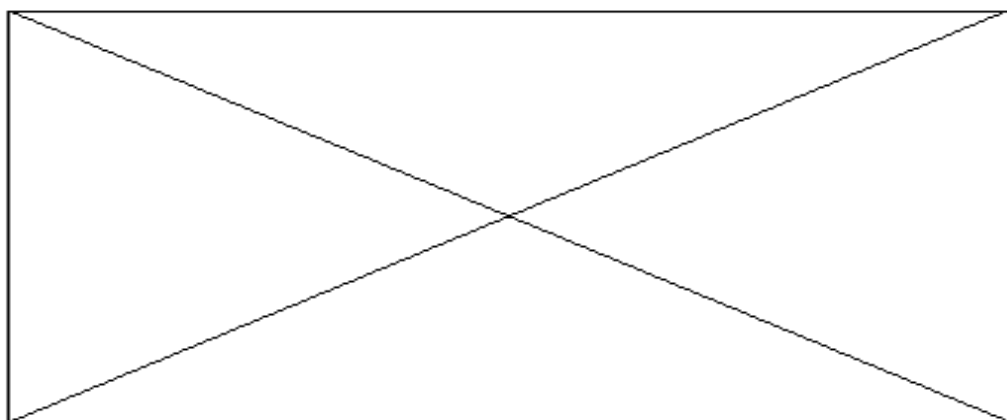
- (KOBIS) 생물자원은 바이오경제의 핵심소재로 경제적·산업적 중요성이 증가하고 있으며, 이러한 생물자원에 대한 통합 관리시스템 구축
 - (근거) 「생명연구자원 확보·관리 및 활용에 관한 법률」에 따라 생명연구자원에 대한 통합정보시스템(KOBIS)을 구축·운영
 - (주요 기능) 범부처 차원의 생물자원정보를 통합·관리할 수 있는 시스템을 구축·운영하면서 생물자원 정보연계 표준에 따른 국가 생물자원정보목록 제시, 생명연구자원 통계, 생명연구자원 검색 등을 제공
 - (구축 방향) 5대 전략을 바탕으로 국가 생물자원과 관련된 공개 자료의 체계적인 수집 및 관리를 위해 KOBIS 구축하고, 개편 TF팀을 구성·운영하여 개선사항 발굴하고, 우선순위를 정하여 개선 중
 - ※ 국내 최고 수준의 생물자원정보 하드웨어 및 소프트웨어 인프라를 구축, 각 분야별 최고 수준의 전문연구 인력이 5대 전략 및 19개 실천과제 수행 통해 '18년 생명연구자원 정보 분야 국가경쟁력 선진 7위권 진입 목표
 - ※ 연구자 대상으로 활용 콘텐츠 강화, 웹포털 고도화 등을 시스템 개발 방향으로 설정

<표 6-5> KOBIS 5대 전략 및 내용

5대 전략	내용
시스템 지속 고도화	• 환경변화(표준안, ABS, 연계기관 추가 등)에 능동적으로 대응을 위한 시스템 재구성 • 연계기관에 대한 물리적 데이터 백업서비스 기능 제공 • 생물자원정보 기탁등록시스템 유지관리 및 등록 지원
분석기반 구축 및 콘텐츠확보	• 기능지향적인 생물종 Index DB를 구축하여 연구의 기본 소재 제공 • 활용도가 높은 분석도구를 재가공해 분석지원 강화 • 데이터 수집과 가공 및 서비스에 필요한 기본 범용 모듈을 개발해 연구개발은 물론 서비스에 활용
정보연계확대	• 범부처 책임기관으로 구성된 책임기관협의회 정보연계 표준안 수정 보완 • 과학기술정보통신부 내 기탁등록보존기관에 대한 정보 연계 지속 추진 및 확대 • 범부처 정보통합시스템과 정보연계 추진
정책계획수립 및 조사	• 범부처의 협조와 검토를 통해 기본 및 시행계획 수립 • 통계집 발간을 통해 국내외 동향과 국내 생명연구자원 구축현황과 실태 파악 • 기존 통계집에 분석내용과 실태조사를 강화하여 생명연구자원 백서로 활용

5대 전략	내용
홍보 및 웹 서비스	• 웹포털 활성화를 통해 양질의 콘텐츠를 확보 및 정보공유 API 제공 • 매월 뉴스레터 발간을 통해 국내외 동향과 주요 이슈에 대한 정보 제공 • 학술발표 및 교육을 통한 연구자들의 활용성 제고

- (Bio-Data) KOBIC은 국내에서 산출되는 생명정보 연구성과물 등록 전담기관으로 연구과제 등록, 데이터 관리 등을 국가과학기술지식정보서비스(NTIS)와 연계한 생명정보 연구성과물 등록시스템(Bio-Data) 운영
 - (근거) 「국가연구개발사업의 관리규정」에 따른 9대 연구성과물 중 생명정보에 대한 기탁받는 시스템(Bio-data)을 구축·운영
 - (주요 기능) 연구과제 정보는 NTIS 공통검색 API를 통해 제공받으며, 등록받은 연구성과물(생명정보)은 NTIS와 연계하여 등록 대행, 분석 및 연계 서비스 제공
 - (구축 방향) 생명정보 연구성과물 등록시스템 개편, 생명정보 연구성과물 데이터 등록 및 검증, 신규 데이터 검증 및 활용 파이프라인 구축



<그림 6-11> 생명정보 연구성과물 등록시스템 모식도

- (GenomeInfraNet, GINs) '14년부터 범부처 차원에서 시작된 '포스트게놈 다부처 유전체사업'에서 생산된 모든 유전체 정보를 통합 관리하는 시스템
 - (근거) 「다부처유전체사업 관리규정」에 따라 다부처 유전체사업에서 창출된 유전체 데이터를 연계
 - (주요 기능) 각 부처 정보센터에서 등록된 유전체 정보 통합시스템인 GINs 구축, GINs를 통해 등록된 유전체를 유전체 연구자 및 산업체에서 활용할 수 있는 서비스 제공
 - (관련 서비스) 유전체정보의 활용을 위해 클라우드 기반의 유전체 정보를 분석하는

Bio-Express 제공, 대용량 사용자 데이터를 고속으로 분석 서비스에 업로드하고 분석 결과를 다운로드하는 고속 전송 시스템 KoDS 제공

- (Bio-Express) 대용량 분석 서버나 분석 기술이 필요한 연구자들을 위하여 Bio-Express 대용량 유전체 데이터 분석 클라우드 서비스 제공
 - (주요 기능) 웹 기반 서비스를 구축하여 누구나 쉽게 접근, 사용이 가능하고 대용량 생명정보 데이터의 저장, 관리 및 활용을 효율적으로 수행하기 위해 하둡(Hadoop) 기반의 클라우드 컴퓨팅 빅데이터 플랫폼을 자체 개발
 - (이용 현황) CLOSHA를 활용한 분석 파이프라인별 이용에서는 RNA Sequencing Pipeline 47.35%, 분석 프로그램별 이용에서는 Whole-genome Sequencing Software Tools 56.37%가 가장 큰 비중 차지
- ※ '17.2월 서비스 이후, '18.7월까지 총 1,716회 서비스 제공

<표 6-6> Bio-Express 대용량 유전체 데이터 분석 클라우드 서비스 소개

서비스명	내용
디스크 캐싱을 이용한 빅데이터 플랫폼	• Linux/Unix 분석 프로그램과 Hadoop용 빅데이터 분석 프로그램 동시 사용 가능 • Hadoop 분산파일 시스템을 사용하여 구축 비용이 싸고 안정성이 높음 • 고속의 대용량 처리를 위하여 최적화된 작업 스케줄링 • 대용량 인프라 시스템/고속 네트워크 망에 최적화된 서비스 제공
CLOSHA 통합 자동 분석 시스템	• 간단한 워크플로우 모델링을 통하여 분석 작업 수행 • Drag&Drop 기반의 편리한 사용자 인터페이스 • 일반 분석 프로그램과 Hadoop 기반의 빅데이터 분석 프로그램을 동시 사용 가능 • 분석 결과 다운로드 및 가시화 기능 • 파이프라인 실행 상태 및 결과 모니터링 기능 • 고성능 인프라 시스템을 활용한 고속/대용량 분석 서비스 제공 • 다양한 종류의 분석 프로그램/파이프라인 제공
고속 전송 시스템 KoDS	• 대용량 바이오데이터의 고속 전송 • 높은 신뢰성과 안정성 제공 • 편리한 사용자 인터페이스 • 네트워크 대역폭의 효율적인 사용 • 재전송 기능/전송이력 제공

□ KOBIC 기반시설 및 전산장비 보유 현황

- (KOBIC 기반시설) 전산장비 운영을 위한 361m² 규모의 서버실 보유
 - 380V 입력전원, 400A 차단기 이중화, 최대 517kW 수용 가능
 - 무정전 전원장치(UPS) 320kVA(30분 정전 보상)
 - 천정·바닥 노즐 8개 하론 가스
 - 온·습도 조절을 위한 항온항습기(210RT)

- \ ○ (전산장비) CPU 3,100 cores, 저장공간 10.7 PB 보유
 - 내부 분석 및 연구지원 900 cores, 외부 분석 및 저장 2,200 cores, 3.8 PB
 - Bio-Express 900 cores, 800 TB
 - 다부처 유전체 정보 데이터 저장 스토리지 2,400 TB
 - 대용량 메모리(4TB) 서버 및 GP-GPU 서버
 - 중요 데이터 백업 저장 스토리지 5,000 TB

2) 개선방향

- 향후, 생명공학 연구자들의 바이오 연구데이터 관리·분석 환경을 지원하고, 데이터 제공자와 연구자간 공동연구를 위한 커뮤니티 활동, 협업 가능한 플랫폼 서비스를 제공
 - 과기정통부 생명연구자원 데이터 수집
 - 과기정통부 생명연구자원 및 연구성과물* 데이터의 수집 및 제공
 - * 국가 R&D를 통해 생성된 결과 중 유전체, 단백질체, 염기서열 정보 등 생명정보 데이터 수집
 - 확보한 생명연구자원 및 연구성과물 데이터의 가공 및 분석
 - 바이오 분야 데이터 허브센터
 - 기탁, 연계 등을 통해 확보한 바이오 분야 데이터 허브센터 역할 수행
 - 글로벌 및 기존 데이터 서비스와 협력 및 연계를 통해 데이터 상호 교류
 - 생명의학 연구수행 지원 플랫폼 구축
 - 생명연구자원 및 연구성과물 데이터 서비스 구축
 - 데이터 분석 환경, 협업 연구 환경, 커뮤니티 서비스 구축 및 지원
- 우리나라는 부처별 주요 자원 중심의 경쟁적·양적 확보에서 벗어나 국가 차원의 전략적·질적 확보를 위해 노력
 - (기반 조성) 기능적 특성이 검증된 자원 중 기업 수요를 반영하여 산업화 가능 자원을 선정하고 유망한 산업으로 꼽히는 세포뱅크 및 산업클러스터 구축을 통한 대량생산 등 산업화 지원
 - 지역별 자원 활용 강화를 위해 산업클러스터를 조성하고 주요 거점기관과 연계하여 산업화 지원 확대

나. 사업추진 전략 및 세부추진 과제

1) 사업추진 목표

목표1: 바이오 연구데이터 연계체계 구축	목표2: 바이오 연구데이터 수집	목표3: 바이오 연구데이터 활용시스템 구축
국가연구데이터 플랫폼과 바이오 연구데이터 간 메타정보 연계 ['22년 연계체계 구축 완료]	과기정통부 생명연구자원 정보 확보 강화 [생명연구자원 '18년 2,668,777건 → '28년 4,350,000건] 생명정보 연구성과물 확보 강화 ['18년 52 TB → '28년 80TB]	생명연구자원 정보기반 멀티오믹스 분석시스템 구축 ['23년 오픈 계획] 정보연계 및 분석 지원을 위한 인프라 확보 CPU: '18년 700 core → '28년 1,000 core 저장공간: '18년 2 PB → '28년 3 PB

<그림 6-12> 바이오 연구데이터 확보·관리·활용 체계 구축

○ 목표 1 : 바이오 연구데이터 연계체계 구축

- 바이오 연구데이터 전문센터와 국가연구데이터 플랫폼 간 메타정보 연계를 위한 기반 구축 및 연계

 연계체계 구축 완료(~'22년)

○ 목표 2: 바이오 연구데이터 수집

- (과기정통부 생명연구자원 정보 확보) 바이오 분야의 연구에서 산출된 바이오 연구데이터를 분야별 연구내용, 연구성과 발표 타입별, 생산된 연구데이터 타입별로 나누어 등록·관리하고, 제공하는 정보의 다변화

 **[생명연구자원 정보]** '18년 2,668,777건→'28년 4,350,000건 (연평균 5% 증가)

- (생명정보 연구성과물 확보) 연구성과물 등록시스템 기반의 연계 체계를 활용하여, 바이오 분야의 연구 방법론, 결과물 등을 수집하고 정제 강화

 '18년 52 TB → '28년 80 TB (연평균 4.4% 증가)

○ 목표 3: 바이오 연구데이터 활용시스템 구축 및 지원을 위한 인프라 구축

- 생명연구자원 정보기반 멀티 오믹스* 분석시스템 구축(개발 중)

\ * 유전체학, 전사체학, 단백질체학 등 생명정보학 연구에 활용되는 집합체 전부

☞ '22년 알파/베타테스트 → '23년 정식 오픈 계획

－ 정보연계 및 분석 지원을 위한 인프라 확보

☞ (CPU) 18년 700 core → '28년 1,000 core (연평균 3.6% 증가, 감가상각 고려 필요)
(저장공간) 18년 2 PB → '28년 3 PB (연평균 4.1% 증가, 감가상각 고려 필요)

2) 사업추진 체계

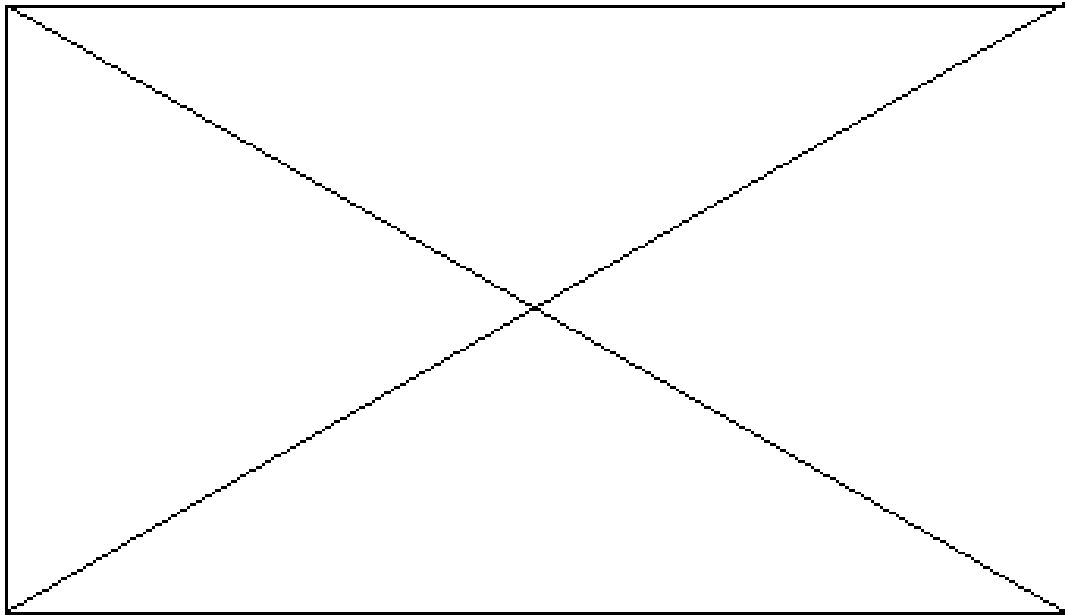
□ 총 사업비(10년, 22,550백만원)

<표 6-7> 연구데이터전문센터 구축 및 운영 단계별 추진 계획

구분	1단계(2021~2023)	2단계(2024~2026)	3단계(2027~2030)
추진전략	연구데이터 전문센터 구축	연구데이터 전문센터 운영	연구데이터 활용체계 구축
	- 생명연구자원 및 연구성과물 정보 확보 및 정제 지속 - 바이오 연구데이터 연계체계 구축	- 생명연구자원 및 연구성과물 정보 확보 강화 및 정제 - 생명연구자원 정보기반 멀티오믹스 분석시스템 구축 - 정보연계 및 분석 지원을 위한 인프라 구축	- 생명연구자원 및 연구성과물 정보 확보 강화 및 정제 - 생명연구자원 정보기반 멀티오믹스 분석시스템 운영 - 정보연계 및 분석 지원을 위한 인프라 구축
예산인력	11.5억~12억/년 (필요인력 15명)	24억/년 (필요인력 20명)	29.5억/년 (필요인력 25명)

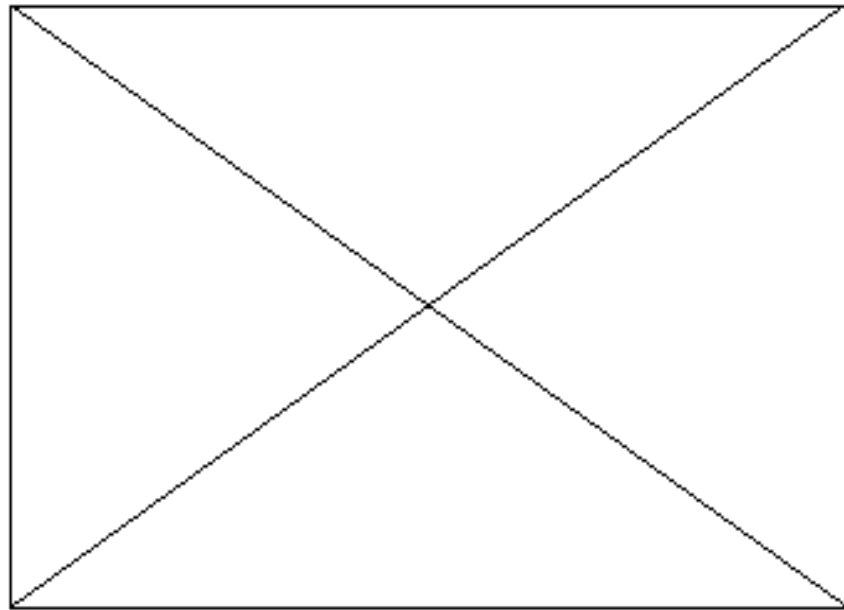
6.3.3.3. 세부과제별 추진 내용

1) (목표 1) 바이오 연구데이터 연계체계 구축



<그림 6-13> 국가연구데이터플랫폼과 전문센터 간 정보연계 체계

- 바이오 연구데이터를 연구데이터 플랫폼에 연계하기 위한 정보연계 체계 구축을 위한 기반 마련
- 바이오 연구데이터 전문센터 운영 모델 구축
 - － (정책적 측면) 바이오 정책 환경변화를 수용하는 정책수요 발굴 기반 및 연구자 중심의 공공서비스 활용체계 마련
 - － (기술적 측면) 바이오 연구환경 변화를 주도하는 신기술 및 정보자원의 효율적 활용을 통한 운영 효율성 제고
 - － (서비스 측면) 정보의 유통, 활용을 지원하는 공공 서비스 생태계 지원
- 연구데이터 전문센터 개념도
 - － 정보서비스를 통한 가치제고, 맞춤형연구지원 및 통합운영서비스를 제공
 - － 연구데이터 등록→연구데이터 관리→연구데이터 활용 지원→연구데이터 전문 검색→연구데이터 분석의 환류체계를 수용하는 데이터 센터 모델 수립



<그림 6-14> 바이오 연구데이터 전문센터의 추진체계

○ 바이오 연구데이터 전문센터 운영 체제 확립

- 기존 운영된 데이터센터의 공통 운영 서비스 요소 및 국내외 운영서비스 사례 분석을 통한 운영모델 구성
- 국가정책 동향 및 변화요소를 고려하여 운영모델 확립
- 운영모델은 정보관리 및 운영을 위한 인프라 구성 측면의 분야와 연구활성화 측면의 서비스 분야로 구성

2) (목표 2) 바이오 연구데이터 수집

○ (과기정통부 생명연구자원 정보 확보) 바이오 분야의 연구에서 산출된 바이오 연구데이터를 분야별 연구내용, 연구성과 발표 타입별, 생산된 연구데이터 타입별로 나누어 등록·관리하고, 제공하는 정보의 다변화

※ 既제공하던 생명연구자원 분량정보, 논문, 특허뿐만아니라, 유전자정보, 유전체·단백체·추출물 정보, DNA 분류, 연구성과물 등 제공 정보의 범위 확대

- 국가생명연구자원 통합정보시스템(KOBIS) 운영 지속

※ KOBIS 접속자 현황 및 장애관련 실시간 모니터링 관리체계 구축, 국가과학기술지식정보서비스(NTIS) 및 국가연구데이터 플랫폼과 정보연계

- 정형데이터 통계정보를 실시간 탐색, 다차원 정보 분석 결과 제공

- ※ 부처별, 자원별, 대구분별, 중구분별, 증식/파생, 실험/정보, 연도별, 신규·누적 등 다양한 변수의 통계정보를 제공
- KOBIS를 통해 다양한 신규 콘텐츠* 지속적 추가
 - * 국내에서 발굴된 생명연구자원의 정보, 유전체 염기서열 분석결과, 국내 생명연구자원을 활용한 특허정보, 생명연구자원 관련된 국내·외 기사 등의 정보를 제공
- 생명연구자원 정보 검증시스템 개발
 - ※ 관계부처 생명연구자원정보센터에서 연계되는 정보의 오류와 충실하지 않은 정보의 증가에 따른 데이터 품질 향상 필요
- 생명연구자원 정보연계 확대
 - ※ 현재, KOBIS는 생명연구자원과 관련된 9개 부처/청 중 8개 부처/청의 생명연구자원 정보를 연계(KOBIS는 6개 부처/청, Genome-InfraNet은 7개 부처/청)
 - ※ 향후, KOBIS와 Genome-InfraNet 간 상호검색 및 연계를 추진할 계획
- (생명정보 연구성과물 확보) 국가연구개발사업 관리규정에 따라 확보·관리되고 있는 생명정보 연구성과물의 확보 강화
 - 생명정보 연구성과물 등록시스템의 정보입력체계 개편('18년 신규 제정 예정인 연구성과물 운영지침에 따라 등록 필수·옵션정보 및 필드 수정)
 - 생명정보 연구성과물 등록 편리성 확대 및 최신 웹디자인 적용을 위한 시스템의 인터페이스 개편
 - ※ 등록 편의 및 가독성 증대를 위한 UX(User Experience) 및 웹페이지 UI(User Interface) 디자인 재구성 필요
 - 연구성과물 등록 전과정을 모니터링 할 수 있는 자동화 시스템 구축·운영
 - ※ 현재는 등록 데이터의 분류체계별 건수만 관리하고 있어, 등록 과정 전체의 모니터링 시스템 구축 필요 → 분류 체계상의 모든 데이터 등록을 받고 있으며, 연구성과물 등록 단계에서는 메타정보만 검증하고 있음. 2차 데이터의 경우, 연구 영역이므로 전담기관에서 불완전한 연구성과물의 판단·검증 곤란
 - 연구성과물 데이터 검증 및 분량 관련 프로세스 및 표준 프로토콜 확립
 - ※ 데이터 검증 및 분량과 관련된 표준 프로토콜은 없으며, 검색과 분량에 대한 상세 메뉴얼 부재
 - 연구성과물 검증 파이프라인을 이용한 업무 수행
 - ※ 신규 3.5세대 NGS 염기서열 분석플랫폼 데이터 검증 파이프라인 구축 및 적용 필요
 - 연구성과물 뿐만아니라, 다양한 생명정보를 수집하여, 연구데이터 분석에 적극 활용
- 3) (목표 3) 바이오 연구데이터 활용 시스템 구축 및 지원을 위한 인프라 구축
 - 국가생명연구자원 통합정보시스템(KOBIS), 연구성과물 등록시스템(Bio-data)와 연계하여 생명연구자원과 다양한 오믹스 데이터를 연계하여, 신약개발, 건강정보 등과 결합한 개방형 융합 분석시스템 구축

- \
 - 비교유전체학, 단백질 구조분석, 유전체분석 파이프라인 및 프로그램 제공 등을 통해 생명정보 분석 파이프라인의 구축 및 제공을 통한 수집 유전체 정보 활용성 증대
 - 생명정보 빅데이터 기반 신약개발, 유전체 정보 활용 맞춤형 의약품개발 등 유전체 정보를 활용한 질환 진단 및 치료제 기술개발 등 지원
- KOBIC에 축적된 연구데이터 및 메타정보 중 공개 가능한 데이터를 선별하여 분리된 별도의 H/W 하드웨어 구축
 - 생명연구자원 및 연구성과물 메타정보 저장을 위한 DB 서버 및 스토리지
 - 연구데이터 저장을 위한 고효율 대용량 스토리지
 - KISTI 연구데이터 플랫폼과 연계를 위한 서버 및 전용 네트워크 스위치
 - KISTI, KOBIC 간 데이터 연계 전용망 구축 및 보안 강화를 위해 물리적인 정보보안 체계 마련
- 바이오 연구데이터 전문센터로서 국내외 관련기관과 협력 및 인력양성
 - 세계적인 생명연구자원 정보센터인 세계생물다양성정보기구(GBIF), 생물자원정보센터(WDCM), 미국 생명공학정보센터(NCBI) 등과 협력을 통해 다양한 연구데이터 확보 추진
 - R, Python 등을 활용한 빅데이터 분석, 머신러닝 등을 적극적으로 이용할 수 있는 바이오 데이터사이언스 육성

6.3.4. 사업운영계획

가. 사업추진 일정

세부과제명	사업기간											
	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30		
바이오 연구데이터 연계체계 구축												
바이오 연구데이터 수집												
바이오 연구데이터 활용시스템 구축 및 지원을 위한 인프라 구축												

나. 사업비 및 인력운영 계획

1) 투자계획 및 재원조달

(단위: 백만 원)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	1,150	1,200	1,200	2,400	2,400	2,400	2,950	2,950	2,950	2,950	22,550
바이오 연구데이터 연계체계 구축	400	450	450								1,300
바이오 연구데이터 수집	750	750	750	750	750	750	750	750	750	750	7,500
바이오 연구데이터 활용시스템 구축 및 지원을 위한 인프라 구축				1,650	1,650	1,650	2,200	2,200	2,200	2,200	13,750

2) 인력 운영계획

(단위: 명)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	15	15	15	20	20	20	25	25	25	25	190
바이오 연구데이터 연계체계 구축	5	5	5								
바이오 연구데이터 수집	10	10	10	10	10	10	10	10	10	10	100
바이오 연구데이터 활용시스템 구축 및 지원을 위한 인프라 구축				10	10	10	15	15	15	15	90

3) 세부과제별 사업비내역

□ 바이오 연구데이터 연계체계 구축

○ 상세 분야별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
바이오 연구데이터 연계 항목 논의 및 연계	400	250	150	—	—	—	—	—	—	—	—
바이오 연구데이터 연계		200	300	—	—	—	—	—	—	—	—
계	400	450	450	—	—	—	—	—	—	—	—

○ 비목별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인건비	소계	300	300	300	—	—	—	—	—	—	900
	투입인력(명)	5	5	5	—	—	—	—	—	—	15
	인건비(년)	60	60	60	—	—	—	—	—	—	180
직접비	소계	51.8	95.8	95.8	—	—	—	—	—	—	243.4
	연구장비·재료비	21.8	65.8	65.8	—	—	—	—	—	—	153.4
	연구활동비	20	20	20	—	—	—	—	—	—	60
	연구과제추진비	10	10	10	—	—	—	—	—	—	30
	공동연구비	—	—	—	—	—	—	—	—	—	0
간접비		48.2	54.2	54.2	—	—	—	—	—	—	156.6
계		400	450	450	—	—	—	—	—	—	1,300

□ 바이오 연구데이터 수집

○ 상세 분야별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
과기정통부 생명연구자원 정보 확보	400	400	400	400	400	400	400	400	400	400	4,000
생명정보 연구성과물 확보	350	350	350	350	350	350	350	350	350	350	3,500
계	750	750	750	750	750	750	750	750	750	750	7,500

○ 비목별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인건비	소계	600	600	600	600	600	600	600	600	600	6,000
	투입인력(명)	10	10	10	10	10	10	10	10	10	100
	인건비(년)	60	60	60	60	60	60	60	60	60	600
직접비	소계	59.6	59.6	59.6	59.6	59.6	59.6	59.6	59.6	59.6	596
	연구장비·재료비	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	19.6	196
	연구활동비	30	30	30	30	30	30	30	30	30	300
	연구과제추진비	10	10	10	10	10	10	10	10	10	100
	공동연구비										—
간접비		90.4	90.4	90.4	90.4	90.4	90.4	90.4	90.4	90.4	904
계		750	750	750	750	750	750	750	750	750	7,500

□ 연구데이터 활용시스템 구축 및 지원을 위한 인프라 구축

○ 상세 분야별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
생명연구자원 정보기반 멀티오믹스 분석시스템 구축	—	—	—	650	650	650	800	800	800	800	5,150
정보연계 및 분석시스템을 위한 인프라 구축	—	—	—	1,000	1,000	1,000	1,400	1,400	1,400	1,400	8,600
계	—	—	—	1,650	1,650	1,650	2,200	2,200	2,200	2,200	13,750

○ 비목별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인건비	소계	—	—	—	600	600	600	900	900	900	5,400
	투입인력(명)	—	—	—	10	10	10	15	15	15	90
	인건비(년)	—	—	—	60	60	60	60	60	60	420
직접비	소계	—	—	—	851.2	851.2	851.2	1,035	1,035	1,035	6,693.6
	연구장비·재료비	—	—	—	800	800	800	1,000	1,000	1,000	6,400
	연구활동비	—	—	—	35	35	35	25	25	25	205
	연구과제추진비	—	—	—	16.2	16.2	16.2	10	10	10	88.6
	공동연구비	—	—	—	—	—	—	—	—	—	—
간접비		—	—	—	198.8	198.8	198.8	265	265	265	1,656.4
계		—	—	—	1,650	1,650	1,650	2,200	2,200	2,200	13,750

6.3.5. 기대효과

가. 경제적 기대효과

- 바이오경제의 핵심소재인 생명연구자원 정보의 효율적인 활용을 통한 바이오산업의 큰 기여
 - 수집된 생명연구자원 및 연구성과물 정보를 활용하여 바이오 연구를 지원하고, 궁극적으로는 바이오 연구 및 산업 발전을 통한 국가경쟁력 향상을 기대
 - － 바이오 연구데이터 확보 및 관리, 분석 방법론 발굴 및 고도화 등을 통하여, 생명공학 관련 연구자 및 산업계를 지원
 - ※ 바이오 빅데이터는 신약개발, 헬스케어 서비스 등에 접목을 위해 시도 활발(전자신문, '18.9.3일자)
 - － 연구데이터 공유 및 공동 연구 저변 확대를 위한 전산 인프라 구축·운영
 - 국가 차원의 체계적인 생명연구자원 관리, 검색 등을 통한 연구개발 및 원천소재에 대한 정보 제공을 통해 산업화과정의 기간 단축 및 효율성을 제고하여 바이오 경제시대의 신산업 창출
 - 기초과학 연구 기본재료인 생명연구자원 및 관련 정보를 연구자에게 제공함으로써 연구 기간 단축, 수입 대체자원 발굴 등으로 경제적 효과 발생
 - ※ 국내 바이오산업에 쓰이는 유전자원의 67%를 해외에 의존(한국경제, '17.4.30일자)
- 바이오 빅데이터 활용으로 인해 데이터 엔지니어, 임상시험 정보보호 기술자 등 신약개발, 의료데이터 분석 영역에서 빅데이터 분석가가 주목받을 전망
 - ※ 빅데이터 분석가, 바이오 새 일감으로 부상(전자신문, '18.9.18일자)
 - 2025년까지 00여명의 일자리 창출근거로 하며, 생명연구자원의 체계적인 연구개발과 투자의 기반 조성 목적

나. 과학기술적 기대효과

- 바이오 연구데이터 현황분석 및 관리를 위한 다양한 활용 인프라 구축
 - 국가 기초과학 인프라 구축을 통해 바이오 연구데이터 무한경쟁시대에서 우위를 차지
 - － 정보 통합검색 및 산업적 활용을 위한 자원별/기관별 정보관리
 - － 바이오 연구데이터의 지속적 확보와 관리를 통한 첨단화를 도모하고, 국제적 연구협력을 통해 바이오 분야 연구의 선진화를 제고
 - 각 부처 유관기관에 저장된 생물학적 정보 네트워크 데이터의 표준화를 통해 고도의 생명연구정보 검색, 공유 및 분석 서비스 제공
 - 각 분야별 표준용어 체계를 기반으로 한 전산코드체계를 확립하여 방대한 정보의 체계적인 관리 가능

6.4. 소재 전문센터 구축



6.4.1. 사업개요

가. 추진배경

1) 소재 데이터 특징 및 현황

- 소재 연구개발 전주기 과정에서 실험과 계산으로 생산되는 모든 정형, 비정형 데이터 및 문헌, 인터넷을 통하여 공개된 소재 데이터
 - 연구데이터는 실험과 시뮬레이션으로 생산된 물리·화학·기계적 성질, 스펙트럼, 미세구조 (조직) 이미지 등 다양한 종류, 형태 및 차원으로 존재함
 - 연구데이터는 연구개발 과정에서 목표특성에 도달하지 못했거나, 예측된 물성을 가진 소재 합성·제조에 성공하지 못한 실패 데이터도 포함

<표 6-8> 소재 연구데이터 예시

구 분	세 부 내 용
합성·제조	• 합성에 사용한 시약 종류(원소성분) 및 양 • 합성방법 (수열합성, 기상증착, 고상반응 등) • 합성조건 (온도, 압력, 시간 분위기 등)
특성평가	• 구조분석 (XRD, SEM, TEM 등) • 특성분석 (FTIR, Raman, XPS, NMR, 인장강도 등) • 응용특성분석 (전기전도도, 열전도도, 가스 흡착률, 센서반응 등)
계산물성	• 결정(분자)구조, 밴드갭, 전하이동도, 전기전도도, 열전도도 등 • 시뮬레이션에 의한 미세구조 변화 및 특성분포 (2, 3차원)

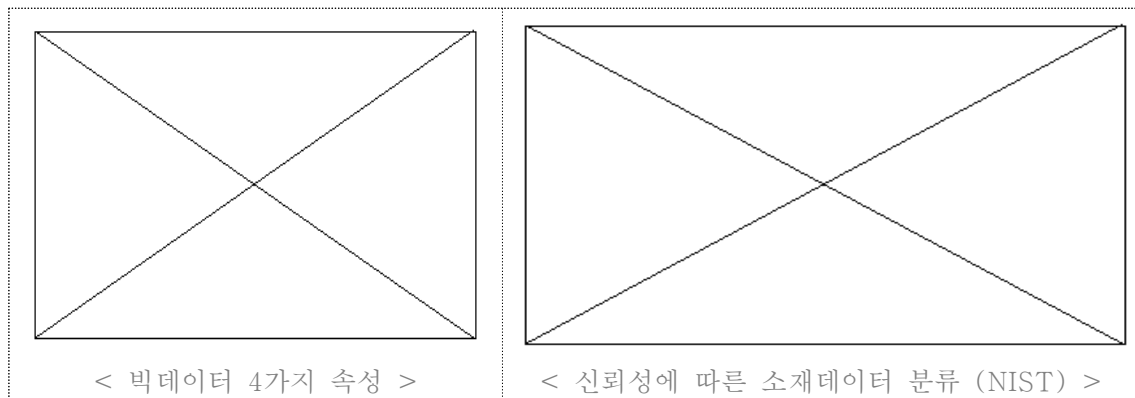
- 융합연구 확산추세에 따라 소재분야 데이터 추세를 독립적으로 추산하기 어려우나 새로운 산업적 수요에 따른 신소재 출현, 출판논문, 특허출원 증가 및 소재 연구방법론의 발전에 따라 데이터의 종류와 양이 크게 증가

※ 소재분야 SCI 논문 추세(웹오브사이언스): 약138,000건('17), 연평균 증가율 7.3%('11-'17)

※ 고속대량(High-Throughput) 실험-계산방법으로 단일 실험-계산에서 데이터 대량 생산

※ 분석·측정방법의 발달(static → in-situ → in operando)에 따른 데이터 크기·차원 증가

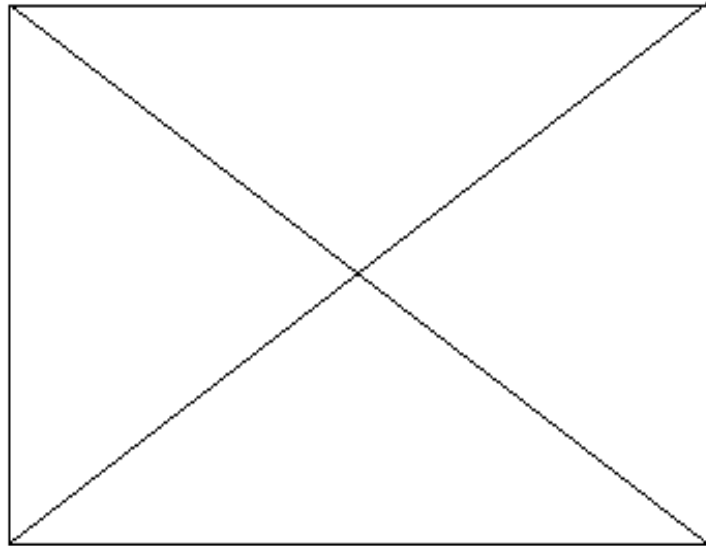
- 소재 연구데이터는 빅데이터의 4가지 속성 중, 데이터의 크기(volume) 및 속도(velocity) 보다는 다양성(variety)과 정확성(veracity)이 지배적으로 나타나 소재를 구성하는 원소특성, 제조공정, 물성 사이의 관계를 일의적으로 파악하기 어려우며, 빅데이터 구축과 데이터과학의 활용한 연구개발이 타 분야 대비 미성숙



- 4차 산업혁명 시대에 진입하면서 인공지능, 빅데이터 데이터 활용기술이 모든 연구분야에서 새로운 연구 패러다임으로 확산되면서 소재분야에서도 데이터 공유·활용이 성숙되어 있는 바이오, 지구과학 분야를 벤치마킹하여 소재 데이터에 대한 니즈 분석, 빅데이터 인프라 및 협업체계 구축과 활용을 지원하기 위한 국가적 정책수립
 - ※ (미, TMS-MRS-ASM) 소재 연구개발과 데이터관리 현황 조사('12-'13), (미, NRC) 소재 연구개발과 빅데이터 리뷰('14), (미, NITRD) 연방 빅데이터 연구개발 전략 계획('16)
- 일반적인 데이터베이스 및 소프트웨어가 저장·관리·분석 가능한 범위를 초과하는 데이터 등 크기(volume)을 강조하던 관점에서 다양한 종류의 데이터로부터 효과적으로 가치를 추출하고 신속한 분석을 지원하는 가치창출 및 활용에 중점
- 상용 및 공공 소재 데이터베이스는 대부분 표준화·규격화된 소재의 데이터 시트 또는 기초 소재물성 정보의 단순 검색·열람기능 제공에 그치고 있어, 소재 데이터 가공·활용에 의해 새로운 부가가치를 창출하는 기능은 극히 제한적
- 대량의 복잡한 소재 데이터에 잠재되어 있는 소재정보에 대한 해석기술이 개발되고 기존과는 질적으로 다른 데이터 분석기술이 소재 연구개발에 활용되면서 장기적 관점에서 국가적 소재 데이터 인프라 정비 및 정부지원의 필요성 대두
 - ※ (일, JST) 데이터과학과 연계·융합에 의한 신세대 물질재료설계연구 촉진('13)
- 소재 연구데이터를 적극적으로 이용하여 신소재 개발에 활용하기 위해서는 소재 연구개발 전 과정에서 모든 연구데이터 수집, 빅데이터화 및 지능정보기술을 활용한 연구데이터 지식자산화와 이를 활용한 연구개발 추진 필요
 - ※ 소재 데이터 플랫폼: 데이터 수집, 체계적인 저장, 공유와 연구개발에서 재활용·재생산 기능을 하는 소재 연구개발 인프라

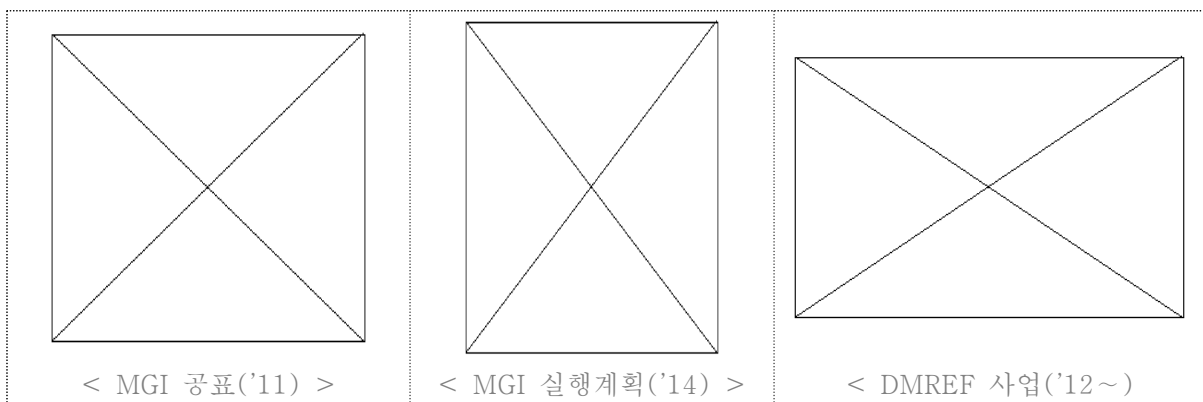
2) 소재 데이터 분야 R&D 환경 변화

- 신소재는 환경·에너지, 정보통신 및 의료 등 모든 산업의 원천으로 과거 수십 년간 각국은 소재 연구개발에 막대한 자원을 투입하여 왔으나, 신물질 발견에서 실용화 과정을 거쳐 시장진입까지 10~30년의 장기간이 소요되는 상태가 지속

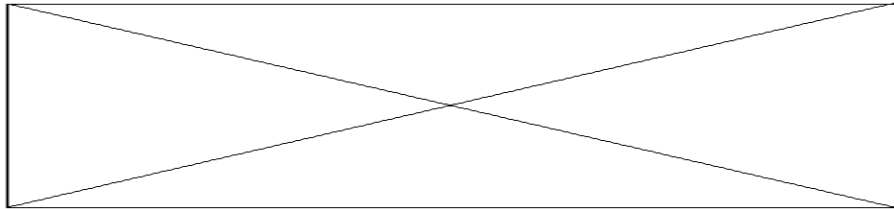


<그림 6-15> 신소재 발견, 개발, 적용 및 재활용 단계

- 소재 특성 측정·분석 기술 발달과 전산재료과학 기반 신소재 탐색·설계 연구의 활성화에 따라 실험과 계산에 의한 소재 연구데이터가 대폭 증가하여, 디지털화된 소재 연구데이터를 활용한 소재 R&D 환경 조성
- 미국에서 2011년부터 시작된 소재계놈연구전략(MGI) 및 2012년 딥러닝을 계기로 일어난 제3차 인공지능 붐에 따라 실험, 이론, 컴퓨터 모델링(시뮬레이션)에 이어 제4의 과학으로 불리는 데이터과학을 활용한 소재 연구개발이 전 세계적으로 확산

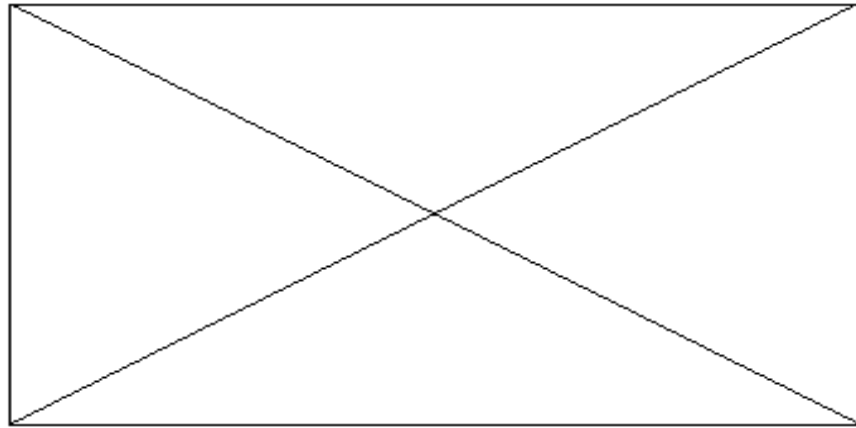


- ※ 소재계능연구전략('11): ①전산재료과학, ②디지털 데이터, ③혁신적 실험방법을 융합한 신소재 연구개발 혁신으로 소재개발 기간 및 비용 50% 절감 (5개년간 5억불 투입)
 - ※ 소재계능연구전략 실행계획('14): ①계산데이터·실험 연계를 위한 연구패러다임 변화, ②계산데이터·실험 통합, ③소재 데이터 접근 환경 개선, ④차세대 소재인력 양성
 - ※ DMREF(Designing Materials to Revolutionize and Engineer our Future): 미국과학재단(NSF)의 MGI 전용 연구사업으로, 연구대상은 소재 전분야(13개 소재군)를 포함
- 컴퓨터과학(데이터과학과 계산과학)과 소재 데이터를 융합하여 재료과학의 제반 문제를 해결하는 소재정보학(materials informatics) 활용에 의해 소재 R&D의 효율화를 추구하는 데이터 구동형 소재 개발방법론이 제4의 소재 R&D 패러다임으로 대두



<그림 6-16> 소재 연구개발 패러다임 변화

- 소재 연구개발에 활용되는 데이터과학은 복잡하고 방대한 데이터로부터 특징적인 패턴을 발견하고 유효한 지식을 추출하는 기술로 데이터 마이닝, 기계학습(머신러닝)이 대표적인 방법
 - “소재정보학”을 도입함으로써 대량의 복잡한 데이터로부터 지식(규칙)을 추출하는 귀납적(데이터구동형) 방법론과 이론·모델에 기초한 연역적(원리구동형) 방법론이 시너지 효과를 창출하여 향후 소재연구에 큰 변화를 이룰 것으로 기대
- 소재 데이터를 기반으로 한 소재정보학 활용 신소재 연구개발 주제
- 다양한 종류·형태의 대량 데이터 분석에 의한 구조-물성 관계 및 기구 해명
 - 대량의 데이터로부터 구조 또는 물성 예측
 - 최적화방법에 의한 물질(구조) 탐색
 - 멀티스케일 모델링
 - 고차원 데이터를 이용한 수리모델의 고정도화
 - 물질공간 또는 해석 데이터 가시화



<그림 6-17> 빅데이터 기반 소재 연구개발 유형

- 데이터 기반 소재 연구개발에는 다양한 분야의 전문가 및 연구그룹의 참여가 필수적으로, 소재정보학을 위한 오픈플랫폼을 구축하고 대학에서 산업체까지 모든 연구자가 용이하게 전 데이터를 활용할 수 있는 총체적 생태계 조성이 필요

나. 필요성

1) 소재 데이터 인프라에 대한 연구자 요구

- 2010년 이후 과학 전 분야에 걸쳐서 빅데이터와 오픈사이언스가 화두가 되면서 해외 소재분야 커뮤니티에서는 소재 데이터의 관리, 공유, 활용을 위한 소재 데이터 인프라 구축에 관련한 다방면의 검토가 진행되어 요구사항이 도출

－소재 데이터 수집, 저장 및 해석

- 소재 데이터(메타데이터, 프로그램 관리) 및 데이터 포맷 표준화, 소프트웨어 환경 설계 및 보급, 연합(federation)/협력(collaboration) 환경 구축
- 데이터 공유 정책에 따른 공개문헌(논문, 출판물)로부터 데이터 생성, 데이터 포맷 결정과 데이터 공개에 대한 인센티브 정책
- 소재 메타데이터/데이터 공유를 위한 출판사의 협력
- 소재 데이터는 다양한 길이 및 시간단위에서 생성되고 공정-구조-특성 관계로 확장되므로 수집 데이터 사이의 관계 및 종속성을 효과적으로 파악할 수 있는 도구
- 효율적으로 데이터 관리 및 데이터 해석 도구를 활용하고 교육자료를 제공할 수 있는 숙련된 소재분야 전문가
- 타 연구분야(IT, 데이터과학, 정보공학)와의 협력 (오픈소스 프로그램 등)

－소재 데이터 공유

- 소재분야 데이터베이스 목록 및 데이터베이스 사이트의 검색, 연결이 가능하고 데이터에 대한 신속한 접근을 가능하게 하는 데이터 레지스트리
- 지적재산권 침해 및 정부규제를 회피하면서 소재 데이터 인프라 참여를 활성화할 수 있는 비즈니스 모델

－소재 데이터 정의 및 큐레이션

- 불확실성, 이력, 출처 등 데이터의 질을 평가하고 이해할 수 있는 표준 절차 및 데이터 큐레이션 작업흐름(workflow)
- 소재 데이터 정의, 온톨로지 및 데이터 변환(translation)

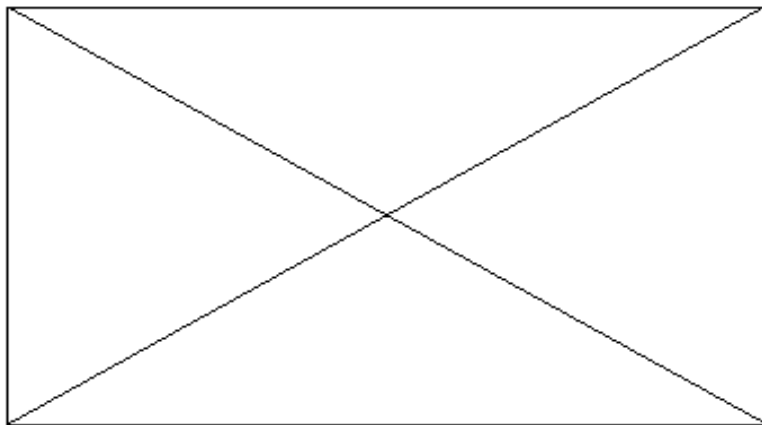
- 소재 연구 및 기술 분야에서 데이터베이스 및 데이터 마이닝 도구의 수요는 소재 분야에 걸쳐 기초 물성, 구조·미세조직에서 실사용 환경에서의 성능에 이르기 까지 망라적으로 존재

－데이터 기반 소재 연구 니즈 (3가지 소재분야로 구분)

- 구조재료: 경량소재, 고온재료, 첨단 고강도강(AHSS) 등
- 기능재료: 촉매, 전자·광전소재, 에너지/전기화학 저장소재, 강상관계 물질 등
- 유연재료: 폴리머 복합재료, 바이오 소재, 유기전자재료 등

－소재 데이터 및 데이터 마이닝 수요

※ 미국 재료학회(MRS), 금속학회(TMS)가 공동으로 실시한 “Big Data Survey(‘13)”설문응답 675건 분포: 북미(64%), 유럽(17%), 아시아(13%)



<그림 6-18> 소재 데이터 수집·활용에 대한 수요

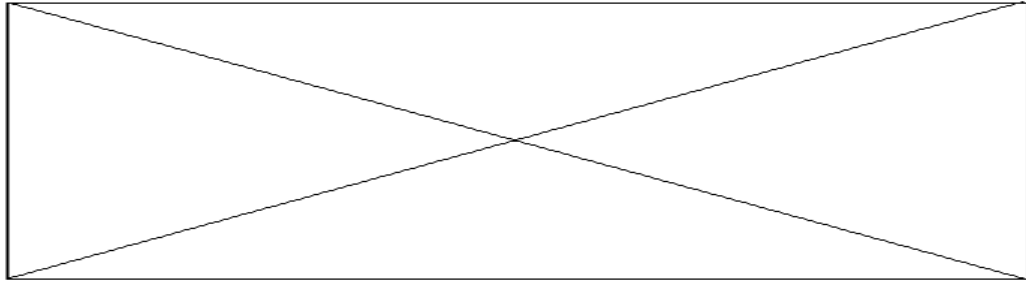
- 물리적, 열적 특성 > 화학적 특성 > 미세조직, 이미지 > 정적 기계적 특성 > 공정·합성 상세 데이터 > 전기·전자 특성 > 동적 기계적 특성 데이터 > 환경, 부식 특성의 순으로 데이터 수집·활용에 대한 수요가 큼

2) 소재 분야 데이터 전문센터 운영 현황

- 결정구조 데이터 저장소(repository)를 제외하고 기존의 소재 데이터베이스 또는 저장소는 소재 연구데이터 수집·가공·공유·활용 체계가 구축되어 있지 않으나, 각국에서는 중장기적인 관점에서 기본적으로 소재 데이터 플랫폼 구축을 진행

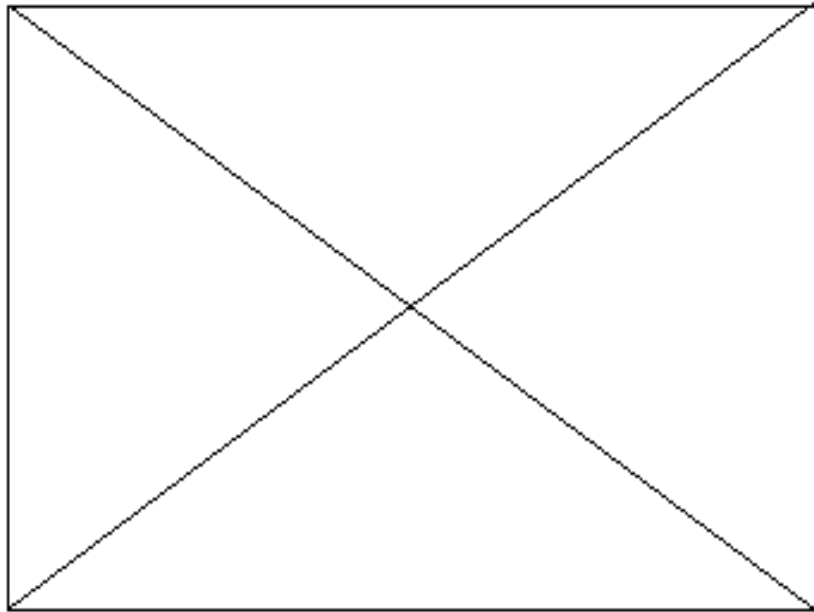
□ 미국 국립표준기술연구원(National Institute of Standards and Technology, NIST)

- 소재측정연구소(Materials Measurement Laboratory, MML), 정보기술연구소(Information Technology Laboratory, ITL)
 - MML은 화학, 생물학, 재료과학 분야의 측량측정기술 연구, 표준화와 참조표준물질 및 데이터(Standard Reference Data, SRD) 제공을 담당. MML의 하위조직인 데이터정보실(Office of Data and Informatics, ODI)은 MML 및 NIST 전체의 생물·화학·재료 과학 분야의 ①연구데이터 정비, ②연구데이터 통합, ③데이터 기반 연구지원 환경구축 진행
 - ODI에서는 연구뿐만 아니라 연구서비스를 주요 업무로 하여, NIST 내의 데이터전문가와 공동으로 FIAR 데이터 원칙에 따라 NIST 데이터전문가와 함께 데이터 발견가능성, 이용가능성, 상호운영성 적정화 지원
 - ※ ODI의 주요 수행업무: ①표준참조데이터 유지(인프라 구축 포함), ②연구데이터 관라·유지(연구데이터관리 계획 작성도구 제공, 연구노트 전자화, 오픈 데이터 레포지토리 실행 등), ③데이터과학 도구 개선(최적 데이터 분석 도구 제시 및 자문), ④글로벌 데이터 커뮤니티 참여(RDA 등)
 - ITL은 미국 정보기술 분야의 대표적인 공적기관으로 계산과학과 고도의 네트워크 기술에서 독자적인 연구개발을 진행하고, 국가계량제도 관리라는 전통적인 임무 틀 안에서 IT 계량연구를 정보기술 및 가상계량 등에 활용. 빅데이터 표준화 관련 “NIST 빅데이터 상호운용성 프레임워크, 표준로드맵” 발표('15)
 - NIST는 소재계량연구전략(MGI)의 핵심적인 수행기관으로 소재 데이터·모델의 통합, 가공(큐레이션), 제공에 관련한 전문성을 활용하여, 상호운용이 가능한 소재 데이터의 수집·표현·탐색의 표준화 및 도구 개발 진행



<그림 6-19> 소재계눔연구전략 데이터 흐름도

※ MGI에서 NIST의 역할: ①핵심 소재 데이터와 모델 공유 프로토콜 확립, ②데이터와 모델의 품질확보 방안 확립, ③소재개발을 가속화하기 위한 신연구방법, 측정법, 능력 확립, ④개발 인프라와 선행 실행사례 제공



<그림 6-20> 데이터 수집·가공·공유 및 모델링 연계 (NIST)

– NIST NMRR(NIST Materials Resource Registry)는 소재 데이터 리소스의 메타데이터 수집, 타 참여기관의 메타데이터와 공유하는 연합데이터베이스 시스템

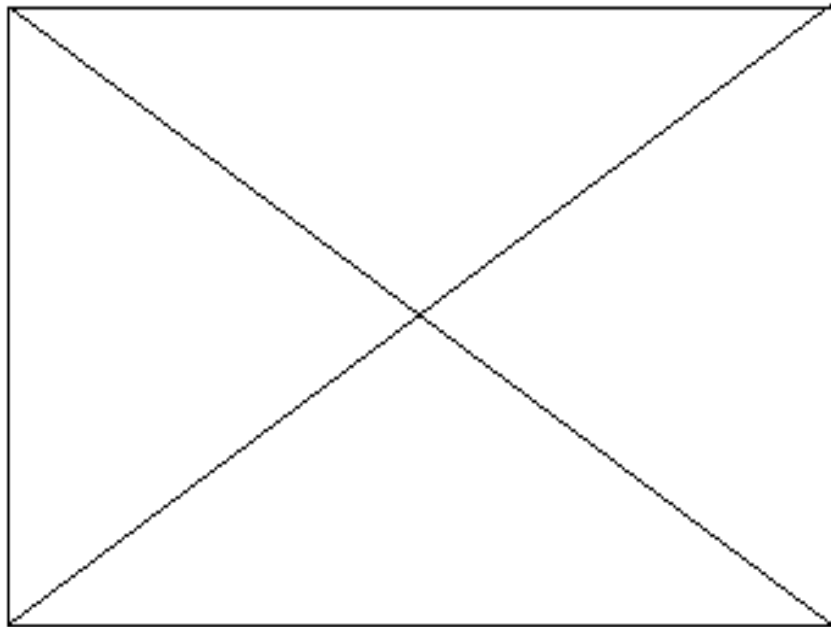
※ 130~140개 data collection이 등록되어 있고, 참여기관이 각자의 메타데이터를 관리, 갱신하여 데이터 품질 유지

– NMRR과 동일한 인프라에 계산재료과학과 소재 데이터 분석 프로그램을 포괄적으로 수집한 ‘MGI Code Catalog’를 통합하여, 데이터 정보분석 소프트웨어 참조 기능 제공

□ 일본 물질재료연구기구(National Institute for Materials Science, NIMS)

○ 통합형 재료개발·정보기반 부문(Materials Data and Integrated System, MaDIS)

- \ - 초스마트사회(Society 5.0) 구현을 모토로 한 제5기 과학기술기본계획('16)에 의거하여 2017년 NIMS 내에 IT·소재 융합연구 촉진 및 이를 지원하는 소재데이터 플랫폼 구축을 목적으로 통합형 재료개발·정보기반 부문 신설
- MaDIS 정보통합형 물질재료 연구거점에서는 응용분야(축전지, 자석 및 열전재료)별 데이터 기반 신소재 탐색·설계 연구, 소재분야에 적용되는 데이터과학 및 인공지능 기술(계산과학·기계학습 연계, 소재물성 예측 알고리즘 등)개발, NIMS 소재 데이터베이스(MatNavi) 및 연구자용 무기재료 데이터베이스(AtomWork-Adv) 구축 수행



<그림 6-21> 통합형 재료개발·정보기반 부문 플랫폼

- 재료 데이터 플랫폼 센터는 기존 과학정보실과 재료정보스테이션이 통합되어 설립되었으며('17), NIMS가 추진하는 혁신적 재료개발능력 강화 프로그램의 하나인 "Materials Research Bank"내에서 재료 데이터 플랫폼 구축과 운용을 담당하고 있으며, 재료 데이터 수집에서 서비스까지 일원화 시스템 구축 진행 중
- (데이터 수집) 실험·계측장비 데이터 수집, 문헌 데이터 수집, 기관 레포지토터리 데이터 수집 등 다수의 데이터원(源)으로부터 고속(high-throughput)·양질의 데이터 수집·등록 시스템 구축 및 정보공학적 방법으로 데이터간 관련성 부여, 저장 데이터의 고부가가치화 수행
- (데이터 서비스) 데이터 가치화 등 각종 데이터 해석 소프트웨어 정비·제공뿐만 아니라 소재개발에서의 문제해결 서비스 및 종래의 도서관 기능, 출판기능, 리포지토리 기능을 연계한 소재 정보 제공 계획

－ NIMS 재료 데이터 플랫폼 센터의 연구개발 내용

- ※ 재료 데이터베이스 그룹: ①지속가능한 데이터베이스 구축(텍스트 마이닝 기술 등)과 품질관리, ②대량 데이터의 정리·통합(메타데이터 기반), ③데이터 수집과 데이터베이스의 직접 연결(인공지능 데이터베이스 개발), ④연구데이터 큐레이션 및 공유재산화
 - ※ 재료 데이터 해석 그룹: ①(수집) 소재 데이터 수집·전송기록, ②(가독화) 소재 데이터의 서식 및 어휘 정형화, ③(가시화) 소재 데이터 도시, ④(고부가가치화를 위한 전처리) 데이터의 수치처리를 통한 참고 데이터의 정보도 이용한 실험 데이터의 신속한 고부가가치화
 - ※ 재료 통합(integration) 그룹: 공정-조직(구조)-특성-성능 관계를 전산 기반으로 수행하는 통합시스템 구축
- 해외에도 기존의 특정한 목적(결정구조 데이터베이스, 제일원리 기반 계산과학 데이터베이스 등)으로 구축된 소재 데이터베이스 및 연구사업에 의해 한시적으로 구축되고 있는 소재 관련 데이터베이스는 다수 존재하나 범용적인 소재 연구데이터의 공유·활용 기능을 수행하는 전문센터는 극소수이고 인프라 구축이 진행 중인 상태
 - 국내에는 미국 NIST, 일본 NIST와 같이 소재 데이터의 수집·공유·활용 기능을 전문적으로 수행하는 전문센터가 구축되어 있지 않으며, 소재분야별(금속, 화학, 세라믹, 섬유) 산업용 소재 정보를 수집하여 검색, 열람 기능을 제공하고 있는 소재정보은행 및 한국표준과학연구원 참조표준물질 데이터베이스가 대표적인 소재 데이터베이스
 - 소재 데이터 수집·공유·활용에는 장기적 관점에서 국가차원의 지원이 필수적으로, 하드웨어 인프라 정비·공유뿐만 아니라 다양한 소재 데이터를 활용 가능한 소프트웨어 인프라의 정비·공유가 시급한 상황

6.4.2. 환경분석

가. 국내외 정책 및 시장 현황

1) 연구데이터 공유·개방 정책

□ 과학기술 전분야

-
- ◆ 오픈사이언스 정책 확산으로 연구개발 사업의 성과물의 의무적 공개·공유
 - ◆ 융합·공공·협력 연구 확대, 데이터 중심 연구개발 확산 등 디지털 연구데이터 이용 새로운 연구 개발 활동 창출 추세
-
- 연구성과과정을 개방하는 오픈사이언스 정책의 전 세계적인 확산에 따라 각국은 공적자금 지원되는 연구개발사업의 성과물을 의무적으로 공개하고 공유하는 정책 수립
 - － 융합연구 및 공동·협력연구 확대, 데이터 중심 연구개발 확산 및 인터넷을 통한 데이터 활용 도구가 증가함에 따라 디지털 연구데이터를 통한 새로운 연구개발 활동이 창출되는 추세
 - － (미국) 백악관 과학기술정책국(OSTP)에서 2013년 ‘연구데이터 관리 및 공유에 대한 지침’을 통하여 오픈사이언스를 선제적으로 추진
 - ※ 공적자금이 투입된 연구성과물(연구데이터)의 연구기관별 관리·공개(12개월의 공개유예 기간 이후 공개) 및 연구과제 제안서에 데이터관리계획(DMP) 제출 의무화
 - ※ 인센티브 정책으로는 ①연구데이터의 디지털 문서 식별자(DOI) 부여, ②DMP 제출에 따른 인프라 부담 경감, ③데이터 인용에 대한 평가체계 확립, ④연구계획서 제출 시 데이터 공유실적 기재, ⑤대학연구기관의 데이터관리자의 연수기회 제공 등
 - － (EU) 디지털 단일시장(Digital Single Market), Horizon 2020 등 주요정책 시행에서 오픈사이언스를 명시하고, 오픈사이언스정책플랫폼(‘16)에 의한 일원화된 정책 시행 및 오픈사이언스 클라우드(European Open Science Cloud) 구축 진행 중
 - ※ Horizon 2020 사업에서 연구데이터 오픈액세스 의무화: 데이터관리에 대한 가이드라인(‘16), 발표논문 및 연구데이터의 오픈액세스에 대한 가이드라인(‘16)
 - FAIR(Findable, Accessible, Interoperable, Reusable) 데이터관리 가이드라인 및 DMP
 - ※ OpenAIRE 프로젝트: Horizon 2020의 프로젝트로 EU의 50개 기관이 참여하여 Horizon 2020의 연구성과 수집·공유·재활용 추진, Horizon 2020 연구성과 논문 오픈액세스 지원

□ 소재 분야 연구데이터 공유 정책

◆ 국가(정부) 주도 소재분야 연구데이터 수집 전략 필요

- 소재 분야의 연구 및 연구데이터는 산업경쟁력에 직결되므로 기밀성과 상업적 가치가 높은 데이터의 공개가 이루어지지 않는 경우가 많고, 소재 데이터에 대한 공개·비공개가 전략적인 국가정책으로 수립
- 미국 표준기술연구원(NIST)는 과학기술정책국(OSTP) 정책('13)에 따라 국비에 의한 연구성과(연구논문 및 데이터)의 공공접근에 대한 무상공개와 생산적인 재이용을 추진하는 "연방자금 지원 연구결과의 공적접근 계획('15)" 발표 및 시행
 - NIST 소속 연구자에 의한 연구성과 및 NIST로부터 전액 또는 일부 연구비를 지원받은 연구성과에 대하여 논문은 출판 후 12개월 이내에 최종판 공개. 연구데이터도 논문과 동일한 형태로 취급
 - ※ 심사과정이 있는 논문은 NIH PMC(PubMed Central) 리포지토리, NIST 보고서와 심사과정이 없는 논문은 정부인쇄국의 govinfo에서 공유
 - 제조업 관련 연구과제는 소재와 연관성이 크므로 대부분의 기업은 데이터 공개에 동의하지 않는 경우가 많으며, 이 경우 데이터 취급에 대한 합의문서 작성
 - NIST는 데이터공개전략(NIST Open Data Strategy) 공개 및 NIST가 관여하는 연구성과에 공공의 접근을 확대하는 것을 명시. 데이터관리계획은 2014년 10월 이후 NIST의 전 조직에 적용되고 있고, 데이터관리계획 작성 지원 도구 진행
 - ※ 공공 공개가 결정된 데이터는 피라미드형 계층구조의 7 종류의 데이터 중에서 표준참조데이터와 출판 데이터이며, 5 종류의 데이터군에 대한 공개는 미정
 - 데이터공개전략(NIST Open Data Strategy) 공개 및 NIST가 관여하는 연구성과에 공공의 접근을 확대하는 것을 명시. 데이터관리계획은 2014년 10월 이후 NIST의 전 조직에 적용되고 있고, 데이터관리계획 작성 지원 도구 진행
 - 데이터 공유 기반으로는 2002년에 구축된 NIST Data Gateway를 기점으로 ①NIST 연구데이터 및 논문검색 서비스, ②NIST 표준참조데이터(Standard Reference Data, SRD), ③무료 및 유료 온라인 데이터베이스 등
 - 미국 소재계놈전략 실행계획의 최우선 순위로 데이터 공유에 관련한 커뮤니티 문화의 변화를 정하고 있고, 이를 위하여 NIST에서는 데이터셋의 구조화 및 재이용을 가능하게 하는 Materials Data Curation System(MDCS) 플랫폼 구축하고 수집·공유한 소재 데이터의 구조화 시 XML 포맷으로 구조화하는 데이터 리포지토리 사이트 Materials Data Repository(MDR) 구축

- 일본 물질재료연구기구(NIMS)에서는 2003년부터 세계 최대 규모의 온라인 재료 데이터베이스인 "NIMS 물질재료 데이터베이스(MatNavi)" 운영하고 있고 2010년 사용자 인터페이스를 전면적으로 개선하고 데이터베이스를 추가하여 사용가능한 데이터베이스 구축을 목표로 하고 있음

<표 6-9> 각국의 오픈사이언스 관련 연구데이터 정책

국가	주요정책
미국	• 정보자유법('66), 미국경쟁력 강화법('10), 백악관 '연구데이터 관리 및 공유에 대한 지침'('13, OSTP)으로 공공데이터의 오픈엑세스 원칙 제시 • 연방정부는 공공 연구성과가 디지털 형태로 공개되어 열람·활용할 수 있도록 연방기관들이 관련 정책을 마련할 것을 요청 • 공적자금이 투입된 연구데이터는 연구기관별(1억달러 이상) 관리·공개*, 연구기획단계에서 데이터관리계획(DMP) 제출 의무화
영국	• 정보자유법에서 공개정보의 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 • '14년 출판법(Copyright, Designs and Patents Act)을 개정하여 기공개된 데이터를 비상업적 목적으로 복사, 데이터마이닝할 수 있도록 허용
독일	• '10년 출판법(Copyright Act, UrhG)을 개정하여, 공적자금으로 지원된 연구성과에 대해 연구자가 출판사에 출판권을 이전한 경우에도 일정기간의 유예기간 경과 후 비상업적·교육적 용도로 2차 출판하는 것을 인정
프랑스	• '16년 디지털공화국법(Law No. 2016-1321 on a Digital Republic)을 통해, 공공연구에서 생산된 학술논문의 경우 연구자가 출판사에 논문을 게재한 경우에도 출판이후 6개월-1년 엠바고 기간 경과 후 자신의 논문을 비상업적 용도로 배포할 수 있는 권리를 부여
호주	• 정보자유법의 공개대상에 모든 정부정책 관련 데이터를 포함, '책임 있는 연구수행을 위한 국가지침'에 연구산출물 관리 규정 마련 • 주 연구위원회(ARC)의 재정지원을 받는 연구기관에 대하여 의무적으로 데이터관리계획 제출 규정('14)
일본	• 과학기술진흥기구(JST)과 일본학술진흥원(JSPS)이 연구지원기관으로 '13년 이후 지원한 공공 연구성과를 공개하도록 하는 오픈엑세스정책 도입 • 과학기술진흥기구(JST)는 '13년 '공공자금으로 지원된 연구의 성과에 관한 오픈엑세스 방침'을 공표, '17년 공공연구로 생산된 연구데이터에 대한 데이터관리계획(DMP) 수립을 권고하는 오픈사이언스정책 발표 • 일본학술진흥원(JSPS)도 오픈엑세스규정을 마련하고 있으나 강제성 없는 권고안이며, 일본의료연구개발기구(AMED)는 최근 일부 프로그램에 DMP를 시범적으로 적용하기 시작('17)

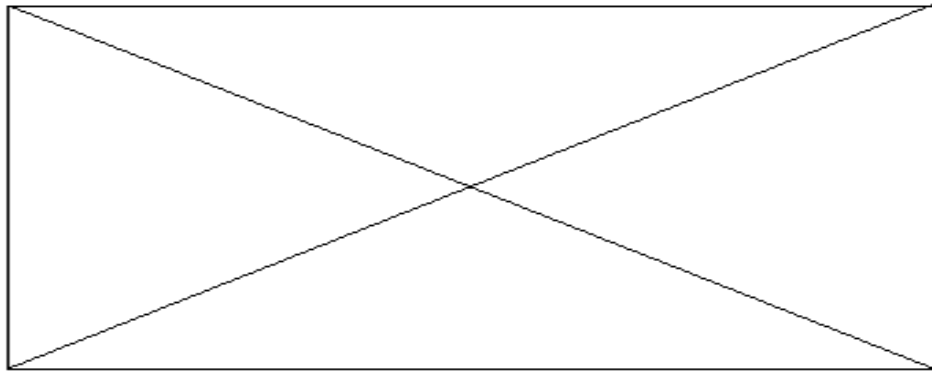
- 2012년 연구 데이터 공유·교환을 촉진하고 데이터 구동형 혁신을 가속화하기 위한 인프라 개발과 커뮤니티 활동을 추진하는 회원(정부기구, 국제기관 및 대학) 중심의 국제기구인 RDA(Research Data Alliance)가 창설

※ 137개 국가에서 7,180명 회원 참여, 45개 기관회원: 대학연구기관(67%), 공적기관(15%), 기업·산업체(11%)

※ 93개 WG(Working Group)과 61개 IG(Interest Group)이 인프라, 정책 및 실행계획, 규격개발과 채택 및

데이터 상호운용성 확보를 위한 활동

- 소재분야에는 1개의 WG "International Materials Resource Registries"와 1개의 IG "RDA/CODATA Materials Data, Infrastructure and Interoperability"이 구성
- 데이터 과학데이터(Scientific Data)에 대한 관심이 고조되면서 글로벌 출판사의 과학데이터 저널을 중심으로 데이터 관련 논문게재가 확산되고 있으며 데이터 공유를 통한 재료과학의 혁신을 가속화하는 흐름에 참여
 - ※ 과학데이터 저널 및 데이터베이스 등을 통해 공유되는 데이터는 바이오 분야(70%>)가 가장 많고, 소재/화학 분야(<10%)는 낮은 편임
- Nature Publishing Group의 오픈엑세스저널인 Scientific Data Journal('14~)은 peer-review 데이터 논문 게재
- Elsevier社は 13개의 소재분야 학술지를 당사의 새로운 소재 데이터 공유정책 (Elsevier's new Materials Science data-sharing initiative)에 참여시키고, 상기 학술지에 논문 투고 시 데이터 공개 및 재활용을 위한 3가지 옵션 제시
 - ※ ① Interactive Plots(iPlots): 그래프 데이터의 데이터 파일을 제출하여 사용자가 데이터 해석 가능, ② 논문과 함께 원시데이터를 함께 제출 시 에디터 및 리뷰어의 검증을 거친 후 다운로드 및 재사용 가능 (Creative Commons CC BY license), ③ 연구결과를 저널에 투고 시 연구결과 및 메타데이터를 소논문 형태로 업로드하면, 소논문은 오픈엑세스저널인 "Data in Brief"에 자동적으로 제출되는 과정(Creative Commons CC BY license)



<그림 6-22> Elsevier 데이터 공유정책에 참여하고 있는 소재분야 13개 학술지

2) 시장현황

□ 소재 데이터 기반 연구개발 및 연구거점 형성

- 최근 데이터(수학·정보)과학 분야에서는 대량의 복잡한 데이터에 잠재되어 있는 정보구조를 추출하는 해석기법이 비약적으로 발전하여 종래와는 질적으로 다른 소재 데이터의 과학적 분석이 가능
- 데이터 구동형 소재 연구개발을 추진하기 위해서는 이를 중추적으로 수행할 데이터 통합·연구개발 센터를 구축하여, 데이터관리 방침, 수집 데이터 종류·범위 결정 및 데이터와 데이터 관리·해석 도구 관리 개발이 필요하며, 데이터 공유 서비스 기능을 위해서 LOD(Linked Open Data) 개념을 도입하고 RDF(Resource Data Framework)/ XML 등 표준화된 포맷을 사용하여 데이터 교환 및 소재정보학 시스템 구축 필요
- 각국에서는 소재계몽연구전략 이후 데이터 구동형 연구개발을 적극적으로 추진하고 있으면 이를 위한 연구거점을 형성

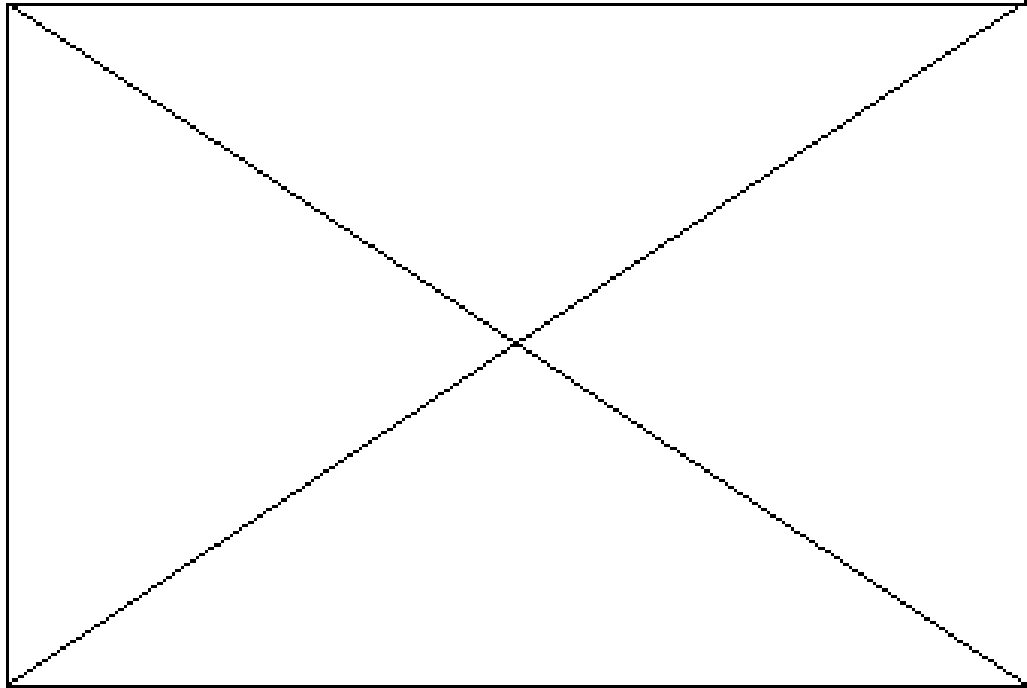
<표 6-10> 각국의 소재 데이터 기반 연구개발 정책 및 연구거점

국가	개요
미국	• 2011년 소재개발의 단기화·저비용화를 목표로한 소재계몽연구전략(MGI) 발표 이후 종래의 실험, 계산뿐만 아니라 데이터의 중요성을 인식하여, 2014년 MGI 전략계획 공표, 2016년까지 약 5억 달러 투자 • 표준기술연구원(NIST), 과학재단(NSF), 에너지성(DOE), 국방성(DoD) 등 관련 프로젝트 추진으로 연구센터 설립 – CNGMD(LBNL, UC Berkeley 등), SUNCAT(스탠포드대, SLAC), CHiMaD(NIST, 노스웨스턴대, 시카고대 등), PRISMS(미시건대) • 소재 데이터 레포지토리를 각 연구센터에 설립
EU	• EU CoE 프로그램으로 NoMAD(Novel Materials Discovery) 프로젝트('15~'18, 500만 유로 투자, 막스플랑크협회 주도). • 스위스 MARVEL(Materials Revolution: Computational Design and Discovery of Novel Materials) 프로젝트 추진('14~'18, 약 200억원 투자, EPFL 주도)
아시아	• (중국) 2016년 국가중점연구개발계획의 일환으로 "재료계몽공학 기술지원 플랫폼(약 500억원/5년)" 개시. 상해, 북경 소재 대학을 중심으로 연구거점 구축 – 상해대 "Materials Genome Institute"('15), 상해교통대 "소재계몽공동연구센터"('16) – 중국과학물리연구소, 북경과기대, 중국강(鋼)연구집단 유한공사 등이 공동으로 "북경재료과학 이노베이션 센터"('17) 설립 • (한국) 2015년부터 계산과학, 조합실험법, 첨단측정 등 ICT 기반 신연구방법론을 활용한 신소재개발사업 개시(현재 23개 연구사업단)

나. 관련 기술동향 분석

1) 소재 데이터 수집·관리·활용 글로벌 동향

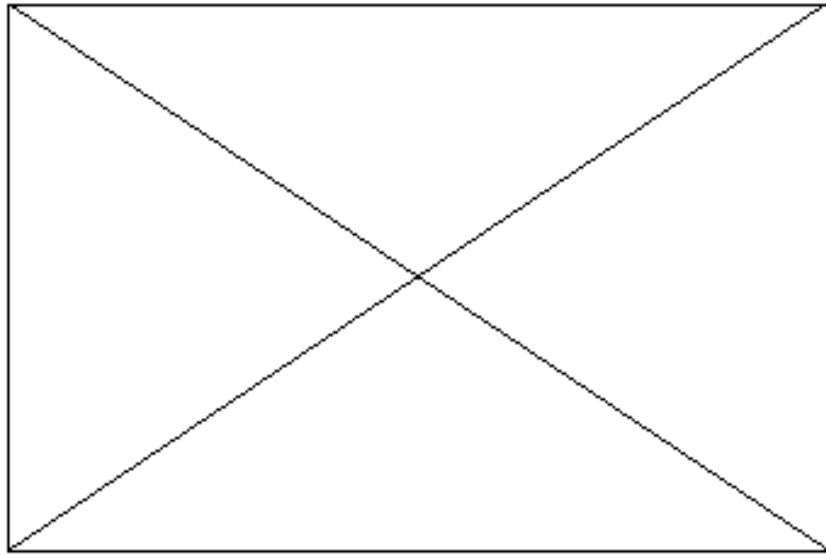
-
- ◆ 데이터 수집·관리의 중요성을 인지하고 국가차원의 지원 및 육성정책 추진
 - ◆ 미국, 일본, 유럽 등 데이터 수집·활용, 플랫폼 구축 등 자국 연구역량 강화
 - ◆ Registry Federation을 통해 상호기록 교환 방식 네트워크 구축 개념 도입 중
-
- 국제 소재 연방 네트워크(Federated Material Registry Network) 관련 글로벌 동향
 - 연합 아키텍처 기반 레지스트리 중재 데이터 검색 프로세스 모델 (천문학 영역의 Virtual Observatory 사용 모델)
 - 레지스트리 생성수준
 - 레지스트리는 데이터 기반 연구에 유용한 데이터 저장소, 아카이브, 웹사이트 및 서비스 같은 리소스에 대한 상위 수준의 메타데이터 (설명) 수집
 - 레지스트리는 메타데이터에 (자원설명) 대한 검색을 가능하게 만듦(기록을 공유하고 통일성 있는 시스템을 구축하여 빅데이터 한계 극복)
 - (특징) 마스터 레지스트리가 아닌 레지스트리 네트워크 보유
 - (작동방법) 레지스트리간 공통의 메타데이터 교환 프로토콜 사용하여 다른 레지스트리의 리소스 설명을 레지스트리 도입 후 알려진 모든 자원에 대한 리소스 설명을 종합적으로 수집 후 다음 컬렉션 검색 가능
 - 글로벌 Federated Registry Framework의 연결수준
 - 마스터 레지스트리를 만들지 않으며, 여러 개의 레지스트리들이 함께 구동할 수 있는 네트워크로 구축
 - 레지스트리 연합은 개방적이며, 모든 조직은 자신의 자원을 세계에 광고하는 수단으로 레지스트리를 호스팅
 - 레지스트리 제공자는 특정자원의 하위 세트에 대한 설명 레코드를 작성하고 큐레이션
 - 소량의 데이터 입력용, 데이터베이스 입력용의 두 가지 형식의 레지스트리가 존재하며 상호연동 가능
 - CHiMaD : 특정 하위 커뮤니티에 대한 레코드 관리 및 사용자 정의된 API를 통해 CHiMaD 센터 회원 조직이 제공하는 데이터 공유
 - NIST: 일반적으로 GUI 인터페이스를 통해 대규모의 커뮤니티와 데이터 공유 가능하며, 작은 단위의 데이터베이스 저장가능



<그림 6-23> Registry 연합 사이의 데이터 상호 교환

□ (미국) 국가적 소재 연구데이터 혁신으로 소재연구 생태계의 연결고리 강화

- (개요 및 경과) 1960년대부터 데이터의 중요성을 인식하여 왔으며, 2011년 MGI 발표 이후 소재분야 빅데이터 연구 패러다임 변화 주도
 - ※ 신뢰성이 검증된 표준참조데이터베이스(SRD) 구축을 시작으로 디지털 데이터를 생산·활용하는 연구 집중 지원
 - ※ 2005년 디지털 데이터 수집: 디지털 데이터 중요성과 파급효과 인식하고 데이터의 체계적 수집 및 관련 교육 시스템 구축 권고
 - ※ 2008년 ICME 활성화: 소재개발 효율성 확대를 위해 NRC(National Research Council)의 ICME(Integrated Computational Materials Engineering) 전략
 - ※ 2011년 MGI 발표를 계기로 단순 참고자료 수준이 아닌 연구데이터의 체계적인 수집·공유·유통 및 활용으로 이어지는 확산적 연구 활성화
- 소재 전 분야 연구데이터 수집·가공을 통한 빅데이터화 및 소재 연구개발 효율성 향상을 위한 다양한 연구를 추진하고 있으며, NIST를 중심으로 협력연구를 통해 데이터 수집·검색 서비스 제공, 데이터 활용 소재개발 및 빅데이터 인프라 구축 추진
 - (NIST 역할) 체계적으로 데이터를 광범위하게 수집할 수 있는 데이터 관리용 양식 프로그램 개발하여 무료 제공(Materials Data Curation System, MDSCS)



<그림 6-24> 미국 NIST의 소재 데이터 수집·가공·공유 시스템

○ (데이터 저장소) NIST Data Repository

－ 범용 데이터와 소재 데이터 관련 2가지 저장소 구축

※ NIST Repositories (<https://nist.matdl.gov>)

- NIST 계산 데이터 파일 저장소
- NIST 실험 데이터 저장소
- NIST 구조재료 선행연구(demonstration) 프로젝트 (ASM 공동) 저장소
- RDA 선행연구 프로젝트(DTR/PID & MGI infrastructure) : 일반인 제한
- 오픈엑세스 저널 (Integrating Materials and Manufacturing Innovation) 사이트

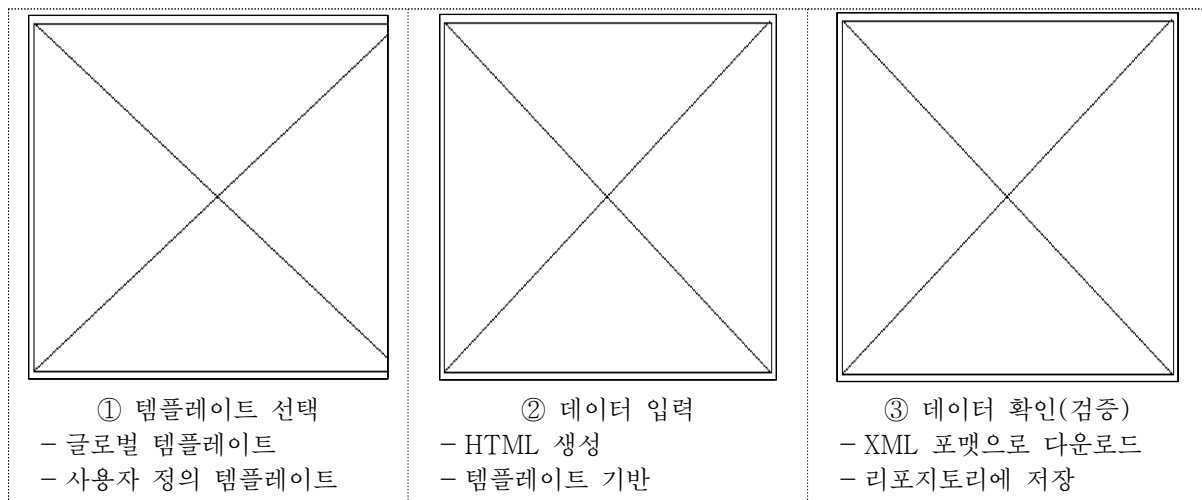
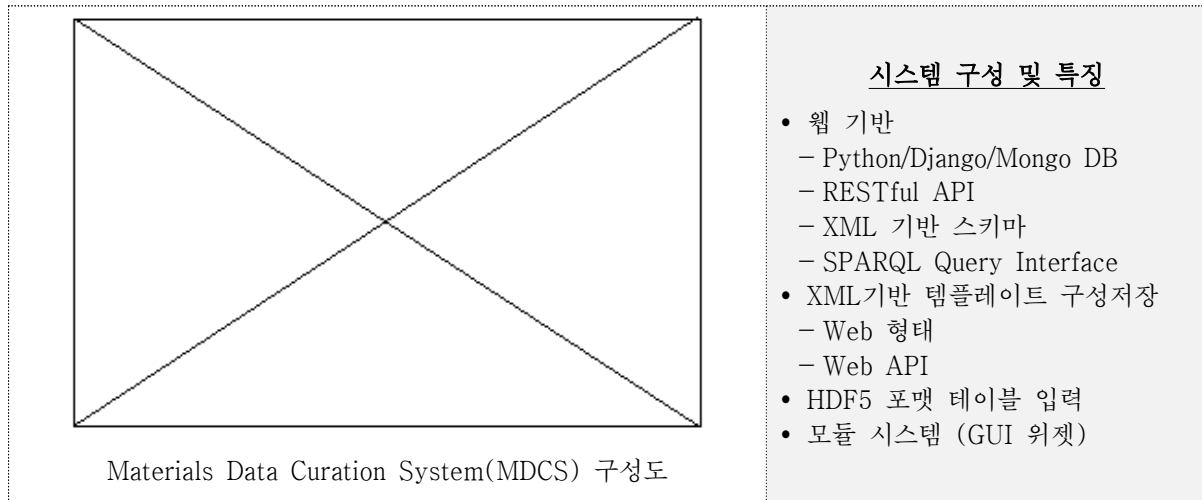
※ NIST Materials Data Repositories (<https://materialsdata.nist.gov>): MGI 관련 수집 데이터

※ 데이터 수집은 출판된 공개 문헌의 입력, 고속(high-throughput) 계산 물성 데이터의 자동수집 및 측정 데이터의 자동수집으로 구성

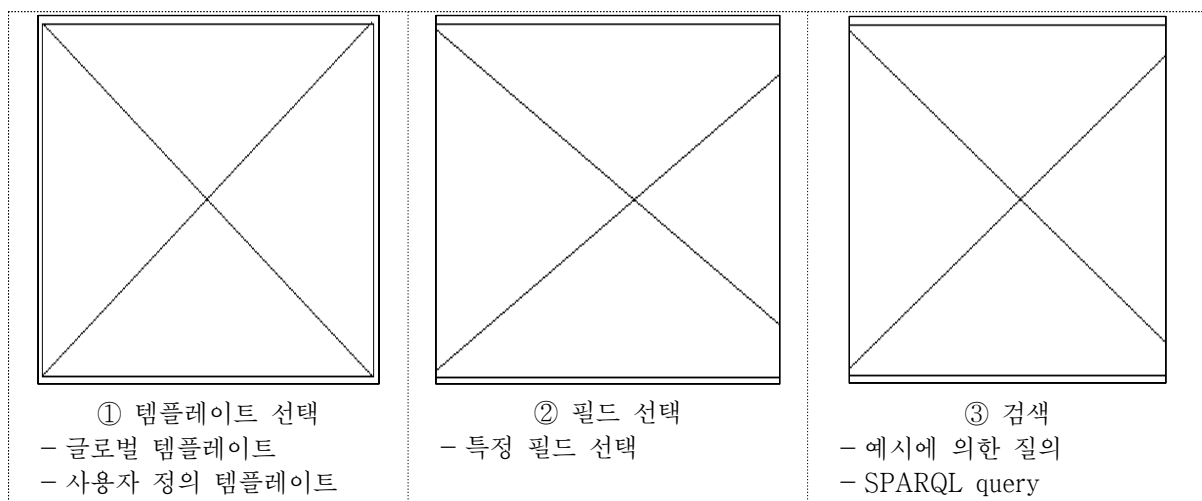
○ (데이터 수집·저장·검색·출력) Material Data Curation System (MDCD)

- － (개요 및 경과) 데이터 수집을 위한 메타데이터의 기록·검색을 위한 포맷 결정을 위해 Material Data Curation System(MDCS) 개발('15)후 지속적 업그레이드 진행
- － (목적) NIST는 GitHub를 통한 무상배포(<http://github.com/usnistgov/MDCS>)를 통해 소재 연구데이터 수집단계에서부터 체계적 관리 유도. 향후 빅데이터 형태 전환 대비, 범국가적 소재 연구데이터 관리 체계 확립 확대 추진

\



<그림 6-25> MDCS 데이터 큐레이션 과정



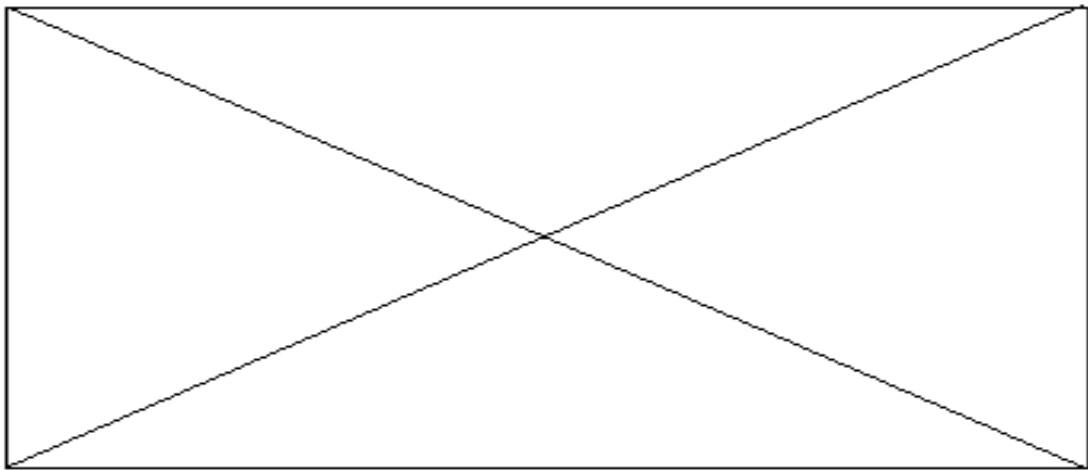
<그림 6-26> MDCS 데이터 검색 과정

○ (데이터 공유) Material Resource Registry (NMRR)

－ (개요 및 경과) 재료과학 데이터 리소스의 메타데이터를 수집하고 타 참여기관의 메타데이터와 공유하는 연합데이터베이스 시스템 구축

※ 130-140개 data collection이 등록되어 있고, 참여기관이 각자의 메타데이터를 관리, 갱신하여 데이터 품질을 유지

－ NMRR과 같은 인프라에 계산재료과학과 소재 데이터 분석에 관한 프로그램을 포괄적으로 수집한 ‘MGI Code Catalog’를 통합하여, 데이터 정보분석 소프트웨어 참조 기능 제공



<그림 6-27> NIST Materials Resource Registry 웹사이트

- 광범위한 재료 연구자 커뮤니티에서 데이터 공유재사용 촉진
- 실험/계산 구분 없이 모든 데이터(주로 문헌자료 기반) 축적
- 데이터 제공 기관, 재료/물성/합성/공정 종류별로 데이터 검색 가능
- 대부분 공공 서비스로 제공하나 일부 데이터는 접근 제한

○ (기타) Citrination Materials-Data Analytics Platform

－ Citrine Informatics社에서 무료로 공개한 재료 데이터 플랫폼

※ 논문, 특허, 기존 DB 등 대량의 다양한 형태의 데이터(정형, 비정형) 망라하고, 실험·계산데이터 모두 데이터베이스 포함

※ 수집된 데이터를 이용해 기계학습을 수행하는 기능을 내장하여 DB 내에 존재하지 않는 소재물성을 예측하는 기능을 제공하고 최종적으로는 연구자들이 원하는 물성의 소재를 탐색하는 “역방향 소재설계”기능 추가

○ 계산과학 기반 소재물성 생성·공유 플랫폼 구축

<표 6-11> 응용분야 소재연구를 위해 구축된 소재 데이터베이스 (허브기관에 해당)

구분	세부 내용	비 고
Materials Project	제일원리 기반(DFT) 소재물성 계산 데이터를 DB화하여 제공	무기물 약 84,000건 밴드구조 약 52,000건 분자 약 22,000건 나노다공질 약 530,000건 등
OQMD Open Quantum Materials Database	제일원리 기반(DFT) 열역학 및 구조 계산 결과 제공	563,000 건
AFLOW Automatic-FLOW for Materials Discovery	제일원리 기반(DFT) 소재물성 계산 데이터를 DB화하여 제공	186만개 화합물 1억8천만 건 계산 물성
NREL MatDB National Renewable Energy Laboratory Materials Database	광전, 광전자, 물분해, 열전재료 등 신재생에너지 분야 소재연구에 특화된 계산물성 DB	
DREAM.3D	3차원 미세조직 데이터 분석·디지털화 및 미세조직 시뮬레이션 초기 데이터로 활용하기 위한 가공 기능	오픈소스 프로그램
GitHub, Inc.	디지털 문서, 데이터 포맷, 개발 프로그램 및 소프트웨어 공유 플랫폼	MGI 관련 다수의 프로그램 저장·보급

<표 6-12> 미국 기관별로 구축된 소재 데이터베이스 사례

데이터베이스	분야	수집 데이터 종류	유료/무료
American Mineralogist Crystal Structure Database	광물, 문헌값	구조 정보	무료
CINDAS High Performance Alloys Database	항공우주 재료 (합금), 계산/실험	열안정성 기계적 강도	유료
DOE Hydrogen Storage Materials Database	수소 저장, 계산/실험	수소 저장 용량	무료
Harvard Clean Energy Project	유기물 태양전지, 계산	태양전지 관련 물성	무료
Thermoelectrics Design Lab	열전소자, 문헌값	열전성능 변수	무료

□ (일본) 소재 데이터베이스 기반 구축 및 연구데이터 활용 촉진

- (개요) NIMS가 운영·관리하는 소재 데이터 통합사이트 "MatiNavi"를 통해 물질·재료데이터베이스 서비스 제공
- (추진경과) 2001년 NIMS에서 Materials Information Technology Station(MITS) 조직 구성, 데이터베이스 인프라 구축하여 2003년 서비스 시작
 - ※ (2003년_데이터 공개) NIMS 인터넷에 Materials Database 서비스 공개
 - ※ (2004년_사이트 구축) 검색엔진을 탑재한 'MatNavi 사이트' 공개
 - ※ (2013년_계산과학 기반 DB) 제일원리 계산과학 기반 소재 물성 DB 구축
 - ※ (2015년_플랫폼 구축) 소재개발에 빅데이터를 활용한 연구사업 "Materials Information Integration Initiative(MI2I)" 개시
- － (특징) 주요 소재 데이터는 발표된 학술논문, 공개문헌을 통해 수집·구축
 - ※ 대부분 물성 값들은 '검색'의 결과로서만 데이터 값을 제공하며, 디지털 데이터로 제공하지 않아 활용하는데 한계가 존재
 - ※ 고분자, 무기 재료, 금속 재료, 초전도 재료 등 11 개 데이터베이스, 4 개의 응용 시스템, 6 종의 구조 재료 데이터 시트로 구성
 - ※ 빅데이터 기반 기계학습을 전제로 구축되어 있지는 않아 텍스트 마이닝을 이용한 논문 정보를 자동으로 디지털화하는 기술개발 추진 중
 - ※ 사용자 등록으로 무료열람이 가능하며, 등록자 수는 12 만명 이상 (해외이용자 28% 차지)
 - ※ 2018년 연구기능을 탑재한 무기물 데이터베이스(AtomWork-Adv) 유료 버전 출시

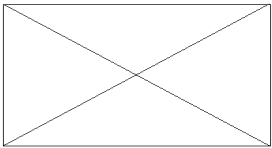
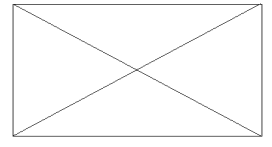
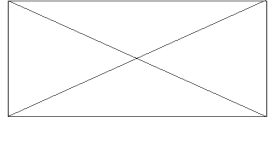
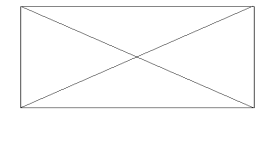
<표 6-13> MatNavi, AtomWork 데이터베이스 주요 내용

명칭	분야 & 특성	데이터 크기
Polymer Database (PoLyInfo)	- 고분자의 구조 및 물성 - NMR 데이터(154 samples)	333,583 개
Inorganic Material Database (AtomWork)	문헌상에 보고된 무기물, 금속소재 결정구조, 물성, 상태도, XRD, 화합물 물성별 (기계적 특성, 광학특성) 검색 및 가시화 기능	28,000 결정구조, 55,000 재료 특성, 15,000 상태도
Computational Electronic Structure Database (CompES-X)	- AtomWork의 구조에 따른 제일원리 전자구조 계산 물성 데이터베이스 - 제일원리 계산 입력값 및 최적화 구조, DOS, 밴드구조, 전하분포 등 - 문헌에 보고된 계산 값의 데이터베이스 - VASP 계산 결과 데이터베이스	80,000 가지 화합물

□ (유럽) 계산과학 기반 소재물성 데이터 생성·공유 기반 구축

- (개요) 소재 연구데이터 개방을 통해 접근성, 활용성 극대화 및 '07년 FP7을 시작으로 계산과학 기반 모델링 연구 활성화
- 계산 나노재료과학 기반 모델링과 정보학 연구 활발히 진행
 - ※ Horizon 2020('14~'20)의 NMP(Nanoscience, Nanotechnology, Materials, and New Production technologies)프로그램, European Materials Modeling Council('14 설립) 등 모델링 기반 소재 연구개발의 적극적 추진
- NoMAD(Novel Materials Discovery)가 대표적인 소재물성 데이터베이스로, 제일원리 계산 주요 프로그램의 입출력 파일 수집·저장
 - ※ NoMAD: Repository, Archive, Encyclopedia, Big-Data Analytics, Advanced Graphics, HPC Services and Infrastructures 등으로 구성(440 여만 건 데이터 구축)
- 소재 계산연구과 계산방법론 로드맵 구축 및 국제협력을 통한 데이터베이스 확장 노력 중

<표 6-14> 유럽 계산재료과학 및 소재정보학 연구 연합체

명칭	내용	기관
NOMAD	- 소재물성 데이터, 빅데이터 분석, 가시화 등 다양한 기능 - 계산연구그룹(8개)과 고성능 컴퓨팅센터(4개)가 참여 중	 EU CoE Center of Excellence
E-CAM	시뮬레이션과 모델링 분야의 소프트웨어, 교육 및 컨설팅을 지원하는 EU의 고성능 컴퓨팅 중핵기관	 EU CoE Center of Excellence
BigMax	구조와 변형, 3D 이미지의 데이터 분석, 패턴·상호관계·인과관계 분석, 기계학습, 메타데이터 등 데이터 및 소재정보학 연구 수행	 독일 막스플랑크 연구소
MARVEL	- 전산재료과학 기반 신소재 설계와 탐색을 위해 2014년 설립 - 고속 시뮬레이션과 소재정보학에 의한 신소재 설계와 탐색 가속화	 스위스 NCCR

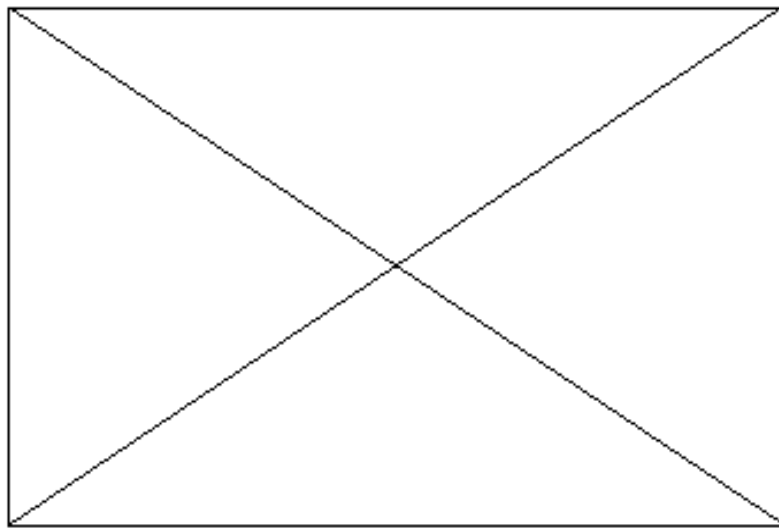
<표 6-15> 유럽 주요 소재데이터베이스 현황

명칭	내용	기관
Pauling File	- 일본 JST, 도쿄대와 함께 시작('95년) - 신소재 설계를 위해 non-organic solid state materials에 대한 phase diagram, 결정구조, 물리적 특성 데이터 등 보유	[스위스] MPDS
AiiDA	계산재료과학 시뮬레이션 자동화용 워크플로우 엔진(AiiDA) 개발·제공	[스위스] EPFL [미국] BOSCH
Springer Materials	- 보유저널 기반으로 소재 연구 지원을 위한 축적된 데이터와 고급 기능 제공 - 소재 종합 데이터베이스, 데이터 시각 자료, 최적 검색(원소 조성, 화학구조 등), 신뢰성 높은 통합 리소스 등 제공	[독일] Springer Materials
CMR (Computational Materials Repository)	- 소재물성 데이터베이스, DB 웹페이지, 페이지 보완(업데이트 등) 17개 프로젝트 수행: 양자화학 기반 소재 물성 계산(예시: 2D materials, Organometal Halide Perovskites 등)	[덴마크] 덴마크공대(DTU)
NoMAD COE (Novel Materials Discovery)	- 소재물성 데이터, 빅데이터 분석, 가시화 등 다양한 기능 - 계산연구그룹(8개)과 고성능컴퓨팅센터(4개)가 참여 중	[EU] CoE NoMAD
ICSD (Inorganic Crystal Structure Database)	- 실험에 기반 한 inorganic compounds 결정구조(187,000개 이상) 보유 - DFT기반 제일원리계산 접목 가능	[독일] FIZ
CSD (Cambridge Structural Centre)	- 실험으로 결정된 inorganic, metal-inorganic 결정 구조의 repository - 약 80만개 이상 데이터 보유	[영국] Cambridge Crystallography Data Centre
ESP (Electronic Structure Projects)	기능소재 발굴을 위한 6만개 이상의 전자구조 계산데이터 확보	[스웨덴] ESP

2) 소재 데이터 기반 소재정보학 활용 연구 동향

◆ 빅데이터와 소재정보학 활용 공정-구조-특성-성능 관계해석 및 소재 R&D 효율화

- (데이터 기반 소재연구) 계산과학 기반 예측과 고속 실증실험 및 평가, 소재 데이터베이스와 AI기술을 통합적으로 연계·활용하여 신소재 탐색·설계하는 소재 연구개발
- (내용) 소재 연구개발의 핵심인 공정-구조-특성-성능(Process-Structure-Property-Performance, PSPP) 관계를 빅데이터와 소재정보학을 활용하여 신속하게 파악함으로써 소재 R&D 효율화



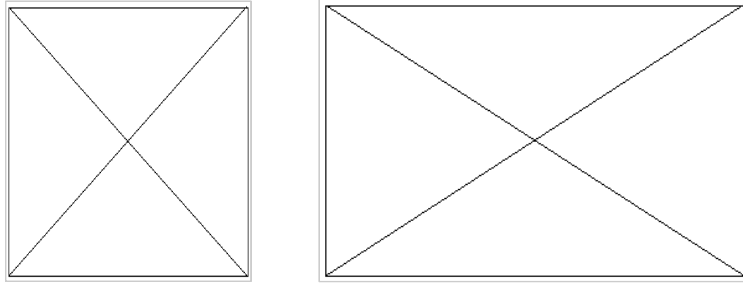
<그림 6-28> 소재 R&D에서 공정-구조-특성-성능 관계

- (기대효과) 기존 재료과학 연구 패러다임으로는 신소재 탐색에서 실용화까지 통상 10~20년이 소요되나, 데이터 기반 소재연구가 성숙되면 반복되는 실험, 계산과정을 대폭 감소시켜 신소재 개발과정 효율화
- (현황) 고속 제일원리 기반 전산과학을 활용한 소재물성 생성, 기 구축된 소재 데이터베이스, 소재 데이터 수집·공유·활용 시스템 구축 및 고속 실험(측정·평가)을 통한 대량의 실험 데이터 생성 및 수집기술을 활용하여 이론·실험·계산과학 및 데이터 과학을 연계한 혁신적 소재개발 연구 확대
- 빅데이터 기반 소재연구는 ①방대한 양의 데이터를 다양한 각도로 시각화하는 플랫폼 개발 연구, ②대용량 고속 스크리닝에 의한 신소재 물성탐색 연구, ③수집 데이터 사이의 상관관계를 통계적으로 학습시키는 연구의 3가지 유형으로 구분

<표 6-16> 빅데이터 기반 소재정보학 활용 연구사례

□ 웹 기반 시각화 플랫폼

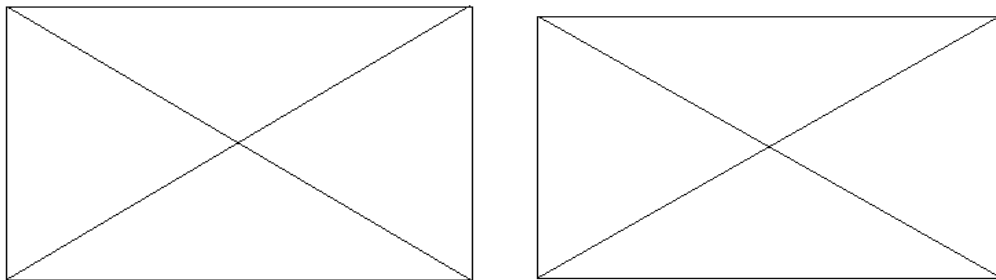
- 출판된 논문에서 열전소재 성능변수(ZT) 데이터를 18,000개 이상 추출하여 데이터베이스화한 후 웹 기반 가시화 플랫폼 구축



열전소재 웹 시각화 플랫폼 (미)UCSB

□ 초고속(High-Throughput) 스크리닝

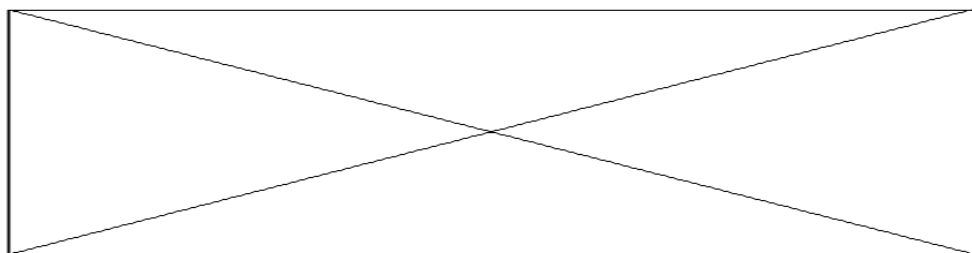
- 700개 이상의 2원계 화합물의 표면 흡착에너지를 제일원리계산으로 예측하여 수소생성 반응성을 예측하고, 실험으로 검증



흡착에너지 기술자에 의한 촉매재료 HT 스크리닝 (미)스탠포드대

□ 통계적 학습법(머신러닝)

- 2원계 전이금속 화합물의 특성 기술자(descriptor)와 CO 흡착에너지 관계를 기계학습 후 미지의 화합물의 흡착에너지 예측



기계학습으로 예측한 촉매재료 CO 흡착에너지 (미)버지니아공대

〈표 6-17〉 응용분야별 소재정보학 활용 신소재 탐색·설계 사례

응용 분야	데이터 종류(계산, 실험)	데이터 크기	접근법	유망 신물질	출처
태양전지	밴드갭, 광흡수율 (제일원리 계산)	260개 I-III-VI chalcopyrite 화합물 (ICSD*)	HT 스크리닝	Cu ₇ TiS ₄ 포함 25개	Phys. Rev. Lett. 108, 068701 (2012)
Solar thermal fuels	이성질화 엔탈피, 에너지 밀도 (제일원리 계산)	430개 이성질성 분자 (CSD**)	HT 스크리닝	물질명 명시되지 않음	Nano Lett, 14, 7046 (2014)
광 물분해	생성에너지, 밴드갭, 밴드에지 (제일원리 계산)	68개 이원계 질화물, 1503개 삼원계 산질화물, 1377개 사원계산질화물 (ICSD)	HT 스크리닝	Ti ₃ O ₃ N ₂ , La ₂ TiO ₂ N ₂ , Li ₅ MoO ₄ N	Energy Environ. Sci. 6, 157 (2013)
	생성에너지, 밴드갭, 밴드에지, 전자이동도 (제일원리 계산)	5400개 페로브스카이트 산화물 또는 산질화물 (ICSD)	HT 스크리닝	MgTaO ₂ N 포함 15개	Energy Environ. Sci. 5, 5814 (2012)
가스 저장/분리	프로펜인/프로필렌 분자의 흡착량, 분리선택도 (몬테카를로 계산)	65,350개 다공성 소재 (ICSD, CSD)	HT 스크리닝	해당 없음	J. Phys. Chem. C 120, 24224 (2016)
	메탄 흡착량 (계산)	137,953개 MOF	HT 스크리닝	NOTT-107 포함 300개 이상	Nat. Chem. 4, 83 (2012)
압전소자	밴드갭, 압전상수, 유전상수, 탄성계수 (제일원리 계산)	987개 Half-Heusler 재료	HT 스크리닝	KMgP 포함 6개	Phys. Rev. Lett. 109, 037602 (2012)
열전소자	열전 성능변수 (zT, S, σ , κ), 자원량 (기출판된 실험논문 100여편)	18,000개	웹기반 시각화 기술	해당 없음	Chem. Mater. 25, 2911 (2013)
	격자 열전도도 (모델링)	79,000개 Half-Heusler 반도체 재료 (AFLWLIB.org)	HT 스크리닝 + 기계학습법	BiBaK 포함 4개	Phys. Rev. X 4, 011019 (2014)
촉매	수소원자 흡착에너지 (제일원리 계산)	700개 이원계 합금	HT 스크리닝	BiPt 포함 5개	Nat. Mater. 5, 909 (2006)
	일산화탄소 흡착에너지, 합금재료 DOS (제일원리 계산)	~400개 이원계 합금 또는 코어-셸 구조 합금	기계 학습법	해당 없음	J. Phys. Chem. Lett. 6, 3528 (2015)
배터리	배터리 용량 및 전압 (제일원리 계산)	60,000개 산화물, 인화물, 황화물 등 (ICSD)	HT 스크리닝	물질명 명시되지 않음	Mater. Res. Soc. bull. 35, 693 (2010)
트랜지스터 (게이트용 유전체)	밴드갭, 유전상수 (제일원리 계산)	1800개 이원계 또는 삼원계 산화물 (ICSD)	HT 스크리닝	c-BeO 포함 20개 이상	NPG Asia Materials 7, e190 (2015)

3) 현황분석을 통한 시사점

□ 소재 연구데이터 수집·관리를 위한 전문센터 구축 필요

- 소재 연구데이터를 활용한 데이터 구동형 소재 연구개발을 위한 국내 인프라 부재
 - (소재 데이터베이스 부족) 2007년부터 산업부 지원으로 구축된 소재정보은행 중심으로 산업용 소재물성 데이터 검색 및 분류 서비스 제공하고 있으나 일부 소재 및 응용분야에 집중되어 한정된 범위 물성 검색만 제공
 - 2013년 과기정통부 창의소재 디스커버리사업을 통해 소재 데이터 기반 소재 R&D 사업이 시작되었으나, 활용 데이터 부족으로 소재정보학으로 발전 미흡
 - (소재 연구데이터 수집 인프라 부재) 현재 국내에서는 DB(데이터베이스)화된 소재 R&D 데이터를 적용할 수 있는 인프라 부재
 - 연구데이터 기반 소재개발을 위해 데이터 수집·공유·활용을 위한 연구데이터 인프라를 구축할 수 있는 자체기술 확보 필요
- 소재 연구데이터 수집 및 데이터 표준화 필요
 - (소재 연구데이터 수집 필요성) 현재 인공지능을 구현하기 위한 기술, 데이터과학 및 IT 등은 높은 수준에 도달해 있으나, 이를 소재연구에 활용하기 위한 소재데이터 부족
 - (소재 연구데이터 표준화 필요성) 데이터 재활용/공유/분석을 위해서는 데이터 표준화를 통한 데이터 저장 및 데이터베이스 구축이 필요하고 수집·활용의 편리성을 위해 데이터베이스 구축단계부터 데이터 수집에 대한 표준화 체계 수립 필수
 - 소재 제조방법 및 특성 평가방법에 따른 데이터 생산·수집·관리를 위한 메타데이터 표준화 필요 (비정형 데이터에 대한 메타데이터 표준화는 필수)
 - 기존 보유데이터 및 신규 생산데이터의 신뢰성 확보를 위한 평가 표준가이드 개발 및 활용 확산 필요
 - (소재연구데이터 표준화 혁신성) 국제수준의 소재 연구데이터 표준화를 통한 범세계적 데이터 표준화 주도권 확보 및 한국 중심의 국제 소재 연방 네트워크(Federated Materials Registry Network) 구성 가능
- 소재 연구데이터 활용을 위한 공유 플랫폼 구축 필요
 - (소재 연구데이터 공유플랫폼 필요성) 소재 응용분야별 소재 데이터 수집 및 공유 플랫폼을 구축하고 범국가적 연구데이터 플랫폼과 연계·활용 가능한 네트워크 구성 필요
 - (소재분야 데이터 공유) 소재분야의 연구는 전통적으로 소규모 연구팀에 의한 데이터 생산과 분석에 익숙하고 지적재산권 확보가 연구성과에서 큰 부분을 차지하여 데이터

\ 재이용·공유에 부정적 인식

- (소재분야 데이터 공유 전망) 최근의 소재 연구데이터 공유 및 빅데이터를 활용한 기계 학습과 분석정밀도 증가는 우수한 연구성과로 나타나고 있고 연구개발 효율화의 핵심요소로 되고 있어 데이터 공유에 대한 인식은 개선되는 상태
- (소재 연구데이터 공유플랫폼 요구조건) 소분야 전문센터에서 응용분야별 소재 데이터 수집·활용을 원활하게 수행하기 위한 유연하고 사용자 친화적인 플랫폼 구축

○ 소재 연구데이터 관리체계 구축 필요

- (데이터의 분류) 데이터는 정형성 또는 산출방법 기준으로 정형, 비정형 또는 실험, 분석, 계산데이터로 구분
 - ※ 정형데이터: 고정된 양식에 저장되는 데이터로 처리·분석 용이
 - ※ 비정형데이터: 고정된 양식에 저장되지 못하는 데이터로 의미 있는 정보추출이 어려움 (데이터 관리·활용 분석을 위한 별도의 처리방법 필요)
 - ※ 실험데이터: 다양한 실험인자(실험환경, 실험장비, 실험방법 등)존재로 구조화·체계화 어려움 (일반적으로 평균값/대표값 저장 및 활용)
 - ※ 계산데이터: 주로 제일원리계산을 통한 데이터로 실험데이터보다 관리 용이
- (데이터 수집·대상) 일반적으로 실험·계산을 통한 긍정적인 결과에 집중 및 예상치 못한 결과, 부정적인 결과는 무시·폐기·삭제하나 소재정보학에 활용되는 데이터는 모든 연구데이터 축적 필요
- (데이터 저장) 데이터에 포함된 정보는 표준화 양식으로 저장되어야 하며 저장방식 설계 시 데이터 타입에 따른 저장기술 및 처리/접근 패턴 (접속빈도, 보안 등) 고려 필요
- (소재 데이터베이스 구축) 다양한 소재 데이터베이스가 존재하나 접근성이 떨어지므로 국가 차원의 저장소 제공 및 신뢰성 제고를 통한 신뢰성을 구비한 소재 데이터베이스 플랫폼 필요
- (소재 데이터 디지털 큐레이션) 소재 데이터를 체계적으로 수집·공유·가공할 수 있는 표준화된 관리 시스템 필요 (예: NIST Materials Data Curation System)
 - ※ 다양한 유형의 소재 데이터의 분류 및 신뢰성 검증 프로세스 자동화 시스템
 - ※ 디지털 큐레이션 센터 제공 단계별 디지털 데이터 라이프 사이클
 - 데이터 개념화: 데이터 형식 파악 및 저장 형식 개념화
 - 데이터 생성: 데이터를 디지털로 변환하고, 관계되는 모든 메타데이터를 첨부 (많은 메타데이터가 존재할수록 정보에 접근성은 향상)
 - 데이터 접근과 활용: 데이터/유저별 데이터베이스 접근 정도를 결정
 - 데이터의 폐기: 데이터베이스에 불필요하거나 신뢰성이 낮은 데이터 폐기

- 데이터의 취합: 디지털화된 자료를 표준화된 저장소로 전송
 - 데이터의 보존: 데이터베이스의 통합성을 유지하기 위한 도구 도입
 - 데이터의 재평가: 데이터 원형과의 관련성 유지여부를 확인하기 위한 데이터 재평가
 - 데이터 보안: 데이터를 표준화된 저장소에 안전하게 보관
 - 데이터 접근 및 재사용: 의도된 대상의 데이터 접근 및 의도되지 않은 자료유출 여부 확인
 - 데이터의 변형: 데이터 형식 변형 필요시 데이터 형식 변형 가능
- 소재 선진국과 비교하여 소재 연구데이터 인프라 구축의 후발주자이기에 데이터베이스 구축과 동시에 소재 데이터베이스에 대한 융합적 활용방법론(머신러닝 방법 등)을 개발하는 Two-Track적 플랫폼 구축 필요
- (소분야 전문센터 데이터 수집) 소재분야 연구데이터를 분야별 특성을 반영하지 않고 수집될 경우 활용성이 매우 떨어지고 데이터 표준화에 장시간이 소요되므로, 소분야별로 체계적인 수집을 통하여 대분야 전문센터의 신속한 데이터 관리 시스템 구축이 가능
 - (소분야 전문센터 데이터 활용) 소재 정보학을 활용한 소재 연구개발로 소재 R&D 효율화 및 4차 산업혁명 시대의 소재 연구개발에 필요한 소재정보분야 융복합 전문가 양성

6.4.3. 추진전략

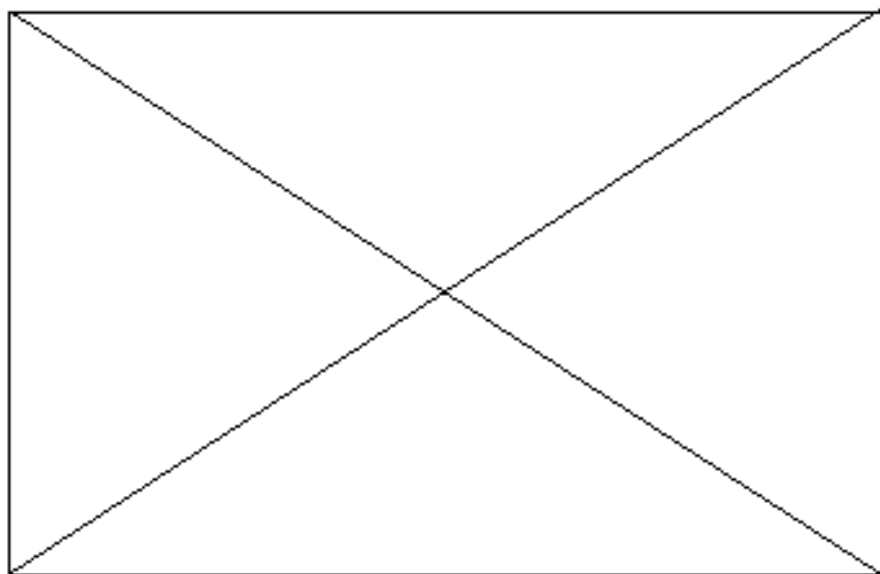
가. 사업추진 방향 및 목표

1) 운영현황 및 개선방향

-
- ◆ 선진국과 비교하여 국내 소재 빅데이터 수집·활용·방안 미흡
 - ◆ 데이터 공유·활용 생태계 구축을 위한 연구자 친화적 데이터 인프라 환경 필요
-

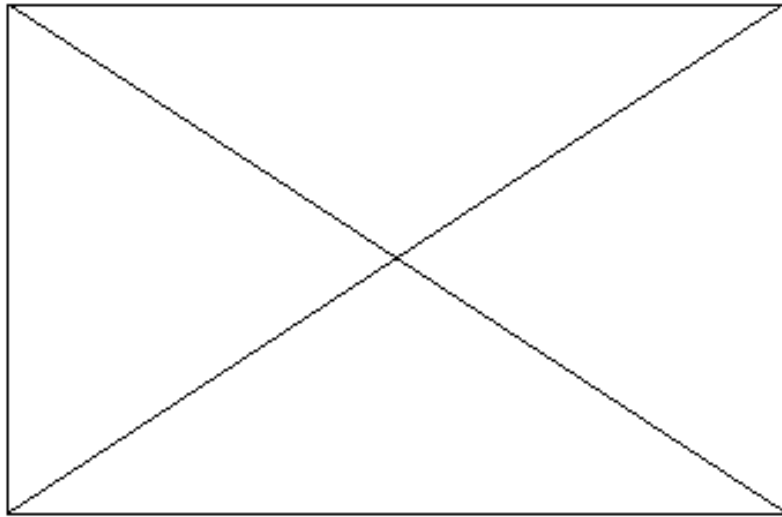
□ 국내 소재분야 데이터 관련 센터 현황

- 소재분야 연구데이터는 대부분 개별 연구자(팀)이 소유하고 있고, 국가R&D사업 성과물은 NTIS(논문, 특허 등)에 등록되는데 그치고 있어, 장기간 연구데이터 수집·등록 시스템이 구축·운용되어 정보를 개방, 활용하고 있는 선진국과는 큰 격차를 보임
- 데이터 분석·해석을 통한 연구개발에 활용 가능한 데이터 부족으로 본격적인 소재정보학으로 발전된 모습은 아직 보이지 못하고 있는 실정
- (국내 기관) 산업소재 물성 데이터 수집·제공하는 소재정보은행, KIST와 화학연구원의 데이터 연구, 표준참조 물성데이터를 제공하는 국가참조표준센터가 구축·운영
- (KIST 빅데이터 선행과제) 소재 대분야 및 소분야 연구센터 구축을 위한 선행과제를 수행 중에 있으며, KIST 내 모든 연구분야에 적용될 수 있는 범용 인프라 구축을 목표로 인프라 실효성을 입증하기 위해 나노촉매 분야를 소분야 전문센터 연구주제로 선정



<그림 6-29> KIST 데이터 기반 R&D 플랫폼 구성도

- － (화학연구원 열전소재) 화학소재 데이터 플랫폼 구축 및 열전 신소재 예측과 개발 사업을 통해 대분야 및 소분야 전문센터 구축에 필요한 기초 사업을 수행 중으로 연구데이터 자동수집 기반 구축, 수집 데이터 분석 지원, 데이터 마이닝, 머신러닝 등에 적용을 목표



<그림 6-30> 데이터 플랫폼을 활용한 데이터 수집 및 활용 (화학연구원)

- － (소재정보은행) 소재물성 DB, 기술 및 시장동향, 특허, 카탈로그, 각종 보고서 등의 산업용 소재 데이터 생성·수집을 통한 핵심 소재정보 보급

- ※ 2007년부터 금속, 화학, 세라믹, 섬유 등 4개 소재정보은행 구성(약 108만건 DB)
- ※ (데이터 구축) 소재물성정보, 지식정보, 기술강좌 등 약 108만건 DB구축
- ※ (시뮬레이션) 소재 제조공정 시뮬레이션 소프트웨어 13개 구축 후 중소중견기업 기술지원

<표 6-18> 소재정보은행 물성 DB 구축 현황 ('16.12)

분야	주관기관	주요 내용	구축 건수		
			수집·가공	생성	합계
금속	재료연구소	• 철강, 알루미늄합금, 마그네슘합금, 분말, 구리소재 등	121,129	90,561	211,690
화학	한국화학연구원	• 플라스틱, 고무, 첨가제, 필름, 접착제, 도료, 복합재료 등	558,352	43,182	601,534
세라믹	세라믹기술원	• 반도체소재, 센서소재, 기계구조재료, 생체소재 등	159,503	35,145	194,648
섬유	다이텍연구원	• PET, 나일론, 아라미드섬유, 탄소섬유, 토목용 섬유 등	56,402	17,560	73,962
합 계			895,386	186,448	1,081,834

- \ - (산업체) 소규모의 ‘Computational Materials Screening’ 방법으로 자체 데이터베이스를 만들어 활용
- ※ LG 화학, 삼성 SDI, 삼성중공업 등 대기업에서 활용
 - ※ 리튬 이차전지 전극 소재 및 전해질 첨가제 등의 소재분야 국한
 - ※ 중소기업에서 소재 빅데이터 활용 신소재개발은 거의 전무한 것으로 판단
- (국가참조표준센터) 해당분야 데이터 수집·생산, 평가 및 참조표준(안) 제정을 통한 신뢰성 있는 물성 데이터 제공
- ※ 2006년부터 금속, 재료 등 8개 분야의 물성 데이터 제공(10개 분야 65개 DB, 3만 8천개 데이터)
 - ※ 미국 NIST는 매년 90만 건 참조표준 데이터 제정보급 및 연간 1,900만 건 다운로드 및 6천 건 패키지 판매(NIST Measurement Service, 2012)
 - ※ (역할) 분야별 데이터 수집·생산 관련 품질 시스템 구축 및 참조표준(안) 제정, 산학연 이용자에게 보급·확산
 - ※ (데이터) 금속, 재료, 바이오 등 10개 분야 65개 DB구축 (총 3만 8천개 데이터, 2016년 기준)
 - ※ (분류) 재현성 있는 데이터, 관측데이터, 측정 데이터로 구분하여 A,B,C그룹으로 분류 및 수집

<표 6-19> 국가참조표준센터 참조표준 데이터 분류 및 데이터 건수

구분	세부분류	데이터 건수
금속	내열/구조강 인장물성	2,410
	철강소재 미세조직	543
	원자력 재료 인장물성	20
	원자력 재료 경도물성	20
	자동차용 강판 동적물성	93
재료	유화중합에 의한 고분자 나노입자 분포도	120
	세라믹 재료의 전기적 특성	33
	반도체 재료 열물성	152
	핵연료재료 물성	17
	원자력구조재료	53

- (출연연 연구데이터 구축 현황)

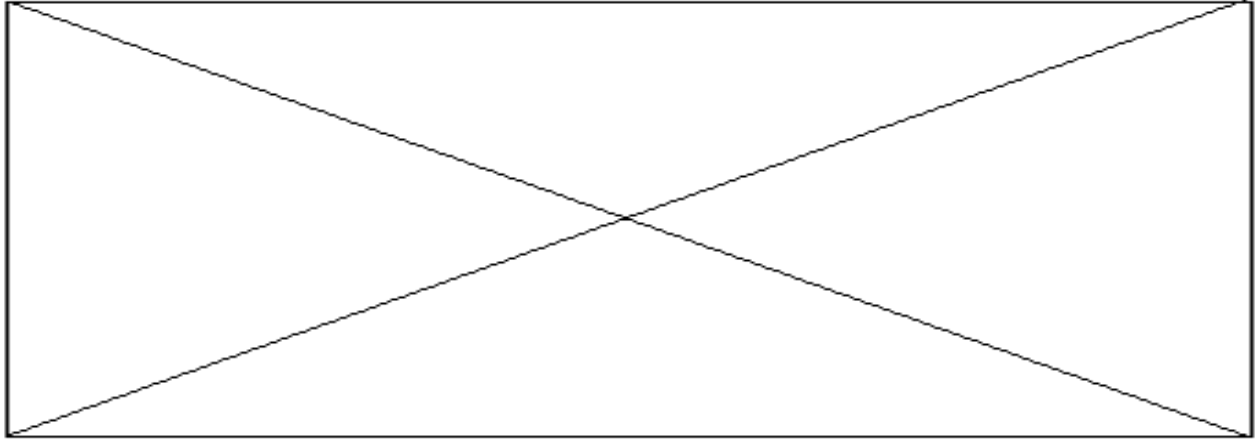
<표 6-20> 출연(연) 소재 데이터 플랫폼 과제 수행 현황

기관명	사업명	분야	사업기간
한국화학연구원	화학데이터 플랫폼 구축사업	유기태양전지 소재	2017-2019
	소재빅데이터 기반 고효율 열전 신소재 예측과 개발	열전소재	2017-2019
한국전기연구원	열전소재 및 에너지 소재 설계 기술 기반구축사업	열전·에너지 등 전기소재 수송물성	2015-2020
한국생명공학연구원	생체적합성 바이오식품 소재 개발 플랫폼 구축사업	천연물 소재	2012-2017
한국과학기술정보연구원	데이터 플랫폼 구축사업	배터리	
한국과학기술연구원	데이터 기반 R&D 환경구축사업	촉매	2017-2018

□ 국내 소재분야 데이터 관련 센터 개선방안

- R&D 과정에서 발생하는 소재 연구데이터를 국가적 자산으로서 관리·활용 할 수 있는 기반과 정책 미흡하고, 소재연구를 통해 생성된 데이터를 연구자만이 소유, 활용하기에는 투입된 국가 R&D비용에 비해 부가적 활용도가 낮으며, 논문게재를 통하여 공공에 공개되는 데이터는 50% 미만으로 예상
- 미래소재 디스커버리 사업 및 출연(연) 주요사업을 통하여 수행되는 계산과학 기반 소재물성 데이터 생성 및 활용연구는 소재 데이터센터의 핵심 요소인 데이터 수집·가공·공유 목적보다는 소재정보학을 적용한 연구개발에 치중
- 소재정보은행사업으로 구축된 산업소재 데이터는 데이터 불확실성 미평가, 메타데이터 결여, 비구조화된 데이터베이스, HTML 기반 데이터베이스로 데이터 공유에 한계가 있으며 소재정보학을 활용한 신소재 연구개발에는 활용이 어려움
- 다양한 소재 데이터를 연구자가 쉽게 저장·관리·활용할 수 있는 데이터 인프라 환경 제공이 필요하고 연구데이터 공유·활용 체계 구축 로드맵에서는 데이터 수집 및 관리제도, 연구데이터 수집 및 관리 지원을 위한 인프라, 공유·활용 플랫폼과 생태계 구축이 모두 고려된 접근 필요

2) 사업추진 목표



□ 목표 1 : 소재분야 연구데이터 대·소전문센터 구축

- 소재 전분야 연구데이터 수집·관리·공유·활용 전문센터 구축
- 계산과학, 에너지·환경, 스마트·IT, 안전(구조) 소재 4개 소분야 전문센터 구축
- 에너지·환경, 스마트·IT, 안전(구조)소재 소분야 전문센터는 응용분야별 과제 수행
— 응용분야별 소분야 전문센터 3개 x 소분야 전문센터 과제 3개/5년 = 18개/10년

대분야 전문센터 1, 소분야 전문센터 4, 전문분야 과제 총 18개

□ 목표 2 : 소재·정보·데이터 융복합 혁신기업 양성

- 소재 R&D 및 소재산업을 위한 소재·정보·데이터 융복합 혁신기업 양성
- 계산과학머신러닝, 실험데이터베이스 혁신기업 양성

혁신기업 2개 (계산과학·소재정보학 전문기업, 소재 데이터베이스 기업)

□ 목표 3 : 소재 데이터 활용 연구개발에 의한 경제적 효과 창출

- 데이터 이(利)활용을 통한 신소재 탐색·설계·개발·적용의 효율화

소재 연구개발 비용 및 개발기간 단축에 의한 경제적 효과 → 10조 원

□ 목표 4 : 정량적 연구개발 성과 목표

- 인력양성, 논문/특허/소프트웨어, 신물질 개발

- 데이터과학 활용 소재 전문가 700명
- 논문/특허/소프트웨어 90편/50건/5개
- 신물질 개발 40개

나. 사업추진 전략 및 세부추진 과제

1) 사업추진 전략

□ 전략 1 : 소재분야 연구데이터 대·소전문센터 구축

- 소재 전분야 연구데이터 수집·관리·공유·활용 전문센터 구축
- 빅데이터 구축에 필요한 표준화, 소재 온톨로지 개발 및 머신러닝 활용을 통한 신물질 개발을 체계적으로 수행하기 위한 응용분야별 소분야 전문센터 구축
 - ※ 미래소재 연구데이터 소분야 전문센터:
 - 계산과학: 고급 계산물성 데이터 생성
 - 에너지·환경: 다공성 소재, 고체화합물 소재 등
 - 스마트·IT: 엑사 스케일 반도체, 멤리스터, IOT 센서 소재 등
 - 안전(구조) 소재: 극한환경 소재, 경량금속 소재, 수소 인프라 소재 등

□ 전략 2 : 소재·정보·데이터 융복합 혁신기업 양성

- 데이터 구동형 소재설계(데이터, 인공지능), 소재 데이터 분석·해석 기반 엔지니어링 및 데이터 패키지를 제공하는 혁신기업 양성
 - ※ Citrination(미): 소재·화학 정보 데이터베이스 서비스, 분석/머신러닝 기술 제공
 - ※ AtomWork-Adv(일): 문헌기반 무기소재의 계산실험 데이터 유료 서비스(NIMS)
 - ※ Granta Design(영): 소재 데이터 기반 소재선정 소프트웨어 및 엔지니어링 제공
- 계산과학·머신러닝 서비스 제공 혁신기업 양성 (1개)
- 실험데이터 기반 데이터베이스 서비스 제공 혁신기업 양성 (1개)

□ 전략 3 : 소재 데이터 활용 연구개발에 의한 경제적 효과 창출

- 미래소재 연구개발 비용 및 개발기간 단축에 의한 경제적 효과 약 10조 창출
 - ※ 미국 MGI 경제적 파급효과(RTI社, NIST 용역): 138~300 조원/년 경제적 효과
 - ※ 한국 경제규모 고려(GDP 미국의 1/20) → 7~15조원/년 예상

□ 전략 4 : 정량적 연구개발 성과 목표

- (인력양성) 현재 국내 데이터과학 활용 가능 소재전문가는 수십여명 이내로 소재 연구 데이터 전문센터 사업 수행 및 교육프로그램 운용으로 소재·데이터 융합기술 보유 소재 전문가 양성 (700명/10년)

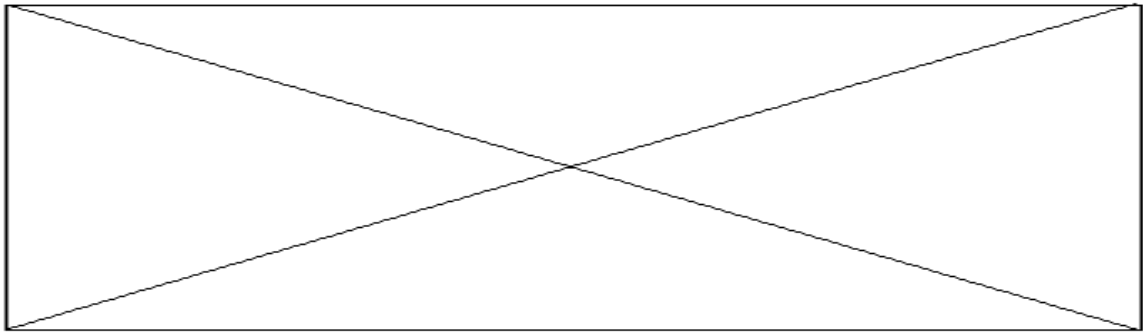
- \ ○ (학술논문) 응용분야별 소분야 전문센터에서 학술지 논문 90편 이상 게재
- (특허) 대분야 및 소분야 전문센터에서 관련 특허 50편 이상 출원
- (소프트웨어) 대분야 전문센터 및 프레임워크 소프트웨어, 소분야 전문분야 개발 소프트웨어 5개 이상
- (신물질 개발) 응용분야별 소분야 전문센터 세세부과제 수행으로 미래소재 신물질 개발 30개 이상

다 세부과제별 추진 내용

1) (세부과제 1) 소재 연구데이터 수집·관리·공유 체계 구축 전문센터

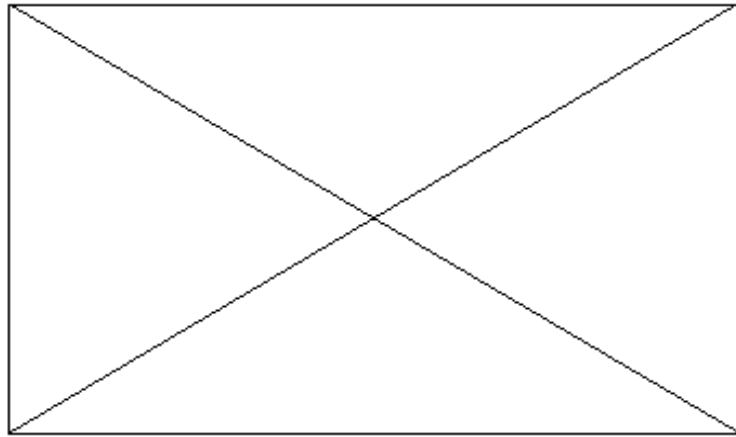
- ◆ 소재 전분야 연구데이터의 표준화를 통하여 효율적인 데이터 수집·관리·활용 플랫폼 개발, 공유 플랫폼을 통하여 융복합 R&D 생태계 구축

<소재 연구데이터 수집·관리·활용·공유 프레임워크>



□ 현황 및 필요성

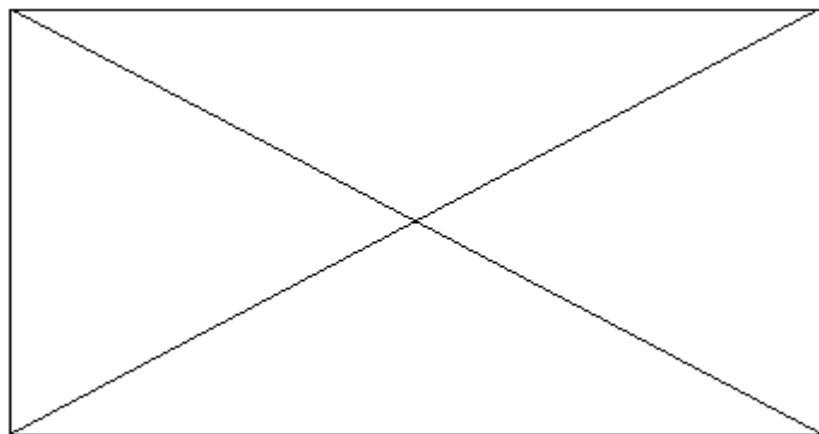
- 국제적으로 실험·이론·계산과학중심의 연구 패러다임이 빅데이터 중심의 데이터기반 연구 패러다임으로 변화함에 따라 기술선진국에서는 이미 소재 관련 데이터들을 수집 및 활용 플랫폼이 성공적으로 개발되고 연구에 활용
- 소재 R&D 전 과정에서 발생하는 연구데이터를 국가적 자산으로서 관리·활용할 수 있는 기반과 정책이 미흡하여, 연구데이터를 연구자만이 소유함으로서, 투입되는 R&D 비용에 비해 데이터의 부가적 활용도 낮음
- 소재 전분야 R&D 과정에서 발생하는 연구데이터는 실험/분석 방법의 종류의 다양성으로 인하여, 수집된 연구데이터의 분류 방법과 그것의 저장 방식에 대한 연구가 필요함
- 최근 서로 다른 분야의 융복합 R&D의 중요성이 부각되고 있으나, 서로 다른 분야의 연구자들 간의 교류 환경이 조성되어 있지 않음. 다양한 분야 간의 연구 교류 및 공유를 통하여 융복합 R&D를 가속화하기 위해서는, “연구 공유 플랫폼”의 개발이 필수적임



<그림 6-31> “연구 공유 플랫폼”을 통한 융복합 R&D 생태계 구축

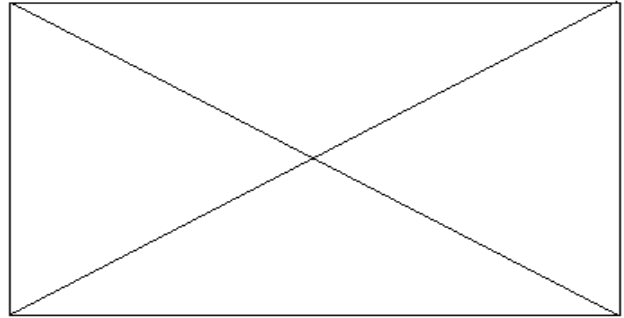
□ 국내외 사례

- (국내) 미래소재디스커버리(2015년~), 나노·소재기술개발사업(2016년~)등 소재 연구데이터를 수집하고자 하는 노력이 있었으나, 연구개발에 치중되어 과제가 진행되어 방대한 양의 연구데이터를 수집하는 기능을 하지는 못하고 있음
- (국내) 최근 한국과학기술연구원을 중심으로 소재 연구데이터 수집 관리 활용 플랫폼 개발 과제가 시작되었으나, 이 기수행과제와는 두 가지 측면에서 차별점이 있음
- (1) 표준화 연구는 수행하지만 실제 데이터 수집은 “전국소재”로 제한
- (2) 공개문헌(논문, 특허, NTIS 보고서 등)에 대한 수집기술 개발연구 미수행
- (미국) 국가주도의 연구데이터 혁신프로젝트(MGI)를 2011년도부터 진행해왔으며, 소재연구데이터 수집/활용 플랫폼 ‘Material Project’을 개발하여, 연구에 활용 중



<그림 6-32> 미국 연구데이터 혁신프로젝트(MGI) 실행계획

- (일본) 국가연구소 NIMS가 관리하는 소재 빅데이터 통합사이트(MatNavi)를 통해 소재 DB를 공개하고 있으며, 빅데이터를 활용하는 플랫폼 ‘MI2I’를 구축/운영 중



<그림 6-33> 일본 NIMS의 소재 빅데이터 통합사이트(MatNavi)와 활용 플랫폼(AtomWork-Adv)

- 글로벌 IT 기업을 중심으로, “사용자 행동기반 데이터” 수집을 통한 빅데이터 구축을 수행하고 있으나, R&D 분야에서 일반 연구자 대상 빅데이터 구축 사례는 전무하며 이를 위해서는 “연구자 행동기반 데이터” 수집·관리 기술개발 필요
- Research Gate, Overleaf, NoMad와 같은 연구공유 플랫폼이 큰 성공을 이루고 있으며, 이러한 “연구공유 플랫폼”은 데이터 교환뿐만 아니라 협업연구를 촉진하는 생태계를 형성

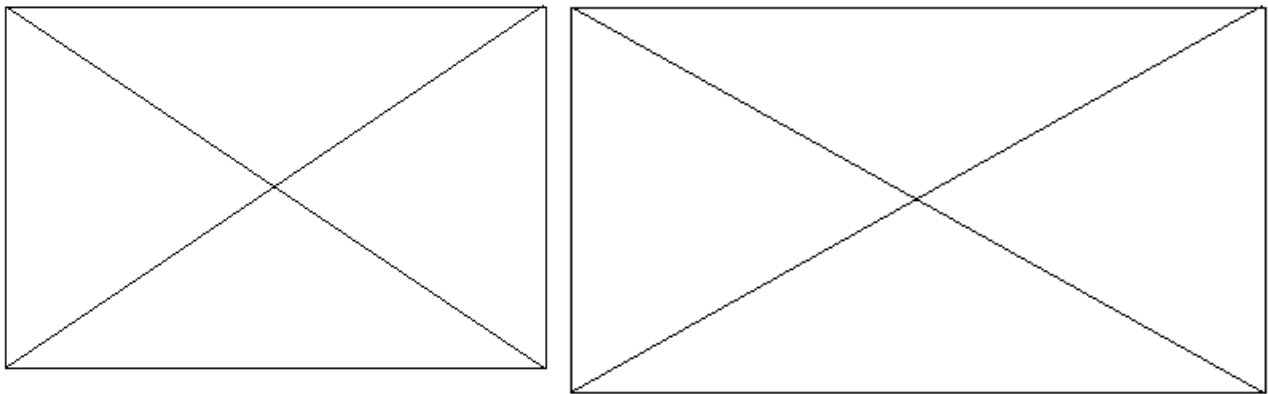
□ 핵심기술

- (인프라) 소재 연구데이터 수집·관리·활용·공유 인프라(하드웨어·시스템) 구축
- (인프라) 계층적 플랫폼 구축을 위한 소재 빅데이터 프레임워크 개발
- (수집) 연구자 행동기반 연구데이터 수집 플랫폼
- (수집) 공개문헌(논문, 특허, NTIS 보고서 등) 수집
- (표준화) 온톨로지 기반 소재 메타데이터 표준화
- (표준화) 비정형 데이터를 저장 관리하기 위한 데이터베이스 설계
- (관리) 연구자 행동기반에 의하여 수집된 연구 빅데이터의 관리 기술
- (관리) 수집된 공개문헌(논문, 특허, 보고서 등)의 관리 및 분류 기술
- (활용) 소재 분야별 데이터 가시화 도구 모듈 기능 개발

- \ ○ (활용) 소재 분야별 특화 기계학습 도구 모듈 기능 개발
- (공유) 글로벌 소재 연방 네트워크 연결(NIST, NIMS 등)
- (공유) 연구 공유 플랫폼 개발

□ 연구내용

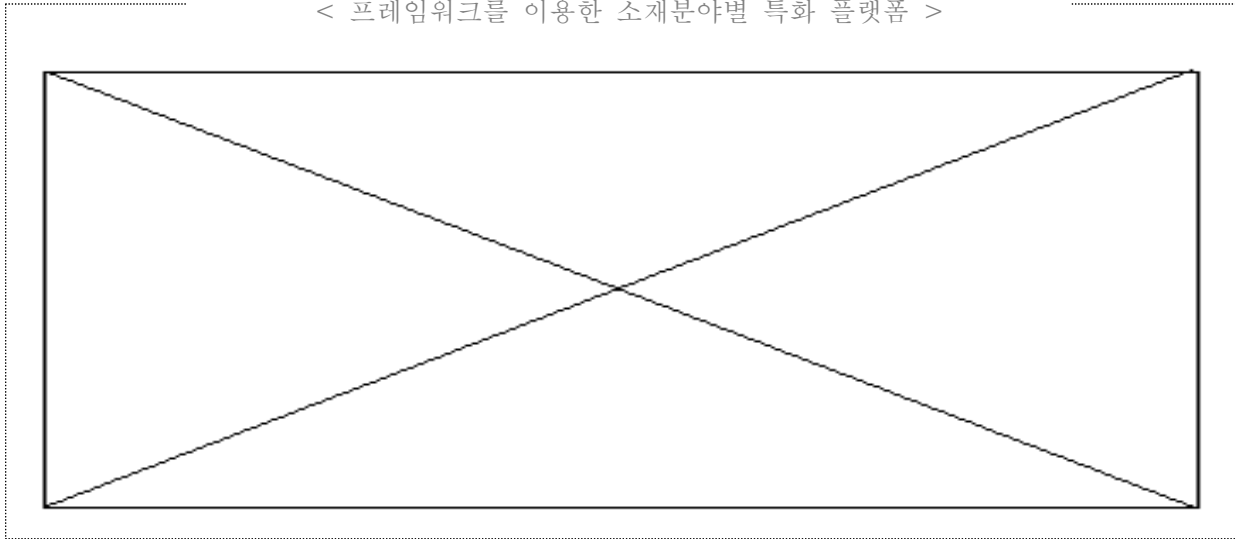
- (인프라) 소재 연구데이터 수집·관리·활용·공유 인프라(하드웨어·시스템) 구축
 - － 10년간 수집된 연구데이터를 포용할 수 있는 서버 하드웨어 구축
 - － 효율적인 연구데이터 관리·활용을 위한 클라우드 분산시스템 구축
 - － MongoDB, Hadoop/Spark 등을 이용한 분산처리 데이터베이스 구축
 - － 구축된 시스템 유지 보수 관리



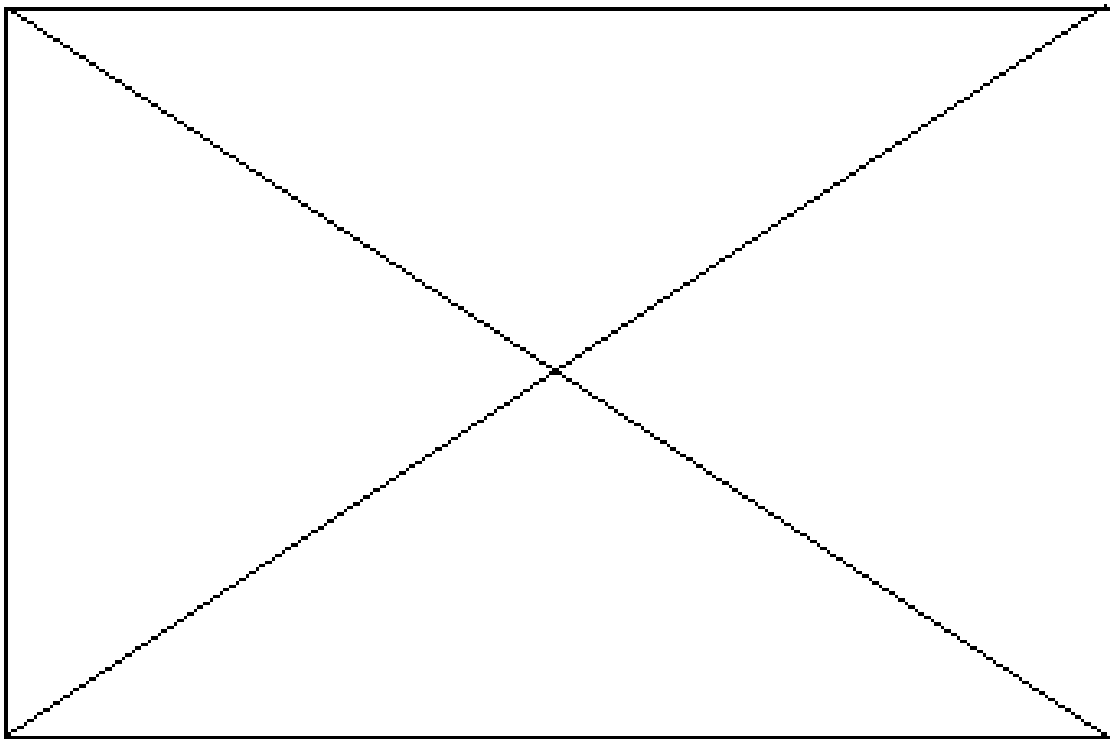
<그림 6-34> 연간 구성해야할 예상 서버 하드웨어(약 5억/년)와 분산시스템 구성

- (인프라) 계층적 플랫폼 구축을 위한 소재 빅데이터 프레임워크 개발
 - － 플랫폼 개발 중 사용되는 재사용이 가능한 수많은 요소기술을 도구 모음 형태로 제공하여, 플랫폼 제작 시 필요 요소 기술을 단순 조립하여 플랫폼 개발 가능
 - － 소재 세부 분야별(각 소전문센터) 맞춤형 플랫폼 개발 목적
 - － 다양한 기능들을 모듈화, 각 분야에 적합한 플랫폼 제작 가능한 프레임워크 개발
 - － 각 소재 분야별 데이터 수집 플랫폼 개발의 요소 기술을 프레임워크 형태로 제공하면, Tier 2의 소재분야별 빅데이터 수집 플랫폼 개발의 효율화 및 수집·활용체계 일관성 확보 가능

< 프레임워크를 이용한 소재분야별 특화 플랫폼 >



－ (예시) 전자정부 표준 프레임워크를 이용해 통일된 각종 정부기관 웹페이지 구축



<그림 6-35> (예시) 전자정부 표준 프레임워크

- (수집) 연구자 행동기반 데이터(비정형 데이터)에서 정형 데이터를 추출하는 기술
 - － 연구자의 “연구행동”이 데이터로 저장
 - － 연구활동에서 발생하는 유·무형 데이터를 모두 수집

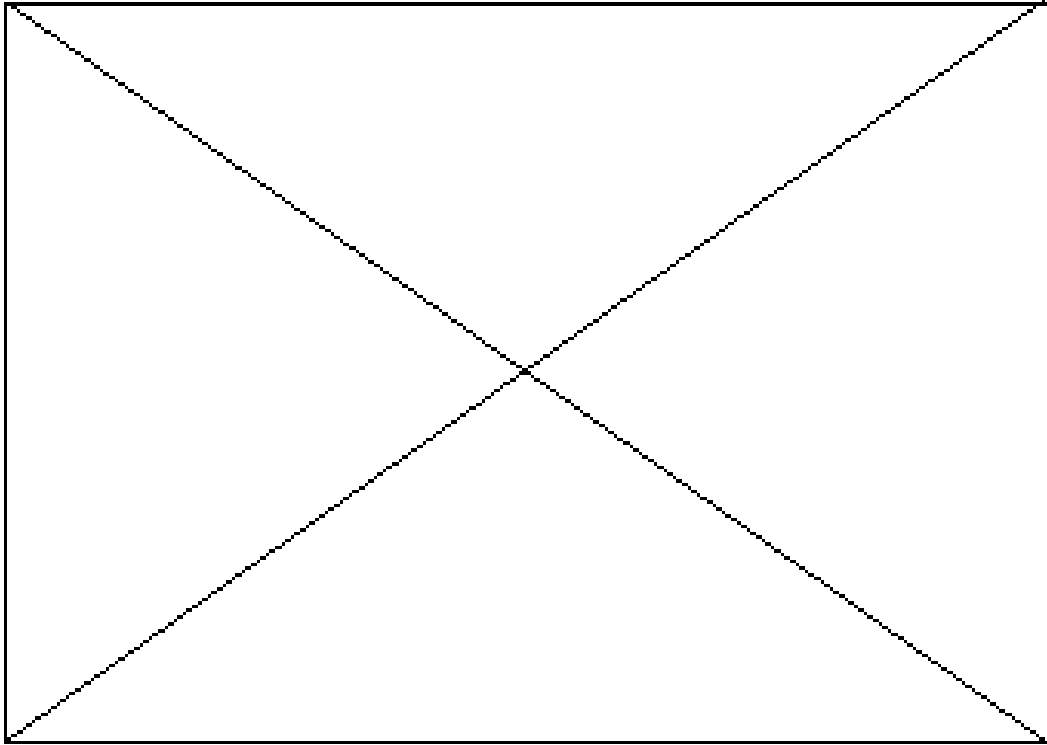
- \ - 연구데이터 수집을 위해 연구자(End-user)에게 추가적인 작업을 요구하는 것은 연구자의 불편과 업무 효율성 저하를 야기, 데이터 수집 플랫폼 활용률 저하
- 정형 데이터뿐만 아니라 비정형 메타데이터(연구노트, 실험정보 등)도 수집 가능

<그림 6-36> 연구자 행동기반 R&D 빅데이터 수집 프로세스

- (수집) 공개문헌(논문, 특허, NTIS 보고서 등) 수집 기술 개발
 - 국내 논문, 특허, 보고서를 각종 국내 학회지·특허청·NTIS과 공조하여 공유
 - ※ 국제 논문, 특허, 보고서는 수집·활용의 법률적 검토가 필요
 - 정부출연연구사업(미래소재디스커버리사업 등) 결과물 수집·활용
 - 정부출연연구사업 결과물, 국내 공개문헌(논문, 특허, 보고서 등) 수집·활용의 법률적 검토
- (표준화) 온톨로지 기반 소재 메타데이터 표준화
 - 소재 분야 연구 전 과정(실험, 분석, 계산, 해석 등)에서 생산되는 데이터의 공통
 - 표준화 설계 (소재분야 전반에 확장 가능한 공통부분의 메타데이터 구조 설계)
 - 소재 세부 분야별(각 소전문센터) 특화된 데이터 표준화 확장 설계
 - 전세계 소재 DB들과의 공유를 위해 소재 연구 빅데이터 구축을 선도하고 있는 RDA산하 IMMR 그룹이 제공하는 'Material Vocabulary'를 참고하여 범세계적으로 적용 가능

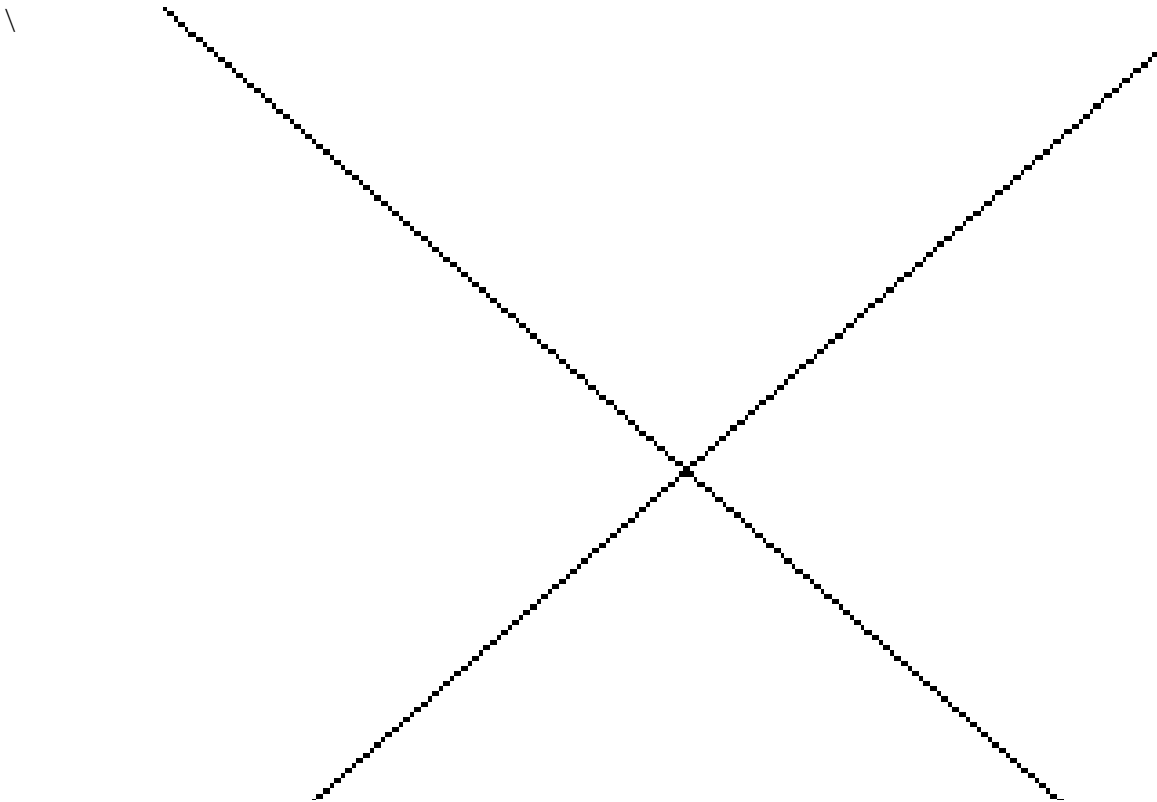
한 메타데이터 표준화

- 데이터베이스 설계를 통하여 정형/비정형 데이터를 효율적으로 저장 관리



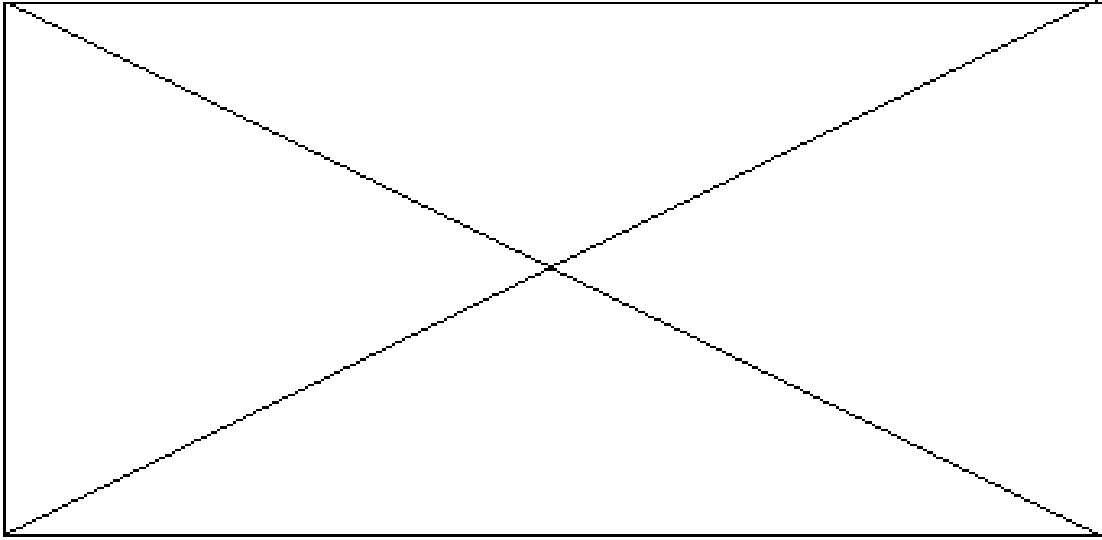
<그림 6-37> RDA 그룹이 제시한 소재 표준 분류(Materials Vocabulary)

- (관리) 연구자 행동기반에 의하여 수집된 연구 빅데이터의 관리 기술
 - 연구자가 자유롭게 작성한 비정형 메타데이터에서 유의미한 데이터를 추출, 메타데이터 표준에 맞게 재배열
- (관리) 수집된 공개문헌(논문, 특허, 보고서 등)의 관리 및 분류 기술
 - 머신러닝 등으로 활용할 수 있도록 수집된 공개문헌(논문, 특허, 보고서 등)에서 정형 데이터 추출 기술 개발
 - (활용) 소재 분야별 데이터 가시화 도구 모듈 기능 개발
 - 연구개발 전과정에서 수집한 빅데이터를 활용하여 기존 소재연구에서 확인하지 못했던 소재 데이터의 경향성 가시화를 통해 연구개발의 명확한 방향성 제시
 - 소재 세부 분야별(각 소전문센터) 특화된 가시화 도구를 위하여 기능을 모듈화



<그림 6-1> (예시) 열전소재 빅데이터 플랫폼에서 제공하는 가시화 도구 (UC Santa Barbara)

- (활용) 소재 분야별 특화 기계학습 도구 모듈 기능 개발
 - 연구자가 소재 세부 분야별(각 소전문센터) 특화된 기계학습을 수행할 수 있도록, 사용자 맞춤형 기계학습 기능 모듈을 개발
 - 아마존 AWS 기계학습 서비스와 같이, 연구자가 자신만의 기계학습 알고리즘 코드를 실행시킬 수 있는 기능을 플랫폼에 구현
- (공유) 글로벌 소재 연방 네트워크 연결
 - API 개발을 통하여 국내·외 연구 DB와의 연동
 - “국가 연구데이터 플랫폼”과의 연동, 국가 연구데이터 플랫폼에서 검색·연구 활동이 가능하도록 메타데이터 제공
 - 국내(소재정보은행, 화학연구원 열전소재, 국가참조표준센터 등)와 국외(NIST, CHiMaD, NIMS 등)의 소재 연구데이터의 메타데이터를 공유
 - 전세계 소재 연구 생태계 연결고리 형성
 - 소재 전문센터가 국내의 저장소(repository) 역할 수행



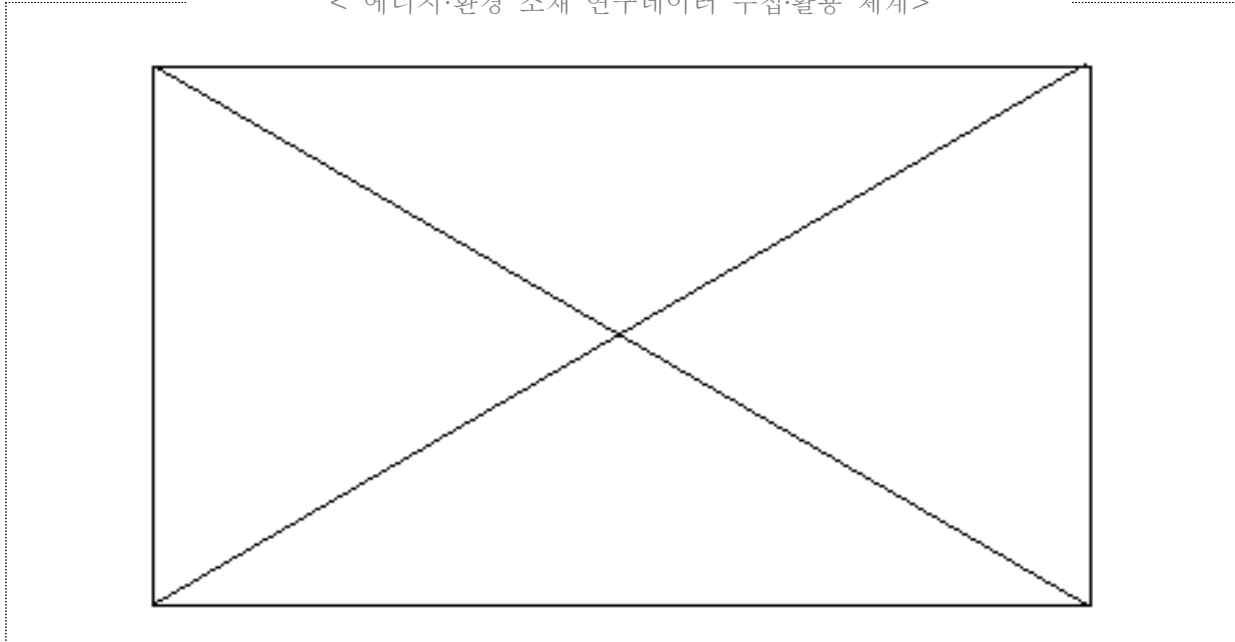
<그림 6-38> 소재 데이터 연방 네트워크에서 당 소재분야 연구데이터 전문센터 역할

- (공유) 연구 공유 플랫폼 개발
 - 사용자 친화적인 연구 공유 플랫폼 개발을 통하여, 연구자의 분야에 관계없이 서로 간의 활발한 교류 및 데이터 공유 환경을 조성하고, 더 나아가 연구 협업 생태계 구축을 위하여 연구데이터 공유와 관련한 다양한 편의 기능 제공
 - (예시) ResearchGate(<https://www.researchgate.net>) 플랫폼은 연구자의 논문 공유 기능과 토의 기능만을 제공함에도 불구하고, 전세계 1500만명 이상의 연구자들이 활발한 교류 및 연구 논문을 공유하고 있음
 - 다양한 분야 연구자 간의 협업 생태계 구축을 통하여, 융복합 R&D 환경 구축

2) (세부과제 2) 에너지·환경 소재 연구데이터 수집·활용 체제 구축

- ◆ 합리적 에너지·환경 소재 연구개발 혁신을 가속화하기 위한 “에너지·환경소재 연구데이터”의 수집·저장 체제 구축 및 빅데이터 기반 소재설계 R&D

< 에너지·환경 소재 연구데이터 수집·활용 체제 >



□ 에너지·환경 소재 데이터 수집 및 관리 기술 개발

- 연구개발 전과정에서 생산된 연구데이터는 분야 및 분석방법 등에 따라 다양한 형태로 존재하며, 공정조건, 분석시기, 방법, 수행주체 등에 따라 동일 소재이더라도 수집된 데이터 간 편차가 존재. 이를 고려하여, 모든 연구데이터 형태를 수용하는 동시에, 에너지·환경 분야를 대표하는 다공성소재, 고체화합물 소재, 이차원 소재에 특화된 데이터 수집, 관리 기술을 개발하는 것이 데이터 인프라 성공의 핵심요소
- 이미 공개된 유사 소재 데이터베이스와의 연동을 통해, 활용 가능한 연구데이터를 늘리고, 이를 통해 타 세세부에서 개발될 빅데이터 기반의 물성예측 모델을 지원하여 그 예측정확도를 향상
- 연구데이터 수집 시 연구자의 부담을 최소화하고 자발적인 연구데이터 공유를 유도하고 자 데이터 수집 과정을 최대한 노출시키지 않고, 연구데이터를 수집하는 과정이 개별 연구자들의 연구 효율성을 높일 수 있도록 지원
- 수집된 연구데이터에 대한 유연하고 강력한 검색엔진을 개발하여, 사용자맞춤형 연구 데이터를 제공함으로써, 연구자의 편이성을 높여 연구개발을 지원

□ 소재데이터 분석 및 소재물성 예측 기계학습 모델 개발

- 소재 데이터베이스가 신소재 개발에 적극적으로 활용되기 위해서는 1차적으로 수집된 연구데이터를 바탕으로 추론할 수 있는 새로운 물성 및 소재정보를 추가적으로 제공하고, 사용자이자 제공주체인 연구자들의 자발적 참여도를 높이기 위해, 소재 데이터에 대한 심화분석과 소재물성 예측에 특화된 기계학습 모델개발 필요
- 화학공간에 광범위하게 분포된 소재 후보군과 소재물성 동시최적화가 필수적인 소재개발에 있어, 빅데이터 기반 소재데이터 분석 및 소재물성 예측모델 필요

□ 연구데이터-소재모델링-기계학습 통합 플랫폼 개발 및 관리

- 에너지·환경 소재 연구데이터 데이터베이스 및 관련 물성예측 기능들을 보다 많은 연구자들이 사용하게 하기 위해서는 웹기반의 플랫폼 구축은 필수적임
- 플랫폼의 필수적인 요소들로는 1) 비전문가도 활용 가능한 영상형 사용자 제어환경(GUI), 2) 연구자가 즉시 피드백을 받을 수 있는 데이터 수집기, 3) 연구데이터-소재모델링-기계학습 기능들 간의 유기적인 연동 4) 플랫폼의 완성도 향상을 위한 베타버전 시험운영 및 시험운영 피드백에 기반한 단계적 플랫폼 최적화 수행, 5) 공개된 플랫폼을 활용한 신규소재 설계 사례 축적을 들 수 있음

가) (세세부 1) 가스 저장·분리용 다공 소재 연구데이터 수집·활용 연구 - 나노포어오믹스(NanoPoreOmics) 플랫폼 개발

□ 현황 및 필요성

- 수소저장, 이산화탄소 및 유독가스 포집 등의 가스저장 및 분리기술 분야의 핵심소재인 다공소재는 에너지·환경 분야에서 폭넓게 활용
- 대표적인 다공소재로는 활성탄소, 제올라이트(Zeolite) 및 금속유기골격체(Metal-organic framework: MOF) 등이 있으며 특히 MOF 소재는 타 다공소재에 비해 높은 비표면적을 지니고 있어 차세대 나노다공소재로 각광을 받고 있음
- 제올라이트 및 MOF계 다공소재는 다양한 형태의 결정구조 및 원소조합 형태를 가질 수 있어 존재 가능한 소재의 경우의 수가 매우 많아 관련 DB의 통합관리 및 인공지능 적용이 절실히 요구됨

□ 국내외 사례

- (미국) 프린스턴 대학을 중심으로 '12년부터 에너지성(DOE) 지원으로 제올라이트 및 MOF 다공소재의 결정구조를 데이터베이스화(DB명: ZEOMICS, MOFomics) 하였고, DOE에서는 다공소재의 수소저장용량을 데이터베이스로 공개
- (미국) 노스웨스턴대학을 중심으로 High-throughput 스크리닝으로 가상의 >100,000 MOF 구조 생성 기술을 개발하였고, 이를 기반으로 관련 소재의 수소/이산화탄소/메탄 가스 저장용량 예측
- (국내) KIST에서 최근('16년도) CSD 및 ICSD 결정구조(~백만개)를 모두 스크리닝하여 65,000종에 이르는 MOF 및 제올라이트계 다공소재 결정구조 데이터베이스 구축

□ 핵심기술

- 문헌(실험/이론) 데이터 수집에 의한 다공소재 구조 및 물성 데이터베이스 구축
- 계산자동화 기술 이용 다공소재의 물성계산 및 데이터베이스 구축
- 소재정보학 기반의 구조-물성 상관관계 규명 및 가시화 기술
- 기계학습 기반의 구조 정보로부터 다공소재 물성 예측 모델 개발
- 기계학습 기반의 다공소재 합성공정 예측 모델 개발
- 모델링-소재 데이터-기계학습 연동 기술 및 웹기반 플랫폼 개발
- 플랫폼 베타버전 시범 운영 및 보완 기술 개발
- 플랫폼 활용 관련 신규 다공소재 설계 사례 연구

□ 연구내용

- (대상 소재) 수소, 천연가스, 이산화탄소 및 유독가스 저장/포집 및 분리용 제올라이트, MOF계 다공소재
- (연구데이터 수집)
 - (물성 데이터) 결정구조, 가스 흡착용량(adsorption isotherm 곡선) 및 선택도, 열·화학적 안정성, 기계적 물성(응력-변형률 곡선) 등
 - (결정구조 및 미세조직 데이터) X-선회절, 미세조직(광학현미경, 주사전자현미경 이미지 등), 스펙트럼(IR, XPS) 데이터 등
 - 다공소재 데이터 수집을 위한 텍스트 마이닝(text mining) 기술 개발

- (계산과학 이용 다공소재 데이터 수집)
 - 다공소재 결정구조 자동생성 프로그램 개발
 - 몬테카를로 시뮬레이션 기반 다공소재 가스 흡착용량 및 선택도 계산 및 DB화
 - 흡착용량 예측정도 향상을 위한 제일원리계산 기반 포스필드(force field) 개발
 - 기존 다공소재 구조-물성 데이터베이스 연동
- (구조-물성 상관관계 규명)
 - Topology Data Science 기반 다공소재 기공구조 정량화/가시화 분석 기술 개발
 - QSPR 방식의 소재 구조와 물성(흡착량, 안정성, 기계적 물성 등) 상관관계 분석 및 인공지능을 위한 feature 도출
- (소재물성 예측 모델 개발)
 - 복잡한 다공소재 결정구조 수치화를 위한 code-based crystallography 기술 개발
 - XRD 구조로부터 다공소재 결정구조 자동화 예측 기계학습 모델 개발
 - 결정구조, 미세조직 정보로부터 물성예측(흡착용량, 안정성 등) 기계학습모델 개발
- (합성공정 예측 모델 개발)
 - 임의의 결정구조로부터 합성/공정 조건 도출을 위한 기계학습 모델 개발
 - 요구물성으로 소재구조 및 합성조건 도출이 가능한 역설계 기계학습 모델 개발
- (모델링-소재 DB-기계학습 연동 기술 개발을 통한 플랫폼 개발)
 - 다공소재 결정구조 모델링 GUI (graphical user interface) 모듈 개발
 - 다공소재 모델링, 소재 데이터베이스, 기계학습 등의 기능별 모듈화
 - 웹 프로그래밍 기술을 이용하여 각 모듈별 연동
 - 계산용 서버(리눅스 클러스터, 클라우드 컴퓨터 등)와의 연동을 통한 물성예측 시뮬레이션 기능 구현 (검색 기능 포함)
- (플랫폼 베타버전 시범 운영)
 - 주기적인 사용자 교육 및 워크샵을 통한 플랫폼 사용자 편의성 제공 및 극대화
 - 설문조사 방식을 통한 사용자 편의성 증대 방안 모색
- (플랫폼 활용 관련 신규 소재 설계 사례 연구)
 - 개발 플랫폼 활용 신규 소재설계 도출 및 연구성과 발표 (신뢰성 검증 목적)
 - 실험검증을 통한 개발 플랫폼의 신뢰성 확보

나) (세세부 2) 전기화학소자용 금속화합물 연구데이터 수집·활용 연구

□ 현황 및 필요성

- 다양한 전기화학 반응(oxidation, hydrogenation, oxygen electrode, gas sensor 등) 기반 소자의 핵심소재군인 금속화합물은 에너지·환경 분야에서 폭넓게 활용
- 전기화학 분야 이외에도 자성, 열전, 유전, 초전도, 압전 등 에너지·환경 연구개발에 널리 적용이 가능하며, 전기화학 물성 데이터베이스를 기반으로 타 연구분야로의 확장이 매우 용이한 원소스멀티유즈 접근이 가능
- 대표적인 금속화합물로는 금속산화물, 금속황화물 등이 있으며, 에너지응용 관련 핵심물성들은 성분, 조성, 화학양론, 샘플형태, 상전이 등 다양한 화학/구조 물성변수에 따라 변화. 조절가능한 물성변수의 다양성에 따라 신규소재 발굴 가능성이 높으므로 관련 연구데이터의 체계적 수집·관리 및 인공지능의 적용이 필수적임

□ 국내외 사례

- (미국) 버클리대학교를 중심으로 진행되고 있는 Materials Project는 2011년부터 8만 종 이상의 무기화합물 결정구조 데이터베이스를 공개하였고, 해당 프로젝트는 배터리 소재 물성에 특화되어 있는 한편, 일부 기초 이론물성도 제공
- (유럽) 스위스연방공대를 중심으로한 Materials Cloud 에서는 대표적 금속화합물인 페로브스카이트 결정구조 9만종에 대한 빅데이터 분석을 수행하여 구조안정성 관련 소재 설계 규칙을 제안
- (일본) NIMS는 무기소재 데이터베이스 (AtomWork) 를 공개하고 있으며, 82,000개 결정구조, 55,000종에 대한 소재물성 등을 제공

□ 핵심기술

- 기존문헌 및 데이터베이스(실험/이론)에서 금속화합물 물성수집 및 관리 기술
- 계산자동화를 통한 소재 결정구조계산 및 관련 물성 수집 및 관리 기술
- 높은 대표성을 띄는 다원계 고체화합물 구조모델 구현 기술
- 주요 실험 데이터 수집 및 분석 기술
- 빅데이터 기반의 핵심물성 예측 기술
- 소재모델링-연구데이터-빅데이터 분석 실시간 연동기술 및 소프트웨어 플랫폼 구축
- 플랫폼 기반 에너지·환경 관련 신규 고체화합물 소재 설계사례 연구

□ 연구내용

- (대상 소재) 다양한 전기화학 반응 기반의 에너지·환경 소자용 다원계 금속산화물, 금속 화합물 등의 고체화합물 소재
- (수집 연구데이터)
 - －(물성 데이터) 전기화학 물성(전기전도도, ASR, surface oxygen exchange coefficient, ORR 활성도), 구조물성, 열화학적 안정성, 기계적 물성, 이온전도도등
 - －(결정구조 및 미세조직 데이터) X-선회절, 미세조직(광학현미경, SEM, TEM 등), 스펙트럼(IR, XPS) 데이터
 - －(계산과학 데이터) 기초물성 데이터(격자상수, 열역학 물성, 밴드갭, 탄성계수 등), 고급 물성 데이터(계면안정성, 계면구조, 박막구조, 전기화학물성)
 - －(빅데이터 기반 데이터) 빅데이터 기반 예측모델로부터 얻어진 신규 물성데이터 및 연구데이터 후처리 데이터
- (계산자동화를 통한 소재 결정구조계산 및 관련 물성 수집 및 관리 기술)
 - －입력 결정구조에 대한 기초물성 데이터 수집·관리 프로세스 자동화 기술 개발
- (높은 대표성을 띄는 다원계 고체화합물 구조모델 구현 기술)
 - －벌크를 포함한 다양한 형태의 소재(박막소재, 표면 등)구조 구현 자동화
 - －대표성 확보를 위해 고급 결정구조 생성 방법론 [Virtual Crystal Approximation (VCA), Special Quasi-random Structure(SQS), 유전알고리즘 기반 소재 예측 (예: USPEX, GASP), CALYPSO 등] 활용 및 최적화 수행
- 기존문헌 및 DB(실험/이론)에서 금속화합물 물성 수집 및 관리 기술)
 - －문헌 연구데이터 수집을 위한 자연어처리 기술 활용 및 최적화
 - －기존 소재 DB를 활용한 효율적인 기초 데이터축적 수행
- (연구 데이터 수집 및 분석 기술)
 - －빅데이터 기반의 데이터 후처리를 통한 연구데이터 확장(미세조직 데이터의 입경분포, 입자크기 분석자동화)
 - －빅데이터 분석에 기반한 기초물성을 함수로 하는 핵심 소재물성 예측모델을 활용한 연구데이터 확장
 - －소재모델링-연구데이터-빅데이터 분석과정의 실시간 연동 기술
- (전기화학 기반 소자용 금속화합물 소재 데이터베이스 플랫폼 개발)

- \ - 소재 시뮬레이션, 소재 데이터베이스, 빅데이터, GUI 등에 대한 세부기능들의 체계적 모듈화/웹 프로그래밍 기술을 이용하여 각 모듈별 연동
- 시뮬레이션용 전산장비(On-site 또는 Cloud)와의 연동을 통한 물성예측 시뮬레이션 기능 구현 (검색 기능 포함)
- (플랫폼 베타버전 시범 운영)
 - 내외부 베타테스트를 통한 플랫폼 성능 최적화
 - 사용자 교육, 워크샵, 설문조사 등을 통한 플랫폼 사용자 편의성 제고
- (플랫폼 활용 관련 신규소재 설계 사례 연구)
 - 개발 플랫폼 활용 신규소재 설계 결과 도출 및 성과 발표 (신뢰성 검증 목적)
 - 실험검증을 통한 개발 플랫폼의 신뢰성 확보

다) (세세부 3) 에너지 변환 이차원 소재 연구데이터 수집·활용 연구

□ 현황 및 필요성

- 2004년 그래핀 제조 성공 이후 실리센(Silicene), 포스포린(phosphorene), BN, MoS2 등 다양한 이차원(2D) 물질들이 발굴되면서, 이차원 물질을 활용한 고효율 에너지 변환 소재 개발이 활성화
- 대표적인 예로 이차원 소재인 BiTe가 이론적 열전성능지수(ZT) ~0.7을 나타내 실험적으로도 가장 우수한 열전물질로 알려졌으며, 최근 물분해에서 MoS2로 백금(Pt) 촉매에 필적하는 성능을 얻어내는 등 이차원 물질을 사용한 연구가 가속화
- 이론적으로 이차원 물질의 수는 수백 종에 이르나 이차원 물질들 간의 이종구조(heterostructure)까지 고려하면 이차원 물질로 가능한 소재 수는 무궁무진한 반면, 이차원 소재 대상 실험연구는 극히 일부 물질에 국한되어 있어 향후 차세대 에너지 변환 및 저장용 이차원 소재 개발을 위해서는 데이터 수집·활용 연구가 필요

□ 국내외 사례

- (유럽) `13년 프랑스, 핀란드, 스웨덴 연구진의 공동연구로 이차원 물질 탐색방법을 개발하여 실험적으로 알려지지 않은 다양한 이차원 물질들을 예측하였고, `18년 덴마크에서는 이차원 소재를 대상으로 기초물성 계산자동화 플랫폼을 개발하고, 이를 통하여 C2DB(Computaitonal 2D Materials Database) 데이터베이스 공개

- (국내) '18년 서울대에서 High-throughput 계산으로 이차원 TMD(transition-metal dichalcogenides) 물질군에 대한 수소흡착에너지를 예측하여 물분해 촉매 성능 예측 데이터베이스 구축

□ 핵심기술

- 기존 문헌(실험/이론)에서 이차원 소재 및 적층 소재의 구조, 물성, 메타데이터 데이터베이스 구축
- 계산자동화 기술 이용 이차원 소재 및 이종구조의 물성 및 안정성 데이터 구축
- 소재정보학 기반의 구조-물성 및 구조-안정성 상관관계 규명 및 가시화 기술
- 기계학습 기반의 구조 정보로부터 이차원 소재 물성 예측 모델 개발
- 기계학습 기반의 이차원 소재 합성 공정 예측 모델 개발
- 모델링-소재 DB-기계학습 연동 기술 및 웹기반 플랫폼 개발
- 플랫폼 활용 관련 신규 이차원 소재 설계 사례 연구

□ 연구내용

- (대상 소재) 열전/촉매/에너지저장용 이차원 소재 및 이차원 이종구조 소재
- (연구데이터 수집)
 - (물성 데이터) 결정구조, 열/화학적 안정성, 합성조건, 열전지수, 열팽창지수, 수소전환 효율(TOF), 물분해 과전압 등
 - (결정구조 및 미세조직 데이터) X-선회절, 미세조직(광학현미경, SEM 등), 스펙트럼(IR, XPS 등) 데이터
 - 이차원 소재 연구 데이터 수집을 위한 텍스트 마이닝(text mining) 기술 개발
- (계산과학 이용 이차원 소재 데이터 수집)
 - 이차원 소재 구조선별 및 생성 자동화 프로그램 개발
 - 이차원 최적 이종구조 탐색 자동화 프로그램 개발
 - 이차원 소재 기초물성(전자구조, 열특성), 분자흡착에너지 계산 자동화플랫폼 개발
 - 기존 이차원 소재 구조/물성 데이터베이스와 연동

\ ○ (구조-물성 상관관계 규명)

- Topology Data Science 기반 이차원 소재의 구조 정량화/가시화 분석 기술 개발
- QSPR 방식의 소재 구조와 물성(흡착량, 안정성, 기계적 물성 등) 상관관계 분석 및 빅데이터 분석을 위한 주요 인자 도출

○ (소재물성 예측 모델 개발)

- XRD 구조로부터 이차원 소재 결정구조 자동화 예측 기계학습 모델 개발
- 단일 이차원 소재정보로부터 이종구조 물성예측(열전지수, 촉매효율, 안정성 등) 기계학습 모델 개발

○ (합성공정 예측 모델 개발)

- 임의 결정구조 정보로부터 합성/공정 조건 도출을 위한 기계학습 모델 개발
- 사용자 요구물성의 소재 구조 및 합성조건 도출 역설계용 기계학습 모델 개발

○ (모델링-소재 데이터베이스-기계학습 연동 기술 개발을 통한 플랫폼 개발)

- 이차원 소재 모델링, 소재 데이터베이스, 기계학습 등의 기능별 모듈화
- 웹 프로그래밍 기술을 이용한 모듈간 연동
- 계산용 서버(리눅스 클러스터, 클라우드 컴퓨터 등)와의 연동을 통한 사용자 친화적 물성 예측 시뮬레이션 플랫폼 구현
- 기존 데이터베이스 검색 및 신규 시뮬레이션 결과와의 연동 자동화 기능 구현
- 플랫폼 오픈베타 버전 시범 운영을 통해 사용자 편의성, 활용성 극대화

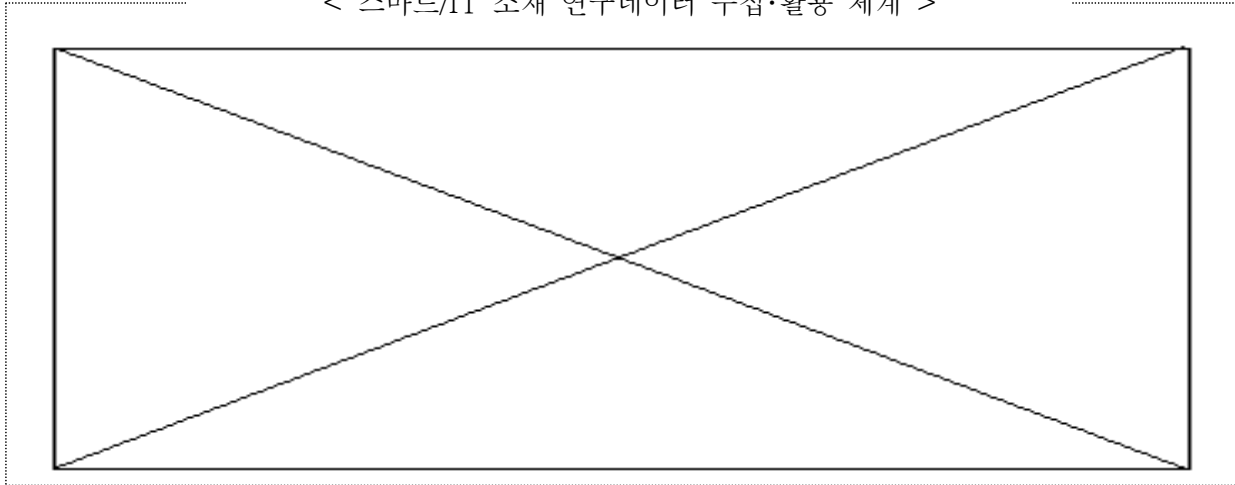
○ (플랫폼 활용 관련 신규 소재 설계 사례 연구)

- 개발 플랫폼 활용 신규소재 설계 결과 도출 및 성과 발표 (신뢰성 검증 목적)
- 실험검증을 통한 개발 플랫폼의 신뢰성 확보

3) (세부과제 3) 스마트/IT 소재 연구데이터 수집·활용 체계 구축

- ◆ IoT/IoH 기반 초연결사회(4차 산업)를 선도하기 위한 데이터 구동형 스마트 · IT 소재 연구개발

< 스마트/IT 소재 연구데이터 수집·활용 체계 >



가) (세세부 1) 차세대 반도체 소재 연구데이터 수집·활용 연구

□ 현황 및 필요성

- IoT/IoH 기반의 초연결사회에서는 데이터의 폭증이 예상되며, 데이터를 효율적으로 처리할 수 있는 Exa스케일 지능형 연산을 위한 새로운 반도체 소재 개발이 필요함
- 기존 Si 기반 반도체 기술로는 고효율 지능형 연산용 시스템의 구현에 한계가 있어 두 뇌모방 인지연산 반도체 기술의 바탕이 되는 신소재 개발이 절실히 요구됨
- 미국·일본·유럽 등 기술선도 국가에서는 최근 소재 데이터베이스 구축을 통하여 R&D 연구데이터 수집·공유·활용 플랫폼 개발이 이뤄지고 있으며, 소재정보학이 신소재 개발 연구의 새로운 패러다임으로 급부상하고 있음

□ 국내외 사례

- (유럽) ICSD, CSD, Chem Spider 등 전 세계적으로 가장 큰 “소재 구조정보” DB가 구축되고 있으며, 소재 데이터를 이용한 신소재 개발 연구가 활발히 이루어짐
- (미국) MGI를 통해 국가 차원의 소재 데이터베이스 구축을 선도적으로 추진 중이며, 빅데이터 분석을 이용한 소재 개발이 에너지 소재, 반도체, 촉매 등 여러 분야에 적용 중임

□ 핵심기술

- Non-Si 기반 고효율·고성능 반도체 신소재 데이터 표준화
- 연구개발 과정에서 발생하는 메타데이터 수집 프로세스 체계화
- 소재 원자 및 화학결합 특성, 미세구조, 공정변수와 반도체 특성 관계 데이터화
- 멀티스케일 계산과학 기반 반도체 물성 예측 기술
- 기계학습 기반 공정-미세구조-특성-성능 (PSPP) 관계 예측 및 공정가이드 기술
- 빅데이터 활용 Non-Si 기반 고효율·고성능 반도체 신소재 탐색 기술

□ 연구내용

- (대상 소재) 초저전력·고집적·고성능 두뇌모방 인지연산 반도체용 Non-Si 소재(금속산화물, 저차원 소재, 화합물반도체 등)
- (연구데이터 수집)
 - －(공정 데이터) 공정이력 (장비, 온도, 습도, 압력, 전구체 종류 등 공정 전 데이터)
 - －(측정 데이터) 미세조직 이미지, 화학조성데이터 (XPS, EDS, EELS, Raman 등), 회절 데이터 (XRD, 전자후방산란회절 등)
 - －(물성 데이터) 전기적 특성(밴드갭, 전하이동도), 기계적 성능(영률, 탄성계수), 물리화학적 특성(열전도도, 화학적 안정성, 표면에너지 등)
 - －(계산 데이터) 제일원리계산 기반 소재 데이터 (원자구조, 결합특성, 밴드구조 등)
- (연구데이터 수집-기계학습 연동 기술 개발을 위한 플랫폼 개발)
 - －연구데이터 수집 체계화 및 데이터 수집 플랫폼 구현 모듈 개발
 - －실험 결과 및 계산과학 결과의 소재 DB 수집, 기계학습 기능별 모듈화
 - －연구자의 접근성 향상을 위한 GUI 및 웹기반 플랫폼 구현
- (연구데이터 플랫폼 활용 신규 소재 설계)
 - －빅데이터, 기계학습을 이용한 PSPP 예측하여 최적 공정가이드 기술 제공
 - －반도체 소재데이터 플랫폼 구축 및 빅데이터 기반 신소재 개발 방법론 확보
- (플랫폼 홍보 및 교육)
 - －플랫폼 사용 매뉴얼 제작 및 배포
 - －주기적인 개발자/사용자 워크숍을 통한 플랫폼 사용자 활용성 극대화

나) (세세부 2) 센서 소재 연구데이터 수집·활용 연구

□ 현황 및 필요성

- 현재 센서생산량이 연평균 1조개 이상이며, 4차 산업혁명 시대 도래에 따라 센서시장은 폭발적으로 확장
- 사물인터넷 (IoT) 시대 진입에 따라 외부환경변화를 인지할 수 있는 센서의 중요성이 커지고 있으며, 특히, 스마트 리빙과 밀접한 관련을 가지는 헬스케어, 환경, 에너지 기술과의 융합이 진행
- 센서로 활용 가능한 소재로서 세라믹, 금속, 고분자 모두 활용 가능하지만 현재 범용으로 가장 널리 활용되는 소재는 실리콘 (Si) 반도체임. 그 외, 고온안전성으로 인해 산화물반도체의 경우 가스센서로서 많이 활용되고 있음
- IoT시대에 이르러서는 실리콘이 가진 한계를 극복하고 다양한 외부환경을 효율적으로 인지하기위한 고선택성, 고감도 센서 소재를 개발하는 것이 중요
- 센서의 성능은 기본적으로 소재 특성에 의해 결정됨. 잠재적인 특정인식인자에 대한 고감도 센서 개발 요구에 신속하게 대응하기 위해서는 다양한 소재들의 센서 특성 데이터를 생산/수집/분석 하는 것이 필요

□ 국내외 사례

- 2000년대 중반 이후 독일 아헨대학의 Simon group은 독일연방정부의 지원으로 high-throughput 실험방법론을 개발하고 이를 통해 가스감지를 위한 센서 소재 선별을 시도하고 있음
- 그 외 미국, 호주, 중국 등에서 활동하고 있는 개별 연구 그룹들이 대용량 실험을 통해 가스센서 소재의 물성 데이터를 생산, 수집 및 분석을 진행하고 있고 이를 통해 센서 소재를 제안

□ 핵심기술

- 다양한 인지인자에 따른 소재의 감지능력 정량화
- 소재 구성 원소 및 구조와 특정인식인자에 대한 감도 사이에 관계 정립
- Data mining과 기계학습을 통한 센서소재 개발 기술
- 대용량 제일원리계산을 통한 소재 특성 예측 및 소재 선별 기술

□ 연구내용

- (인지인자) 빛, 가스, 온도, 습도 등 스마트 리빙에 필요한 인자들을 효율적으로 감지할 수 있는 센서소재 데이터 확보
- (연구데이터 수집)
 - － 기국가연구과제를 통해 생산 및 보고된 연구 데이터 수집 및 향후 국가과제를 통해 소재데이터 직접 생산 및 수집
 - － 감지대상 맞춤형 소재 물성 데이터 베이스 정립

<표 6-21> 센서 소재 연구데이터 수집 예시

인지인자	관련물성
빛	빛 노출시 저항 변화 및 반응 속도, 밴드갭 에너지, 파장별 빛 흡수도
가스	가습 흡착시 저항 변화 및 반응 속도, 가스와 센서의 상대적 양자 에너지 레벨, 가스 흡착 에너지
온도 및 습도	온도 및 습도 변화시 저항 변화 및 반응 속도, 열팽창 계수

- － 극한 환경에 적용하기 위한 소재의 열역학 및 기계적 성질(강도, 연성, 인성, 휘어짐) 데이터 확보
- － (계산 데이터) 제일원리계산에 의한 저비용 고효율 물성데이터 확보 및 소재의 밴드갭 에너지, 기계적성질 계산 데이터, 열역학 데이터 생산 및 수집
- (연구데이터 활용)
 - － 스마트 리빙을 위한 센서소재 데이터 기술 (플랫폼+데이터활용) 확립
 - － 기계학습 연동 센서용 신소재 설계 기술 개발

다) (세세부 3) 바이오/의료 소재 연구데이터 수집·활용 연구

□ 현황 및 필요성

- 바이오/의료 소재는 의학/약학/생물학 등 이종학문과의 필수적 다학제 융합연구, 동물실험/임상실험 및 소재활용 단계에서의 필수적 인허가 등 다양한 장벽이 존재하여 최근까지 선진국 주도로 R&D가 추진됨
- 한편 전 세계적 고령화와 생활수준개선에 따라 바이오/의료 소재산업은 급속히 성장하고 있으며, 또한 우리나라의 경우 기존 주력산업의 정체상황을 극복하기 위한 가장 전망 있

는 미래성장동력 중 하나로 인식하여 국가적 차원에서 활발한 지원이 이루어지고 있음

- 바이오/의료 소재 분야 후발국가로서 신속히 글로벌 리더로서의 위치를 점하고 세계 바이오의료시장의 주역으로 발돋움하기 위해서는 소재분야 연구 데이터를 공유/활용하는 것은 매우 효과적임
- 바이오의료 소재는 소재의 구성 및 종류가 다양하고, 물리화학적 특성 외에도 생물학적 특성과 의학적 기능 등 막대한 소재정보에 대한 평가 및 확인이 요구되므로 소재정보의 공유 및 활용에 따른 파급 효과가 매우 큼(연구개발 기간 단축, 실용화 기회 확대, 의료 사고 방지 등)

□ 핵심기술

- 소재 데이터 수집 및 마이닝 기법 개발
- 열역학 계산, 고분자 구조 계산을 활용한 비실험적 방법에 의한 소재 데이터 생성
- 수집 데이터에 대한 집단지성을 활용한 검증을 통한 데이터 신뢰도 향상

□ 연구내용

- (소재 개요 및 특성) 바이오/의료 소재는 기존 소재 외에 정형화/표준화하기 어려운 생체유래물질 등을 포함하는 등 범위가 매우 넓고, 동시에 주변 환경(media 환경 또는 생체환경)에 따라 특성/기능/효과가 변화하는 매우 복잡한 특성을 가짐
- (연구데이터 수집)
 - － 일반적인 소재정보 뿐만 아니라 생체적합성, 의료용 소재로서의 치료 기능, 오감증강 기능, 센싱 기능 등 다양한 관련 정보를 수집하고 데이터베이스 구축 및 실제 활용을 위한 인허가 정보 수집
- (연구데이터 생성, 정보화 및 활용)
 - － 논문, 특허, 검증이 필요한 연구과제 결과보고서, 언론 공개정보, 상용 소프트웨어 생성 정보 등 소재정보 수집 source를 개발하고, source별 취합 데이터 신뢰성 제고 방안 도출
 - － 금속, 세라믹, 고분자, 복합소재 및 생체유래 물질 등 다양한 바이오/의료 소재 관련 정보의 효과적 수집 기법 및 시스템 개발
 - － 기존 데이터 활용을 목적으로 수집된 데이터의 발췌/분석/표준화/저장을 위한 일반화된 체계 구축
 - － 수집 데이터의 신뢰성 검증 및 신뢰성 향상 방안 도출

표 6-22 > 바이오/의료 소재 연구데이터 수집 예시

정보 분류		상세 정보 내용
소재구성		구성원소, 조성, 결정구조, 제조법 및 제조 조건 등
소재특성	기본 물성	기계적 물성(강도/연신률/영률 등) 및 전기화학적 물성(내식성, 전기전도도 등)
	생체적합성	세포독성, 면역독성, 감작성, 유전독성 등 ISO10993 요구 정보
	생체/의료기능	조직재생/혈관화 촉진 기능, 세포 분화조절 기능, 질병 치료 기능, 오감증강 기능, 약물전달 기능, 바이오시그널 및 바이오마커 검지 기능 등
인허가 정보		전임상, 임상, 식약처 인허가, CE 허가, FDA 허가 획득 정보

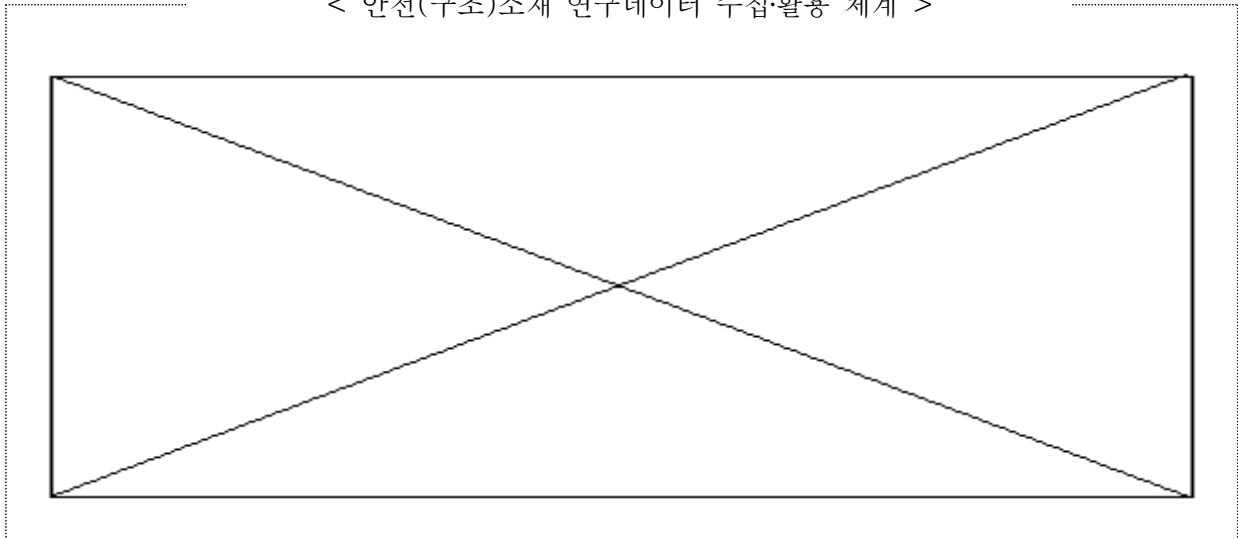
○ (연구데이터 활용)

- 구축 데이터를 활용하여 머신러닝을 통해 생체소재의 특성예측
- 바이오의료용 신소재 최적 설계 방안 도출을 위한 가상공학 플랫폼 개발

4) (세부과제 4) 안전(구조) 소재 연구데이터 수집·활용 체계 구축

- ◆ 국가 주력산업의 기반인 구조 소재의 연구방법론 혁신을 위한 안전(구조)소재 연구데이터 수집·활용 체계 구축 및 데이터 구동형 신소재 개발

< 안전(구조)소재 연구데이터 수집·활용 체계 >



가) (세세부 1) 극한환경 소재 연구데이터 수집·활용 연구

□ 현황 및 필요성

- 에너지 자원의 고갈, 우주항공개발을 통한 신산업 독점 및 안보 위협, 국민 안전을 위협하는 대기 및 환경오염, 첨단무기 개발 등을 대상으로 하는 극한연구는 국가안보(에너지, 우주항공, 환경, 국방)와 직결되어 있으며, 극한 측정기술과 극한 소재는 극한연구의 활용을 위해 가장 핵심적으로 확보해야할 기술
- 선진국은 극한 기초과학을 지속가능 성장을 위한 국가 미래성장 동력의 초석으로서 인식하고, 4차 산업혁명에 대한 성공적인 대응과 국가생존과 관련된 미래 신산업 발굴이라는 측면에서 국가 주도하에 극한소재에 대한 연구를 장기적으로 진행

□ 국내외 사례

- (미국) 2015년부터 MURI(Multidisciplinary University Research Initiative) 프로그램의 일환으로 초고온에서 사용 가능한 소재개발을 진행하고 있으며 단순히 신소재 개발에 그치지 않고 데이터 기반 소재개발 방법론을 도출하는 것으로 목표로 진행

- \ ○ (일본) 2014년부터 연 370억원 규모의 SIP-SM4I 프로젝트를 통해 세라믹 코팅 및 복합소재를 포함하는 항공우주 소재개발 연구를 진행하고 있으며, 국립재료연구소(NIMS)는 1966년부터 1,000여기 이상의 크리프 시험기를 운용하여 고온 크리프 물성 데이터베이스를 축적

□ 핵심기술

- 극한소재 온톨로지 설계 및 데이터 표준화 기술
- 극한환경 구현/측정, 소재 원소특성, 미세조직 및 역학물성 관계 가시화 기술
- 극한환경 실험적 측정 및 전산재료과학 기법을 이용한 미세조직 및 소재물성 예측 기술
- 기계학습을 이용한 성분-미세조직-특성 관계 해석 및 역설계 기술

□ 연구내용

- (대상 소재) 극고온, 극저온, 극고압, 극자장 및 극부식 등의 극한환경에 적용 가능한 금속, 세라믹 및 유무기복합 소재
- (연구데이터 수집)
 - －(환경 데이터) 온도, 압력, 자장, 산도 등의 환경 조건
 - －(물성 데이터) 기계적 특성 및 성능(강도, 연성, 인성, 크리프, 피로), 열적/물리화학적 성질(결정구조, 상전이점, 밀도, 열팽창률, 비열, 열전도도, 전기전도도) 및 부식/산화 성능
 - －(미세조직 데이터) 미세조직 이미지(광학·전자현미경) 및 회절데이터(에너지산란분광, 전자후방산란회절 등)
 - －(계산 데이터) 제일원리계산 기반 소재물성 데이터, 계산열역학 데이터(상태도 및 미세조직 예측 데이터 등)
- (연구데이터 활용)
 - －극한소재 연구데이터 플랫폼 구축과 데이터 기반의 소재설계 방법론 확보
 - －수집 데이터 및 기계학습을 활용한 극한환경용 신소재 도출 및 실험적 검증

나) (세세부 2) 경량소재 연구데이터 수집·활용 연구

□ 현황 및 필요성

- 구조용 소재는 소재설계에서 실용화 적용까지 장기간(10~20년)이 소요되어 2,000년 이후 전산재료공학 기반 소재 모델링 기법, 2010년 중반 이후에는 전산재료공학과 데이터과학의 융합에 의한 연구방법론 개발로 연구개발 혁신을 가속화하는 단계
- 미래형 자동차, 항공우주, 국방 및 로봇 등 미래주력산업의 경쟁력은 철강 대비 비강도 특성이 우수한 경량소재 기술 확보에 좌우되며, 우리나라에서는 경량소재 세계시장 확대에 대응하기 위하여 경량소재를 9대 국가전략프로젝트의 하나로 추진
- 4차 산업혁명의 대두에 따라 적층조형공정에 적용되는 경량신소재 개발 및 전자파 간섭 저감, 스텔스 기능 등 전자기 특성제어 복합소재에 대한 니즈 증가
- 기술선진국에서는 소재 데이터베이스 구축을 통하여 연구개발 결과의 체계화 및 소재 신뢰성을 확보하여 왔으며, 최근 실험, 데이터와 시뮬레이션의 연계에 의한 정량적 구조-특성-성능 관계에 입각한 경량소재 개발 기술이 급속히 부상하고 있음

□ 국내외 사례

- (미국) ‘16년부터 DOE 지원으로 경량소재 R&D를 위한 국가연구소 콘소시엄인 LightMAT를 결성하고, 5개 국가연구소에서 데이터 수집, 가공, 활용을 통한 생태계 구성 및 계산재료공학과 연계로 통한 경량소재 연구개발 혁신 추진
- (일본) ‘14년부터 내각부 주도 전략적 이노베이션 창조 프로그램(SIP)의 혁신적 구조재료과제(’18년, 34억엔)에서 소재 데이터베이스를 활용한 재료-정보공학 연계에 의해, 차세대 경량재료 개발을 위한 역설계 기반 신소재 설계 기술 개발
- 구조소재의 경우 기초물성(물리화학) 데이터, 미세조직 데이터에서 실사용 환경에서의 성능 데이터까지 다양한 종류의 데이터가 산발적으로 존재하여, 이러한 요인을 극복하기 위하여 전이학습(Transfer Learning), 순차적 학습(Sequential Learning), 합성곱신경망(Convolutional Neural Network) 등 다양한 기계학습 방법이 구조소재 연구데이터 해석에 적용되고 있음

□ 핵심기술

- 경량 구조용 합금 온톨로지 설계, 메타데이터 및 데이터 표준화
- 소재 원소특성, 물리화학적 인자, 미세조직(구조)와 특성관계 가시화

- \ ○ 기계학습 기반 멀티스케일 정형 및 비정형 미세구조(조직) 데이터 정량화
- 제일원리 기반 소재물성 예측 기술 및 데이터 동화에 의한 소재물성 추출 기술
- 고엔트로피합금 등 불규칙 구조 다원계 합금의 제일원리계산 데이터 생성
- 기계학습 기반 공정-미세조직-특성-성능(PSPP) 관계 해석 및 역설계 기술

□ 연구내용

- (대상 소재) 미래형 자동차, 우주항공, 로봇 등 미래 산업용 경량합금(알루미늄, 타이타늄 합금) 및 복합재료
- (연구데이터 수집)
 - (물성 데이터) 기계적 특성(강도, 연성, 인성), 기계적 성능(크리프, 피로) 물리화학적 성질(확산계수, 밀도, 열전도도, 전기전도도, 전기화학적 성질 등)
 - (전자기적 특성) 전자파 차폐/흡수 특성 데이터(유전율, 투자율, 임피던스 등), 전자기 특성 데이터(보자력, 자속밀도, 최대 에너지적 등)
 - (미세조직 데이터) 미세조직 이미지(광학·전자현미경) 및 회절데이터(에너지산란분광, 전자후방산란회절 등)
 - (계산 데이터) 제일원리계산 기반 기초 소재물성 데이터, 계산열역학 데이터(상태도 및 미세조직 예측 데이터 등), 시뮬레이션 데이터(전산모사 기반 미세조직 예측 데이터, 유한요소해석 데이터 등)
- (연구데이터 활용)
 - 구조용 경량합금 데이터 플랫폼 구축과 데이터 구동형 소재연구 방법론 확보
 - 전자기 특성 제어용 경량 복합소재 특성예측 인공지능 및 가상공학 기술 개발
 - 수집 데이터, 기계학습을 활용한 경량 신소재 도출 및 실험적 검증

다) (세세부 3) 수소인프라 소재 연구데이터 수집·활용 연구

□ 현황 및 필요성

- 수소 경제 사회가 도래함에 따라 수소전기차, 수소스테이션, 연료전지, 개질기 및 수전해기 등 수소를 사용하는 인프라 소재에 대한 기술 확보가 매우 시급한 상황이며, 기획재정부에서는 빅데이터, 인공지능과 더불어 수소 분야를 3대 전략투자 분야로 선정하고 2019년부터 집중적으로 예산을 투입할 예정
- 미국 및 유럽 등 기술선진국에서는 수소인프라 소재에 대한 데이터베이스 구축을 통하

여 기존 소재 특성을 체계적으로 정리하고 새로운 소재 개발에 활용하고 있으며, 특히 최근 급격히 발전한 전산모사 기법과 연결하여 직접적으로 관찰이 어려운 소재 내의 수소 거동을 이해하고 예측하는 성과 창출

□ 국내외 사례

- (미국) 2,000년부터 DOE 지원으로 Hydrogen & Fuel Cells 프로그램이 출범하여 국립연구소를 중심으로 수소 생산, 이송, 저장 및 이용에 관련된 다양한 소재에 대한 데이터 수집이 이루어지고 있음
- Sandia National Laboratories를 중심으로 수소인프라 소재에 대한 데이터베이스 (HydPark) 구축 및 코드제정 및 표준화가 진행되어 왔으며, 최근 이 데이터베이스는 DOE Fuel Cell Technologies Office 산하의 Hydrogen Storage Materials Database로 편입되어 운용중
- (유럽) 2014년부터 FCH JU 파트너쉽을 통해 Horizon 2020 프로그램의 50여개의 수소 및 연료전지 과제를 통해 확보된 수송 및 에너지 저장에 관련된 소재 데이터를 공유
- (일본) 2,000년대 중반부터 신에너지·산업기술종합개발기구(NEDO)의 수소선단과학기초연구사업, 수소제조·수송·저장 시스템 등 기술개발사업 등을 통하여, 수소제조·수송·저장 시스템 등에 사용되는 금속재료 개발 및 국제표준화규제 제정을 위한 평가기술 개발 및 소재 데이터 취득에 관한 광범위한 이론(계산) 및 실험연구 수행

□ 핵심기술

- 수소인프라소재의 온톨로지 설계 및 데이터 표준화 기술
- 소재 내부 수소의 물리화학적 특성인자, 미세조직 및 역학물성 관계 가시화 기술
- 전산재료과학 기법을 이용한 소재 내부 수소 거동 및 소재 물성 예측 기술
- 축적된 기초데이터와 실험결과를 연계할 수 있는 데이터 플랫폼 기술

□ 연구내용

- (대상 소재) 수소인프라의 압력용기, 배관/밸브/펌프, 연료전지 분리판, 수소저장 및 분리막 소재
- (연구데이터 수집)
 - (수집방법) 기존 정부지원 연구사업(수소에너지프론티어사업, 수소연료전지사업, 기후변

- \ 화대응사업, 미래소재디스커버리사업 등) 지원과제 연구데이터 확보, 국내외 논문/특허/데이터베이스 데이터 수집 및 high-throughput 실험을 통한 데이터 확보
 - (물성 데이터) 물리화학적 성질(수소 용해도, 확산계수, 투과도, 결정구조, 반응엔탈피, 흡착에너지, 열전도도) 및 기계적 성질(경도, 항복강도, 인장강도, 연신율, 파괴인성)
 - (수소 흡탈장 데이터) PCT(압력-조성-온도), 상온흡탈장, 열분석 데이터
 - (미세조직 데이터) 미세조직 이미지(광학·전자현미경) 및 회절데이터(에너지산란분광, X-선회절, 전자후방산란회절 등)
 - (계산 데이터) 제일원리계산 및 분자동역학 기반 물리화학적 성질 데이터, 계산열역학 데이터(평형분압, 엔탈피, 엔트로피 및 상태도)
- (연구데이터 활용)
 - 수소인프라소재 데이터 플랫폼 구축과 데이터 기반 소재 설계 방법론 확보
 - 수집 데이터 및 기계학습, 전산모사 연계를 통한 수소인프라 신소재 도출

6.4.4. 사업운영계획

가. 사업추진 일정

세부과제명	사업기간											
	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30		
• 소재 연구데이터 수집·관리·공유 체계 구축												
• 에너지·환경소재 데이터 수집·활용 체계 구축												
– 다공성 소재 연구데이터 수집·활용 연구												
– 금속화합물 소재 연구데이터 수집·활용 연구												
– 이차원 소재 연구데이터 수집·활용 연구												
• 스마트/IT 소재 데이터 수집·활용 체계 구축												
– 차세대 반도체 소재 연구데이터 수집·활용 연구												
– 센서 소재 연구데이터 수집·활용 연구												
– 바이오/의료 소재 연구데이터 수집·활용 연구												
• 안전(구조)소재 데이터 수집·활용 체계 구축												
– 극한환경 소재 연구데이터 수집·활용 연구												
– 경량 소재 연구데이터 수집·활용 연구												
– 수소인프라 소재 연구데이터 수집·활용 연구												

나. 사업비 및 인력운영 계획

1) 투자계획 및 재원조달

(단위: 백만 원)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	2,850	4,200	5,400	5,150	4,700	4,400	4,000	4,000	3,100	2,200	40,000
소재 연구데이터 수집·관리·공유 체제구축	1,500	1,500	1,500	1,400	1,400	1,400	1,300	1,300	1,300	1,300	13,900
에너지·환경 소재 연구데이터 수집·활용 체제 구축	450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700
스마트/IT 소재 연구데이터 수집·활용 체제 구축	450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700
안전(구조)소재 연구데이터 수집·활용 체제 구축	450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700

2) 인력 운영계획

(단위: 명)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	39	55	73	73	71	71	71	69	51	35	608
소재 연구데이터 수집·관리·공유 체제구축	21	21	21	21	19	19	19	17	17	17	192
에너지·환경 소재 연구데이터 수집·활용 체제 구축	6	12	18	18	18	18	18	18	12	6	144
스마트/IT 소재 연구데이터 수집·활용 체제 구축	6	10	16	16	16	16	16	16	10	6	128
안전(구조)소재 연구데이터 수집·활용 체제 구축	6	12	18	18	18	18	18	18	12	6	144

3) 세부과제별 사업비내역

□ 세부과제 1: 소재 연구데이터 수집·관리·공유 체계 구축 전문센터

○ 상세 분야별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
연구데이터 수집·관리·공유 체계 구축	1,500	1,500	1,500	1,400	1,400	1,400	1,300	1,300	1,300	1,300	13,900
계	1,500	1,500	1,500	1,400	1,400	1,400	1,300	1,300	1,300	1,300	13,900

○ 비목별 예산

(단위: 백만 원)

구분 \ 연도		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인건비	소계	350	350	364	371	341	348	348	335	335	335	3,477
	인건비(책임)	22	22	22	23	23	24	24	25	25	25	
	투입인력(책임)	7	7	7	7	7	7	7	7	7	7	
	인건비(선임이하)	14	14	15	15	15	15	15	16	16	16	
	투입인력(선임이하)	14	14	14	14	12	12	12	10	10	10	
직접비	소계	995	995	983	890	916	910	823	835	835	835	9,017
	연구장비·재료비	650	650	638	545	506	500	413	395	395	395	5,087
	연구활동비	145	145	145	145	150	150	150	160	160	160	1,510
	연구과제추진비	100	100	100	100	120	120	120	140	140	140	1,180
	공동연구비	100	100	100	100	140	140	140	140	140	140	1,240
간접비		155	155	153	139	143	142	129	130	130	130	1,406
계		1,500	1,500	1,500	1,400	1,400	1,400	1,300	1,300	1,300	1,300	13,900

□ 세부과제 2: 환경·에너지 소재 연구데이터 수집·활용 체제 구축

○ 상세 분야별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
다공소재 소재 연구데이터 수집·활용 연구	450	450	400	400	300	300	300	300			2,900
금속화합물 소재 연구데이터 수집·활용 연구		450	450	400	400	300	300	300	300		2,900
이차원 소재 연구데이터 수집·활용 연구			450	450	400	400	300	300	300	300	2,900
계	450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700

○ 비목별 예산

(단위: 백만 원)

구분 \ 연도		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인건비	소계	100	200	304	310	316	320	326	334	226	114	2550
	인건비(책임)	22	44	66	67	68	70	71	73	49	25	555
	투입인력(책임)	2	4	6	6	6	6	6	6	4	2	48
	인건비(선임이하)	14	28	43	44	45	45	46	47	32	16	360
	투입인력(선임이하)	4	8	12	12	12	12	12	12	8	4	96
직접비	소계	309	618	878	827	685	590	493	485	320	159	5364
	연구장비·재료비	169	338	473	439	340	273	203	198	130	65	2628
	연구활동비	35	70	103	101	91	83	75	74	49	24	705
	연구과제추진비	30	60	92	92	84	74	65	63	41	20	621
	공동연구비	75	150	210	195	170	160	150	150	100	50	1410
간접비		41	82	118	113	99	90	81	81	54	27	786
계		450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700

□ 세부과제 3: 스마트/IT 소재 연구데이터 수집·활용 체계 구축

○ 상세 분야별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
차세대 반도체 소재 연구데이터 수집·활용 연구	450	450	400	400	300	300	300	300			2,900
센서 소재 연구데이터수집·활용 연구		450	450	400	400	300	300	300	300		2,900
바이오/의료 소재 연구데이터 수집·활용 연구			450	450	400	400	300	300	300	300	2,900
계	450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700

○ 비목별 예산

(단위: 백만 원)

연도		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
구분												
인건비	소계	100	200	304	310	316	320	326	334	226	114	2550
	인건비(책임)	22	44	66	67	68	70	71	73	49	25	555
	투입인력(책임)	2	3	5	5	5	5	5	5	3	2	44
	인건비(선임이하)	14	28	43	44	45	45	46	47	32	16	360
	투입인력(선임이하)	4	7	11	11	11	11	11	11	7	4	88
직접비	소계	309	618	878	827	685	590	493	485	320	159	5364
	연구장비·재료비	169	338	473	439	340	273	203	198	130	65	2628
	연구활동비	35	70	103	101	91	83	75	74	49	24	705
	연구과제추진비	30	60	92	92	84	74	65	63	41	20	621
	공동연구비	75	150	210	195	170	160	150	150	100	50	1410
간접비		41	82	118	113	99	90	81	81	54	27	786
계		450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700

□ 세부과제 4: 안전(구조)소재 연구데이터 수집·활용 체제 구축

○ 상세 분야별 예산

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
극한환경 소재 연구데이터 수집·활용 연구	450	450	400	400	300	300	300	300			2,900
경량 소재 연구데이터 수집·활용 연구		450	450	400	400	300	300	300	300		2,900
수소인프라 소재 연구데이터 수집·활용 연구			450	450	400	400	300	300	300	300	2,900
계	450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700

○ 비목별 예산

(단위: 백만 원)

연도		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
구분												
인건비	소계	100	200	304	310	316	320	326	334	226	114	2550
	인건비(책임)	22	44	66	67	68	70	71	73	49	25	555
	투입인력(책임)	2	4	6	6	6	6	6	6	4	2	48
	인건비(선임이하)	14	28	43	44	45	45	46	47	32	16	360
	투입인력(선임이하)	4	8	12	12	12	12	12	12	8	4	96
직접비	소계	309	618	878	827	685	590	493	485	320	159	5364
	연구장비·재료비	169	338	473	439	340	273	203	198	130	65	2628
	연구활동비	35	70	103	101	91	83	75	74	49	24	705
	연구과제추진비	30	60	92	92	84	74	65	63	41	20	621
	공동연구비	75	150	210	195	170	160	150	150	100	50	1410
간접비		41	82	118	113	99	90	81	81	54	27	786
계		450	900	1,300	1,250	1,100	1,000	900	900	600	300	8,700

6.4.5. 기대효과

가. 경제적 기대효과

□ 직접효과(정량적 측면)

- 연구데이터 유실방지, 공유 및 지식자산화에 의한 연구개발 효율화
 - － 연구데이터 재활용을 30% 이상 확보
- 데이터 구동형 소재 R&D에 의한 신소재 개발 기간단축 및 비용절감
 - － 소재 전분야(금속소재, 열전소재, 촉매재료, 에너지재료 등)에서 연구데이터 기반 신규 소재군 탐색, 물성예측 등을 통해 고비용 실험연구 최소화 등, 소재 R&D에 소요되는 시간/비용/위험요인 등을 50% 이상 획기적으로 절감
 - － 연구자/기관에 다양한 연구 데이터를 공유, 중복 연구 방지 및 R&D 비용 절감
 - － 현재 세계 메모리 반도체 시장은 약 110조 이상으로 인공지능(AI), 사물인터넷(IoT), 가상현실(VR) 등 4차 산업의 수요 증가와 함께 매년 10% 이상 성장 전망. 유전체 이용 차세대 메모리 기술 확보로 이익 창출 가능
 - － BCC Research에서 발간한 보고서에 의하면 IoT 및 wearable 기술수요 증대와 맞물려 전체 센서시장은 매년 급격한 성장세를 보이고 있으며 2023년에는 2,834억 달러 규모를 나타낼 것이라고 예상. 데이터 기반 센서소재 설계기술 확보로 신규시장 선점
 - － 전세계적으로 50조원에 이르는 극한소재 시장의 10 %를 국내 기업 점유 가능
 - － 2030년 세계적으로 400조원에 이를 것으로 예상되는 수소 및 관련 인프라 시장의 20 %를 국내 기업 점유 가능

□ 직접효과(정성적 측면)

- 데이터 기반 R&D 생태계 구축/확산을 통한 소재분야 산업 경쟁력 향상
 - － 연구활동에 의한 모든 데이터 및 연구 성과물(논문, 특허, 기술보고서) 수집·관리·활용 플랫폼 서비스를 통해 데이터 기반 연구개발 환경을 공유하고 확산시킴으로써 소재분야 산업의 경쟁력 향상
- 4차 산업혁명 대응 원천소재 관련 경제적 경쟁력 확보
 - － 에너지·환경, 스마트/IT, 안전 등 4차 산업혁명 환경 및 각종 사회적 현안의 변화에 대응가능한 원천소재 개발 역량 향상으로 관련 산업의 수출, 수입대체 경쟁력 확보 및 사회·경제적 기여

□ 간접효과(정량적 측면)

- 다양한 분야로의 신소재 개발 수요 및 투자 확대
 - －소재 개발주기의 획기적 단축, 다양한 분야의 신소재 개발수요 창출로 정부 및 민간의 투자 확대 예상, 관련 장비 산업 창출에 기여
- 환경·에너지소재의 타 분야로의 확대 적용
 - －DB의 예측물성 확장을 통한 타 연구분야 소재개발 지원: 고체화합물과 다공성소재는 에너지·환경 분야 이외에 다양한 연구분야에 적용됨 (예: 고체화합물은 반도체, 다강체, 강유전체 등에 적용가능). 소재물성의 확장을 통해 국가 R&D 효율증대 및 비용절감에 기여
- 스마트/IT 소재의 타 분야로의 확대 적용, 일자리 창출
 - －소재, 공정개발 및 빅데이터 관련 반도체 소재 전문인력 양성으로 일자리 창출
 - －소재 물성데이터 대상 Digital Twin 및 Digital Thread 구축 경험을 타 산업에 적용하여 국가적 스마트 제조 경쟁력 강화
 - －현재 메모리 반도체 시장은 국내 기업과 미국 기업의 과점 체제로 두 국가의 시장 점유율이 90% 이상이나 중국의 추격이 거센 상황에서 메모리 산업부문 경쟁력 격차 유지 및 확대
 - －센서기술은 의료, 국방, 우주, 에너지, 환경 등의 다른 산업과 관련이 크기 때문에 간접적 센서소재 설계 기술이 지니는 간접적인 경제적 효과가 큼

□ 간접효과(정성적 측면)

- 소재 지식 서비스 시장 선점
 - －신뢰성 있는 플랫폼 제공을 통해 다양한 소재의 성능 측정, 평가, 인증의 지식 서비스 시장 선점 가능
- 응용 산업기술 분야 경쟁력 강화 및 신규 시장 창출
 - －다양한 산업에 적용되는 응용기술 발전의 기반이 되기 때문에 산업기술 분야에서 국가 경쟁력 강화와 위상 제고
 - －수소 및 연료전지 분야 신규 시장 창출 및 관련 산업 육성과 산업화 기반 마련

나. 과학기술적 기대효과

□ 직접효과(정량적 측면)

- 소재 연구데이터를 이용한 소재개발의 구심점
 - － 소재연구데이터의 국가 자산화, 체계적인 관리와 확산
 - － 소재 연구데이터와 연구성과물 데이터를 체계적으로 수집·관리하는 연구데이터센터 구축함으로써 분산적인 정보유통을 중앙집중화하여, 연구센터를 기반으로 한 양질의 자료제공과 신속한 정보교환 가능, 소재개발의 구심점 역할
 - － 의미를 기반으로 하는 지식구조를 정리한(소재 온톨로지 기반) 검색 플랫폼을 제공해 실제 연구자들이 연구진행과정에서 필요한 영역들을 시각화하여 이용자들의 정보욕구 충족시킴으로써 지적산물들을 공유하고 확산
- 빅데이터 기반 기계학습을 통한 신소재 탐색 능력 향상
 - － 빅데이터 분석을 활용한 효율적 소재 설계 방법론 확보 및 공정-구조-특성-성능 상호 예측모델 도출을 통한 신소재 탐색 능력 향상
 - － IT, ITO 소자의 중요 특성은 낮은 전력 소비량, 높은 동작성능임. 차세대 메모리의 경우 기존 메모리 반도체보다 100배 낮은 전력 소비량, 780배 빠른 동작성능을 가질 것으로 예측
 - － 소재 종류 및 데이터의 다양성/복잡성으로 인해 연구개발 및 실용화 기간이 매우 긴 바이오의료용 소재의 개발기간 획기적 단축 가능

□ 직접효과(정성적 측면)

- 차세대 신소재 원천기술 선점
 - － 소재연구데이터 센터의 소재 빅데이터를 활용한 효율적인 소재 설계, 머신러닝 기반 데이터 추출 알고리즘 개발 등 데이터 생산의 새로운 방법론 개발을 통한 차세대 소재 개발 가속화, 소재 개발에서 사업화까지 소재 연구 개발을 효율적으로 수행하여 차세대 소재에 대한 원천기술 선점
- 국제 사회에서의 소재 연구 빅데이터 기술 선도
 - － 국외 연구 데이터 기관/서비스 등과 연계하여 범국가적인 소재 연구 연맹 체제 구축, 국제 사회에서의 빅데이터 프로토콜/표준화 선도
- 융복합 R&D 및 열악한 연구 분야 혁신적 신소재 개발 도출
 - － 국내 연구과제의 계산재료공학 연구데이터를 수집하고, 재활용하게 함으로서 중복 계산

- \ 으로 생기는 계산 하드웨어 자원의 낭비 없이 계산연구의 효율성을 증대 시킬 수 있으며, 협동연구 촉진
 - 선진국 대비 기술수준이 매우 열위한 국내의 우주, 항공, 해양 및 국방 분야의 과학기술 연구 기반 구축에 기여하고 수소인프라 보급에 핵심 역할을 하는 혁신적 신소재 도출에 기여

□ 간접효과(정량적 측면)

○ 빅데이터 기반 인력 양성

- 빅데이터기반의 새로운 소재연구개발 패러다임 변화를 수행할 수 있는 국내 연구인력 양성 기회 제공

□ 간접효과(정성적 측면)

○ 데이터 기반 R&D 인프라 구축 및 다양한 소재산업 경쟁력 제고

- 기술 인프라 발전: 대용량 데이터에 기반을 둔 의사결정이나 전략수립에 있어 지식정보 자원을 구축하는 기관이 정보자원의 상호운영성과 공유, 일관된 검색 등을 체계적으로 구축할 수 있도록 지침 제공
- 데이터 기반 연구가 활성화 되면서 데이터를 기반으로 소재를 개발하는 소재정보학의 새로운 연구분야를 발전시켜, 소재 개발의 시간을 단축시킬 수 있음.
- 공개 소재 물성 DB에만 의존하지 않고, 실질적으로 소재개발에 필요한 계산 데이터를 국내 연구자들이 공유하고 활용함으로써, 국내 소재산업의 경쟁력 제고
- 사회적 이슈인 대기오염(미세먼지) 및 지구온난화와 같은 환경문제 개선 및 지구온난화 문제의 주범인 온실가스의 획기적 저감과 신재생에너지 보급에 기여
- 바이오의료용 소재의 부작용(체내 이식 후 파손, 단기독성/장기독성 등)에 대한 사전 예측을 통해 의료 사고 예방

6.5. 인공지능 전문센터 구축

6.5.1. 사업개요

가. 추진배경 및 필요성

□ 4차 산업혁명시대의 도래에 따라 인공지능의 중요도가 더욱 높아지고 있음

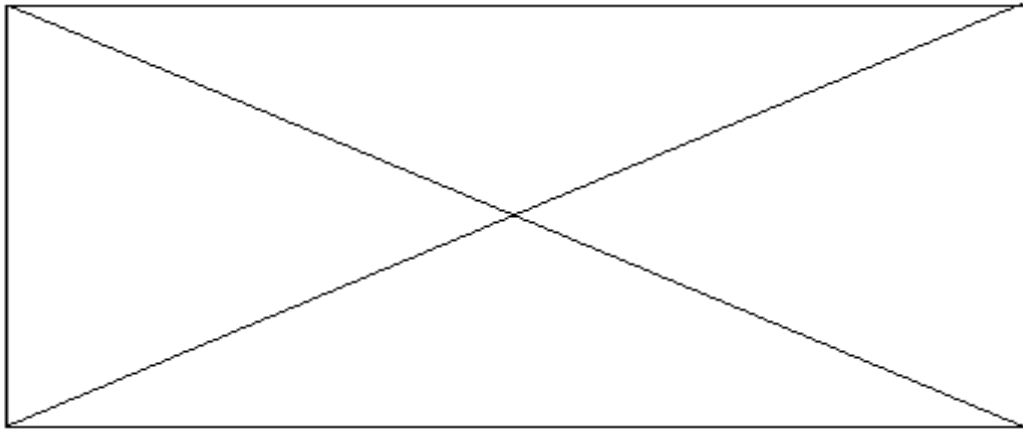
- AI(인공지능)는 인지, 학습 등 인간의 지적능력(지능)의 일부 또는 전체를 ‘컴퓨터를 이용해 구현하는 지능’을 의미
- AI는 단순 신기술이 아닌 4차 산업혁명을 촉발하는 핵심 동력
 - － 증기기관, 전기 등과 유사한 범용기술로서의 특성(①다른 분야로 급속히 확산, ②지속적 개선 가능, ③혁신을 유발하여 경제사회에 큰 파급효과를 미침)을 보유
- AI는 파괴적 기술혁신을 통해 산업구조의 변화를 야기하고, 사회·제도의 변화까지 유발할 것으로 전망
- AI는 가까운 미래의 경제·사회 대변혁의 핵심 동력으로 작용하여 각국 경제 성장에 비약적인 파급효과를 실현할 전망
 - － 세계 AI 시장 규모는 2015년 약 1,270억 달러에서 2017년 약 1,650억 달러로 연평균 14.0%의 고성장을 지속할 것으로 전망

□ 학계·산업계는 연구개발을 위한 AI 학습용 데이터가 부족함을 지속적으로 주장

- AI는 알고리즘만으로는 작동하지 않으며 스스로 인식, 이해할 수 있으려면 기계가 이해할 수 있도록 가공된 대량의 훈련용 데이터 구축 및 학습이 반드시 필요
- AI를 학습시키기 위해서는 일정량 이상의 데이터가 필요하고 데이터 품질에 따라 성능이 결정되며, 분야별 데이터의 확보도 필요
 - － 중국은 영상인식 등 대량의 데이터를 생성·활용하는 분야를 중심으로 미국을 빠르게 추격 중
 - － 미국인 일상 데이터로 학습한 AI 제품·서비스를 한국인에게 적용시 오류발생 가능성 증가
- AI 기술연구, 응용서비스 개발의 활성화를 위해 무엇보다도 AI를 학습시킬 데이터, AI 알고리즘 및 모델, 고성능 컴퓨팅 파워 등 핵심 인프라의 제공 및 활성화가 시급

□ AI 학습을 위한 기본 데이터의 공개와 공유 부족

- 우리나라는 공공데이터를 중심으로 개방 건수는 증가하고 있으나 AI R&D에 필요한 AI 학습용 데이터 수요를 반영한 활용도 높은 데이터 개방은 여전히 부족
- 공공데이터 개방 1위('15 OECD), 데이터(디지털정보) 교역은 44위('16 맥킨지)



<그림 6-39> ODB 기계학습용 대용량 데이터셋 제공 분야수('16)

※ ODB(Open Data Barometer)는 15개 분야별(지도, 국가 선거결과 등)로 기계학습용 대용량 데이터 셋의 다운로드 가능 여부를 평가, 우리나라는 국가 선거결과만 제공(1개)

- 실제 활용 가능한 데이터와 컴퓨팅 인프라가 부족한 가운데, 공유 활용률도 저조하고 산·학·연의 개방형 협력 역시 미약
- 공개 공공데이터(24,588개) 중 0.3%(68개)가 기계학습 가능 데이터이며 기계학습용 대용량 데이터셋 제공평가에서는 56위 기록
- 학습용 데이터, 알고리즘, 컴퓨팅 파워는 AI 혁신의 핵심원천이나, 공공·민간의 자발적 축적 개방을 통한 활성화 노력이 미진한 가운데, 정부의 'AI 허브' 구축('18.1.)에도 불구하고, 활용 가능한 데이터의 양·질이 아직 부족하며 민간 확산을 위한 제도적 장치 역시 미흡

□ 산업데이터에서 연구데이터로 자연스럽게 연결되는 생태계 시스템 필요

- 해외는 다양한 분야에서 AI 학습 데이터 및 학습 환경 등 기반 인프라가 활성화
 - DBpedia, WordNet, ImageNet, Microsoft COCO 등 학습용 데이터 축적에 많은 노력 투자
 - IBM 왓슨도 학습을 위해 위키피디아, DBpedia, WordNet, YAGO ontology 등 지식기

반 데이터 활용

- 반면, 국내의 경우 양질의 AI 기술개발 및 학습을 위한 데이터 개방 및 활용은 사실상 전무
 - － AI 연구개발에 필요한 학습용 데이터 제작 관련 노하우 부족과 시간·비용적 부담으로 중소·벤처 기업, 스타트업, 학교, 연구소가 자체 구축하기에는 한계
 - ※ 해외는 이미지넷(美, 100만건), DBpedia(獨, 458만건) 등 대규모 데이터셋을 공개한 반면 국내는 AI 학습에 필요한 데이터 공개에 부족한 상황이며 우리나라 컴퓨팅 파워는 전세계의 0.8% 수준에 미치지
- 지능정보산업 창출을 위해서는 양질의 데이터 확보가 핵심 경쟁 요소로 정부차원에서 다양한 연구용 데이터 등을 활용한 주요 분야별 데이터 인프라 구축·활용 지원 필요

6.5.2. 환경분석

가. 국내외 정책 및 시장 현황

□ 정책현황

- 현재 세계 각국은 미래 먹거리인 AI를 개발하기 위해 국가차원의 전략을 수립하고 각종 정책 사업을 추진하고 있음
- (한국) 최근 정부는 국가의 혁신성장 전략투자 분야 중의 하나로 AI를 선정하고('18.8월) 관련 투자의 확대를 추진 중
- (미국) AI 암흑기에도 지속적인 장기 투자를 통해 뇌과학 등 기초·원천기술을 확보*하고, AI 핵심 기술개발과 인재양성**에 주력해왔음
 - * ('90, NIH) Decade of the Brain, ('96, NSF) Learning and Intelligence System, ('04, OSD) Joint Robotics Program: Master plan 2004, ('08, DARPA) SyNAPSE, ('13.4, Whitehouse) Brain Initiative
 - ** 국가 AI R&D 전략계획('16.10, '15년 11억\$ 투자), AI, 자동화 그리고 경제('16.12) 등
- 연방정부는 AI 기술 혁신가로서 원천기술 개발 후 민간에 이양, 혁신적 제품·서비스* 상용화를 전인하며 시장주도의 원동력 역할을 수행
 - * 국방부 CALO 프로젝트('03~'08, 2억\$ 투자) 中 '음성개인비서 연구 부문'을 독립시켜 벤처기업 Siri 설립('07.12) → 애플이 2억 달러 인수 후 아이폰4S에 탑재('11.10)
- 개방·경쟁형 기술개발 방식(Challenge)을 선제적으로 도입, 선도적 AI 기술 경쟁력 확보에 정책적 역량을 집중
 - * (challenge.gov) 정부가 도전형 과제를 제시하고 민간은 챌린지 방식으로 경쟁, 자율주행차·이미지 인식 알고리즘 개발 등 총 819개 해결('12~'18.2)
 - * (DARPA Challenge) 자율주행차('04~'05), 로보틱스('07), 사이버해킹('16) 등
- (중국) AI를 국가 전략산업으로 인식, 정부 주도의 대규모 투자*·인력양성**을 추진하는 한편, 선도기업을 지정, 특화플랫폼을 육성***
 - * 민관협력 '차세대 AI발전계획위원회'를 설립('17.11), 3년간 1,000억위안(약 18조원) 투입
 - ** 중·미 대학 공동 AI인력양성을 포함한 '중국대학 AI 인재 국제육성계획'

발표('18.4)

*** (바이두) 자율주행차, (텐센트) 의료·헬스, (알리바바) 스마트시티, (아이플라이텍) 음성인식

– '30년까지 AI 핵심산업 1조 위안(약 180조원), 연관산업 10조 위안 규모의 시장육성을 목표로 국가 AI 기술혁신에 집중

※ '인터넷 플러스 AI 3년 협동 실시 방안('16)', '차세대 AI 발전계획('17.7)' 등

※ 중국은 AI 분야에서 세계 1위 선도국가로 도약할 수 있는 환경과 능력 보유(Forbes, '17)

– AI 융합을 강조, '20년까지 AI 적용 제품·응용범위*를 확대하고, AI 전반의 핵심기술**향상을 도모, 글로벌 시장 선점에 주력

* 커넥티드카, 로봇, 드론, 의료영상 진단시스템, 영상식별, 음성 인식·번역, 가전

** 스마트 센서 제품 개발, 신경망 칩 양산, 개방형 플랫폼 구축

○ (일본) 저성장·고령화 극복을 위한 국가 경제·사회 혁신의 수단으로 AI 기술경쟁력 확보의 중요성을 강조하며 AI 기술혁신을 가속화

• 일본재흥전략('16.4), AI 산업화 로드맵('17.4), 신산업 구조비전('17.5)

– 특히, 국내외 최고 AI 연구자 결집을 위해 연구거점을 마련(혁신지능통합연구센터, '16.4)*하고, 개방형 AI R&D 플랫폼**을 구축

* 이화학연구소를 중심으로 산학연 파트너십 기반의 기술개발 프로젝트인 AIP(Advanced Integrated Intelligence Platform)를 추진, 3개 연구그룹(총 52개 팀) 및 기업협업센터로 구성

** 195억엔(약 2천억원)을 투자하여 데이터 기반의 AI 클라우드 인프라(ABCI) 구축('18.완료목표)

○ (영국) AI의 진보에 따른 경제적, 윤리적, 사회적 영향에 대한 조사

– 하원 선택위원회(The Lord 's Select Committee of AI)는 '영국의 AI : ready, willing and able'라는 장문의 보고서를 발표했음

– 10개월에 걸친 조사를 통해 기업이 데이터를 독점할 가능성을 검토하고, 데이터 셋 감사에 대한 새로운 접근 방식을 장려

○ (인도) 시장 진출을 높이고 데이터 수집에 소요되는 시간과 비용을 줄이기 위한 국가 AI 마켓 플레이스를 만들고, 전반적인 인력이 기술을 습득 할 수 있도록 지원 추진

– COREs(Centres of Research Excellence in AI) 및 ICTAIs(International Centers for Transformational AI)에서 윤리위원회 컨소시엄을 구성

– 개인 정보 보호, 보안 및 윤리에 대한 부문 별 가이드라인을 개발

□ 법제도 현황

- (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 연구 수행을 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)

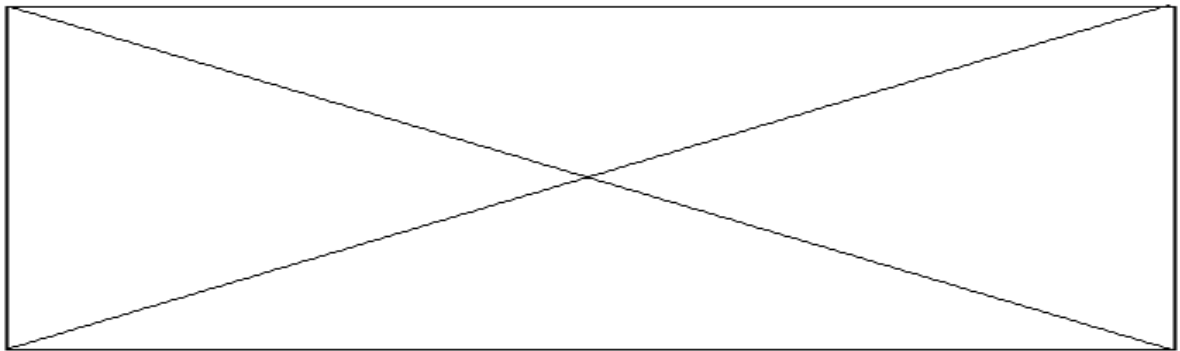
□ 시장현황

- Govini의 보고서에 따르면, 2017년 비공개 예산에서 미국의 펜타곤은 AI와 이를 지원하기 위한 빅데이터·클라우드 컴퓨팅과 같은 분야에 약 74억 달러 지출
- 기업의 경우 공동활용(범용) AI플랫폼의 개방/제휴를 통해 사업영역을 확대하거나 사업자간 연합생태계를 조성, 시장 지배력 강화 가속화
- 영국, 미국 등 주요국은 공공데이터를 중심으로 산업적 활용을 촉진하고 민간 데이터 거래중개 시장도 활발
 - － AI 학습 및 산업적 활용을 위해 LoD 등 고품질 데이터 개방 및 공유를 위한 데이터 맵/인프라(lod-cloud.net) 운영 중
 - ※ (IBM) 왓슨이 금융 의료 등에 적용되면서 해당 전문지식을 지속 축적, 왓슨생태계 구축
 - ※ (아마존-MS 연합) 알렉사 검색에 MS Bing 검색엔진 활용, 알렉사를 통한 MS 오피스에 접근과 동시에 MS 코타나를 통한 아마존 전자상거래 플랫폼으로 활용 기대

나. 관련 기술동향 분석

- 그간 두 번의 AI암흑기*에도 불구하고, 데이터 축적·컴퓨팅 파워 진전·알고리즘(딥러닝) 진화 등으로 최근 AI 부흥기에 진입

- * 1차('70년대): 메모리·처리속도 등이 미구축, 실패로 AI에 대한 지원 중단
- 2차('00년대): 전문가시스템의 高유지비용·업데이트 한계·오류 등으로 AI회의론 확산



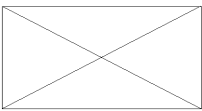
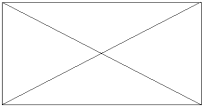
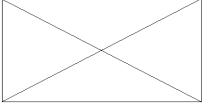
<그림 6-40> AI 기술 발전 전망(출처: AI R&D 전략(과기정통부, '18)

- 엣지 컴퓨팅 등 AI chip이 가까운 미래에 상용화될 것으로 예상되어, 웨어러블 등 초소형 디바이스 자체에서도 AI이 구현될 것으로 전망
 - 향후에는 고도화된 뇌연구를 바탕으로 현재의 딥러닝 기반 AI의 한계*를 돌파하는 차세대 AI 알고리즘으로 발전
 - * 딥러닝 개발 이후 AI 부흥기에 진입하는 중이나, 딥러닝 자체의 한계(①데이터 가공(레이블)에 단순노동이 상당히 소요되고, ②분야별로 데이터가 축적되어 있어야 하며, ③정보처리과정의 Black Box로 인해 투명성·신뢰도가 낮음)가 존재하는 상황
 - 특히, 원시데이터 자체로 학습이 가능한 비지도 학습 AI, 분야별 상호 학습이 가능한 전이학습 AI 등으로 제2의 AI 부흥기 도래가 예상
 - AI 발전 속도는 점차 가속화되어, 향후 10여년 간의 변화는 AI 개념이 등장한 1950년대 중반 이후 현재까지 약 60년 간의 변화를 압도할 것으로 전망
- 도입 단계의 AI기술이 실험실 연구수준을 넘어 상용화 단계로 발전하면서, 폭발적 시장 성장*과 他산업의 지능형 융합을 견인
- * AI 세계 시장 규모(IDC, '17.11) : ('16) 78.1억\$ → ('21) 522억\$, CAGR 34%
 - 글로벌 ICT기업은 AI 플랫폼을 기반으로 타산업과의 협력을 통해 개방형 생태계 구축에

Ⅰ 주력*, 핵심시장 선점을 가속화

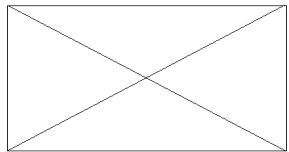
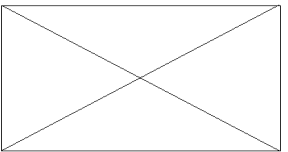
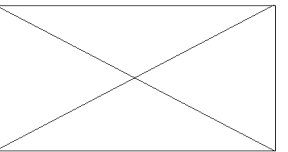
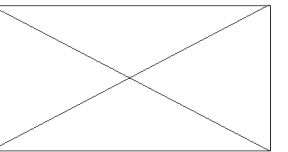
* (아마존 알렉사) 가전제품, 자동차 등의 지능형융합을 통해 전방위적으로 시장 확대

- 제조, 자동차 등 他 산업과의 지능형 융합으로 차별적 부가가치 창출은 물론, 기존 산업분야의 획기적 변화를 촉발

산업분야	발전 전망	적용 사례	
자동차 	부분 자율주행 → 완전 자율주행	(테슬라)오토파일럿 (자율주행 3단계) 구현('16)	(벤츠) 자율주행 5단계 AV 상용화('21)
제조 	공장 내 자동화 → 공장 간 연계 자동화	(아디다스) 신발 주문-제작까지 1일 소요('16)	(보쉬) 전세계 250개 공장 연결('20)
에너지 	건물 에너지 효율화 → 국가 전력망 효율화	(딥마인드) 구글 데이터센터 냉각효율 40% 개선('16)	(GE) 전력망의 전력배분 효율성 제고('17~)

□ AI 기술은 자체의 기술혁신을 넘어 과학기술과 산업기술 영역의 성장을 가속하는 혁신의 조력자(Enabler) 역할을 수행

- 장기간 연구가 필요분야(물리실험, 신약개발 등)에 AI를 활용, 시간비용을 획기적으로 단축, 新제품서비스의 연구개발 촉매제 역할을 수행

(물리) 美 페르미연구소	(천문학) NASA-구글	(신소재) 도요타	(신약) 아툼와이즈
중성자 물리연구 분석('15) (분석시간: 수개월→하루)	외계행성 탐색 적용('17) (제2의 태양계 발견)	연료전지 촉매제 개발('17~) (신형차 CO2배출량 감축)	에볼라 치료약 후보 발굴('17) (소요기간: 수년→하루)
			

□ 기술 트렌드가 급속도로 발전하고 있음에도 불구하고 국내 연구 인프라에 대한 연구 및 기술 추세는 선진국에 비해 미흡

- AI 기술개발 및 산업응용을 위해 필수적인 학습데이터의 부족

- 구축을 위해 큰 비용이 소모되는 연구과제 진행을 위해 생성되고 축적된 연구데이터가 과제 종료 이후 효용성을 잃는 경우가 대부분 → 연구데이터 공유 플랫폼 구축 필요
- 특히, AI 분야의 연구데이터는 세 가지로 정의/분류 가능
 - 특수 분야에 국한되지 않고 전 분야에서 공통으로 요구되는 AI 기술 존재(범용 AI, e.g. 인지/행동 등) → 공통으로 활용되는 AI 기술 개발에 사용될 수 있는 연구데이터 (문서, 이미지, 음성 등)
 - AI 기술 개발의 궁극적 목표는 실제 인간과 같이 행동/사고하는 강인공지능(Strong AI, AGI : artificial general intelligence) → 강인공지능 개발을 위한 기초연구(뇌과학 등)에서 구축/활용된 데이터
 - AI 기술은 다양한 산업분야에 적용되고 있으며(통테일 성격이 강함), 타 전문센터에 소속되지 않는 연구 존재 → 기타 AI 분야 연구데이터
- 각기 다른 성격의 AI 분야 연구데이터를 관리할 수 있는 플랫폼 구축을 위한 과제 필요
- 인터넷에 연결된 IoT(Internet of Things)들의 내장된 센서를 통해 방대한 새로운 데이터를 수집하고 이를 통해 더 새로운 서비스와 효율성을 개선, 경쟁력 있는 비즈니스 모델의 가능성을 열어두고 있음
 - 초기 IoT가 센서·단말, 무선통신 및 원격제어 등을 중심으로 연결을 통한 서비스에 초점이 있었다면, 최근 기술 영역이 빅데이터, AI, 가상물리시스템 등의 기술과 융합되면서 이제는 데이터 분석, 예측, 자율제어를 통한 새로운 가치창출을 위한 목적이 되고 있음

< 참고 : 해외 공개 AI 학습추론용 데이터셋 현황 >

- (이미지) 구글 Open image*, MNIST(필기체 숫자), Caltech256, STL-10(스탠포드대), Street-View-House-Number(구글), MS-COCO, NORB, PASCAL VOC, COIL-100(360도 사진), CIFAR-100, LSUN, ImageNet 등 15종 이상 수천만장
- * 구글 오픈 이미지 : 6,000개 카테고리, 900백만개 이미지 및 객체 어노테이션 제공
- (비디오) 구글 YouTube-8M(800만개 영상, 5십만시간, 19억 프레임의 영상 데이터 공개, 2016)
- (텍스트) 20 newsgroup, Reuters News dataset, Penn Treebank, UCI's Spambase, Broadcast News, Text Classification Datasets, SQuAD, Billion Words dataset, Common Crawl, Google Books Ngrams, Yelp Open Dataset 등 자연어 이해 등을 위한 12종 데이터셋 수억 어절
- (질의응답) Maluuba News QA Dataset, Quora Question Pairs, CMU Q/A Dataset, Maluuba goal-oriented dialogue, bAbi(페이스북) 등 질문과 답변의 병렬 데이터 5종 수십만 개
- (음성) 2000 HUB5 English(바이두), LibriSpeech: Audio books data, VoxForge, TIMIT, CHIME(소음환경하 음성), TED-LIUM 등 6종 공개
- (페이스 이미지) Labelled Faces, UMD Faces, CASIA, MS-Celeb, Olivetti, Multi-Pie(CMU), JACFEE, FERET, The Yale Face Database 등 10여종 수백만장 얼굴 이미지
- (감성) Multidomain sentiment analysis dataset, IMDB, Stanford Sentiment Treebank 등 수종의 감성분석용 텍스트 데이터셋
- (기타) 구글 오디오셋(200만개 소리 데이터), 우버Trip Data, Core50(연속 이미지), 케글 데이터셋, HeadCT scan 등 다양한 분야 및 형태의 데이터셋이 계속 공개 중

6.5.3. 추진전략

가. 사업추진 방향 및 목표

1) 운영현황 및 개선방향

□ 운영현황(NIA)

○ 데이터관리 현황

- 인공지능 관련 국가R&D사업 등을 통해 구축된 기계학습용 데이터셋을 “AI 오픈 이노베이션 허브”포털을 통해 개방·공유 중

<표 6-23> AI 학습용 데이터 구축·제공 현황

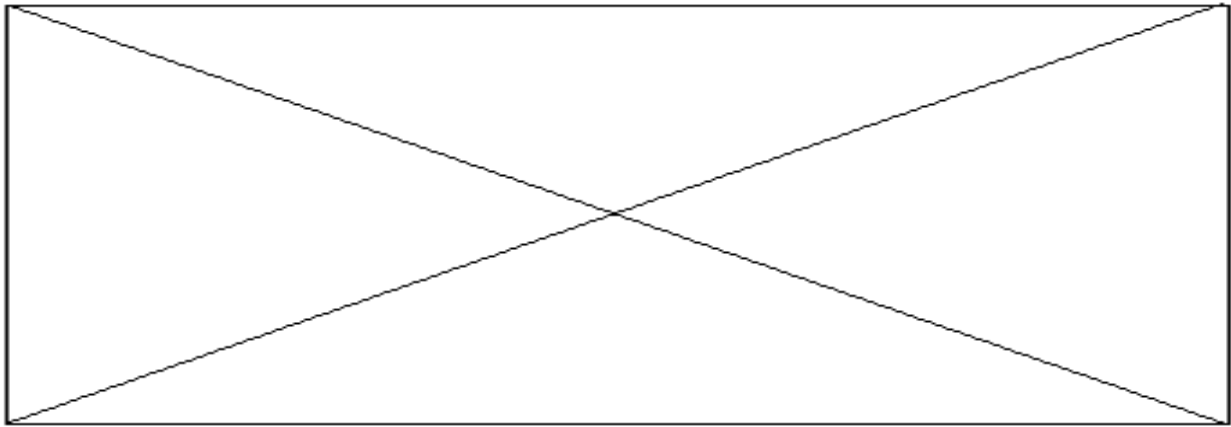
분야	데이터 종류	파일 형식	수량
법률	법령용어	json, N-triple	11,000건
	생활용어	json, N-triple	106,252건
	법률 지식베이스	json, N-triple	3,000건(표제어) 330,000건(트리플)
특허	특허 지식베이스	RDF	1,013,075건(표제어) 6천만건 이상(트리플)
	기술용어	RDF	120,687건
	도면태그정보	RDF	2,338만건
일반상식	일반상식 지식베이스	json	161,651건(표제어) 750,000건(트리플)
	일반상식 질문-답변-제시문	json	100,000건(MRC 데이터)
한국형 이미지	한국인 얼굴 이미지	jpg	6,480,000장
	한식 이미지	jpg	150,000장

○ 추진방향

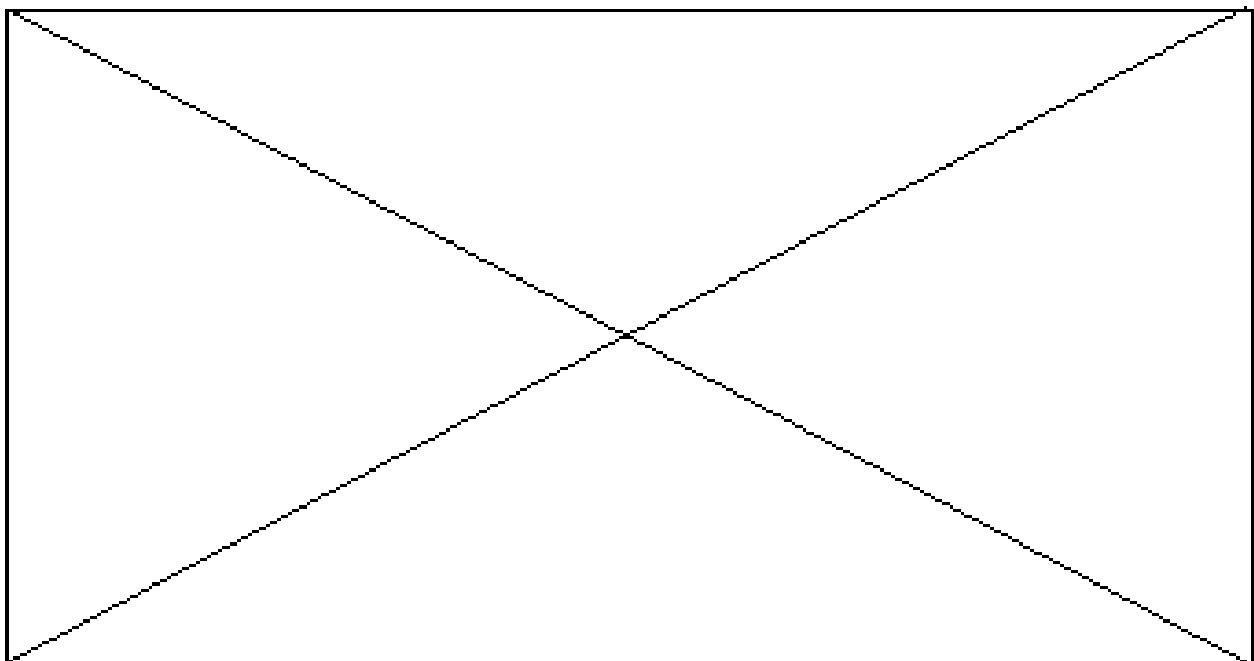
- 인공지능 국가 R&D 과정에서 연구수행기관(출연연·대학·기업 등)이 축적한 기계학습 데이터를 취합·공개·활용

※ 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



<그림 6-41> AI 오픈이노베이션 허브 개념도



<그림 6-42> AI 오픈 이노베이션 허브 시범구축·운영안('18년)

□ 개선방향

- 현행 AI 오픈 이노베이션 허브의 기능을 확대하되, 그 기능과 역할을 조정하여 AI 연구 개발 데이터의 구축·연계·융합을 지원하는 인프라로 발전
- 컴퓨팅 파워의 KISTI 연구용 슈퍼컴퓨팅센터 연계
 - － 현행 GPU기반 컴퓨팅 체계에서 높은 연산량이 요구되는 데이터와 즉시적 분석이 필요한 데이터에 대하여 슈퍼컴퓨팅센터와 연계하여 정보 분석 제공
 - － 대량의 AI데이터 처리를 위해 슈퍼컴퓨터 5호기(세계 10위권 성능) 서비스 개시('18.3 분기)에 발맞추어, 컴퓨팅 자원을 AI분야에 할당하도록 추진
- 국가연구데이터플랫폼에 표준화된 어휘로 제공되는 연구용 데이터셋 인프라를 제공하여 AI 연구데이터 허브화 추진
 - － 국내에서 구축된 온톨로지 등 관련 정보를 외부 정보와 연계하여 국내에서 구축된 AI 데이터로 활용토록 제도 마련
 - － 공공데이터 포털에서 만들어진 데이터에서 AI 학습에 적합한 데이터를 추출해야 하며, 이를 문서 등에서 표시하는 표준화 체계가 마련되어야 함
- AI분야 공개SW 인프라 및 커뮤니티 제공
 - － 구현되는 소스나 데이터(전략 데이터를 제외)에 대하여 배포 체계를 오픈소스화 하되 커뮤니티를 활성화 하고 데이터+소스코드를 이용한 머신러닝 해커톤 등 다양한 대회를 개최하여 민간/공공의 AI 데이터 활용도를 고취
 - － 현행 공공, 민간, 연구 데이터 수집 체계에 있어 각 정보 체계간의 호환성을 확보하고 데이터의 품질을 보장하기 위한 서비스 및 프로세스 체계 마련
 - － 마스터 데이터의 저장 및 배포, 시험적 분석후 고성능 컴퓨팅이 필요한 정보에 대한 국가 연구데이터 플랫폼 자원 활용 지원 체계 및 운영 커뮤니티
 - － 타 정책기관의 공개SW 개발지원센터 및 과제와 협업하여 AI 알고리즘 개발에 대한 기술 개발 지원 필요
- 현행 알고리즘 공개 범위에 대하여 공공 AI R&D 결과물을 민간이 시제품 개발에 자유롭게 활용하도록 Open API 형태로 제공 개수 확대
 - Open API 기술 활용 지원을 위한 전담 조직(과제 수행기관 內)을 운영하여 공공 AI SW기술의 지속적인 활용·확산을 지원
- AI 데이터 표준화 및 품질 보증 및 체계에 대한 지속적인 연구 생태계 마련
 - － AI학습에 적합한 이미지의 크기와 정밀도, 학습에 요구되는 메타데이터의 정의에 대한 표

\ 준 가이드에 대한 연구

- AI학습에 적합하게 구성되기 위한 메타데이터의 표준과 관련 정보 연구
- 데이터의 라이선스에 대한 사항과 신뢰성 있는 AI 학습에 필요한 데이터를 제공하는 인증 기관으로서 그 인증에 대한 기준을 제공하는 방안 마련
- 개인정보 및 민감정보를 거르고 이를 제도화하며, 공개하는 가이드라인을 제공하는 방안 마련
- 글로벌 지향으로 국내 활용이 아닌 해외 확장을 감안한 연구데이터 연계체계 마련
 - 해외의 경우 오랜 시간 AI 기반 데이터를 구축해 왔으며, 해당 데이터들이 토대가 되어 현재 AI 학습 데이터의 토대가 되었음
 - 적극적 해외 데이터의 수용에 덧붙여 국내 데이터를 확장하는 전략으로 추진(해외에서 구축하고 이미 활발하게 활용되고 있는 데이터를 국내에 신속히 도입)
 - 해외의 공개되어 있는 양질의 지식 기반 데이터의 국내 내재화(한글화 및 한국에 적합한 데이터로 연계)
 - ※ 이미지넷의 경우 프린스턴 워드넷과 메타데이터에 대한 연계가 있어서 쉽게 구축이 가능했음
 - ※ 이미지넷 + 한국어 워드넷(KorLex 같은)과 연계점만 만들어줄 경우 이미지넷의 국내 AI 학습에 활용할 수 있음
- 한국 환경 및 한국어에 특화된 데이터를 구축하되 전략 데이터로 구성하여 국내 기업에만 제한적으로 제공하고 현행 지도 반출과 같이 엄격한 심사 기준 적용
 - ※ 누구에게나 제공될 경우 구글, 애플, 아마존 등 다국적 기업의 컴퓨팅 파워를 감안할 때 국내기업들에 비해 빠른 습득이 예상
 - 한국어 정보의 최적화 학습 결과데이터를 국내기업/연구자에만 제한적으로 제공하여 추가적인 컴퓨팅 파워가 필요하지 않도록 제공
 - 향후 AI Chip과 같이 AI 연산을 필요로 하는 핵심 칩에 기본 데이터로 제공되도록 함으로써 Edge AI와 같은 미래 유망 기술 분야에 활용되도록 구성
 - ※ AI Chip은 인공지능 기능을 구현한 주문형 반도체로서 차세대 기술로 예상됨
 - ※ Edge AI는 현재 클라우드 컴퓨터에서 이루어지는 AI 학습 및 분석 기능이 스마트폰과 같은 끝단 장비에서 이루어진다는 개념으로써 가트너 그룹은 향후 5~10년 사이 도래할 것으로 예측
- 국내 사물인터넷 인프라와 연계되는 지식 수집체계 마련
 - 리빙랩이나 스마트시티에서 연계되는 다양한 사물 데이터를 수집하여 연계하는 방안과 이의 지원 체계 마련
 - CCTV, 음성데이터, 위치정보, 소상공인의 매출 트렌드 정보 등 다양한 형태의 정보와 함께 개인정보 비식별화 조치가 반영된 정보의 제공

- 도전형 과제를 통한 소분야 전문센터 마련과 데이터 거래 활성화 마련
 - － 연구데이터 인프라의 데이터 분석도구, 데이터 검색에 세부 분야로 참여할 최소 10개 분야의 온라인 기반 소분야 연구센터 지정(산학연 연계 과제로 추진)
 - － 각 연구데이터는 데이터 거래 허브를 만들어 직접적인 데이터 유통이 이루어지도록 유도
 - － 연구자 간 자율경쟁 기반으로 AI 기반 경제·사회문제를 해결할 수 있도록, 온라인 플랫폼 구축(2021~)

2) 사업추진 목표

□ (최종목표)“연구공유형 AI데이터 허브 구축·활성화”을 통하여 국가 AI경쟁력 혁신

목표	1단계 (2021~2023)	2단계 (2024~2026)	3단계 (2027~2030)
학습데이터셋	1,000개	2,000개	3,000개
허브 수용률	20%	40%	80%
데이터표준모델	8종(기준모델)	2종(보완모델)	5종(통합모델)
인력양성	1,500명	3,000명(누적)	5,000명(누적)

□ 목표1: 인공지능 학습용 데이터셋 6,000개 제공

- 1단계(1~3차년도): 2억개의 데이터를 포함한 분야별 데이터셋 1,000개
- 2단계(4~6차년도): 5억개의 데이터를 포함한 분야별 데이터셋 2,000개
- 3단계(7~10차년도): 10억개의 데이터를 포함한 분야별 데이터셋 3,000개
 - ※ 정부는 기업들이 필요로 하는 인공지능 학습용 데이터를 2022년까지 1억 6천건을 구축 할 계획
 - ※ 단계별 구축 목표는 평가에 따라서 조정될 수 있음

□ 목표2: 인공지능 관련 연구데이터의 클라우드 허브 플랫폼 수용률 80% 달성

- 1단계(1~3차년도): IT분야 중심으로 20% 달성
- 2단계(4~6차년도): 전체산업응용 중심으로 40% 달성
- 3단계(7~10차년도): 일반연구분야 확산을 통해 80% 달성

□ 목표 3: 타분야 연구데이터가 참고가능한 데이터 표준모델 10종류 수립

- 1단계(1~3차년도): 소분야 전문센터를 중심으로 네 가지 지능분야 별 데이터 기준모델 2종류씩 총 8종류의 표준모델 구축
- 2단계(4~6차년도): 1단계 8종류의 표준모델의 단점을 보완한 두 종류의 모델 추가, 누적 10종류의 표준모델 구축
- 3단계(7~10차년도): 2단계의 표준모델 중 중복되는 모델을 통합하여 5종류로 단순화
- 데이터의 복잡도를 고려하여 단일 표준화 지향

□ 목표 4: AI기술 이론과 실데이터 경험을 보유한 AI데이터 전문가 5,000명 양성

- 1단계(1~3차년도): 단기 교육(6개월)과 표준화 경험 중심의 연 500명 전문인력 양성
- 2단계(4~6차년도): 단기 교육(6개월)과 표준화 경험 인력 및 장기 교육(1년)과 연구 경험을 가진 연 500명의 전문인력 양성
- 3단계(7~10차년도): 장기 교육(1년)과 연구 경험 중심의 연 500명의 전문인력 양성

나. 사업추진 전략 및 세부과제

1) 사업추진 전략

□ AI는 향후 경제·사회 전반을 혁신할 근본기술이라는 점을 고려, 국민 삶의 질과 국가 경쟁력 제고를 위해 AI 기술력 확보가 필수

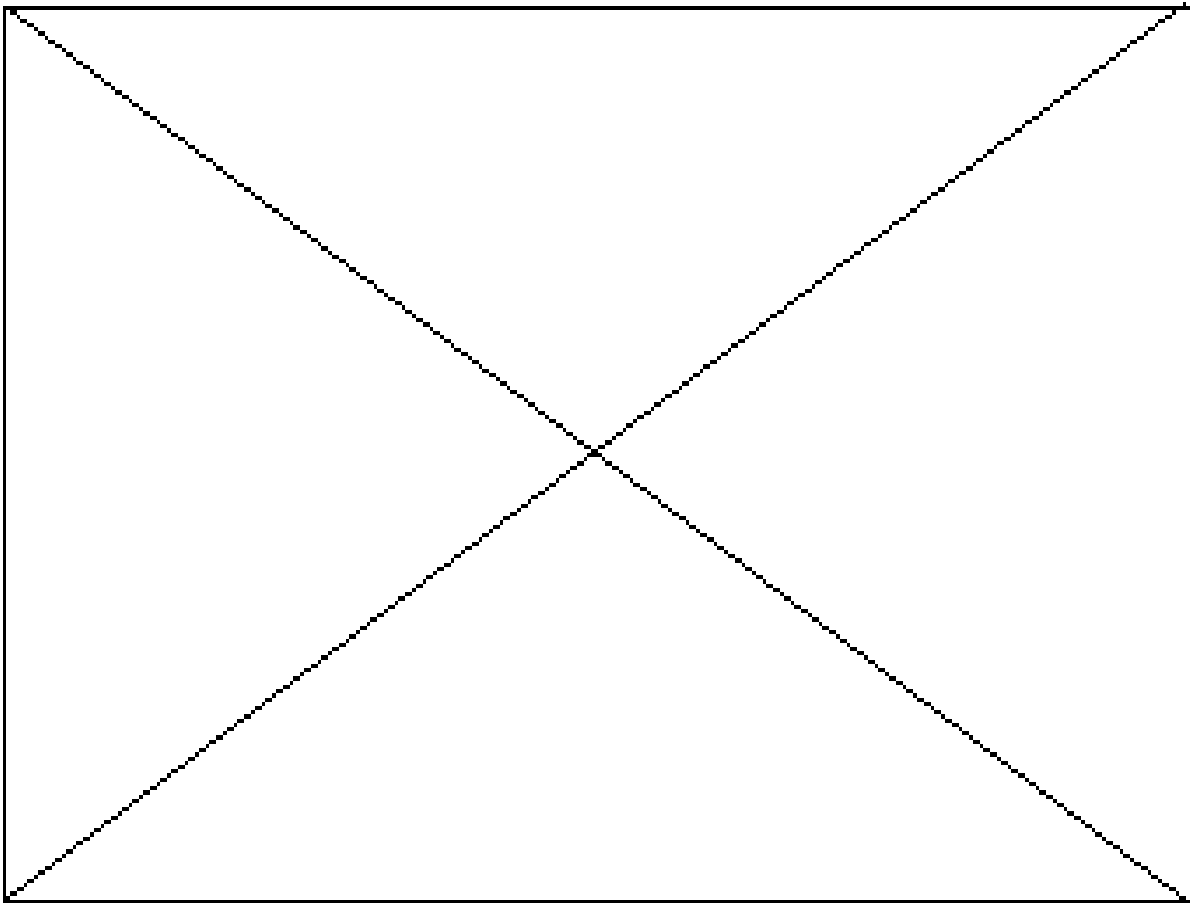
- 우리 AI 기술력은 美·中에 다소 취약하나, ICT산업(네트워크 등)이 두루 발전되어 있어 AI를 개발·활용할 수 있는 여건 양호, 여타 국가들과 비교시 상대적으로 유리한 위치
- 기본적인 AI 알고리즘이 대부분 오픈소스로 공개되어 있고, 많은 산업 분야가 AI 적용을 시작하는 단계이므로 전략적 접근 시 경쟁력 확보 가능

※ 특히 AI는 Data 학습을 근간으로 하는 바, 특정 분야 Data를 확보·학습시킨 AI는 특정 산업 영역에서 충분한 경쟁력을 가질 수 있으며, 타 기술·산업의 혁신에도 적용 가능

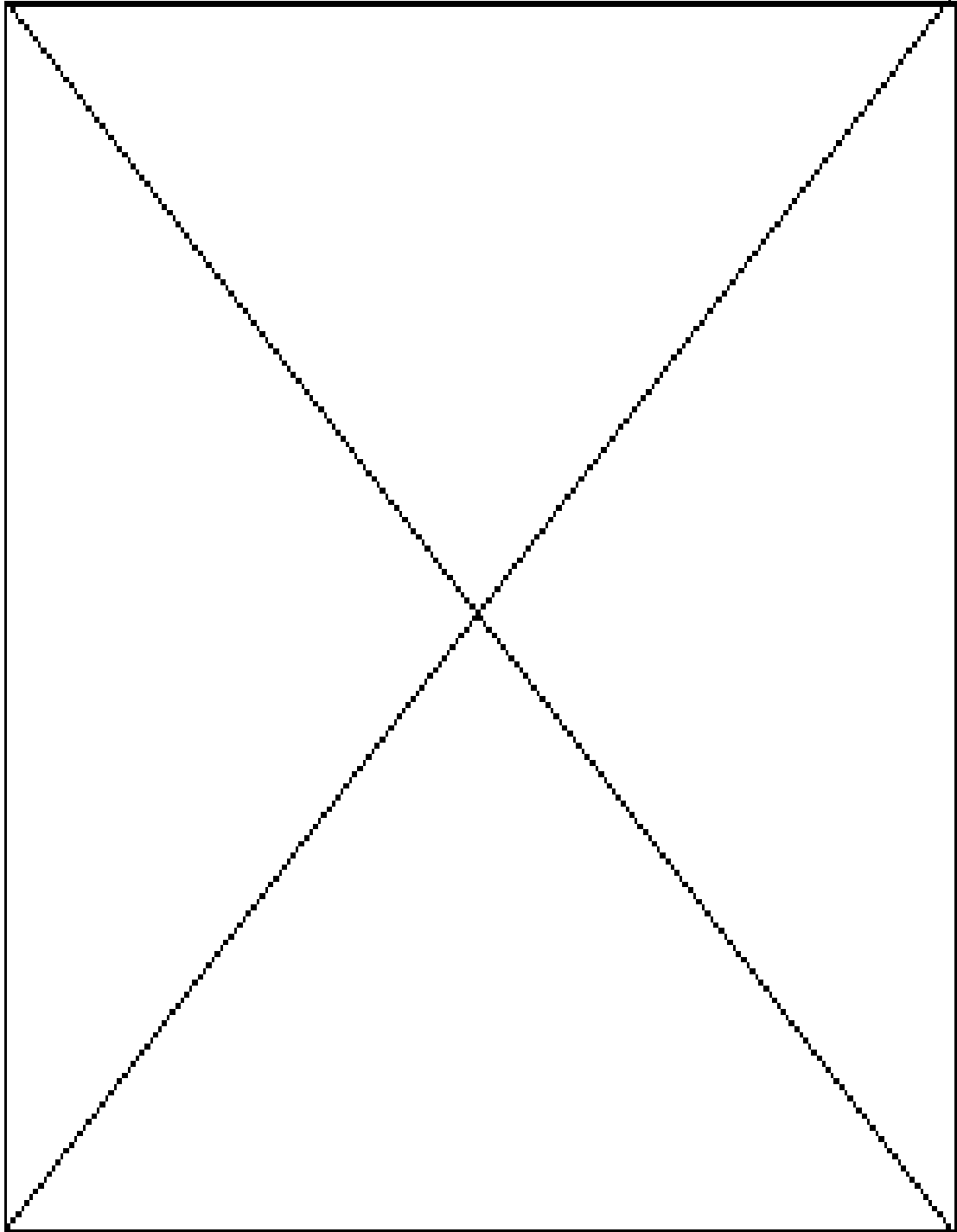
< 우리나라 AI 경쟁력의 SWOT 분석(출처: AI R&D 전략(과기정통부, '18)>

강 점 (S)	약 점 (W)
○ OS 없이 ICT 강국으로 도약한 경험 ○ 풍부한 데이터와 AI활용을 준비하는 다수의 고도화된 산업군 ○ C-P-N-D 분야별 역량을 갖춘 기업 ○ 기술수용성이 높은 소비자층	○ AI 분야 기술경쟁력 취약 ○ 양·질적으로 부족한 AI인재 ○ 기계학습용 데이터와 컴퓨팅파워 부족 ○ 개방·협력형 비즈니스 마인드 부족
기 회 (O)	위 험 (T)
○ 이제 막 개화한 AI시장 ○ AI 알고리즘의 대부분 오픈소스 공개 ○ 범용 AI플랫폼 없이도 수익창출 가능 ○ AI기반 지능화로 他산업에서 성과창출 기대	○ 美·中 등 선도국과의 기술격차 심화 ○ 글로벌 기업의 국내진출 가속화

- 각 세부 전략에 따라서 서비스 구축, 기반 마련, 생태계 확보의 세부 과제로 구성할 수 있음
- 제도화 및 인력양성, 지원 센터 등을 설립하는 정책, 거버넌스 전략을 토대로 상호 세부과제의 영역을 나누고, 하드웨어 인프라를 제공할 수 있는 저장소 마련
 - 데이터 인프라와 서비스 인프라를 나누고, 이것이 실제 소분야별로 연구가 이루어지며 중장기적으로 Git/커뮤니티와 연계할 수 있는 인프라를 제공
 - 공유형 인프라를 제공하되, 데이터의 보안성 및 개인정보 식별, 클라우드 인프라에 대한 활용과 지원체계를 마련하여 실질적인 AI연구데이터 공유가 이루어지도록 지원함



<그림 6-43> AI연구 데이터 추진과제 구성

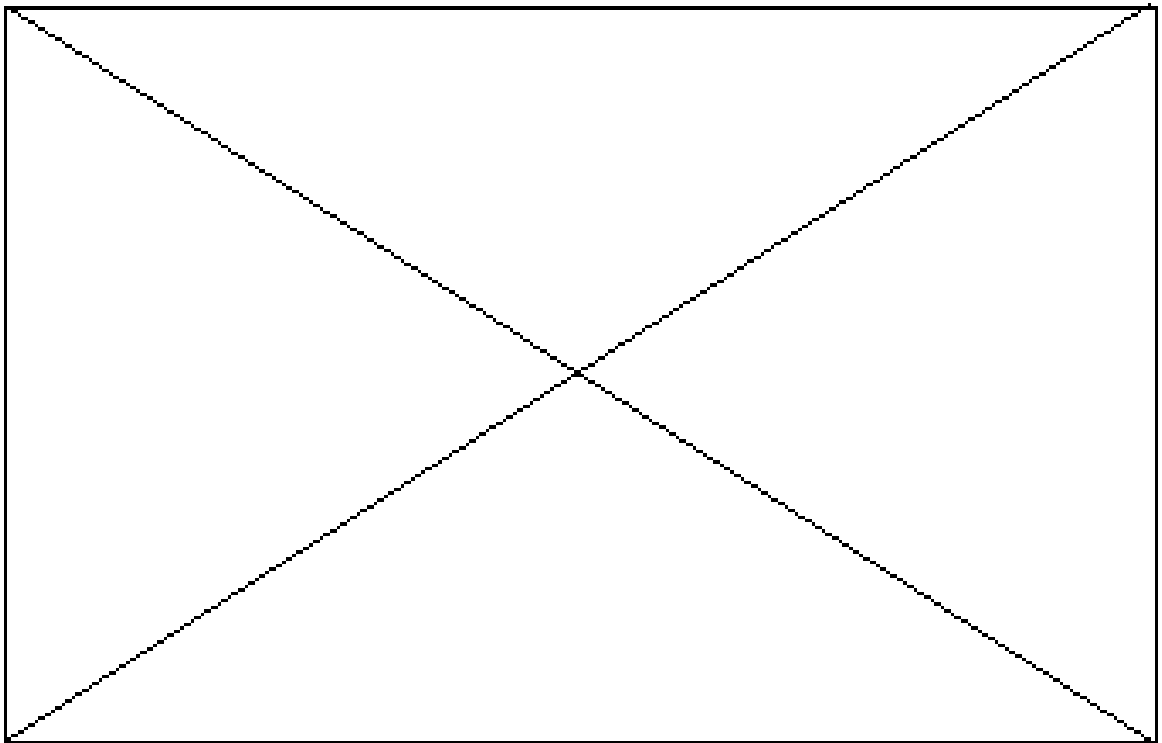


라. 세부과제별 추진 내용

1) AI데이터 서비스

□ 개요

- 정부 R&D 과제를 수행한 연구소, 학교, 기업이 연구를 위해 구축한 데이터를 쉽게 등록, 처리, 검색 활용 할 수 있도록 데이터 오픈 플랫폼 조성



□ 국가 AI 연구데이터 등록 지원, 관리 및 운용

- (데이터 등록) 정부 R&D 수행기관에서 구축한 데이터를 등록
 - － 정부 R&D 과제 수행을 통해 구축한 데이터 등록 인터페이스 구축
 - － 정부 R&D 저장 빅데이터 시스템 구축
 - AI 학습용 빅데이터 저장 시스템으로 확대 구현
 - 고속 처리 및 대용량 데이터 처리 지원
 - － 데이터 인증 및 심의 프로세스 구축
 - 잘못된 데이터(데이터 내용 또는 메타 정보의 오류 등)나 보안성 위험이 있는 데이터 등의 업로드 및 데이터 통합을 막기 위해 데이터 인증 및 심의 프로세스를 구축 및 운영

○ (데이터 관리) 데이터 관리 기술 연구

- 소프트웨어정의 스토리지 기반 분산 빅데이터 관리 정책 연구

- 데이터의 연관성 기반 분산 관리 정책 정의
- 데이터의 형상 추적 및 수명관리 정책 정의

- 데이터 수집 및 처리 모니터링 기술 연구

- 데이터 수집 및 처리 이벤트 발생 실시간 감지 기법 정의
- 모니터링 이력 관리 및 검색 효율성 향상 기법 정의

□ 국가연구데이터 통합 및 전처리 기술 개발 및 지원

○ (데이터 통합 및 전처리) 정부 R&D 수행기관에서 구축한 데이터 전처리 및 통합

- 빅데이터 이질성 유형 조사 분석: 정형 데이터와 반정형 데이터, 비정형 데이터(텍스트, 이미지, 음성, 대규모 그래프 등)의 각 유형별 특징을 데이터 처리 절차, 목적, 저장 및 처리기법 별로 유형화

- 빅데이터 유형별 분석을 통한 메타데이터 모델 연구: 데이터의 출처, 관련된 데이터, 콘텐츠, 네트워크 관계 등을 고려한 메타데이터 계층 구조 및 관련 속성 분석 및 새로운 데이터 유형 발생을 고려한 자가 확장 가능한 빅데이터 메타데이터 모델 정의

- 데이터 유형별 동적 데이터 필터링 기술 연구

- 데이터 유형별 처리 기술 분석
- 데이터 유형별 데이터 필터링 패턴 및 전략 분석
- 패턴 기반 동적 데이터 필터링 모델 정의

- 데이터 유형별 동적 데이터 정제 기술 연구

- 데이터 유형별 데이터 정제 기술 분석
- 데이터 유형별 데이터 정제 패턴 및 전략 분석
- 패턴 기반 동적 데이터 정제 모델 정의

- 데이터 유형별 동적 데이터 축소 기술 연구

- 데이터 유형별 데이터 전처리 기술 분석
- 데이터 유형별 데이터 축소 패턴 및 전략 분석
- 패턴 기반 동적 데이터 축소 모델 정의

- 데이터 분석 효율성 향상을 위한 연관성 기반 색인 기술 연구

- 데이터 분석 요구사항과 주제 연관성 기반 데이터 색인 기법 정의

□ 국가 AI 연구데이터와 국가연구데이터플랫폼 네트워크 연계

- (데이터 목록 수집) 정부 R&D 수행기관에서 구축한 데이터 목록을 실시간 수집하여 한국정보화진흥원의 K-ICT 빅데이터센터에서 관리
 - 공개SW 기반의 데이터 플랫폼을 구축하여 K-ICT 빅데이터센터 데이터셋을 효율적으로 운영·관리하고 타 플랫폼과의 연동 수행
 - 개방형 LOD(Linked Open Data) 플랫폼을 구축하여 다른 연구분야에서 만든 정보와 호환성 및 상호 운용성 확보
- ※ CKAN(Comprehensive Knowledge Archive Network)은 국제적으로 많이 활용되는 오픈소스 기반 플랫폼 (146개 공공, 학술단체 등 활용중)
- (데이터 맵 제공) 수집한 데이터 목록을 웹에서 검색할 수 있도록 데이터 맵을 구축·제공하여 데이터 간 연계·활용 촉진
 - 데이터 간의 연관성 분석을 통한 연관성 분석 기술 연구 추진
 - 데이터셋 간 의미적 연관성을 시각화하여 검색·활용할 수 있도록 기능 적용
 - 데이터 목록 검색 기능 제공
 - 융합 연구를 위한 연관 데이터 목록 검색 기능 제공
 - 데이터셋 별로 연관 데이터 목록을 추천해 Mash-up이 쉽고 빠르게 이루어질 수 있도록 함
 - 타 연구자들이 해당 데이터셋과 어떤 데이터셋을 함께 사용했는지의 정보도 제공
 - 데이터셋과 관련된 연구 목록 및 관련 연구 성과 제공
 - 기존 연구자들이 해당 데이터셋을 이용해 연구한 목록과 논문, 보고서 등의 관련 연구 성과를 제공해 빠르고 효율적으로 관련 정보를 취득할 수 있도록 함
 - 구축한 통합검색 및 데이터맵의 기능을 K-ICT 빅데이터센터 인프라(데이터 수집, 분석, 시각화 등), 홈페이지 등에서 활용 할 수 있도록 기능 연계 개발
- ※ K-ICT 빅데이터센터 인프라와 연계해 효과적인 확산이 가능해짐

2) AI데이터기반 구축

□ 연구데이터 관리시스템 클라우드 구축

- 연구데이터 관리시스템 구축을 위한 ISP 수립(2020년)
 - 연구데이터 법제도 반영 및 연구데이터 관리를 위한 환경분석 및 협업지원을 위한 기능내역 도출
 - R&D 연구 데이터 관리방안 도출 및 데이터 융·복합 시스템 기능 구현방안 수립
 - 연구기관간 데이터 연계 및 공동활용을 위한 서비스 모델 및 사용유형 개발

- 연구데이터 관리시스템 서비스 관리자를 위한 인프라, 플랫폼, 소프트웨어 서비스 구축 및 관리·운영 정책 개발

○ 연구데이터 관리시스템 구축 및 운영(2021년~)

- 지속적으로 증가하는 연구데이터를 효율적으로 관리하기 위해 클라우드 기반 시스템 구축
- 연구자의 연구데이터 관리 효율성 향상을 위한 연구자별·협업연구 기관별 다양한 포털 서비스 환경 구현
- 연말 및 연구과제 마감 시기에 따른 데이터 이용 폭주환경에 대비한 대응 체계 마련
- 인프라 운영 및 솔루션 서비스 제공을 위한 HW환경 구축
- 데이터 관리 기능 향상을 위한 데이터 정제, 데이터 셰어링 등의 다양한 기능을 포함한 데이터 활용 중심의 시스템 구축

※ 데이터 관리시스템 기능 : 데이터 셰어링, 데이터 관리, 데이터 정제, 사물인터넷 데이터 처리, 데이터 업테이트, 데이터 백업, 검색최적화, 파일접속, 클라우드 데이터 싱크 등

○ 연구데이터 관리시스템 사후관리 및 민원 대응체계 마련(2021년~)

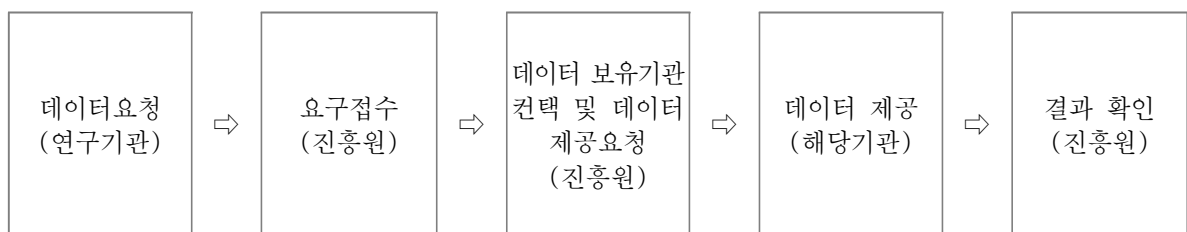
- 본 시스템의 효율적 운영 및 연구 인력에 대한 서비스 향상을 위한 시스템 이용에 대한 컨설팅 및 교육
- 한국정보화진흥원을 중심으로 24시간 연구지원 콜센터 운영을 통한 연구자의 민원처리 및 요구사항에 대한 대응체계 마련

□ 실시간 필요 연구 데이터 지원 시스템 구축

○ 정부부처 및 공공기관이 보유한 데이터에 대한 연계방안 연구(2020년)

- 공공데이터 포털 및 빅데이터 포털을 통한 연구자에게 제공할 수 있는 대책 마련 TF 운영
- 연구과정에서 필요한 연구데이터에 대한 중계 역할 수행 및 데이터 전처리 역할 지원방안 수립

< 연구데이터 이용에 따른 지원체계 업무흐름도 >



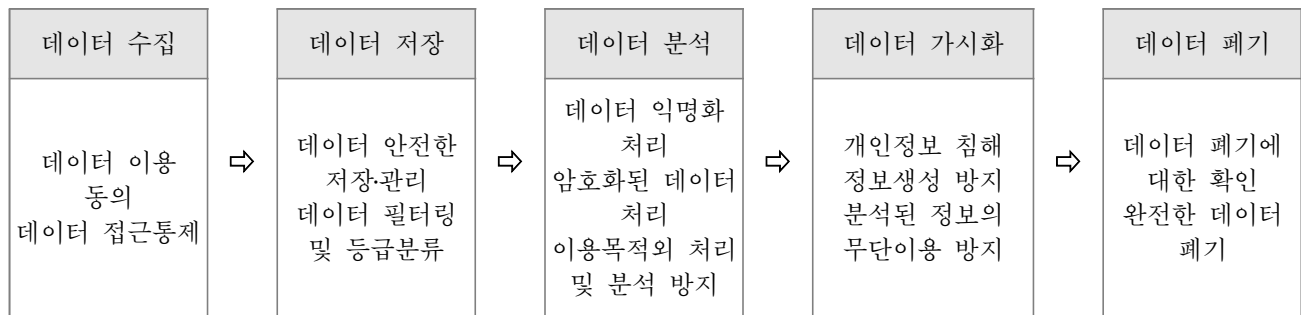
- 연구신청과정에서 필요한 공공데이터에 대한 요구 및 제공 업무 절차 확립

- \ ○ 실시간 센싱 데이터 제공을 위한 수집 및 배포 시스템 구축·운영(2021년)
 - R&D 연구과제 수행시 필요한 실시간 데이터 수집 시스템 구축
 - 실시간 데이터 관리기관과의 협의체 구축 및 운영
 - 민간의 데이터 생성기관(카드사, 이동사 등)과의 데이터 연계 협력체계 구축
 - R&D 연구기관의 센싱 데이터 설치 및 데이터 수집 지원센터 운영
 - 실시간 IoT데이터에 대한 개인정보 보호체계 마련 및 비식별 조치 시스템 기능 구현

□ 연구데이터의 효율적·안정적 관리체계 마련 및 협력 네트워크 기반 조성

- 연구데이터 생애주기별 보안 관리체계 마련(2020년~)
 - 연구데이터에 대한 수집·저장·분석·가시화·폐기 단계별 보안 요건 마련 및 시스템 기능 구현

< 연구데이터 단계별 보안 요건>



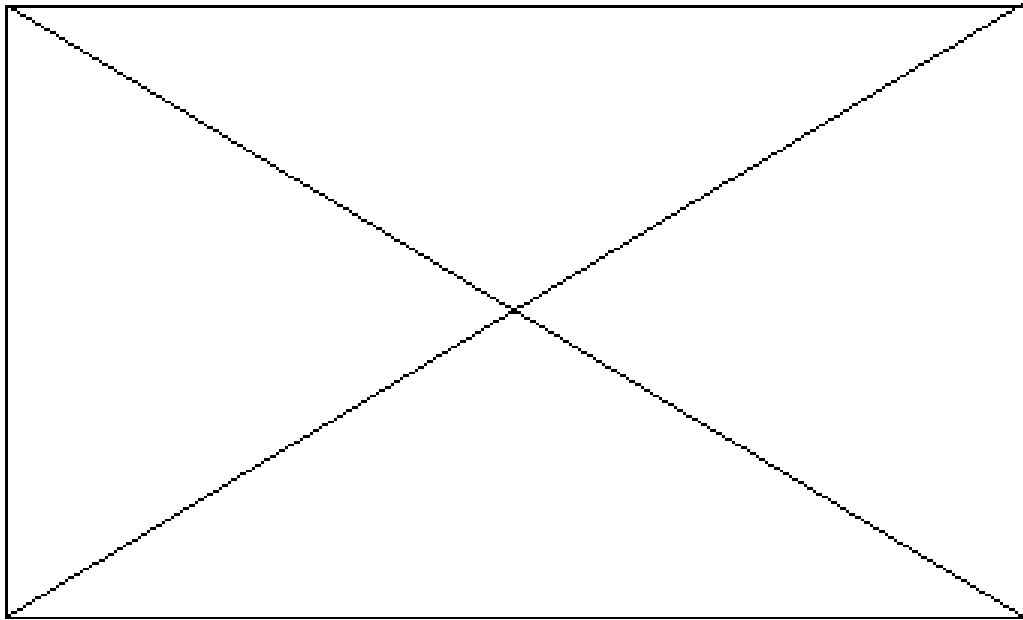
- 데이터 관리 및 데이터 이용의 법·제도 정비에 따른 시스템 보안 정책 수립 및 운영
- 연구데이터의 개인정보 처리 및 익명화를 위한 필터링 기능 지속적 개선
- 연구데이터의 외부 무단유출 방지 및 오남용 방지를 위한 보안관리 솔루션 도입·운영
- 연구데이터 보호를 위한 정보보호관리체계(ISMS) 인증 획득
- 연구데이터 기반의 연구협력 네트워크 구축 및 운영(2020년~)
 - 연구데이터의 공동활용, 연구결과 공유를 위한 협력 네트워크 활용을 위한 커뮤니티 시스템 구축
 - 연구데이터 전처리 및 활용 지원을 위한 전문교육 및 전문가 지원 프로그램 운영

3) AI 연구생태계 구축

□ 중장기적 AI연구생태계 활성화를 위한 제도 마련, 플랫폼 구축 및 인력 양성

- (단기) AI 연구데이터 거버넌스 정립, (중기) 개방형 AI연구 플랫폼 구축, (중장기) AI데이터 전문인력 양성

< 개방형 AI연구생태계 구조 >



□ AI 연구데이터 거버넌스 정립

- AI 연구데이터의 공유 및 AI연구 생태계의 활성화를 위해 법제도, 표준화, 품질인증, 가치측정, 이용체계의 AI 연구데이터 거버넌스 정립
- (법제도) 연구데이터 공유를 위한 정책 수립 및 법제도 개선
 - － (제도화) 국가지원 AI연구의 데이터 제출 제도화
 - 데이터의 중요성이 대두되는 시점에서 연구자들로부터 데이터를 능동적으로 확보하기 위해 연구데이터 제출 제도화
 - 국가 연구과제의 지원 금액 중 일부분을 데이터 공유를 조건으로 하는 가이드라인 마련
 - 연구데이터 공유에 대한 보상 정의
 - － (활성화) AI 데이터 공유 활성화를 위한 민감데이터에 대한 법제도 완화
 - 데이터의 개인정보 및 라이선스 등과 같은 법적 이슈를 자문하고 데이터 사용 시의 가이드라인을 제공하는 ‘민감데이터 전문기관’ 마련

- \
 - 특히, IRB 등 기존 유사 기관에서는 민감데이터 보호에 중점을 두었다면 본 기관에서는 데이터 활성화에 중점을 두어 운영
 - (개인정보) 의료데이터나 개인정보데이터 등 국내법상 공유가 어려운 민감데이터의 공유를 위해 법제도 개정
 - (라이선스) 이미지 데이터 등의 라이선스 이슈를 해결하기 위한 자문 지원 및 가이드라인 수립
 - 비상업적 사용에도 라이선싱이 필요한 경우 연구 필요성을 검증한 후 라이선싱 보조금 지원
- (보호) 보호가 필요한 연구에 대해서는 제도적 보호장치 마련
 - 국방 등 보호의 필요가 있는 연구 주제에 대한 데이터의 경우 공유 지연(Embargo)을 위한 제도 마련
 - 연구데이터의 보호가치를 측정하는 가이드라인 수립
- (표준화) 연구데이터 공유를 위한 데이터 표준모델 마련 및 표준화기법 개발
 - (표준모델) 효과적인 연구데이터 공유를 위한 데이터 표준모델 마련
 - 전체 연구데이터 중 AI의 지능분야 별 대표 데이터셋 선정 후 대표 데이터셋의 유형/주제/용량 별 군집화를 통한 공통 형태 도출
 - 공통 형태의 추후 보완을 통해 총 5~10종류의 데이터 표준모델 수립
 - 표준모델에 대한 분석을 기반으로 AI데이터의 필요 요건(학습을 위한 최소 인스턴스 수 등) 정의
 - (표준화기법) 과학적인 연구데이터 공유를 위한 데이터 표준화기법 개발
 - 데이터의 종류에 따라서는 단일 형태로 표준화하는 것이 불가능하거나 표준화 과정에서 과도한 데이터의 변형 혹은 정보의 손실이 발생
 - 원본데이터(Raw Data)는 유지하면서 타연구자의 필요에 따라 적합한 표준모델로 변환을 지원하는 표준화기법의 개발을 통해 데이터 보호
 - 대표적인 대용량 데이터(e.g. 시계열 데이터, 텍스트 데이터)의 형태를 정의하고 각 형태에 맞춘 '표준모델 변환 S/W 개발' 연구 수행
 - 장기적으로는 다양한 원본데이터 형태를 지원하기 위해 'AI 기반 표준화기법 개발' 연구 수행
- (품질인증) 연구데이터의 품질인증 지표 마련
 - 연구데이터 공유를 강제하였을 경우 과제 평가만을 목적으로 낮은 품질의 데이터 공유하는 경우 발생
 - 연구데이터의 공유적합성과 성능정확성을 기준으로 품질평가 및 인증을 위한 가이드라

인 개발

- (공유적합성) 공유를 염두에 두고 연구데이터가 수집되었는지 확인
 - 과제 시작 전 데이터 표준모델 중 해당 과제에 적합한 모델을 선택하게 하여 과제 수행 과정에서 데이터 수집 및 정제를 단일화하고 그 결과 데이터 표준모델과의 적합도를 하나의 품질 지표로 사용
 - 과제에서 수집된 데이터가 표준모델에 Mapping이 불가능한 경우에는 표준화기법의 적용 가능성을 품질 지표로 사용
- (성능정확성) 특정 과제의 산출물인 연구데이터 및 연구결과물의 조합으로 기대하는 성능이 도출되는지 확인
 - 데이터가 의도한 바에 따라 수집되었으며, Noise가 적고, 잘 정제되었는지 판단하기 위해 동일 연구에서 개발된 AI 방법론을 토대로 성능 평가
 - 다수의 반복 실험 결과 정량적으로 목표한 성능을 얻었는지 판단
 - (중기) AI연구가 초기부터 성능정확성 지표에 맞춰 진행될 수 있도록 지원하는 ‘연구 결과의 성능 평가 프레임워크 개발’ 과제 진행
- (가치평가) 연구데이터의 공유를 위한 가치측정 기술 개발
 - 연구데이터의 가치를 측정하여 가치에 합당한 대우를 제공하여 양질의 연구데이터를 생산하도록 유도하는 제도 마련
 - 근본적으로 연구데이터에 대한 권리는 나라에 귀속되지만 연구 플랫폼 내에서의 연구데이터의 권리를 연구자에게 부여하는 가이드라인 구축
 - 연구데이터를 거래할 수 있는 ‘블록체인 기반 데이터 마켓’을 연구 플랫폼 내에 구축하여 높은 가치의 연구데이터는 마켓 내에서 상업적으로 거래하고 거래 수익을 연구자에게 연구비로 반환하는 제도 마련
 - 데이터의 활용범위, 수집의 난이도, 품질을 토대로 연구데이터의 가치를 측정하는 기술 개발
 - 항목별 정량적인 평가를 통해 품질을 분류하는 가이드라인 마련
 - (중기) 마켓 내 거래 데이터를 토대로 데이터의 가치를 자동적으로 측정하는 ‘연구데이터 가치 측정 AI모델 개발 연구’ 진행
- (이용체계) 연구데이터가 효과적으로 공유 및 활용되기 위한 이용체계 구축
 - (공유) 데이터 공유의 활성화를 위해 일반적으로 사용할 데이터를 찾을 때 확인하는 정보의 확보
 - 데이터를 제공하는 시점에서 필수 정보를 함께 입력하도록 하는 가이드라인 개발
 - 데이터의 기본정보로부터 데이터의 주제/분야를 파악할 수 있는 ‘데이터 분류 및 토픽

\ 추출 연구' 수행

- (활용) 데이터의 활용도를 높이기 위한 서비스 툴 개발
 - 쉽게 타 연구데이터와 연계 사용 및 테스트가 가능하도록 설계
 - GUI 기반의 데이터 조합 및 가시화 툴 개발

□ 개방형 AI연구 플랫폼 구축

- AI연구생태계 활성화를 위해 AI 연구자들의 모여 소통하고 정보를 얻을 수 있는 플랫폼 구축
- (데이터공유) 블록체인 기반 데이터 마켓 개발
 - AI데이터 공유를 활성화하고 데이터 제공과 특히, 활용범위가 넓은 양질의 데이터를 구축한 연구자에 대한 보상을 지원하기 위한 상업적 마켓 운영
 - 단, 데이터 자체의 소유권은 국가에 귀속되어야 하므로 연구자는 해당 마켓에서 발행하는 코인 형태로만 수익 창출 가능
 - 코인을 활용하여 다른 데이터를 구매하는데 사용하거나 일정 액수 이상은 연구과제 형태로 제공하는 수익화 가이드라인 구축
- (방법론개발) Korean Data Challenge
 - 한국어/한국말로 구현된 데이터에 대한 방법론들은 비교적 수요가 적은 언어라는 이유로 발전이 더디다는 문제점이 있음
 - 이를 해결하기 위해 국내 활용 필요성은 높으나 수익 등의 이유로 방법론 개발이 활성화되고 있지 못한 데이터셋 및 문제를 선정하고 이를 해결하는 방법론의 성능을 겨루는 'Korean Data Challenge'를 운영
 - 높은 상금과 혜택(e.g. NIA 서류면접 면제, 국가과제 신청시 가산점 부여)을 제공하여 해당 분야의 국내외 연구가 활성화될 수 있도록 유도
- (연구공유) 데이터 중심의 개방형 연구 커뮤니티
 - AI 연구 결과와 데이터를 공유하고 소통할 수 있는 커뮤니티 'AI Expert Center' 개발
 - 연구자 개인페이지와 공유된 데이터 페이지를 마련하여 데이터를 중심으로 연구자들이 소통하고 해당 데이터를 재활용하여 추가적인 연구가 활발하게 진행되도록 설계

□ AI데이터 전문인력 양성

- AI기술 이론과 실데이터에 대한 감각을 모두 보유한 AI데이터 전문가 양성
 - 수요대비 공급이 부족한 AI데이터 전문가 양성으로 AI연구생태계 활성화

- AI기술에 대한 이론 교육과 함께 AI데이터 거버넌스 구축 및 국내외 연구과제의 AI데이터 수집/정제 수행에 인력을 대여해주는 제도를 운영하는 'AI데이터 전문가 양성 센터' 설립

○ (교육) AI기술 이론 교육

- 전문인력 양성 생애주기 중 초기, 말기에 AI기술 이론에 대한 중점적인 교육프로그램을 진행
 - 6개월 커리큘럼 - 초기 2개월, 말기 1개월에 집중 교육 실시
 - 1년 커리큘럼 - 초기 2개월, 말기 2개월에 집중 교육 실시
 - 초기 집중 교육은 AI/ML에 대한 기초 이론 교육, 말기 교육은 고급 이론 교육 및 실무경험을 정리하는 교육 수행
- 전문인력 양성 생애주기 중 중기의 실무 투입 과정에서도 세미나 형식으로 인력 양성 단계에 맞는 AI/ML 기법에 대한 지속적인 교육 수행

○ (실무) 표준모델/표준화기법 개발 및 국내외 연구과제에 인력 대여

- (표준화) 초기 표준모델 정착 이전에 데이터의 표준화를 지원하는 인력을 투입하여 해당 인력의 데이터 전문가로의 성장 유도
 - 표준모델 및 표준화기법 개발을 지원하는 인력은 다수의 AI 연구데이터를 경험함으로써 AI연구 분야에 대한 이해도를 높일 수 있을 뿐만 아니라 표준화기법 개발에 직접적으로 참여함으로써 프로그래밍 역량과 AI 기법 활용 역량을 증진할 수 있음
 - 표준모델/표준화기법 뿐만 아니라 제도적 전문성을 갖추 수 있도록 교육 프로그램 운영
 - 양성된 인력은 지속적으로 데이터 표준화를 위한 다양한 활동에 투입
- (연구) AI연구에서 필수적인 데이터 구축과정을 지원하는 경험을 통해 해당 인력의 데이터 전문가로의 성장 유도
 - 국가지원 연구를 수행 중인 연구자 중 데이터 구축 시 보조 인력을 필요로 하는 경우 신청과 선발 과정을 통해 인력 공급
 - 국내 연구뿐만 아니라 국외 연구자에게도 인력 지원기회를 열어두고 해외 체류비 등을 지원하여 글로벌 실무 경험을 활성화

라) 소분야 AI연구데이터 챌린지 센터 설립

□ AI연구데이터 지원 및 운영/유통센터(가칭)설립

- AI연구데이터 지원과 관련 행정실무를 총괄하는 유통센터 설립
- 데이터에 대한 표준 어휘 및 온톨로지, 데이터 아키텍처 유통 실무 총괄

- \ - 표준화된 데이터의 유통과 관리를 지원하기 위한 RDF 트리플과 온톨로지의 유지 관리
 - 신규 데이터세트의 관리성을 확보하기 위한 어휘, 전문용어 관리체계 마련(과학기술정보연구원의 기본 어휘데이터 유통체계 연계)
- 연구자들에 대한 AI연구데이터 표준 컨설팅 및 교육
 - AI데이터 전문가 양성센터 운영 지원 및 통계 집계, 정책 지원
 - 전문가 양성센터를 위한 교육데이터 설계 및 코스웨어 준비
- AI 전문인력 양성을 위한 기본 학습데이터 제공
 - 인력양성 및 측정, 평가용 학습데이터의 구축과 확산 배포
 - 표준 테스트 데이터세트의 검증 및 품질 관리
- 클라우드 인프라의 운영
 - AI빅데이터 저장소 관리와 하드웨어 소요 예산 측정
 - 타분야 연구데이터 연계에 대한 OpenAPI 기반의 메시징 인프라 구축 지원
 - 슈퍼컴퓨팅 연구와 연계될 수 있는 인프라 마련
- 표준모델 관리센터 운영
 - 표준화된 데이터의 실제 정합성, 현행 데이터의 구조 설계와 확산체계 마련
 - 오픈소스 형태로 제공되는 정보의 최신성 유지
 - 연구데이터 전처리 기술에 대한 표준화와 상호 검증 방법론 제공
- 커뮤니티 운영 및 지원(AI Expert Center)
 - AI연구데이터 밋업(meet up)과 같은 커뮤니티 기반의 자유로운 의사소통 채널 마련과 행사 경비 지원

□ AI연구데이터 챌린지

- 연구과제를 수행하고 있고 표준화된 데이터를 통하여 만든 AI 데이터를 실제로 데이터 학습 결과를 대회로 연계
- 초중고 뿐만 아니라 기업을 대상으로 연구데이터를 통한 사회현안 해결 및 과학적 난제 해결을 위한 챌린지 개최
 - 양자데이터 분석, 기계학습, 리빙랩 데이터 분석 등 다양한 분야의 활용성을 검증하고 연구데이터의 품질 검증
- 구축 알고리즘에 대한 공개SW형태의 확산 및 커뮤니티 구성 독려

- 단순히 대회를 개최하는 것이 아니라 지속적으로 대회를 준비하고 커뮤니티를 통해 함께 연구할 이들을 서로간에 규합함으로써 성취감을 가질 수 있게 지원
- SW중심대학 등 다양한 SW관련 연구기관들을 통하여 AI 연구데이터에 대한 SW측면의 접근법을 적극 제공 및 커뮤니티 활동 안내
- 높은 상금 및 보상체계로 대중적 홍보성 확보
 - 기업별 공채 혹은 특채 기회 제공이나 창업이 가능한 수준의 상금 보상체계를 통하여 동기 유발 제공

6.5.4. 사업운영계획

가. 사업추진 일정

세부과제명	사업기간									
	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
AI 데이터서비스										
AI데이터기반										
AI연구생태계 구축										
소분야 AI연구데이터 챌린지 센터										

나. 사업비 및 인력운영 계획

1) 투자계획 및 재원조달

(단위: 백만원)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	10,000	10,000	11,500	10,500	7,500	7,500	5,500	5,500	5,500	5,500	79,000
AI 데이터서비스	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000	40,000
AI데이터기반	4,000	4,000	4,000	3,000							15,000
AI연구생태계 구축	2,000	2,000	2,000	2,000	2,000	2,000					12,000
소분야 AI연구데이터 챌린지 센터			1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	12,000

2) 인력 운영계획

(단위: 명)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	6	6	7	7	7	7	5	5	4	4	58
AI 데이터서비스	3	3	3	3	3	3	3	3	3	3	30
AI데이터기반	2	2	2	2	2	2					12
AI연구생태계 구축	1	1	1	1	1	1	1	1			8
소분야 AI연구데이터 챌린지 센터			1	1	1	1	1	1	1	1	8

3) 세부과제별 사업비내역

□ AI데이터 서비스

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
국가연구데이터 등록 지원, 관리 및 운용	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	10,000
국가연구데이터 통합 및 전처리 기술 개발 및 지원	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	10,000
국가연구데이터와 빅데이터 네트워크 연계	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	20,000
계	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000	4,000	40,000

□ AI데이터기반 구축

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
클라우드 기반 연구데이터 관리 시스템	2,000	2,000	2,000	2,000	—	—	—	—	—	—	8,000
실시간 데이터 수집/배포 시스템	1,000	1,000	1,000	1,000	—	—	—	—	—	—	4,000
연구데이터 관리 및 보안체계	1,000	1,000	1,000	—	—	—	—	—	—	—	3,000
계	4,000	4,000	4,000	3,000							15,000

□ AI 연구생태계 구축

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
AI 연구데이터 거버넌스 정립	500	500	500	500	500	500	—	—	—	—	3,000
개방형 AI연구 플랫폼 구축	1,000	1,000	1,000	1,000	1,000	1,000	—	—	—	—	6,000
AI데이터 전문인력 양성	500	500	500	500	500	500	—	—	—	—	3,000
계	2,000	2,000	2,000	2,000	2,000	2,000					12,000

□ 소분야 AI연구데이터 챌린지 센터 운영

(단위: 백만 원)

구분 \ 연도	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
챌린지지원 센터 운영	—	—	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	8,000
챌린지 개최	—	—	500	500	500	500	500	500	500	500	4,000
계			1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	12,000

6.5.5. 기대효과

가. 경제적 기대효과

- 인공지능 기술의 산업확산으로 2030년 기준 최대 신규매출 90조원, 원가절감 210조원, 소비자잉여 270조원으로 총 570조원의 경제효과가 발생할 것으로 예측(NIPA,'17)
 - ※ GDP와 지능정보화 가능성을 기준으로 선정한 15대 산업의 130개 활용사례별 경제 효과를 합산하여 2030년까지 경제효과 산출
- AI 개발에 필요한 기계학습용 데이터, 컴퓨팅 파워, 우수(특화) 알고리즘 개발 환경이 원활히 개방·활용되는 연구개발 생태계 구축
- 자율적 경쟁과 협업이 가능한 AI 기술혁신의 장이 마련되고, 우수 인재 발굴 및 연구 확산 독려
- 교통·환경·의료·사회복지 문제 해결, 생활안전과 편의성 증진 등 전반적으로 삶의 질을 향상하는 방향으로 영향을 줄 것으로 기대
 - 만성질환, 희귀난치성 질환, 퇴행성 뇌·신경 질환, 공기오염, 수질 오염, 흉악범죄, 안전 사고, 기상재해, 에너지 빈곤, 교통혼잡, 교통안전, 자살, 교육격차 등 문제 해결
- AI 개발 필수 인프라의 본격 구축·제공을 통해 인공지능 경쟁력의 조기 확보와 시장 선점에 기여
- 희소성이 높은(수집 비용이 비싼) 연구데이터의 재활용을 통하여 중복 투자 및 중복 AI 연구데이터의 품질 향상
- 경우에 따라 핵심 연구데이터의 기술 거래를 통한 부가가치 창출

나. 과학기술적 기대효과

- AI데이터 서비스를 통한 직접적 효과
 - － 데이터의 부족으로 만족스러운 성능을 얻기 어려웠던 범용 AI 기술의 성능 향상
 - － 데이터 수집 역량 부족으로 진행되지 못한 연구들이 진행되어 연구 성과의 증대
- AI 개발을 위한 데이터 구축 등에 소요되는 비용과 시간을 절감하고 기술 공유를 통해 국내 최고 또는 세계 수준의 기술을 보다 신속히 개발 가능
 - － 연구데이터 공유·활용을 통해 국내 연구 개발 생산성 증대*

\ - 칸막이를 넘어 연구 주체 간·분야 간 융합·협력 연구* 강화

※ LIGO 연구단 '17년 노벨상 수상(16개국 900명 협업), CERN 힉스입자 발견(85개국, 8,000명 협업) 등

- 장기간 데이터 축척이 필요한 분야(물리실험, 신약개발 등)에 활용, 데이터 구축 시간 비용을 획기적으로 단축, 신제품 서비스의 연구개발 촉매제 역할을 수행

※ 연구데이터 공유·활용을 통해 첨단소재분야 연구기간 최대 10배 단축 가능(NRC, '08)

- 인력 및 제도적 관점에서 추가적인 효과 기대 가능

- 연구데이터 활용능력 및 과학적 추론 능력을 겸비한 데이터 전문 과학자 양성

- 동일 데이터셋을 이용해 연구들간의 객관적인 성과 비교 가능

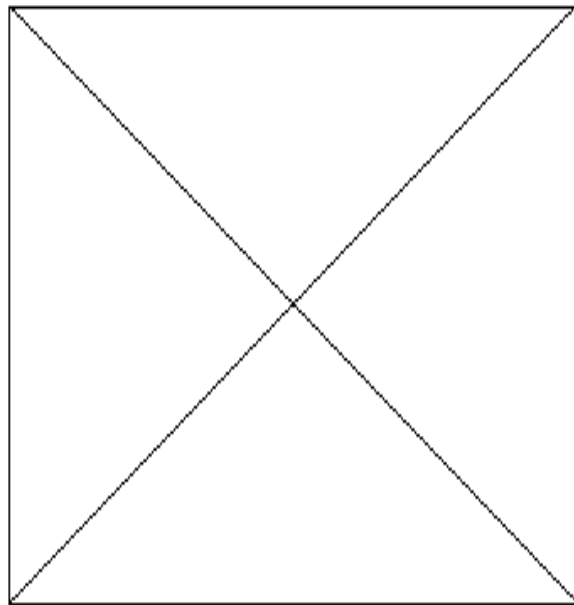
- 데이터 구축에 들어가는 시간을 줄이고 알고리즘 고도화에 더 많은 연구 시간 배분이 가능하며 이로 인한 연구 성과 증대 기대

6.6. 전문센터 우선분야 선정과정

6.6.1 우선분야 선정개요

□ 우선분야 선정 방법

- 국가연구데이터 공유·활용 우선분야 선정을 위해 대상 분야를 결정하고, 평가지표 선정 및 평가(정량평가 및 중요도 반영)를 통해 분야별 우선순위 도출



<그림 6-44> 우선분야 선정 과정

6.6.2 대상 분야 선정

- 정부의 연구개발 기준과 연구개발 중점 분야 및 국내·외 데이터 전문센터 사례 검토 등을 통해 연구데이터 공유·활용을 위한 11개 분야 선정

<표 6-24> 대상 분야 선정결과

1	2	3	4	5	6	7	8	9	10	11
ICT·SW	생명·보건의료	에너지·자원	소재·나노	기계·제조	농림수산·식품	우주·항공·해양	건설·교통	환경·기상	원자력·핵융합	재난안전

6.6.3 평가지표 선정 및 정량평가

□ 평가지표 선정

- 우선분야 선정을 위한 평가요소를 국가 전략적 중요성, 연구데이터 활용성, 파급효과 등으로 구성하여 과학기술 분야의 공공성, 기술성, 시장성 고려
- 국가전략적 중요성
 - － 국가 정책 또는 상위계획과의 부합성, 사업추진 시급성 등을 평가하는 항목으로 과학기술 분야의 공공성을 고려하는 지표
- 연구데이터 활용성
 - － 대상 분야가 생산하는 데이터양과 연구데이터 공유의 필요성, 활용 가능성 등을 평가하는 항목으로 과학기술 분야의 기술성을 고려하는 지표
- 파급효과
 - － 과학기술분야 파급성, 시장(산업) 파급성, 사회적(고용유발) 파급성 등을 평가하는 항목으로 과학기술 분야의 기술성 및 시장성을 고려하는 지표

<표 6-25> 우선분야 선정을 위한 평가지표

대분류	중분류	세부 평가지표
국가 전략적 중요성 (공공성)	국가전략 부합성	· 국정과제, 과학기술기본계획, 부처별 전략계획 및 그 외 중·장기전략 등 반영여부에 따른 1점~5점 구간으로 평가
	사업추진 시급성	· 해외 공공에서 추진 중인 연구데이터 공유·활용 분야 포함(활성화) 여부에 따른 1점~5점 구간으로 평가
연구데이터 활용성 (기술성)	연구데이터 생산량	· 분야별 연구데이터 생산량을 5분위수로 구분(1점~5점)하여 평가
	수요자 공유요구	· 분야별 연구데이터 활용의사(분야별 응답비율)를 5분위수로 구분(1점~5점) 하여 평가
	타 분야 융합 가능성	· 분야별 기술이 속해 있는 전략기술 대분류의 결점 수에 따라 1점~5점 구간으로 평가
파급효과 (시장성)	과학기술적 파급성	· 분야별 기술이 속해 있는 포트폴리오 영역에 따라 1점~5점으로 구분(1점~5점)하여 평가
	시장 파급성	· 분야별 타 산업분야 파급효과(생산유발)를 5분위수로 구분(1점~5점) 하여 평가
	사회적 파급성	· 분야별 고용유발효과를 5분위수로 구분(1점~5점) 하여 평가

□ 분야별 정량평가

- 항목별 평가방법을 적용하여 정량평가 결과 도출

<표 6-26> 정략적 평가 결과

구분	국가전략적 중요성		연구데이터 활용성			파급효과			계
	국가전략 부합성	사업추진 시급성	연구데이 터 생산량	수요자 공유요구	타 분야 융합 가능성	과학기술 적 파급성	시장 파급성	사회적 파급성	
ICT·S/W	5	3	3	5	3	5	3	3	30
생명·보건	4	4	4	5	3	5	2	5	32
에너지·자원	5	4	1	1	5	2	1	1	20
소재·나노	5	4	2	4	4	2	4	1	26
기계·제조	5	2	3	3	4	1	5	3	26
농림수산·식품	5	2	2	4	3	5	3	2	26
우주·항공·해양	5	3	4	2	3	2	5	3	27
건설·교통	5	1	1	1	3	1	4	3	19
환경·기상	5	4	5	3	2	3	1	4	27
원자력·핵융합	4	2	5	3	5	1	1	1	22
재난안전	5	1	1	1	3	1	2	5	19

6.6.4 평가지표별 중요도 산정

- 전문가 설문조사 및 AHP 분석을 통해 평가 항목별 중요도 산정

□ 전문가 설문조사

- 전문가 설문조사
 - － 설문조사 일시 : 2018년 10월 15일 ~ 2018년 10월 26일
 - － 설문조사 대상 : 분과위원회 위원 14명

□ 항목별 중요도(가중치) 산정

- 항목별 중요도(가중치) 산정결과 국가전략 부합성(0.278)이 가장 높게 도출됨

<표 6-27> 항목별 가중치 산정 결과

대분류(A)		소분류(B)		항목별 가중치(C)
평가지표	가중치	평가지표	가중치	
국가전략적 중요성	0.385	국가전략부합성	0.723	0.278
		사업추진시급성	0.277	0.107
		소계	1.000	—
연구데이터 활용성	0.415	연구데이터생산량	0.166	0.069
		수요자공유요구	0.512	0.212
		타분야융합가능성	0.322	0.134
		소계	1.000	—
파급효과	0.200	과학기술적파급성	0.509	0.102
		시장파급성	0.248	0.050
		사회적파급성	0.244	0.049
		소계	1.000	—
소계	1.000	—		
합계				1.00

6.6.5. 분야별 우선순위 도출

- 정량평가 결과에 항목별 중요도를 반영하여 분야별 우선순위 도출
- ICT·SW, 생명·보건의료, 농림수산물·식품, 환경·기상, 소재·나노 등이 상위 5위로 도출됨

<표 6-28> 분야별 우선순위 도출 결과

순위	분야명	1. 국가전략적 중요성		2. 연구데이터 활용성			3. 파급효과			총점
		국가전략 부합성	사업추진 시급성	연구데이터 생산량	수요자 공유요구	타 분야 융합 가능성	과학기술적 파급성	시장 파급성	사회적 파급성	
1	ICT·SW	1.392	0.320	0.207	1.062	0.401	0.509	0.149	0.146	4.19
2	생명·보건의료	1.114	0.426	0.276	1.062	0.401	0.509	0.099	0.244	4.13
3	소재·나노	1.392	0.426	0.138	0.849	0.534	0.204	0.198	0.049	3.79
4	농림수산물·식품	1.392	0.213	0.138	0.849	0.401	0.509	0.149	0.098	3.75
5	환경·기상	1.392	0.426	0.345	0.637	0.267	0.305	0.050	0.195	3.62
6	기계·제조	1.392	0.213	0.207	0.637	0.534	0.102	0.248	0.146	3.48
7	우주·항공·해양	1.392	0.320	0.276	0.425	0.401	0.204	0.248	0.146	3.41
8	원자력·핵융합	1.114	0.213	0.345	0.637	0.668	0.102	0.050	0.049	3.18
9	에너지·자원	1.392	0.426	0.069	0.212	0.668	0.204	0.050	0.049	3.07
10	건설·교통	1.392	0.107	0.069	0.212	0.401	0.102	0.198	0.146	2.63
11	재난안전	1.392	0.107	0.069	0.212	0.401	0.102	0.099	0.244	2.63

데이터 및 컴퓨팅활용 기반의 R&D 인재육성

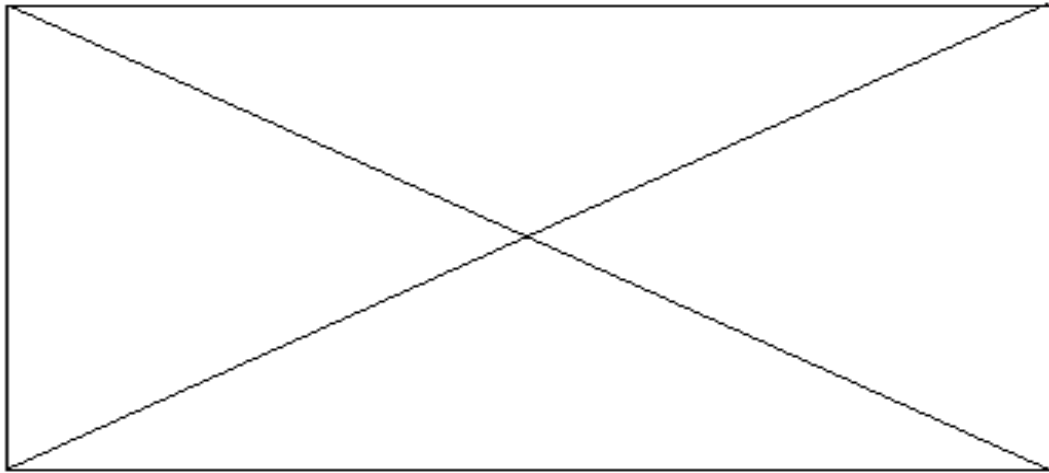
- 7.1. 과제개요
- 7.2. 환경분석
- 7.3. 추진전략
- 7.4. 세부과제별 추진내용
- 7.5. 기대효과

7.1 과제 개요

7.1.1 추진 배경

□ 디지털 기술과 연구 환경 변화에 따른 연구데이터 활용의 중요성

- 4차 산업혁명의 핵심기술 기반인 인공지능, 빅데이터, 사물인터넷(IoT) 등은 대규모 데이터의 고속 처리 및 저장·관리·분석과 이를 수행할 수 있는 인력과 컴퓨팅 환경이 중요
 - － 데이터 수집 및 실시간 전달(IoT-Mobile), 효율적 저장 및 분석(Cloud, Big Data)에서의 생산성 향상이 모든 산업분야의 핵심 경쟁력으로 대두
- Open Science 환경 연구개발 분야 또한 연구데이터를 수집, 저장, 처리, 분석 및 관리 하는 전문 역량과 제반 환경확보가 경쟁력 있는 연구 성과를 달성하는 핵심전략으로 인식



<그림 7-1> 연구데이터 활용을 통한 연구방식 사례

□ 연구데이터 활용 인력의 부족과 인재 양성 프로그램 미흡

- R&D에서도 데이터가 경쟁력이 되어가는 환경에서 국내 연구데이터 전문가는 산업 수요에 비해 양적으로 부족하고 질적으로도 크게 미흡하여 연구현장에서 활용 가능한 인력 확보에 어려움
 - － 국내의 경우, 단기교육에 기반한 기술자(Technician)를 주로 배출
 - － 2017년도 데이터 산업 총 인력은 294,753명이며, 그중 데이터 직무인력은 77,105명이 고 전체 인력에 대한 부족율은 5.9%(4,865명)으로 조사* 됨

- 부족인력 중 데이터 분석가는 7.6%(394명), 데이터 사이언티스트는 11%(158명) 임
 - * ‘2017 데이터 산업 현황 조사’(한국데이터 진흥원, 과학기술정보통신부)
- 미국 등 선진국은 R&D와 인력양성 동시 추진으로 고급인재 육성
 - ’18년까지 분석전문가 14만~19만 명, 관리인력 150만 명이 부족할 것으로 전망 (McKinsey, 2011). Data X 프로그램 등을 통해 학제간 교류, 공동연구로 세분화된 전문가 육성 중
- 기초과학 분야 컴퓨팅 및 데이터 인력양성을 위한 전문기관 극히 부족
 - 컴퓨팅(계산과학) 분야 전문 인력양성 전문기관 부족
 - 대학(서울대, 연세대, 서울시립대) 및 출연(연) (KISTI, KIST, 화학(연) 등) 계산과학 조직 보유

7.1.2 필요성

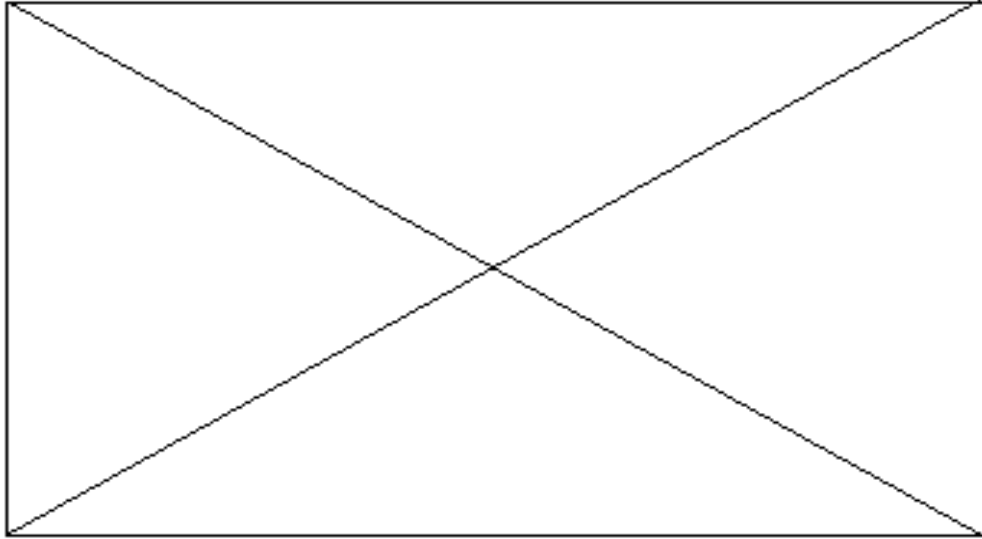
□ 연구데이터 활용 활성화를 위한 장기적 인재양성 프로그램 개발 시급

- 데이터에 기반한 연구 패러다임 변화에 맞춰 데이터 과학에 대한 전문지식과 활용 역량을 확보하기 위한 중장기 인재양성 프로그램 필요
 - 컴퓨팅과 데이터를 자유롭게 활용하여 연구데이터 수집, 활용을 체계적이고 효과적으로 수행할 수 있는 인력양성을 위한 교육과정 필요
 - 데이터 과학에 필요한 전문교육 프로그램을 체계적으로 운영하고 있는 타 연구(또는 교육) 기관과의 협력과 연계가 필요
- 연구데이터 수집, 활용을 체계적이고 효과적으로 하기 위하여 연구데이터 교육을 일원화하고 관련기관과의 연계를 위한 브랜드 구축으로 성과, 서비스, 활용 사례 공유 등 연구데이터 교육의 활성화 필요

□ 연구데이터 공유·활용 전문센터의 기반 강화를 위해 전문인력 확보 필요

- 인프라 운영, 데이터 관리 및 병렬분석 환경, 효율적 전송, 시스템 모니터링, 장애분석, 보안 등에 대응 가능한 실무형 관리자 양성 필요
- 연구데이터 및 일반 (빅)데이터 시장수요의 급증에 대비 ICT 기반의 데이터 개발인력의 수급*이 절대적으로 부족한 부문을 시급히 해결하기 위한 인재육성을 시장에서 요구
 - * 정보화진흥원의 조사 데이터 개발인력 공급률은 39.2% 이며, 데이터 엔지니어 공급률은 27.3%

- 데이터 기반 연구 및 전문센터의 지속적인 구축 예정에 따라 수요가 크게 증가할 것으로 예상되는 ICT 기반 데이터 관리·운영·서비스 인력 양성



<그림 7-2> 주요데이터 직무별 인력 현황 및 수요*

* **한국데이터진흥원(2015), 「2015 데이터산업 현황조사」

□ 연구데이터 활용 교육을 위한 콘텐츠 및 서비스 구축 필요

- 정제되지 않은 복잡한 데이터보다 교육 활용에 적합한 교육용 데이터를 분야별로 축적하여 제공할 수 있는 데이터뱅크 구축 및 서비스 필요
 - SW 기반의 교육 및 콘텐츠 활용을 위해 ICT기반의 교육환경을 구축하여 쉽게 언제, 어디서나 접근하여 활용할 수 있는 교육·연구 융합 환경*을 구축
 - * 교육과 연계해 연구 및 문제해결을 즉시 수행할 수 있는 환경
- 데이터·컴퓨팅 활용 교육·실습을 위해 수준별 교재를 개발하고 교육과정에 데이터 처리 SW*, 광범위한 교육용 데이터** 등을 구축
 - * 데이터 처리 S/W는 빅데이터 처리기술, 클라우드 기술 등의 데이터 수집/저장/운영하는 S/W
 - **교육용 데이터는 보고서·논문 등 연구개발 성과물, 국내외 공개데이터, 계산데이터 등
- 학생 및 연구자들이 쉽게 데이터 분석을 배우고 활용할 수 있도록 데이터 분석에 대한 접근성을 제고할 수 있는 교육, 실습 인프라 및 서비스 구축 필요

7.2. 환경분석

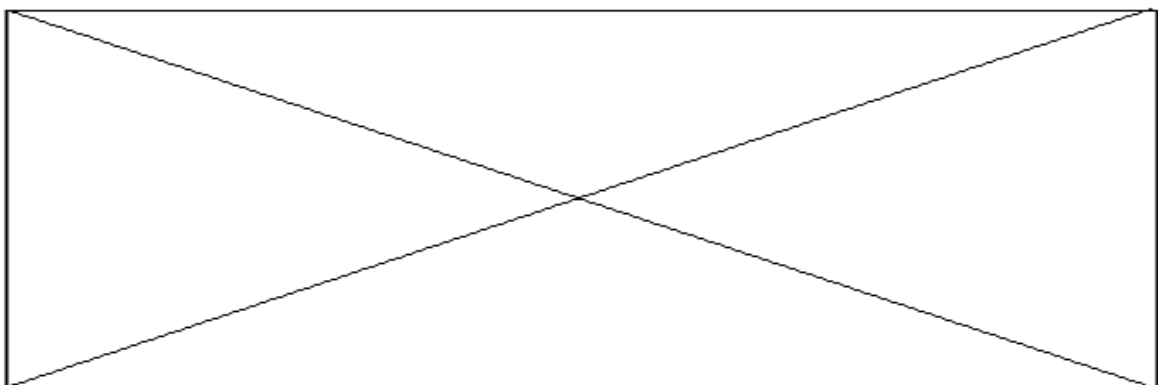
7.2.1. 연구데이터 활용 인재양성 프로그램

□ 연구 분야별 전문지식 중심 인재양성 과정

- 국가생명연구자원정보센터(KOBIC) 교육지원
 - － 국가연구개발사업 등에서 생성된 생명정보생명연구자원 정보를 연계 및 수집하여 생명 정보를 통합하고 분석할 수 있는 연구지원 서비스를 제공
 - － 이와 관련된 제반 교육과 BIT 기반의 새로운 융합 분야에 대한 교육
- UST의 교육시스템을 활용한 교육
 - － 학제 간 신생 융합 기술분야의 현장 경험 교육과 연구 활동을 통해 핵심·원천기술의 발전 과 산업기술 혁신을 선도하는 실천적이고 창의적인 인력양성
 - － 타 연구기관 또는 민간업체에서 수행 중인 과제에 참여하고 연구실에서 근무하도록 하는 Lab Rotation 제도를 시행하여 다양한 분야의 연구현장을 경험

□ 연구데이터 처리/분석을 위한 교육플랫폼 중심의 인재양성 과정

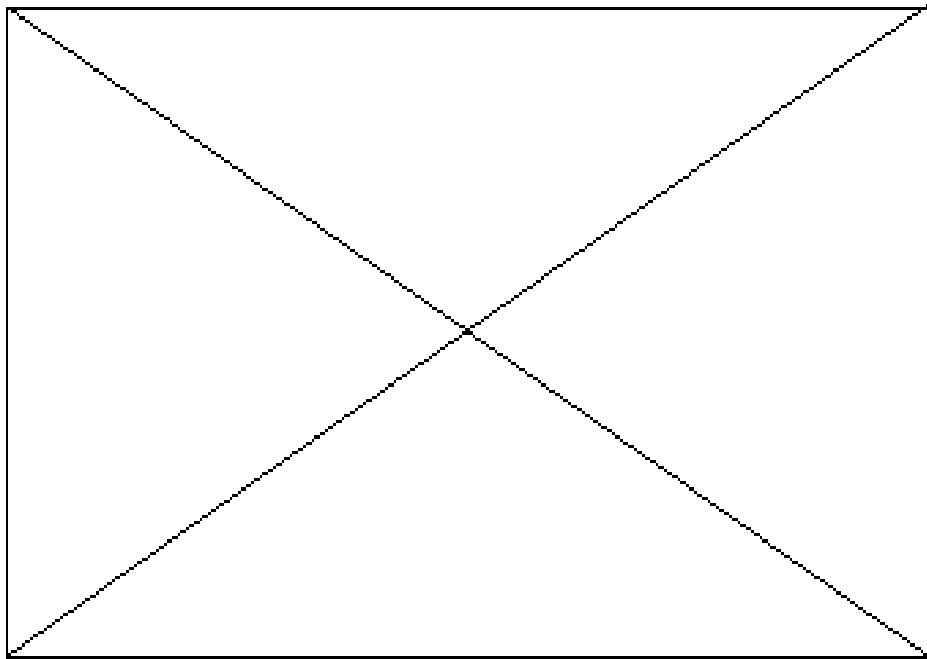
- 한국과학기술정보연구원의 과학데이터스쿨
 - － 국가차원의 연구개발 데이터 공유 활용 체제 구축에 따른 데이터와 컴퓨팅 활용 인재양성 으로 4차 산업 시대의 새로운 산업에 대응하기 위한 전문 인력 수요 대응
 - － 국가 슈퍼컴퓨팅 육성 기본계획에 따른 인력양성 체제 구축



<그림 7-3> 과학데이터스쿨(KSDS) 사업내용

○ 한국과학기술정보연구원 EDISON

- － 최신 과학기술 적응력 제고를 통한 계산과학공학 우수인재 양성을 위해 이공계 교육·연구용 계산과학공학 시뮬레이션 플랫폼 개발 및 국내 SW 실용화 기반 구축
- － 계산과학공학 플랫폼, 계산과학공학시뮬레이션 SW 및 콘텐츠, 전문 응용분야 계산과학공학 시뮬레이션 SW 및 콘텐츠 등을 제공하고 관련 교육시행



<그림 7-4> EDISON의 사업추진 체계 및 협력관계

7.2.2. (빅)데이터활용 인재양성 프로그램

□ 공공기관의 데이터 활용 인재양성 교육과정

○ 한국데이터진흥원 빅데이터 아카데미

- － 현업 DB전문가를 대상으로 체계적인 빅데이터 교육을 제공, 국내 빅데이터 기술 경쟁력 강화 및 글로벌 빅데이터 시장 선도를 목표로 운영
- － 빅데이터 프로젝트 진행 또는 예정인 산업계 재직자 또는 각 교육과정별 관련 경력 3년 이상을 대상으로 직무별, 유관 기관별, 지역별 프로그램 운영

○ 한국정보화진흥원(NIA) 공공데이터 교육 및 컨설팅 사업

- － 공공데이터혁신 우수사례 발굴·확산, 교육 및 컨설팅 등 공공데이터 생활화를 지원

- 공공데이터의 효율적인 제공을 위해 통합 제공 시스템인 ‘공공데이터포털’을 운영·관리하고, 국가중점 데이터 개방, 공공데이터 제공기반 구축·품질관리, 공공데이터를 활용한 창업과 사업화 등을 지원하여 국민들의 공공데이터 이용의 편의성 지원
- 대전 데이터 사이언티스트 아카데미(D-DS Academy)
 - 미래 산업 맞춤형 인재인 데이터 사이언티스트 교육을(Data Scientist) 통해 창의 융합 인재 양성을 목표로 운영
 - 인공지능 관련 개념 및 지식 습득, 공공기관 및 기업의 실제 데이터 기반 문제해결 프로젝트 수행, 커리어 코칭 및 전문가 인사담당자와 교류기회 제공 등의 프로그램으로 데이터 사이언티스트를 꿈꾸는 39세 미만의 청년을 대상으로 운영
- 정보통신기술진흥센터 혁신성장 청년인재 집중양성 프로그램
 - 인공지능, 빅데이터, AR/VR 등 4차 산업혁명 선도 8대 분야의 우수인재를 집중 양성하여 국가 기술경쟁력 강화 및 산업체 인력수요를 해소

□ 학계의 데이터 활용 인재양성 과정

- 고려대학교 빅데이터 사이언티스트 인력 양성 사업
 - 빅데이터로 부터 유용한 가치를 발굴하고 이를 기반으로 차세대 스마트 대한민국을 선도할 역량을 갖춘 글로벌-통합형-창의적 빅데이터 사이언티스트 인력 양성
- 연세대학교 빅데이터 (Strategic Management of Big Data Analytics : SMBA) 석사과정
 - 기업의 빅데이터 니즈에 맞추어 최적으로 설계된 프로그램. 계약프로그램은 기업에서 빅데이터 전문가를 키우려 할 때 가장 알맞은 프로그램으로서 기업업무를 가장 잘 아는 인재를 빅데이터 전문가로 양성
 - 다양한 분야의 전문가들과 외부 실무 전문가 들이 모여서 만들어진 융합 프로그램으로서 빅데이터에 대한 분석뿐 아니라 기업의 빅데이터 를 통한 의사결정에 도움이 될 수 있는 인재 양성이 목표
- 서울대학교 빅데이터 연구원 SNU BIG DATA ACADEMY
 - IT 계열 업무 담당자들에게 빅데이터 분석에 필요한 고급 분석 기술을 교육
 - 마케팅/고객 분석 업무 담당자들에게 data driven 마케팅을 위해 필요한 분석방법론과 R 프로그래밍 등을 교육함으로써 각 산업 분야에서 빅데이터 융합의 활성화를 도모

7.2.3. 해외선진 인재양성 프로그램

□ 미국

- XSEDE (The Extreme Science and Engineering Discovery Environment)
 - 국가 차원의 통합된 슈퍼컴퓨팅 교육 및 인력양성 시행과 현업에 활용 가능한 실무중심으로 대학(원)생, 일반인 대상 교육 수행을 통하여 과학적 발견을 가속화하는 디지털 연구서비스 생태계 구축을 목표
 - XSEDE 프로젝트는 일리노이 대학을 중심으로 시카고대학·UC버클리·버지니아 대학 등 총 18개 대학 및 연구기관이 협력하고 미국 전역의 7천여 연구수행자(그룹)에 슈퍼컴퓨터로 처리된 데이터 제공
- VSCSE (Virtual School of Computational Science and Engineering)
 - 전산과학 커리큘럼 개발 및 제공하는 가상 여름학교로 매년 10여개 대학 및 데이터 활용 슈퍼컴퓨팅센터 참여하고 화상회의를 통해 지리적으로 분산된 학습그룹을 가상교실에 소집하며 병렬컴퓨팅·매니코어 프로세스 알고리즘 등의 교육시행
- RRSCS (Ralph Regula School of Computational Science)
 - 대학(원)생, 중·고등학생을 대상으로 OSC(Ohio Supercomputing Center)에서 실시하는 HPC 인력양성 교육 프로그램으로, 오하이오주 14개 대학이 참여하고 있고 수료생이 산업계에 바로 도움이 될 수 있는 실무 중심의 슈퍼컴퓨팅 교육을 제공하고, 교과과정 수료 시 학위 부여 및 학위 연계 인정

□ 유럽

- EPCC (The University of Edinburgh, 영국)
 - 대학(원)생, 연구진을 대상으로 HPC 및 데이터공학 온라인 교육과정 및 HPC 파트너쉽 박사과정을 운영(EPSRC DTP)
 - HPC Short Courses for the UK를 타 11개 대학 및 1개의 연구소와 컨소시엄 구축하여 대학원생 및 연구원 대상 HPC 교육 프로그램(10일)을 운영
- PRACE (Partnership for Advanced Computing in Europe)
 - 대학(원)생, 연구진, 일반인(업계 전문가)을 대상으로 슈퍼컴퓨팅 교육 및 컴퓨터공학 인력양성 프로그램을 범유럽 차원에서 체계적으로 제공
 - 2017년 기준 벨기에, 영국, 스위스, 독일, 폴란드, 터키, 이스라엘 등 총 24개 국가 내

슈퍼컴퓨터 관련 기관이 참여 중

- 고급연수센터(PATC) 교육과정은 PRACE의 핵심 교육 프로그램으로서, 유럽의 컴퓨터 공학자에게 세계 최고 수준의 교육 및 훈련 기회를 제공하고 있고, 2017년 기준 6개 국가의 슈퍼컴퓨터센터에서 총 51개의 교육 프로그램 제공

□ 일본

○ HPCI (High Performanc Computing Infrastructure)

- HPCI는 지역 내로 제한되어 있는 HPC 사용자가 전국의 HPC 리소스를 효율적으로 활용할 수 있는 프레임워크 제공하여 컴퓨터계산공학 커뮤니티 확대 및 육성하여 성과를 극대화시키는 것이 목표
- RIKEN, 규슈대, 오사카대, 교토대, 나고야대, 도쿄대, 토호쿠대, 홋카이도대, 쓰쿠바대, ISM, TIT, JAMST 등 총 12개 대학 및 연구기관이 참여
- HPCI 웹사이트에서 K 슈퍼컴퓨터 사용 가이드라인을 제공하며, 정기 교육프로그램보다는 비정기적 워크숍·계절학교 운영으로 교육을 진행

○ 일본 문부과학성

- 차세대 슈퍼컴퓨터 인재육성 방안 중 총 5가지 유형(고도계산공학(2), 고도계산기공학, 계산공학중개, 계산공학지원)의 계산과학(HPC) 인재 육성 방안을 제시하며 슈퍼컴퓨터 이용 및 시스템 개발 능력이 뛰어난 인재를 육성하기 위한 생태계 조성에 주력

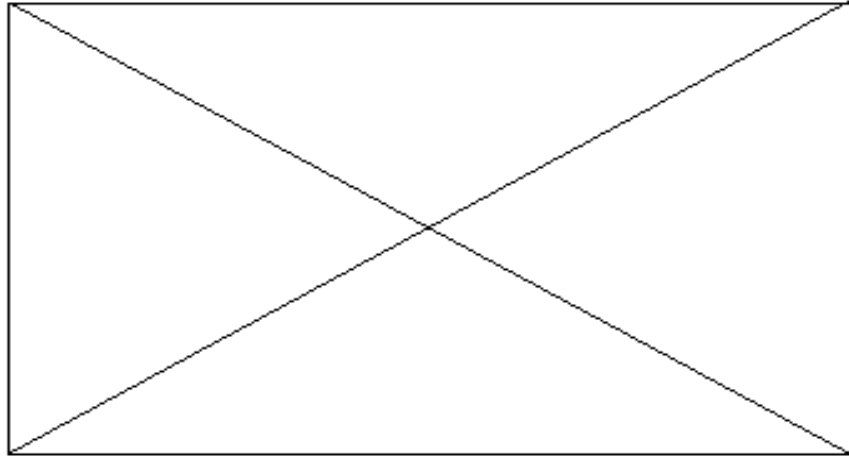
□ 중국

○ National High Performance Computing Center (NHPCC, 중국국방과기대)

- 학부생 및 석·박사 대상 병렬 알고리즘 프로그래밍 전문 과정 운영 중이며 병렬 알고리즘 이론, 알고리즘 설계 및 분석, 알고리즘 구현 및 알고리즘 적용 등 4단계 과정으로 구성
- 중국국방과기대(USTC)는 컴퓨터공학 학부, 석사, 박사 등 3가지 학위 과정을 제공하고 국내외 대학 및 연구소와 연계하여 석·박사 과정 운영

7.2.4 연구데이터와 (빅)데이터 활용의 절차적 특징

□ R&D 연구수행의 LifeCycle*



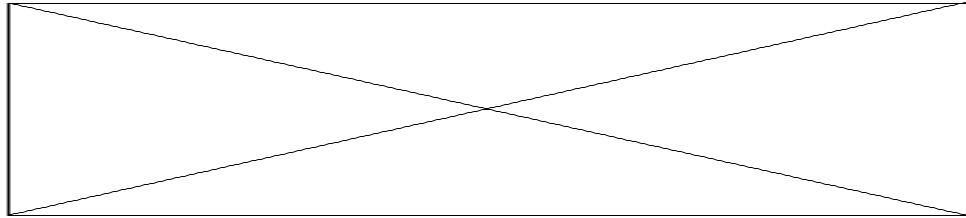
<그림 7-5> 한국과학기술 R&D전주기 과정

* *과학기술분야 R&D 전주기 연구 - 국내 생명 및 나노과학기술 연구자를 중심으로 - 참조

- (아이디어 생성 및 개발) (1) 아이디어 생성 및 초기가설 개발, (2) 실험기자재 파악, (3) 실행가능성 테스트 단계 등의 연구 활동으로 구성
- (연구지원비 확보) (1) 연구실행을 위한 연구지원비 파악, (2) 연구팀구성, (3)연구지원비 확보를 위한 연구계획서 작성
- (실험 및 분석단계)연구지원비 확보와 함께 연구프로젝트에서 제안한 연구 아이디어를 본격적으로 실험하는 연구실행단계로, 크게 (1) 실험과 (2) 결과 분석 순으로 진행
- (성과창출 단계)연구결과보고서, 원천기술, 특허출원, 학술논문, 상품화 등의 성과물 창출
- (평가)연구과제의 종료시점과 함께 연구제안서에 약속한 연구 성과물을 제대로 산출했는지 그 결과를 구두 발표와 서면으로 연구비 지원 기관에 보고

□ 연구데이터 생성 및 활용 절차에 필요한 역량 및 교육과정

- 연구데이터 생성 및 활용 시 데이터 수집, 저장, 분석 및 운영의 상세절차



<그림 7-6> 연구데이터 생성 및 활용 LifeCycle

- 연구데이터는 R&D 절차중 실험 및 분석단계에서 시뮬레이션, 실험, 관찰을 통해 데이터가 생성되고 수집, 저장, 분석, 시각화 등의 절차에 따라 활용

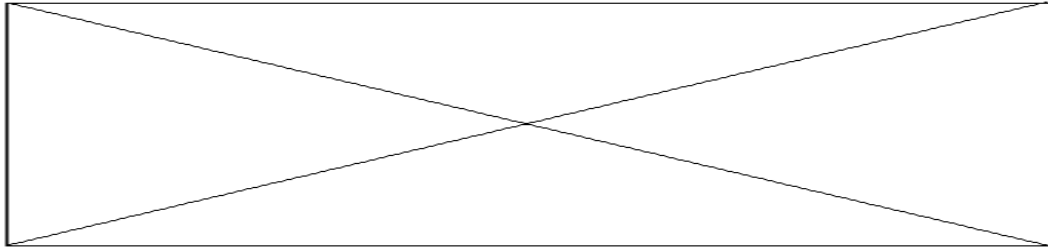
<표 7-1> 연구데이터 활용절차에 필요한 역량 및 교육과정

활용절차	역량	교육내용*
연구기획	연구데이터 정책 및 적용사례와 기술환경을 이해하고, 연구 및 융합연구 활용분야와 기술동향에 대한 지식을 기반으로 빅데이터 아키텍처 설계 및 향후 발전 방향에 대한 전략 도출	• 연구데이터 정책 및 연구과정의 이해 • 연구데이터 플랫폼이론 • 연구데이터 아키텍처 구성
연구데이터 수집	연구데이터 플랫폼에 대한 지식과 연구데이터 수집유형, 수집방법, 수집기술을 이해하고, 사용할 수 있는 역량	• 연구데이터 플랫폼 이론 • 연구데이터수집
연구데이터 저장	연구데이터를 관리할 수 있는 플랫폼 특징에 대한 지식과 저장유형, 저장방법, 저장기술을 이해하고, 사용할 수 있는 역량	• 연구데이터 플랫폼 이론 • 연구데이터 저장 및 관리 • 연구데이터 플랫폼 구축 및 활용
연구데이터 분석	연구데이터 분석 Tool을 활용하여 예측모델, 분석모델, 최적화 모델 등을 설계. 분석모델을 이해하고, 분석모델 결과를 해석하여 연구결과의 의미를 부여할 수 있는 역량	• 연구데이터 마이닝, 비정형 연구데이터 마이닝 • 연구데이터 예측분석 • 연구분야별 전문지식, 전문연구분야 Analytics • 분석 모형 및 성과평가
연구데이터 시각화	연구데이터 분석모델 결과를 해석하고, 논리적으로 표현할 수 있는 역량. 시각화 도구를 이용하여 분석결과를 화면으로 구현할 수 있는 역량	• 논리적 자기표현 • 기초 통계이론 • 연구데이터 시각화 프로그래밍
연구데이터 운영	연구데이터 플랫폼을 이해하고, 백업 및 복구기술 등을 적용하여 시스템의 안정적 운영을 유지할 수 있는 역량	• 연구데이터 플랫폼이론 • 연구데이터 플랫폼구축 및 활용 • 연구데이터 아키텍처

* 교육내용은 빅데이터역량 모델(NIA:BCM)과 UST 등의 교육과정을 참조

□ (빅)데이터 활용 절차에 필요한 역량 및 교육과정

- 일반 (빅)데이터 활용을 위한 데이터 수집, 저장, 분석 및 운영 절차



<그림 7-7> 일반 (빅)데이터 활용 LifeCycle

- 데이터 및 빅데이터를 구축하고 활용하기 위해서는 기획단계에서부터 수집, 저장, 분석, 시각화 등의 절차에 따라 활용 및 운영

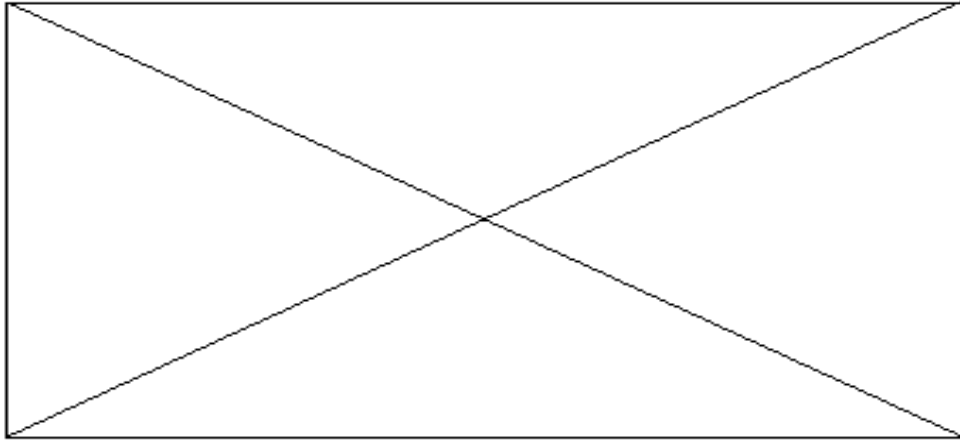
- 일반 (빅)데이터 활용 시 필요 역량과 교육내용

<표 7-2> (빅)데이터 활용절차에 필요한 역량 및 교육과정

활용절차	역량	교육내용*
환경분석/ 데이터 기획	빅데이터 정책 및 적용사례와 기술환경을 이해하고, 비즈니스 활용분야와 기술동향에 대한 지식을 기반으로 빅데이터 아키텍처 설계 및 향후 발전 방향에 대한 전략 도출	• 빅데이터 정책 및 비즈니스 이해 • 빅데이터트랜드, 빅데이터 플랫폼이론 • 빅데이터 아키텍처 구성,
데이터 수집	빅데이터 플랫폼에 대한 지식과 빅데이터 수집유형, 수집방법, 수집기술을 이해하고, 사용할 수 있는 역량	• 빅데이터 플랫폼 이론 • 빅데이터수집
데이터 저장	빅데이터 플랫폼에 대한 지식과 저장유형, 저장방법, 저장기술을 이해하고, 사용할 수 있는 역량	• 빅데이터 플랫폼 이론 • 빅데이터저장 및 관리 • 빅데이터플랫폼 구축 및 활용
데이터 분석	통계 및 수학지식을 기반으로 통계패키지를 활용하여 예측모델, 분석모델, 최적화 모델 등을 설계. 분석모델을 이해하고, 분석모델 결과를 해석하여 통계적 의미를 부여할 수 있는 역량	• 데이터마이닝 • 비정형데이터마이닝 • 빅데이터 예측분석 • 업종 특화지식, Business Analytics • 분석 모형 및 성과평가
데이터 시각화	빅데이터 분석모델 결과를 해석하고, 논리적으로 표현할 수 있는 역량 시각화 도구를 이용하여 분석결과를 화면으로 구현할 수 있는 역량	• 논리적 자기표현 • 기초 통계이론 • 빅데이터 프로그래밍
데이터 운영	빅데이터 플랫폼을 이해하고, 백업 및 복구 기술 등을 적용하여 시스템의 안정적 운영을 유지할 수 있는 역량	• 빅데이터 플랫폼이론 • 빅데이터 플랫폼구축 및 활용 • 빅데이터아키텍처

* 교육내용은 빅데이터역량 모델(NIA:BCM)을 참조

□ 연구데이터와 (빅)데이터 활용의 필요역량의 차이



<그림 7-8> 일반데이터와 연구데이터 활용 역량의 차이영역

- 데이터 분석절차와 일반연구수행 절차를 비교하면, 절차 및 처리 프로세스 간에는 큰 차이가 보이지 않으나, 연구수행 시에는 데이터 분석과정에서 해당분야의 전문지식이 필요
 - － 수집대상 데이터가 Text데이터 중심의 논문정보가 많다는 점은 분야별 연구데이터 분석 및 시각화 과정에서 데이터 분석의 특화된 역량이 요구됨
- 데이터의 수집, 연계, 저장, 운영은 일반데이터와 연구데이터의 클라우드, 빅데이터 처리 등의 기술적 차이는 없으나 표준화, 품질 측면에서 데이터 분야의 전문적인 지식은 필요

□ 연구데이터와 (빅)데이터의 교육기관 차이

<표 7-3> 연구데이터와 빅데이터 교육기관 비교

단계	기획	수집	저장	전 처리	분석	시각화	운영
연구데이터 활용 교육	×	×	×	KISTI의 KSDS	KISTI의 KSDS, EDISON	KISTI의 KSDS, EDISON	×
(빅)데이터 활용 교육	빅데이터 아카데미 등	SDS멀티 캠퍼스 등 기술교육 위주의 전문교육기관* 중심		—	빅데이터 아카데미, 한국SAS 등 전문기관 중심	연세대학교, 고려대학교 등 연구조직 중심	LGCNS, SK C&C, 삼성SDS 등 IT 전문기업 중심

- 연구데이터 활용은 빅데이터 활용과 비교할 때 전 처리, 분석, 시각화에서 전문 역량이 요구되고 있고, 교육기관 또한 연구데이터 활용교육은 전처리, 분석, 시각화에 집중되어 있음
- 전문교육기관은 데이터 전 과정을 기관의 목적이나 사업영역에 맞게 제공. 대표적으로 공공기관인 한국DB진흥원외에 삼성SDS멀티캠퍼스, 한국SAS, 한국EMC를 비롯 DW/BI 솔루션 업체와 시각화Tool 솔루션 기업

<표 7-4> 전문교육기관의 데이터 교육현황

구분	데이터수집	데이터처리	데이터분석	데이터시각화	데이터운영
한국DB진흥원		○			○
한국EMC					○
한국SAS			○	○	
삼성SDS멀티 캠퍼스	○	○	○	○	○
DW/BI솔루션업체	○	○	○	○	
시각화Tool업체				○	

7.2.5 시사점

□ 연구데이터 활용 전문인력 공급측면

- 전반적인 데이터 전문 인력은 수요에 비해 공급이 부족한 실정이고, 특히 오픈사이언스 환경하의 R&D 혁신을 위해서는 체계적이고 집중적인 연구데이터의 전문인력 양성이 필요
- 해외의 경우 기업이 보유하는 실 데이터를 활용한 교육과 연구데이터 실험실을 이용한 실무 중심의 교육으로 전문교육기관, 정부연구기관, 대학 연구소 중심으로 활발한 인력 육성 체계를 구축

□ 교육과정 및 교육기관의 운영측면

- 일반 (빅)데이터 활용 교육과정은 기본과정부터 고급과정까지, 데이터 구축부터 운영까지 전 과정이 기업체, 학교 등의 기관 중심으로 개설되어 활발하게 인력을 육성
- 연구데이터 활용 교육과정은 특성상 분석을 위한 전처리, 분석, 분석을 위한 시각화에 집중되어 있고 교육과정 및 교육기관도 일부 해당 연구 분야의 기관 및 전문 연구기관 중심으로 인력 육성을 수행

- 연구데이터 활용 역량제고를 위해서는 실험환경을 통해 실무중심의 체계적인 교육과정을 개설이 필요하고, 교육과정 발굴과 교육운영을 위한 전문 조직운영이 필요

□ 교육을 위한 인프라 및 서비스 측면

- UST, KISTI 등의 과학분야 연구 및 교육 기관은 연구데이터를 활용하기 위한 교육 및 실험 등을 가능하도록 하는 연구데이터 활용 플랫폼* 중심으로 서비스를 구축 및 운영, 이를 기반으로 특정분야의 커뮤니티를 형성에 지속적으로 협업 및 융합 연구를 할 수 있는 환경을 구축

* 연구데이터 활용 플랫폼 : 분석도구 등을 자동으로 서버에 탑재할 수 있는 인프라를 기반으로 공학 시뮬레이션 S/W, 콘텐츠를 제공. 연구자가 사이버 연구 환경에서 활발히 연구 활동을 수행하고 연구자간 협업 및 융합연구도 가능하도록 하는 기능제공

- 연구데이터 수집/저장/운영 등 일반 (빅)데이터 관리와 동일한 역량이 필요한 영역에서 데이터를 다루는 실무 기술 교육은 연구기관 및 연구자에 대해 미흡한 실정
- 교육용 데이터는 각 출연연, 대학연구소 등 개별 기관 또는 연구자 중심으로 관리되고 있으나 체계적인 교육과정과 병행해서 구축 운영되지는 못함
- 해외의 경우 프로젝트를 통한 분석결과를 오픈 소스 소프트웨어 플랫폼을 통해 제공

7.3. 추진전략

7.3.1 추진방향 및 목표

가. 추진방향

□ 연구데이터 활용인력의 수요 및 활용 유형을 분류하여 인재육성의 방향수립

- 데이터 기반 R&D 응용 및 ICT 기반 데이터 운영·관리 분야 등으로 인재육성 필요
 - － ICT기반 데이터 관리 및 운영 분야는 활용기반 강화를 목적으로 초·중급 중심으로 인력을 육성, R&D응용분야는 성과창출을 목적으로 석·박사 중심의 고급 인력을 육성하는 인재양성 체계수립

<표 7-5> 연구데이터 활용 유형과 교육분야

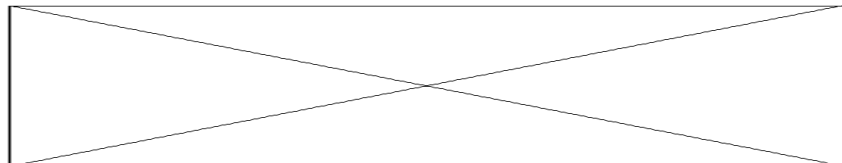
구분 \ 유형	활용 기반강화 형		성과창출 형	
	산업체	대학	국가연구기관	연구자
R&D 응용*	공통**	공통/응용	공통/응용	공통/응용
기반 운영 및 개발	공통/시스템			

* 응용은 주요 도메인 분야 응용·문제해결을 위한 역량으로 알고리즘, 수학적 모델링, 수치해석 등

** 공통은 컴퓨터 및 정보과학으로 계산 및 메모리장치, 대용량 데이터처리, 초고속 연결네트워크 등

□ 연구 Domain의 전문지식과 데이터 처리 및 분석 역량을 동시에 갖춘 고급 인재육성을 목표

- R&D인재 육성의 최종 방향은 Domain별 데이터분석이 가능한 고급분석가를 양성
- 데이터분석 전문 연구인력이 보유해야 할 역량은 빅데이터에 기반한 데이터분석의 기본 지식과 해당 연구 분야 별 Domain 업무지식의 보유

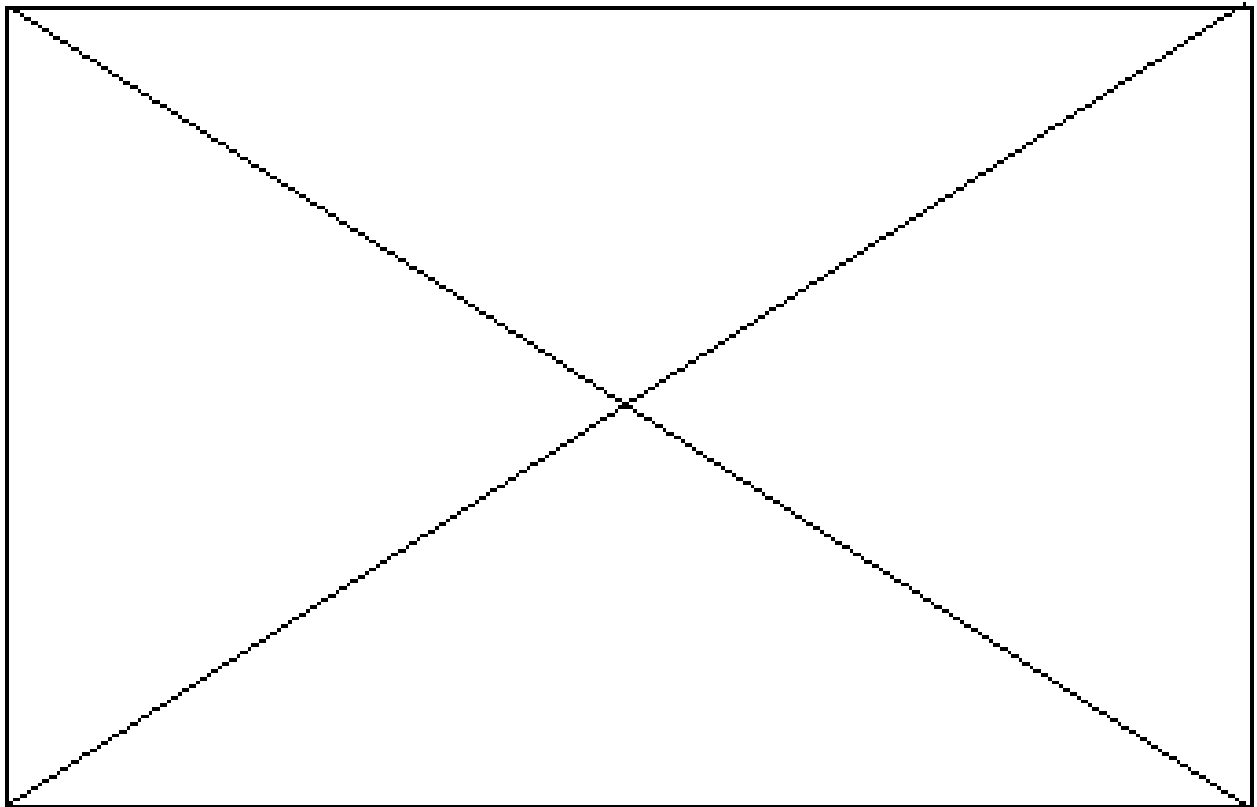


<그림 7-9> 연구데이터 전문활용 역량

□ 교육대상자의 범위 및 역량에 따라 차별화된 교육과정 수립

- 활용기반을 강화하거나 연구데이터에 대한 기본지식과 역량을 제고 하기위해서 데이터 전문교육기관의 교육과정을 활용
- 전문 Domain 지식을 활용하여 성과를 창출하는 고급 인력에 대해서는 연구분야 데이터 분석이 가능한 고급 분석가 양성을 전문적으로 수행하는 연구데이터 전문양성과정 활용

나. 추진목표



□ 연구데이터 수집/저장/운영 등 ICT기반 활용역량 강화

- 국가기관, 학계, 산업계, 공공분야의 초·중급 연구원 및 기술자 대상 연구데이터 수집/저장/운영 등에 관련 된 데이터 활용을 위한 공통 기술교육

□ 연구데이터 전처리/분석/시각화 등 분야별 R&D 응용역량 강화

- 국가기관, 학계, 기업 및 공공분야 연구소 등의 석·박사 등 고급인력 대상 연구데이터 전처리/분석/결과 시각화 등에 관련 된 분야별 전문 활용교육
- 교육 진행을 위해 데이터 교육지원센터를 운영하고 교육과목, 피교육생 교육진행을 관리 하는 지원시스템 구축

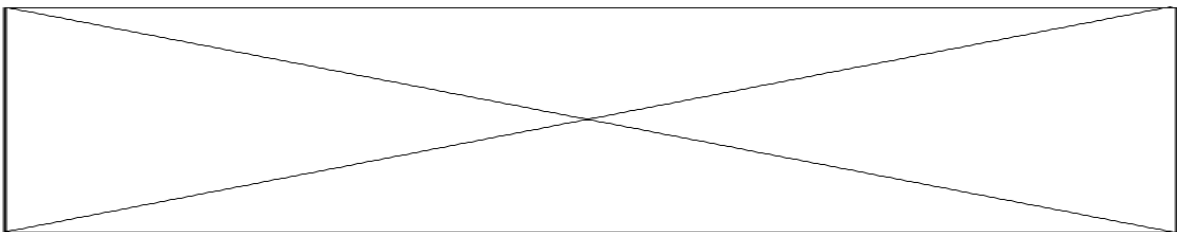
□ 연구데이터 활용교육을 위한 교육 인프라 및 서비스 구축

- 교육용 데이터 리포지토리와 동영상 강의와 연계된 시뮬레이션 SW 및 데이터 분석툴, 연구데이터 처리 및 분석을 위한 S/W 기술 교재 등을 기존 계산과학공학 플랫폼인 EDISON에 탑재하여 활용

7.3.2 추진전략 및 세부과제

가. 추진전략

- 전략1. 연구데이터와 (빅)데이터 활용 및 운영 과정을 통합적 관점에서 설계하고 연구데이터와 컴퓨팅을 활용하는 직무와 역량을 정의
- 연구데이터 활용 절차와 (빅)데이터 활용절차를 유사한 분야 통합관점에서 7개의 연구데이터 및 컴퓨팅 R&D Task*를 설계



<그림 7-10> 연구데이터 및 컴퓨팅 활용 R&D Task

- * 시뮬레이션, 실험, 관찰 등은 데이터 생성단계로 생성이후 수집단계부터 연구데이터 활용단계
- 7개의 연구데이터 및 컴퓨팅 R&D Task의 상세 Task를 도출하고 연구데이터 관리자, 연구데이터 운영자, 연구데이터 분석가로 직무를 구분

<표 7-6> 연구데이터 활용 직무 및 상세 Task

직 무	연구데이터 활용 R&D TASK	상세 Task
연구데이터 관리자	연구데이터 수집	데이터수집
		데이터표준화
	연구데이터 저장	데이터저장
		데이터통합
연구데이터 운영자	연구데이터 운영	시스템운영
		데이터용량관리
연구데이터 분석가	연구데이터 전 처리	요건정의
		분석모델링
	연구데이터 분석	모델구현
		결과해석
	분석결과 시각화	데이터표현
		시각화구현

□ 전략2. 연구 분야별 분석활용 교육은 전문 교육인프라 활용

- 7개의 연구데이터 및 컴퓨팅활용 R&D Task 중 연구데이터 전처리 --> 연구데이터 분석 --> 분석결과 시각화 영역의 교육
- KISTI 과학데이터 스쿨(KSDS) 교육 인프라 활용
 - 과학기술분야에서의 연구데이터·컴퓨팅 활용을 위한 전문 교육 프로그램 제공
 - 연구데이터 및 HPC 활용 전문 인력 양성을 위한 교육 전담 조직 설치·운영
 - 빅데이터, 사물인터넷, 딥러닝 등 교육 프로그램 개발, 국내외 전문 교육기관(학교, 산업체, 학회)과의 협력을 통한 커리큘럼 연계와 활용
 - 연구데이터·컴퓨팅 활용 온라인 교육 환경(KSDSweb) 및 콘텐츠(KSDSbank) 구축
 - 연구데이터 및 HPC 교육을 위한 통합 웹사이트 구축
 - 교육 콘텐츠(KSDS_MOOC 포함), 강사진, 온라인 강좌 데이터베이스 및 리포지토리 구축
 - 청소년, 대학생, 일반인 대상 국민 참여 교육 프로그램 제공
 - 청소년 캠프 및 경진대회: 국가슈퍼컴퓨팅 청소년 캠프, 국가슈퍼 컴퓨팅 경진대회
 - 대학생/일반인 캠프: HPC DATA SCHOOL, AI School, 해커튼 대회

<표 7-7> KSDS 교육과정 사례

분 야	과 정
시스템	리눅스 시스템, 슈퍼컴퓨터 시스템
최적화	최적화, 디버깅, 수치해석 라이브러리 활용
병렬 프로그램	Vectorization, 병렬 컴퓨팅, GPU, OPENMP, MPI, OpenCL, Hybrid (OpenMP+MPI)
응용 S/W	ABAQUS, Autodesk Moldflow, EDEM, Ansys CFX, Ansys, Fluent, CFX, SC/Tetra, LS-Dyna, Solidworks, 가시화, ANSYS Workbench, 슈퍼컴퓨터 활용, International School, Summer School, Intel SW College, HPC School (PLSI), HPC School
프로그래밍 언어	Python, Fortran, PGAS, Coarray, UPC
기타	KISTI 내/외부 교육, 비정기 교육

- UST 교육 인프라 활용
 - UST는 단순 지식 전달형의 수동적인 교육이 아닌, 현장 연구 중심의 자기 주도형 교육을 통하여 창의적 인재를 양성하는 것을 목표
 - 다학제적 성격의 전공을 중점적으로 개설하고, LAB ROTATION 등을 통해 다양한 분야의 연구현장을 경험하도록 함

– 강의식 교육과정을 최소한으로 줄이고, 연구현장 참여*를 통해 학점을 이수하도록 하는 연구현장 중심교육

* 지도교수의 연구실 외에 타 연구기관 또는 민간업체에서 수행 중인 과제에 참여하고 연구실에서 근무하도록 하는 Lab Rotation 제도를 시행하여 다양한 분야의 연구현장을 경험할 수 있도록 함

– 연구현장의 리더로서 필요한 소양과 사회 전반에 관한 안목을 기르기 위한 교양과목*을 필수적으로 이수하도록 함

* 개설과목: 기술경영학, 기술경제학, 기술정책학, 연구기획관리론, 벤처창업론, Technical Communication, Technical Writing, Project Management 등

□ 전략3. ICT기반 데이터관리 기술교육은 전문교육기관을 활용하는 위탁교육

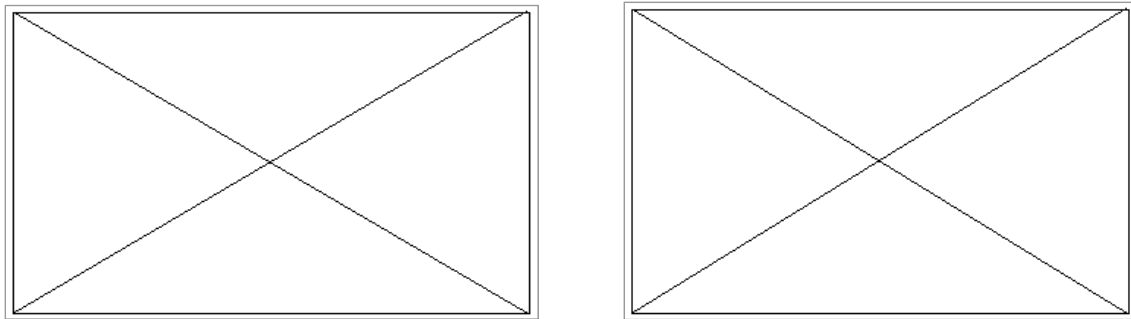
- 7개의 연구데이터 및 컴퓨팅활용 R&D Task 중 데이터 수집 --> 데이터 저장 --> 데이터 운영 영역의 교육
- 전문교육기관 교육인프라 활용
 - 빅데이터의 개념에 대한 교육과 데이터 수집, 처리, 분석, 시각화, 데이터 운영에 이르는 전 과정을 제공하는 기관 활용
 - 대학의 경우 국내 15개 대학에 대한 빅데이터/데이터 교과과정 중 교과과정의 주요 사례에 부합하는 대학 중심으로 교과과정 활용

<표 7-8> 데이터 관리기술 전문교육기관 사례

전문교육기관		데이터수집	데이터저장	데이터운영
공공 기관 및 기업	한국DB진흥원		○	○
	한국EMC			○
	삼성SDS멀티 캠퍼스	○	○	○
	DW/BI솔루션업체	○	○	
대학	정보미디어경영대학원테크노MBA(한국과학기술원)		○	
	빅데이터(SMBA) 석사과정 (연세대학교)	○	○	○
	컴퓨터정보통신대학원빅데이터융합학과(고려대학교)	○	○	○
	정보통신대학원빅데이터 학과(성균관대학교)	○	○	○

□ 전략4. 계산 과학공학 플랫폼인 EDISON의 교보재 체계 및 인프라 활용

- 교육·연구용 계산과학공학 플랫폼을 보급하여 ‘11년부터 웹 기반 EDISON 플랫폼을 활용하여 7개 계산과학공학 분야(전산열유체, 나노물리, 계산화학, 구조동력학, 전산설계, 전산의학, 도시환경) 인력양성 중
- － 슈퍼컴퓨터 활용, 입출력 자동화, 시각화 등 핵심기술 탑재 및 공개하여 국내외 51개 대학, 417종 SW 개발, 1,237개 교과목에서 4만 여명 활용 중



<그림 7-11> EDISON 교보재 사례

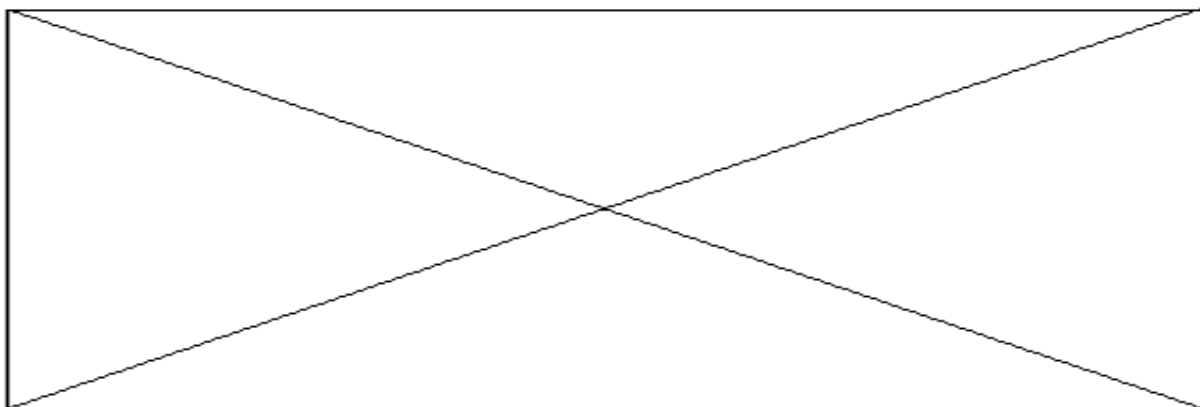
- 계산·데이터·인공지능을 융합하여 학습할 수 있는 EDISON-AI 플랫폼 개발
 - － 데이터 관리/처리/분석/시각화 기술 개발 및 기계학습 기능 탑재
 - － 소재 분야 50여만 건 데이터 탑재 및 기계학습 활용 신소재 개발 지원하고 기계, 항공, 화학 등 과학기술 전분야로 확대 추진
- 언제, 어디서나 기계학습을 활용하여 데이터를 직접 분석·예측·학습할 수 있는 플랫폼 제공
 - － 계산 + 데이터 + 인공지능을 융합 및 활용 가능한 인재 양성

<표 7-9> EDISON 교육플랫폼의 환경

구분	내용
활용환경	실시간 데이터 분석용 웹 기반 iPython notebook 개발환경 제공
분석 알고리즘	• 공개 기계학습 툴(Tensorflow, Pandas, Keras, H2O 등) • 기계학습 알고리즘(CNN, RNN, SVM 등) 지원
데이터	• 소재물성 및 제조업 등의 기초, 공학분야 데이터 • 국내외 공개 데이터, 전문연구 데이터 및 문헌 추출 데이터 등
인프라	• 기계학습에 최적화된 기계 학습 인프라 구축(GPU 클러스터)

나. 세부과제 구성

☐ 세부과제 도출



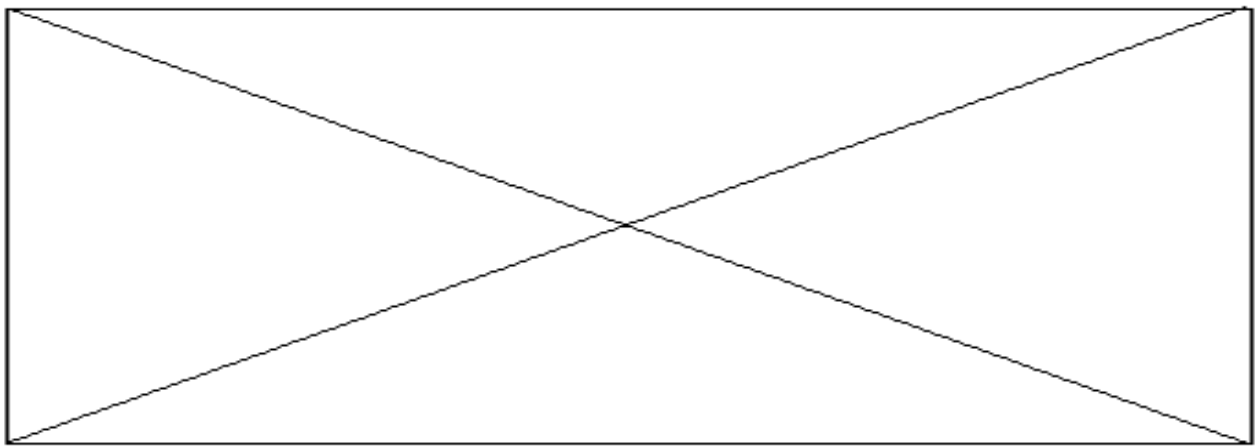
☐ 세부과제 구성

구분	세부 과제	개 요
교육과정	연구데이터 관리를 위한 ICT기반 기술교육	연구데이터 수집, 저장, 운영 등 데이터 관리를 위한 수집/저장 등의 IT 기술 교육
	연구데이터의 R&D 응용을 위한 분야별 전문 활용교육	연구데이터 전처리, 분석, 시각화 등 연구분야별 전문 분석 활용을 위한 교육
교육환경	연구데이터 기반기술 및 전문활용 교육용 데이터뱅크 및 서비스 구축	교육과정을 지원하기 위한 교육용 데이터, 교육용 기술 S/W 등을 교육과정에 맞게 구축

7.4 세부과제별 추진 내용

7.4.1 연구데이터 관리를 위한 ICT 기반기술 교육

가. 교육목표



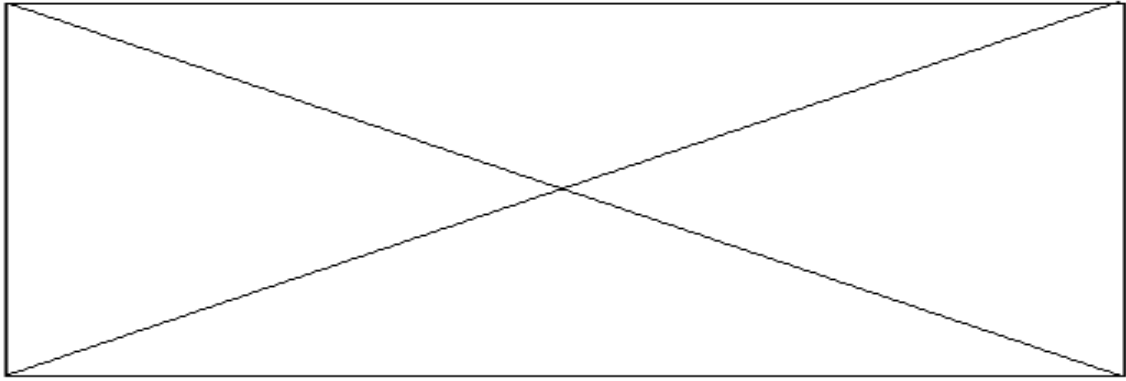
나. 교육 대상 및 범위

□ 교육대상

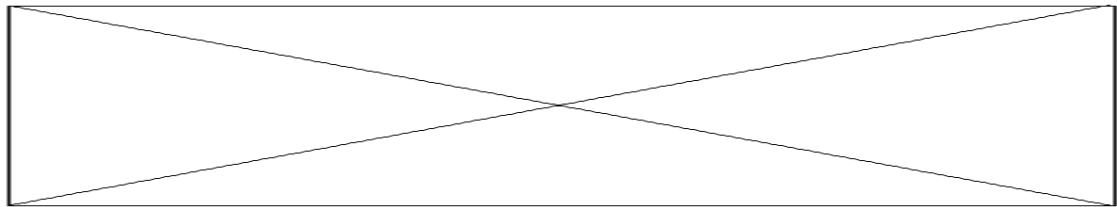
- 국가 연구기관 및 대학등 학계의 초·중급 연구원, 기업등 산업계의 산업기술 개발인력
공공분야 연구인력을 대상으로 초·중급 데이터 관리 및 운영 인재육성

<표 7-10> ICT 기반기술 기관별 교육 대상

국가 연구기관	대학 등 학계	기업 등 산업계
석·박사급 고급인력	석·박사급 고급인력	산업기술 개발인력
데이터처리 중급연구원	데이터 처리 중급연구원	공공분야 연구인력
데이터 처리 초급연구원	데이터 처리 초급연구원	



□ 교육범위



<그림 7-12> 연구데이터 활용 Task 중 ICT기반기술 교육 범위

- 연구데이터 및 컴퓨팅 활용 R&D Task 중 데이터 관리영역인 연구데이터 수집과 연구데이터 저장, 데이터 운영영역인 연구데이터 운영 Task에 대한 교육

직 무	연구데이터 활용 R&D TASK	상세 Task
연구 데이터 관리	연구데이터 수집	데이터수집
		데이터표준화
	연구데이터 저장	데이터저장
		데이터통합
연구 데이터 운영	연구데이터 운영	시스템운영
		데이터용량관리

<표 7-11> ICT 기반기술 교육 교육범위

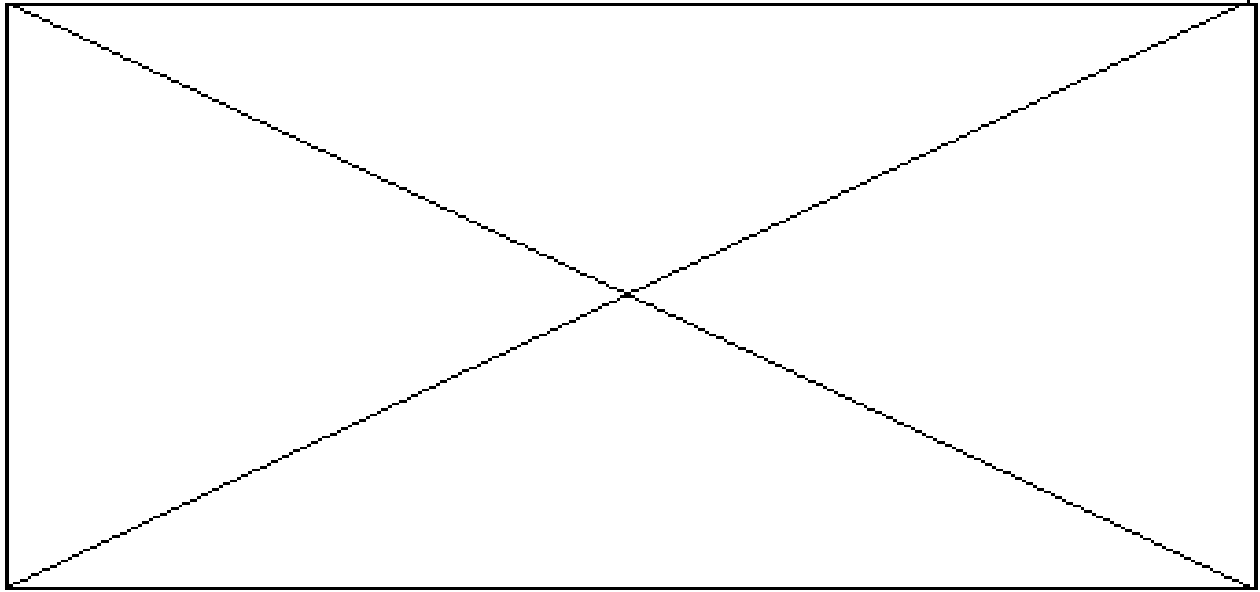
다. 교육내용 및 교육방법

□ 교육 내용

구분		교육 내용
직무	단계	
데이터 관리자	기본	빅데이터 플랫폼과 하둡 이해
		정형/비정형 데이터 수집 및 수집기술의 이해
		빅데이터DB의 이해
		하둡 분산파일 시스템 구조와 활용
	심화	빅데이터 기반의 데이터처리도구 이해와 활용
		하둡 에코시스템 설치와 활용
		데이터표준화 및 메타데이터관리 솔루션 이해와 적용
		빅데이터 아키텍처 설계방법론
		DW 이해와 구축
데이터 운영자	기본	빅데이터 보안분석
		빅데이터 환경의 개인정보보호대책
		빅데이터 플랫폼과 하둡 이해
	심화	하둡 에코시스템 설치와 활용
		빅데이터 아키텍처 설계방법론

□ 교육 방법

- 데이터분석을 위한 기본 지식(데이터 수집, 저장, 운영 등)과 스킬에 대해서는 전문교육기관의 교육과정을 이용할 수 있도록 전문기관을 활용한 위탁교육을 수행
- 교육 지원 내용
 - － 지원 내용 : 1인당 연간 최대 100만원 이내 데이터 교육 강좌 수강료 지원
 - － 지원 방식 : 쿠폰 지급 또는 전자 바우처 등
- 바우처를 활용한 데이터 교육수행 절차



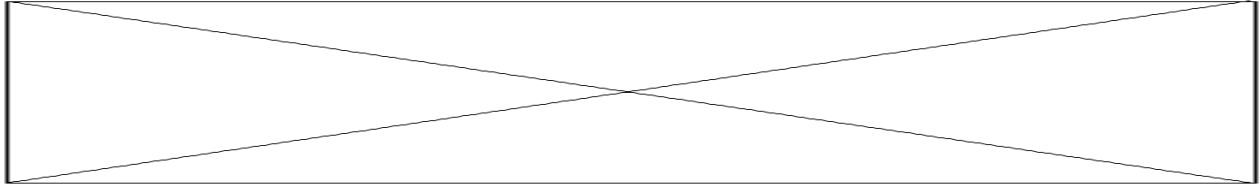
<그림 7-13> 바우처를 활용한 교육수행 절차(체크카드 활용 예)

- 단계별 상세 수행내역

<표 7-12> 바우처를 활용한 이용절차

단계	주체	상세 수행내역
① 바우처지원시스템 가입및지원신청	바우처 이용자	최초1회에 한하여 데이터교육운영 지원시스템 가입 ※ 개별가입이 불가능한 지원자의 경우 데이터교육운영센터를 통해 가입·지원신청 • 연간 지원한도 및 지원방식 등 안내·확인 • 바우처 지원신청(정보제공 동의포함) ※ 시스템을 통한 개별접수, 데이터교육운영센터를 통한 접수 등 • 가입자 연간 교육계획입력(선택)
	데이터 교육기관	최초 1회에 한하여 데이터교육운영 지원시스템 가입 • 바우처사용기관등록 • 제공강좌정보입력
② 자격확인 등 지원여부확정	데이터교육운영센터 (가칭)	데이터교육운영 지원시스템을 통한 자격확인→지원여부확정 • 신청자에 지원확정 여부 및 지원금액, 지원방식 안내
③ 교육신청및 교육비지원	바우처 이용자	지원 대상자는 데이터교육운영센터에 교육신청(등록교육기관은 홈페이지 에서확인) • 평생교육기관은 교육 신청자를 시스템에 입력 • 데이터교육운영센터는 전자바우처 형태로 지원대상자에게 바우처지급(연간 100만원)
④ 교육이수 및 이수결과확인	데이터교육기관	위탁기관인 데이터교육 수행기관이 교육 이수 여부(출결 사항)를 시스템에 입력
⑤ 최종학습관리	바우처 이용자	데이터교육운영센터의 바우처 지원시스템에 교육 및 학습결과 입력 • 바우처지원시스템에서 잔여지원한도 확인 후 추가교육 신청

- 데이터 교육기관 등록 절차



<그림 7-14> 교육기관 등록절차

- 데이터 교육기관 신청절차는 데이터교육 바우처 홈페이지를 통해 등록 하며, 6개의 단계에 걸쳐 진행 되고 등록 여부에 대한 통보는 문자나 이메일로 처리

라. 연차별 교육계획

□ 교육 일정

구분	'20	'21	'22	'23	'24	'25	'26	'27	'28	'29
데이터 관리자										
데이터 운영자										

□ 교육 대상

(단위 : 명)

구분	'20	'21	'22	'23	'24	'25	'26	'27	'28	'29	계
예상 연구원수	398,679	417,550	437,315	458,015	479,696	502,402	526,183	551,090	577,176	604,497	-
증가 연구원수	18,019	18,871	19,765	20,700	21,680	22,706	23,781	24,907	26,086	27,321	223,837
교육대상	1,802	1,887	1,976	2,070	2,168	2,271	2,378	2,491	2,609	2,732	22,384

마. 소요예산

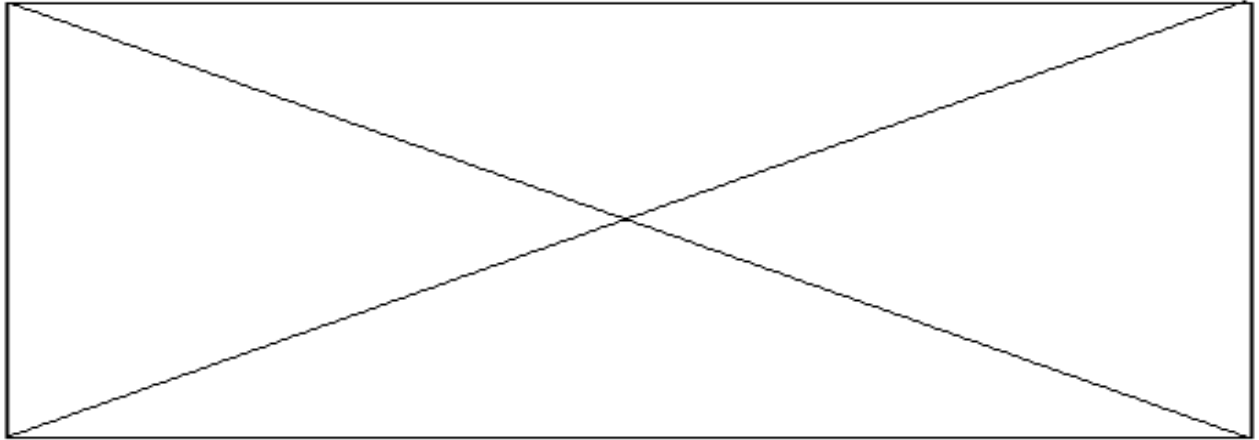
□ 연차별 교육비용

(단위 : 백만원)

구분	'20	'21	'22	'23	'24	'25	'26	'27	'28	'29	계
비용	1,802	1,887	1,976	2,070	2,168	2,271	2,378	2,491	2,609	2,732	22,384

7.4.2 연구데이터의 R&D 응용을 위한 분야별 전문 활용 교육

가. 교육목표



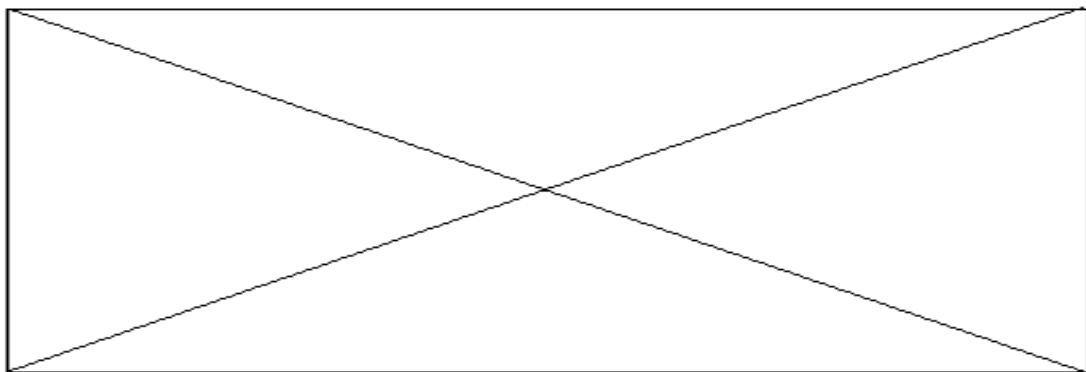
나. 교육 대상 및 범위

□ 교육대상

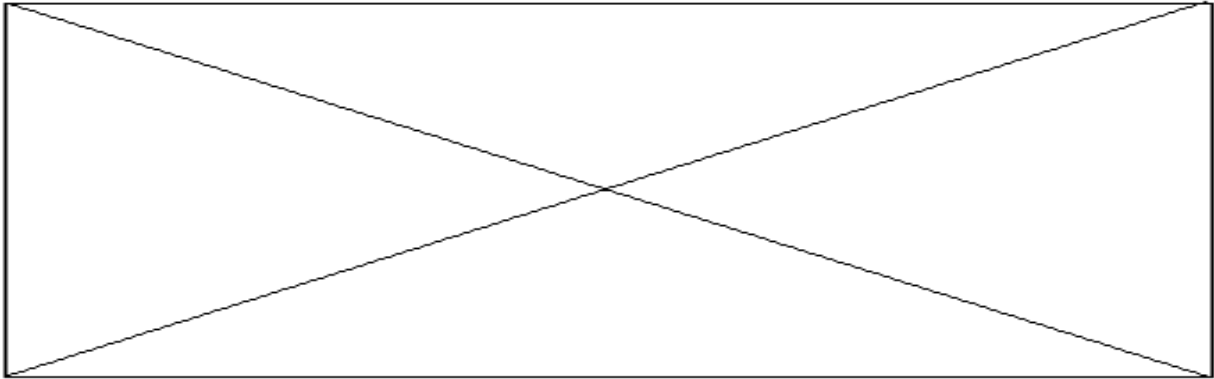
- 국가 연구기관 및 대학 등 학계의 석·박사급 고급인력을 대상으로 분야별로 고급 인재육성데이터 관리 및 운영, 기업 등 산업계에 대해서는 필요 시 고급인력 교육 수행

<표 7-13> 분야별 전문 활용 기관별 교육 대상

국가 연구기관	대학 등 학계	기업 등 산업계
석·박사급 고급인력	석·박사급 고급인력	산업기술 개발인력
데이터처리 중급연구원	데이터 처리 중급연구원	공공분야 연구인력
데이터 처리 초급연구원	데이터 처리 초급연구원	



□ 교육범위



<그림 7-15> 연구데이터 활용 Task 중 분야별 전문활용 교육 범위

- 연구데이터 및 컴퓨팅 활용 R&D Task 중 데이터 관리영역인 연구데이터 전처리, 연구데이터 분석, 분석결과 시각화 Task에 대한 교육

직 무	연구데이터 활용 R&D TASK	상세 Task
연구 데이터 분석	연구데이터 전 처리	요건정의
		분석모델링
	연구데이터 분석	모델구현
		결과해석
	분석결과 시각화	데이터표현
		시각화구현
	HPC 활용 및 운영	HPC 활용 및 지원
		시뮬레이션 및 결과 도출

<그림 7-16> 분야별 전문활용 교육범위

다. 교육내용 및 교육방법

☐ 교육 내용

구분		교육 내용
직무	단계	
데이터 분석가	공통	분석적 사고 향상
		HPC 활용 지원
		R프로그래밍과 분석
		상용 Tool 활용법
		BI 이해
		비정형데이터 이해
		데이터마이닝
		Forecasting
		비즈니스통찰력과 커뮤니케이션
		대용량 데이터와 인포그래픽 이해
		시각화 Tool 이해와 활용
		분석모델 평가방법의 이해
		성과평가시스템
		데이터 관련 이해관계자 관리
		설득과 커뮤니케이션
	분야별	응용 분야(바이오, 물리, 사회학 등)별 알고리즘
		수학적 모델링 및 시뮬레이션
		수치해석 기법
		성능 최적화 및 병렬화
		과학적 가시화 기법
		프로그래밍 언어 교육

☐ 교육 방법(KISTI KSDS)

- 연구데이터 활용 분석 실무능력 향상 교육 중심
- 데이터 전문과학자 고급인력 양성

○ 분야별 연구현장의 수요를 기반으로 실무중심의 커리큘럼 운영

- 연구기관, 학교 등 연구 현장의 업무와 학교 내 교육의 이질성을 지양하고 현장 실무능력을 배양할 수 있는 교육 과정 구성 및 교육 교재 마련

- 급격한 기술변화에 대응하기 위해 단순 지식 습득에서 창의, 문제 해결 역량 등을 중심으로 한 교육 체계 구성
- 산·학·연의 인력 수요/공급을 위한 소통 채널을 구성하고 인력 양성에 대한 라이프 사이클 구축
 - 연구기관, 학교, 등 연구 현장이 요구하는 실무능력이 즉각적으로 교육 현장에 반영될 수 있는 환경 마련하고, 신기술 및 최첨단 기술 교육을 전수함으로 재교육 효과를 증대하여 R&D의 디지털 역량 및 경쟁력을 제고
- 획일화된 교육이 아닌 4차 산업의 성장 동력과 연계하여 분야별 맞춤형 교육을 통한 전문성을 확보

<표 7-14> <연구데이터 활용 분석에 필요한 분야별 교육 훈련 예시>

분야	관련 교육훈련					
기초 학문	통계학	수리경제학	수학	컴퓨터과학	경영학	인문/사회 과학
데이터플랫폼	유닉스 등 서버환경	윈도우 등 네트워크 환경	클라우드 컴퓨팅 환경	코볼, C+ 등 프로그램 언어	하둡 등 빅데이터 관련	기타..
분석 소프트웨어	엑셀 등 기본 분석 SW	SAS, R 등 통계 SW	SQL 서버 등 DB SW	파이선 등 빅데이터 SW	슈퍼컴퓨터 활용 SW	기타...
분석 알고리즘	평균 비교 등 기초 통계 분석	시계열분석, 회귀분석, 업리프트 모델링 등 예측/추정	시각화, 군집분석, 요인분석, 주성분분석 등 원인 분석	의사결정나무, 신경회로망 등 분류분석	텍스트마이닝, 소셜네트워크 분석 등 관계망 분석	앙상블 기법, 기계학습 등 예측 최적화 분석
데이터 활용 관련 역량	인사이트 발견 등 영역 전문성	발표, 컨설팅 기법	고객 등과의 의사소통 기법	인사이트 현업 적용 등 문제해결 역량	기타	

- 교육 방법의 변화에 따른 다양화를 수용하여 교육의 질을 높이고 균형 잡힌 교육 수행

<표 7-15> KSDS 교육의 변화

(TO-BE) 기존 교육	(AS-IS) 변화된 교육
이론식	실습식, 참여식
강사 중심	수강생 중심
대형 교육장 중심	소형 교육장 중심
집합 교육	온라인 교육
통합 교육	성별, 연령별, 경력별 교육

- 실무능력 평가 및 인증제도(자격증) 시행
 - － 실무능력 평가를 위한 표준화된 평가 지표 마련
 - 평가 과목 선정 및 과목별 교육 이수 시간 제정
 - 집합 교육 및 온라인 교육의 적절한 배분을 통해 평가
 - － 데이터 활용 실무능력 자격증 제도 마련
 - 대학교와 연계하여 자격증 제도를 운영할 수 있으며, 학점 연계도 고려
 - 격증 취득자에 대하여 직무 연관성이 있는 기업에 취업시 가산점 부여

<표 7-16> 데이터 관련 자격증 예

자격증명	기관	과목
ADsP (데이터분석 준전문가)	(국가공인) 한국데이터진흥원	데이터 기획 데이터 분석
ADP (데이터분석 전문가)	(국가공인) 한국데이터진흥원	데이터 기획, 데이터 분석 데이터 시각화
Data Science Certificate (데이터 과학 자격증)	하버드 평생 교육원 Harvard Extension School	통계학, 파이썬/R 빅데이터/딥러닝, 데이터 사이언스
Certification of Professional Achievement in Data Sciences	콜롬비아 대학교	알고리즘, 확률과 통계 기계학습, 데이터분석 및 시각화

- 실무중심의 교육환경 구성
 - － HPC 실습을 위한 고성능 컴퓨팅 노드를 교육용 노드로 별도로 구성하여 상시 교육에 활용할 수 있는 체제 마련
 - HPC 클러스터 및 병렬 컴퓨팅 실습, 고성능 GPU 병렬 컴퓨팅 실습
 - － 데이터 생성, 데이터 수집, 데이터 분석 및 활용 등 데이터의 전주기 사이클을 실습할 수 있는 데이터 생태계 환경 구성
 - 드론을 사용한 이미지 데이터 생성 및 수집, 각종 센서를 사용한 IoT 데이터 생성 및 수집, 스마트 시티 등 도시환경 시뮬레이션 환경 구축
 - － 실제 업무에서 사용하는 소프트웨어와 하드웨어 인프라를 구축하여 실습 교육을 진행함으로써 교육 과정 이후 해당 업무에 바로 적용할 수 있는 체계 확립
- 기존 KISTI 데이터/컴퓨팅 활용 교육 프로그램 확대
 - － 현재 분산되어 운영하고 있는 데이터·컴퓨팅 활용 교육 프로그램을 통합하여 수요도 및 중요도가 높은 교육 위주의 교육 프로그램으로 재구성
 - － 고급, 심화과정으로 구분하여 이론과 실습을 조절한 맞춤형 교육 환경 제공

- 지역캠퍼스(중소기업지원센터*)와 연계·협력하여 각 지역별 필요한 전문분야를 파악하고 인력양성을 위한 맞춤형 교육 프로그램을 설계하며, 필요한 경우 온라인 시스템을 이용하여 교육 진행
 - * 부산울산경남지원, 대구경북지원, 호남지원, 충청지원, 수도권지원 연계
- UST·과기특성화대학에 연구데이터 활용 능력 배양을 위한 전문 교과 개설 추진
 - UST 강좌 개설 추진하고 향후 4대 과기특성화대 확산 추진하고 대학과의 컨소시엄을 구축하여, 연구 데이터 활용교육 교과목을 개설 및 학점 연계가 가능하도록 구성
 - 대학과 협력하여 석·박사 대상으로 학점 연계를 통한 실습 교육을 제공
- 데이터 과학 관련 분야 전문자격인증제도 구축
 - 인공지능, 드론 등 데이터 활용 여러 분야에서의 인증시험을 통해 한국과학기술정보통신부에서 인증하는 난이도별 국가 공인 자격증*발급
 - * 한국데이터진흥원 주관 빅데이터 분석 (준)전문가 자격증 ADP: Advanced Data (Semi) Professional으로 ADSP자격증 보유자에 한해 ADP 시험 응시 자격조건이 주어짐.
 - 자격증 취득한 실무자의 경우 기업/공공기관 내에서의 개인평가 시에 가점으로 활용될 수 있도록 국가 차원에서 장려하며, 전문 강사로 초빙하여 후속 세대의 전문가 양성을 위한 기회 제공
- 산업 트렌드를 고려한 고급 인력 양성
 - 고급 인력 양성을 위해 센터를 설립하고 센터와 지역을 구분하여 센터차원으로는 단기, 중기, 장기 교육과정을 개발하여 4차 산업시대의 트렌드에 맞는 산업을 중심으로 분야 및 난이도별 교육 프로그램 설계
 - 각 지역의 중소기업지원센터*와 연계하여 그 지역의 특화된 산업 위주 및 현안을 해결할 수 있는 교육 과정 개발
 - * 부산울산경남지원, 대구경북지원, 호남지원, 충청지원, 수도권지역 연계
 - 각 지역에서 진행하는 교육은 오프라인을 중심으로 하되, 이동이 어려운 지역의 경우 온라인 교육을 제공
 - 학교 또는 공공기관에서 단체 집중 실무 교육을 원하는 경우 온라인 교육을 지원하거나 강사가 직접 방문하여 교육을 제공

<표 7-17> 분야별 산업 트렌드를 고려한 고급 인력 양성

구분	내용
센터	• 4차 산업 시대와 관련된 분야로 각 분야별 인력수급 예측하여 종합관리체계 • 단기, 중기, 장기 교육과정을 각각 개발하여 중장기 전략체계 • 전문가 인력 풀을 구성하여 기업/공공기관에서 인력수급 필요 시 전문가추천 또는 채용연계 • 전문가 인력 풀을 통해 교육 강사로 초빙을 할 수 있으며, 정기적 모임을 활성화하여 최신 기술 동향 공유 및 교육 훈련 프로그램 개발에 참여
지역	• 각 지역별 특화된 산업 분야위주의 인력 수급 예측하여 관리체계 • 다양한 산업 분야의 실무진들을 위한 전문가 양성 • 대학(원)생들의 취업연계 위한 기업에서 바로 활용 가능한 분야 예측 및 관련 교육 프로그램 개발

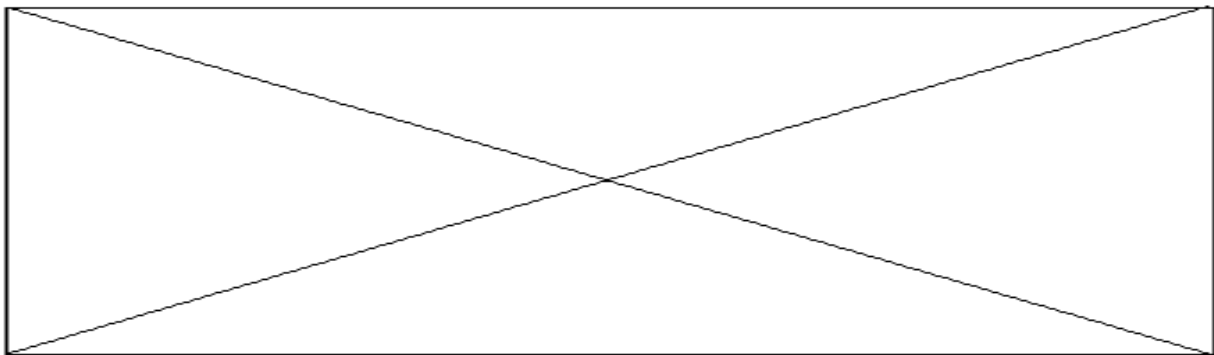
- 과학기술 연구에 데이터 분석·활용 접목을 가속화하기 위한 인공지능·빅데이터 분야 공학연구센터 설치하여 전문 인재 양성

－ 대구경북과학기술원*과 연계·협력하여 사용되지 못하고 버려지는 방대한 양의 빅데이터를 최대한 활용가치가 높은 데이터로 정제하고 관리하기 위해 데이터 제공 및 슈퍼컴퓨터 등의 자원을 지원

* ‘연구개발 지원을 위한 인공지능·빅데이터 연구센터’ 공고(‘18.3월 공고, 8월부터 7년간 연구개시 예정) 대구경북과학기술원에서 주관하는 암흑데이터 극한 활용 연구센터와 연계

□ 교육 방법(UST)

- EDISON 교육플랫폼과 연계한 연구분야별 전문교육 시스템



<그림 7-17> EDISON과 연계된 UST 연구분야별 전문교육

○ 전문 분야별 Domain 지식습득 및 IT 과정

분야	전문 분야별 교육과정		
이학	생물분석과학 나노계측과학 광물·지하수자원학 방사화학 및 핵비확산 천문우주과학 측정과학	의학물리학 한의생명과학 해양생물학 해양환경과학 핵융합과학 극지과학	방사선종양의과학 기초과학 생명과학 바이오-메디컬 융합 의학화학 및 약리생물학 생명공학
IT	과학기술경영정책 빅데이터과학 과학기술정보과학	화학소재 및 공정 ICT 생산기술	나노-정보 융합
BT	인체 및 환경 독성학 해양생명공학 방사선동위원소응용 및	생명공학 식품생명공학 생명공학	바이오-메디컬 융합
NT	나노메카트로닉스 나노-정보 융합	전기기능소재공학 화학소재 및 공정	
ET	원자력 및 방사선안전 에너지변환공학 가속기 및 핵융합 물리공학 신형원자력시스템공학 양자에너지화학공학 환경에너지기계공학 플랜트기계공학	신에너지 및 시스템기술 재생에너지공학 신에너지 및 시스템공학 자원순환공학 석유자원공학 물리탐사공학 스마트도시·건설융합	에너지-환경 융합 신소재 공학 교통시스템공학 선박해양공학 화학소재 및 공정 생산기술
ST	무기체계공학 항공우주시스템공학		

○ 연구데이터 및 (빅)데이터 관련 교육과정

구분	교육과정	
	기초과정	심화과정
빅데이터 과학	• 빅데이터 과학 개론 • PYTHON 기초 프로그래밍 • JAVA 기초 프로그래밍 • 빅데이터 분석을 위한 R 프로그래밍 • 데이터 마이닝 II • 데이터베이스 이론 및 실습 • 빅데이터 시각화 • 기계학습 I • 빅데이터 수리과학 I • 빅데이터 기반 사회현안 해결 사례 및 기술 연구	• PYTHON 고급 프로그래밍 • JAVA 고급 프로그래밍 • 데이터 마이닝 II • 분산데이터베이스 이론 및 실습 • 기계학습 II • 빅데이터 수리과학 II • 고급 수치해석 및 초병렬 컴퓨팅 알고리즘 • 생명의료 빅데이터 처리 및 활용 • 과학 빅데이터 처리 및 활용

□ 데이터 교육지원센터(가칭) 운영 및 지원시스템 구축

- ICT 기반 기술교육과 분야별 전문활용 교육 진행을 위해서 가칭 데이터 교육지원센터를 과기정통부 산하에 설립하고 교육 운영
 - － 교육 수강생 교육 신청 및 진행, 종료관리
 - － 바우처 교육 위탁 기관 관리 및 바우처 관리

(단위 : 명)

구분	'20	'21	'22	'23	'24	'25	'26	'27	'28	'29
총괄	1	1	1	1	1	1	1	1	1	1
ICT 기본기술 교육	2	2	2	2	2	2	2	2	2	2
분야별 전문활용 교육	2	2	2	2	2	2	2	2	2	2
계	5	5	5	5	5	5	5	5	5	5

- ICT 기반 기술교육과 분야별 전문활용 교육 진행을 위한 교육운영 지원시스템 구축

서브시스템 명		단위업무 기능명
로그인		회원가입
		아이디찾기
		비밀번호변경
교육과정정보	교육과정정보	교육과정 간편등록
		교육과정 상세등록
일반교육 진행관리	교육신청	신청가능여부 확인
		양식다운로드
	학습자 사용관리	파일첨부
		신청서 작성
		방문접수 현황내역
		교육과정별학습자 정보조회
		교육과정별학습자 정보등록
		학습자출석 및 과정이수 정보 등록 및 삭제
		학습자 정보조회
	교육수강생 신청관리	약관동의
		본인 인증
		자격 검증

서비스시스템 명		단위업무 기능명
		교육 신청
		회원가입
		학습계획서 작성
		이용자 신청내역 조회
	교육 대상자 선정관리	교육 대상자 선정결과 확인
		교육 신청내역 조회
바우처교육 진행관리	바우처교육기관신청	기관기본정보등록
		업무담당자정보등록
		증빙자료등록
	바우처교육기관신청현황확인	교육기관신청현황
		기관정보수정
		연도별 수강료증빙자료 등록(업로드)
	바우처접수신청	신청가능여부 확인
		양식다운로드
	바우처 학습자 사용관리	파일첨부
		신청서 작성
		바우처 계좌 연계
		방문접수 현황내역
		교육과정별학습자 정보조회
		교육과정별학습자 정보등록
		학습자출석 및 과정이수 정보 등록 및 삭제
		학습자 정보조회
	교육수강생 바우처신청관리	약관동의
		본인 인증
		자격 검증
		바우처 신청
		회원가입
		학습계획서 작성
		바우처이용자 신청내역 조회
		바우처이용자 선정결과 확인
	바우처 선정관리	바우처 신청내역 조회
		바우처 잔액관리
	바우처 잔액관리	바우처잔액정보확인
수강내역관리	수강내역관리	수강내역 확인

- ICT기반 기술교육 및 분야별 전문활용 교육을 운영 및 진행하는 가칭 데이터 교육지원 센터를 운영하고 이를 지원하고 피교육생 및 바우처 등을 관리하는 지원시스템 구축

□ 데이터 처리 및 분석 등의 기술교재 구축

- 연구기관, 전문센터 등의 연구데이터 처리 및 분석, 운영을 위한 ICT 기반 기술교육용 S/W, Tool, 패키지 등에 대한 교보재 구축
- 교보재는 EDISON 플랫폼에 탑재되어 수강자가 활용할 수 있도록 S/W, Tool 등을 구축하고 별도의 이론 교육을 할 수 있는 교재 제작
- S/W, Tool, 패키지, 교재 등은 국가연구데이터 플랫폼, 대소분야 전문센터 구축 시 테스트베드 및 산출물 활용

연구데이터 활용 Task	교보재 내용	교보재
연구데이터 수집	연구데이터 연계기술	S/W Tool 패키지 교재
	메타데이터 연계기술	
연구데이터 저장	클라우드 기반 분산,병렬 저장 기술	
	데이터 저장 및 연계 보안기술	
	데이터표준 수립 및 품질관리 기술	
	연구데이터 저작권 관리 기술	
연구데이터 분석	연구데이터 검색 기술	
	BI 서비스 기반기술	
	AI 서비스 기반기술	
	MI 서비스 기반기술	
	DI 서비스 기반기술	
	가시화 기반 기술	
	Batch & Realtime 분석 기술	
	Workflow 서비스 기술	
연구데이터 운영	협업 서비스 및 관리 기술	
	커뮤니티 생성 및 관리기술	
	고가용성 기술	
	어드민/아키텍처 관리 기술	
	보안 관리 기술	
	가상머시 관리 기술	

라. 연차별 교육계획

□ 교육 일정

구분		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
공통											
분야별	ICT·SW										
	생명·보건의료										
	소재·나노										
	농림수산·식품										
	환경·기상										
	기계·제조										
	우주·항공·해양										
	원자력·핵융합										
	에너지·자원										
	건설·교통										
	재난안전										
데이터교육지원센터 지원시스템구축											

□ 교육 대상

(단위 : 명)

구분	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
예상 연구원수	141,161	144,258	147,423	150,657	153,963	157,341	160,793	164,320	167,925	171,609	—
증가 연구원수	3,030	3,097	3,165	3,234	3,305	3,378	3,452	3,528	3,605	3,684	33,479
교육대상	303	310	316	323	331	338	345	353	361	368	3,348

□ 인력운영 계획

(단위 : 명)

분야별		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계		33	16	16	16	17	17	17	17	17	17	183
데이터 교육지원 센터	소계	10	11	11	11	12	12	12	12	12	12	115
	총괄	2	3	3	3	4	4	4	4	4	4	35
	ICT기반기술 교육	3	3	3	3	3	3	3	3	3	3	30
	분야별 전문활용 교육	5	5	5	5	5	5	5	5	5	5	50
데이터교육지원센터지원시스템구축		13	2	2	2	2	2	2	2	2	2	31
교재작성		10	3	3	3	3	3	3	3	3	3	37

7.4.2.5 소요예산

□ 연차별 교육비용

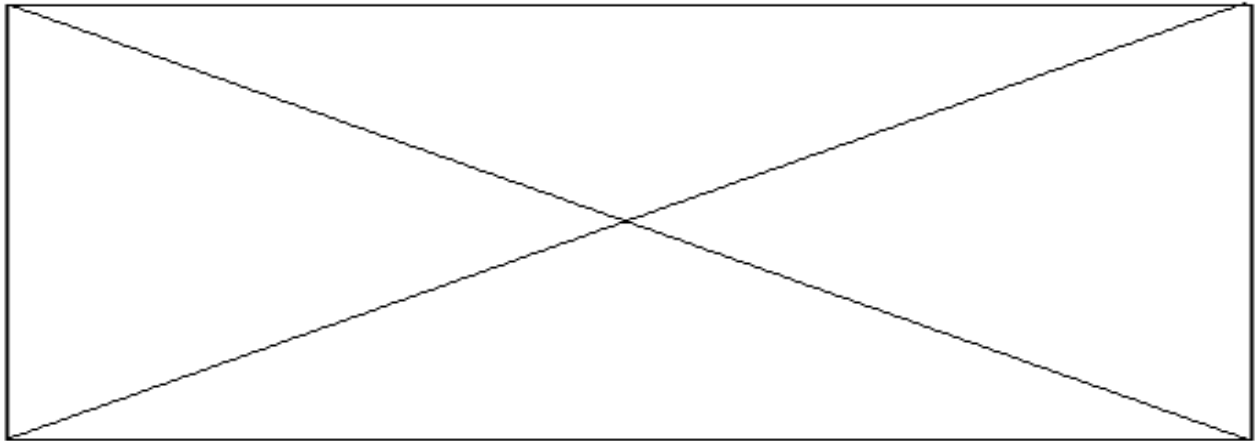
(단위 : 백만원)

구분	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	1,678	1,129	1,143	1,147	1,061	1,076	1,090	906	921	937	11,088
교육비용	606	619	633	647	661	676	690	706	721	737	6,696
데이터교육지원센터 지원시스템구축	992	450	450	450	350	350	350	150	150	150	3,842
교재작성	80	60	60	50	50	50	50	50	50	50	550

7.4.3 연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축

가. 구축목표

□ 목표



□ 주요 구축내용

- 전문분야 교육 커리큘럼에 적합한 교육용 데이터셋 수집, 정제, 개발 및 구축
- EDISON 플랫폼 기반 교육 인프라 및 서비스 구축
 - － 교육용 데이터 구축, 교육용 데이터 분석 기능 추가 개발, 사용자 지원 및 운영
- 전문분야 교육 커리큘럼, 일반 데이터 처리 및 분석 교육에 적합한 CX-Learning 교육 교보재 제작
 - － 동영상 강의, 교안, 실습 환경, 메모장 등을 융합한 반응형(Interactive) 교보재 개발
- ICT기반 기술교육 및 분야별 전문활용 교육을 운영·진행하는 (가칭) 데이터 활용 교육 지원센터를 운영·지원하고 피교육생 및 바우처 등을 관리하는 지원시스템 구축

나. 과제내용

□ 전문분야별 교육용 데이터셋 구축 및 분석 튜토리얼 개발

- 전문분야 교육 커리큘럼에 적합한 교육용 데이터셋 수집, 정제, 개발 및 구축
- 개발된 데이터셋 맞춤형 데이터 분석 튜토리얼 개발
- 교육용 데이터셋의 목적에 적합한 전용 분석 SW 개발

<표 7-18> 교육용 데이터셋 개요

교육용 데이터셋의 특성 및 요구사항	교육 전용 데이터셋 개발이 필요한 이유
• 한정된 교육 시간 이내에 분석이 가능한 데이터셋 사이즈를 가져야 함 (수분 이내) • 작은 사이즈임에도 의미있는 분석 결과를 충분히 도출 가능한 데이터 • 오류를 최소화할 수 있는 정제된 데이터셋 • 분야별 특성을 충분히 반영 (예: 소재 분야의 경우 소재의 3차원 구조의 특성을 충분히 반영할 수 있는 데이터셋)	• 좌측 설명된 교육용 데이터셋의 특성으로 인하여 구축된 R&D용 데이터셋을 그대로 활용하기 어려움 • 해외 데이터셋, 국내 구축 대용량 R&D 데이터셋 중 의미가 충분히 반영된 일부의 데이터만 추출하고 정제를 통한 교육용 데이터셋 개발 필요 * 교육용 데이터셋의 개발은 ICT 개발자 단독으로는 불가능하며, 관련 응용분야 연구자, 교육자 등 다분야 전문가의 융합이 필수적임

□ EDISON 플랫폼 기반 데이터 교육용 콘텐츠 서비스 기능 및 문제해결 환경 개발, 서비스 운영 및 사용자 지원

- EDISON 플랫폼 기반 교육용 콘텐츠 서비스 기능 개발
 - － 데이터셋 전용 검색, 전처리, 조회 기능 개발
 - － EDISON 플랫폼 기반 교육용 튜토리얼 전용 검색, 조회 기능 개발
 - － EDISON 플랫폼 기반 튜토리얼 실행환경 개발 및 구축
 - － EDISON 플랫폼 기반 데이터셋 분석 전용 SW의 전처리 및 가시화 기능 개발

○ EDISON 플랫폼 기반 교육용 문제해결 환경 개발

* 데이터셋 공유는 물론, 데이터셋 분석에 가장 효율적인 SW를 학생들이 경쟁적으로 개발·제시하고 성능을 비교할 수 있는 공유, 협력, 경쟁의 장을 제공 (해외 Kaggle 사이트 유사 기능 제공)

- － 문제해결 프로젝트 및 팀 관리 기능 개발
- － 협력을 위한 포럼 및 연구주제 관리 기능 개발

- 사용자 랭크, 점수, 프로파일 관리 기능 개발
- 데이터셋 및 코드 관리 기능 개발
- 교육 코스, 태그, 조직 관리 및 블로그, 구인구직, FAQ 기능 개발
- EDISON 플랫폼 기반 교육용 콘텐츠 서비스 운영 및 사용자 지원

서브시스템 명		단위업무 기능명
계정 관리		계정 정보 관리 : 조회 / 수정 / 삭제
		계정 API 토큰 생성 / 삭제
		Quota 조회
교육 코스 관리		코스 리스트
		코스 상세 조회
		코스 생성, 수정, 삭제
		코스 알람
프로젝트 관리	프로젝트 현황 조회	참여 프로젝트 현황 조회
		프로젝트 검색
		범주별 프로젝트 검색
	프로젝트 관리	프로젝트 뷰
		프로젝트 오버뷰 관리
		프로젝트 튜토리얼 관리
		프로젝트 평가기준 관리
		프로젝트 규칙 관리
		프로젝트 타임라인 관리
		프로젝트 Prize 관리
		접근 URL 관리
		결과 제출 관리
		평가 관리
		체크 리스트 관리
		서약서 관리
	프로젝트 활용 데이터 관리	데이터 조회
		데이터 접근 API
		데이터 다운로드
		데이터 컬럼 관리
		데이터 오버뷰 관리
		데이터 CSV 뷰

서브시스템 명		단위업무 기능명
	프로젝트 활용 코드 관리	프로젝트 활용 코드 리스트
		프로젝트 활용 코드 검색
		프로젝트 활용 코드 결과물 기준 검색
		프로젝트 활용 코드 언어 기준 검색
		프로젝트 활용 코드 유형 기준 검색
		프로젝트 활용 코드 키워드 기준 검색
		프로젝트 활용 My 코드 조회
		프로젝트 활용 Favorites 코드 조회
		Favorites 코드 설정
		신규 코드 생성
	프로젝트 활용 포럼 관리	프로젝트 활용 포럼 리스트
		프로젝트 활용 포럼 리스트
		프로젝트 활용 포럼 검색
		프로젝트 활용 포럼 정렬
		프로젝트 활용 포럼 Follow
	프로젝트 랭크 및 점수 관리	프로젝트 활용 랭크 및 점수 리스트
	팀 관리	팀 생성
		팀 멤버 초대
		팀 멤버 요청
		팀 멤버 승인
		팀 합병
		팀 리더 변경
		팀 정보 변경
		팀 삭제
	프로젝트 결과 관리	프로젝트 결과 제출
		코드로 부터 결과 제출
포럼 및 연구주제 관리	포럼 관리	분류별 포럼 조회
		포럼 연구주제 리스트
		포럼 연구주제 생성
	연구주제 관리	연구주제 상세 뷰
		연구주제 정렬
		연구주제 follow / unfollow
		연구주제 좋아요 추가
		내 연구주제 조회
		연구주제 Quote / 링크

서브시스템 명		단위업무 기능명
	커멘트 관리	커멘트 작성 / 수정 / 삭제
		커멘트 좋아요 추가
		커멘트 리스트 뷰
사용자 랭크 및 점수 관리		프로젝트 기여 랭크 및 점수 리스트
		코드 기여 랭크 및 점수 리스트
		포럼 기여 랭크 및 점수 리스트
사용자 프로파일 관리		프로젝트 기여 랭크 및 점수 조회
		참여 프로젝트 리스트 조회
		기여 코드 점수 조회
		기여 코드 리스트 조회
		참여 포럼 점수 조회
		참여 포럼 리스트 조회
		사용자 Contact
		사용자 Follow
		사용자 follower 조회
		My 프로파일 관리
		My 조직 관리 : 생성 / 수정 / 뷰 / 삭제
데이터셋 관리	데이터셋 생성, 검색	신규 데이터셋 생성
		데이터셋 접근 URL 관리
		데이터셋 정보보호 관리
		데이터셋 검색
		데이터셋 정렬
		데이터셋 태그 관리
		My 데이터셋 조회
		Favorite 데이터셋 조회
	데이터셋 조회	데이터소스 상세 뷰
		개별 파일 분석 뷰
		데이터셋 히스토리 뷰
		데이터셋 오버뷰 뷰
		데이터셋 코드 뷰
		데이터셋 연구주제 뷰
		데이터셋 활동 뷰
	데이터셋 연관 연구주제 관리	연구주제 상세 뷰
		연구주제 정렬

서브시스템 명		단위업무 기능명
		연구주제 follow/unfollow
		연구주제 좋아요 추가
		내 연구주제 조회
	데이터셋 연관 액티비티 관리	액티비티 통계
		코드 통계
		토론 통계
	데이터셋 연관 조직 관리	조직 할당
		데이터셋 연관 조직 리스트
코드 관리	공개 코드 조회 및 검색	공개 코드 리스트 조회
		공개 코드 리스트 정렬
		공개 코드 카테고리 별 검색
		공개 코드 결과물 형식 별 검색
		공개 코드 사용 언어 별 검색
		공개 코드 타입 별 검색
		공개 코드 키워드 별 검색
		공개 코드 타이틀 검색
		내 코드 조회 및 검색
		즐거찾기 코드 조회 및 검색
	코드 생성 및 관리	스크립트 타입 코드 생성
		스크립트 타입 코드 작성
		스크립트 타입 코드 실행
		스크립트 파일 업로드
		스크립트 파일 다운로드
		노트북 타입 코드 생성
		노트북 타입 코드 작성
		노트북 타입 코드 실행
		ipynb 노트북 파일 업로드
		ipynb 노트북 파일 다운로드
		코드 커밋(Commit)
		코드 커밋 로그 관리
		코드 콘솔 화면 제공
		코드 수정
		코드 삭제

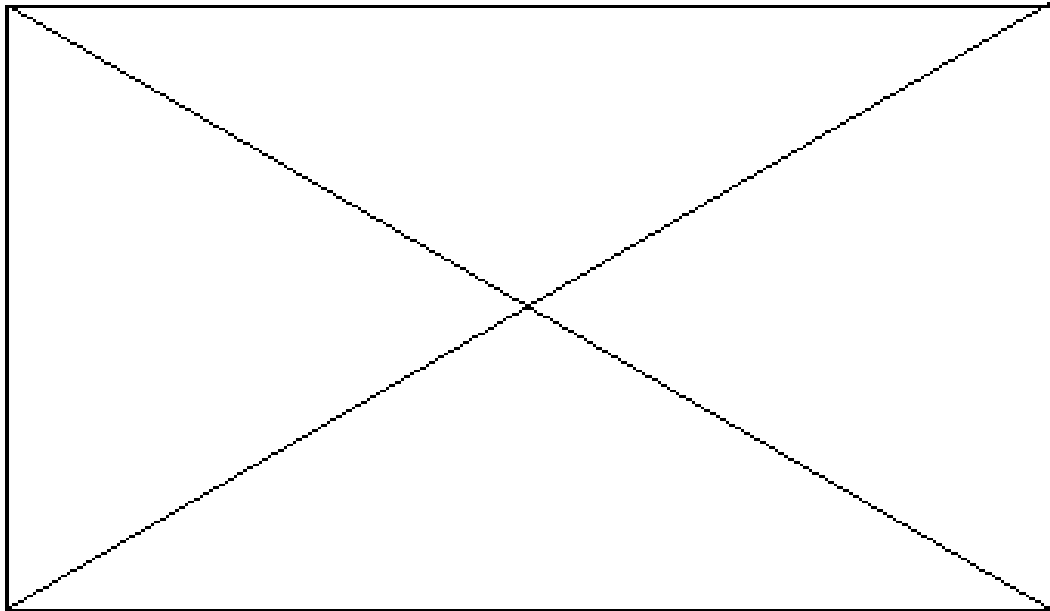
서브시스템 명		단위업무 기능명
		코드 복제
		코드 평가 점수 관리
		코드 조회 횟수 관리
		코드 복제 횟수 관리
		코드 코멘트 관리
		코드 마지막 실행 날짜 관리
		코드 버전 관리
		코드 키워드 관리
		코드에 사용된 데이터셋 관리
	코드 환경 설정	Public/Private 공유 기능
		유저 별 공유 기능
		사용 언어 설정
		코드 버전 정보 관리
		코드에 기 업로드 된 데이터셋 추가
		코드에서 새 데이터셋 업로드 및 추가
		추가된 데이터셋 리스트 조회
		추가된 데이터셋 미리보기
		코드 GPU 사용 설정
		사용할 도커(Docker) 이미지 설정
		코드가 사용 중인 리소스 정보 조회
	코드 부가 기능 관리	코드 노트북 조회
		코드 코드 조회
		코드 코드 다운로드
		코드에 사용된 데이터셋 조회
		코드 결과물 조회
		코드 코멘트 조회
		코드 코멘트 정렬
		코드 코멘트 등록
		코드 코멘트 평가
		코드 커밋 로그 다운로드
		코드 생성용 도커 이미지 다운로드
		코드 복제 횟수 관리
		코드 버전 조회
		코드 버전 별 비교
		코드 키워드 추천 및 자동 등록

서브시스템 명		단위업무 기능명
태그 관리		전체 태그 리스트
		개별 태그 뷰
		태그 정렬
		태그 검색
		태그 연계 프로젝트 리스트
		태그 연계 데이터셋 리스트
		태그 연계 코드 리스트 조회
		유사 태그 조회
조직 관리		조직 생성
		조직 초대
		조직 요청
		조직 승인
		조직 정보 변경
		조직 삭제
구인구직 관리	구인 관리	Job 등록 및 삭제
		Job 게시 기간 관리
		Job의 SNS 등록 및 관리
		Job 게시판 팔로워에게 이메일 전송
		메인 사이트에 Job 배너 게시
	구직 관리	Job 지원
		Job 리스트 조회
		Job 리스트 급여/경력/직군 별 검색
		Job 리스트 타이틀 검색
		Job 게시판 팔로잉 기능
블로그 관리	블로그 관리	블로그 작성
		블로그 수정, 삭제
		블로그 조회
	블로그 부가 기능 관리	블로그 코멘트 관리
		블로그 구독자 관리
		블로그 카테고리, 키워드 설정
		블로그 카테고리, 키워드 별 검색
		블로그 타이틀 검색
기타		온라인 매뉴얼
		CLI

서브시스템 명	단위업무 기능명
	REST API
	프로젝트, 데이터셋, 코드 통합 검색
	최신 프로젝트, 데이터셋, 코드 조회
	Notification 보내기, 보기
	FAQ 작성/수정/삭제/카테고리 설정

□ 교육용 CX-Learning 교보재 개발

- 웹사이트에서 전문분야 이론의 이해를 돕기 위해 기존의 전자교재, 강의영상, e-Book 등의 교재 형태뿐만 아니라 데이터·컴퓨팅 시스템에 기반한 각종 컴퓨터 시뮬레이션, 데이터 분석 활동을 병행할 수 있는 융합형 CX-Learning 교보재 개발
 - * CX-Learning : CX는 Customer Experience와 Computational Experience의 중의적 줄임 표현. 텍스트 중심의 이러닝을 넘어, 시뮬레이션 SW + 텍스트 + 데이터 분석 + HW를 결합한 과학기술분야 전문 이공계 융합 교육을 의미



<그림 7-18> CX-Learning 예시 : 강의영상+교안+실습+메모장 융합

- 전문분야 데이터 활용·분석 및 컴퓨터 시뮬레이션을 포함한 교육 커리큘럼에 적합한 차세대 CX-Learning 교육용 콘텐츠(교보재) 제작

다. 연차별 구축계획

□ 구축일정

사업내용	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
전문분야별 교육용 데이터셋 구축 및 분석 튜토리얼 개발										
EDISON 플랫폼 기반 데이터 교육용 콘텐츠 서비스 기능 및 문제해결 환경 개발, 사용자 지원 및 운영										
교육용 콘텐츠 서비스 기능 개발										
교육용 문제해결 환경 개발										
교육용 콘텐츠 서비스 운영 및 사용자 지원										
교육용 CX-Learning 교보재 개발										

□ 인력운영 계획

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	24	24	24	24	24	18	18	18	18	18	210
전문분야별 교육용 데이터셋 구축 및 분석 튜토리얼 개발	12	12	12	12	12	12	12	12	12	12	120
EDISON 플랫폼 기반 데이터 교육용 콘텐츠 서비스 기능 및 문제해결 환경 개발, 사용자 지원 및 운영	8	8	8	8	8	2	2	2	2	2	50
교육용 CX-Learning 교보재 개발	4	4	4	4	4	4	4	4	4	4	40

7.4.3.4 소요예산

□ 소요예산

(단위 :백만원)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	3,500	3,500	3,500	3,500	3,500	2,200	2,200	2,200	2,200	2,200	28,500
전문분야별 교육용 데이터셋 구축 및 분석 튜토리얼 개발	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	1,500	15,000
EDISON 플랫폼 기반 데이터 교육용 콘텐츠 서비스 기능 및 문제해결 환경 개발, 사용자 지원 및 운영	1,500	1,500	1,500	1,500	1,500	200	200	200	200	200	8,500
CX-Learning 도구개발 및 콘텐츠 교보재제작	500	500	500	500	500	500	500	500	500	500	5,000

7.5 기대 효과

□ 기술적 기대효과

- 이론과 텍스트 중심의 교육을 전문기관과의 실습중심 교육으로 과학데이터 전문교육 패러다임 변화를 선도하고 4차 산업시대에 부합하는 데이터 전문과학자 양성으로 국가 데이터 산업육성과 발전에 기여
- 슈퍼컴퓨팅 이용의 저변 확대
 - － 입문자 및 초·중급 교육 등 체계적인 교육 커리큘럼을 통해 다양한 분야와 연령층에서 슈퍼컴퓨팅을 활용할 수 있는 기회 제공
 - － 체계적인 교육으로 현장 실무 능력을 갖춘 전문인력 교육을 통해 신산업 분야의 일자리 창출
- EDISON등 데이터 분석을 위한 플랫폼 활용과 적용 교육을 통해 다양한 데이터 분석기술 습득 및 연구에 활용
- 바우처제도를 통해 데이터 분야별 다양한 전문 프로그램을 통한 다양한 데이터분석용 전문 실습 교육을 통한 데이터분석을 위한 기본 지식 습득
 - － 데이터 분석 실습 및 교육
 - － 데이터 정제 및 전처리 실습 및 교육
 - － 데이터 연산조직 실습 및 교육
 - － 데이터 시각화 실습 및 교육

□ 경제적 기대효과

- 바우처 제도를 통하여, 데이터 분석을 위한 다양한 분야의 기본 교과과정과 전문 강사를 공동 활용하여, 데이터 교육을 위한 비용절감(강의 개설, 강의 운용, 강사운용 등)
- 취업 후 민간에서 재교육하는 비용 절감으로 국가 산업 경쟁력 향상

□ 사회적 기대효과

- 대학과 출연연구소의 센터와 연계하여 실무 적용 및 활용 중심의 빅데이터 교육을 실시함으로써 현장 전문 인력 양성 및 배출

- 4차 산업혁명 시대의 도래에 따른 특성화된 실무중심의 인력양성으로 사회 전반에 필요한 인재를 적재적소에 활용 가능
- 교육현장과 연구와 산업현장 간의 공간적·시간적 단절을 극복하고, 교육에서 연구나 산업으로 진입하는 기간을 단축시켜 실무능력 뛰어난 전문 인력을 조기에 확보하고 양성

사업운영계획 및 소요예산

- 8.1. 사업추진체계 및 역할분담
- 8.2. 과제·소요예산 및 재원조달
방안

8.1. 사업추진체계 및 역할분담

8.1.1. 사업추진체계

- (추진체계 수립의 주안점) 본 사업의 목표 및 특성, 세부사업 간 연계성 확보 등을 고려
 - 관련 규정과 주관부처 및 산하·유관기관 간 역할구조를 바탕으로 효과적·효율적 추진체계 수립
- (국가연구데이터 플랫폼) 구축 및 운영은 6대 분야의 핵심기술군 단위로 연구단 구성
 - 요소기술별 세부과제가 병렬적으로 구성되기도 하고, 상호연계가 필요하기도 하여 연구단별 성공적 과제수행을 위한 수행전략 수립 필요
 - 특히, 요소기술은 기술 트렌드 변화, 시장 및 산업의 변화 등에 민감한 특성을 가지고 있어 변화관리 체계를 수립할 필요
 - 연구단 선정 시 세부과제의 핵심 기술군에 대한 이해, 수행전략, 수행역량, 연구단장 전문성 등을 종합적으로 고려
- (대·소분야 전문센터) 구축 및 운영 지원은 전문센터별로 연구단을 구성
 - 기술의 융복합적 특성, 전문센터 역할의 유사성을 감안하여 연구자간 협력을 도모할 수 있는 혁신 네트워크 구축 지원 필요
 - 연구단 선정 시 전문분야에 대한 이해, 수행전략, 수행역량, 연구단장 전문성 등을 종합적으로 고려
 - 1단계 바이오, 소재, 인공지능 분야 3개 분야, 2단계 4개 분야, 3단계 4개 분야 등 총 11개 분야의 전문센터 구축을 위한 연구단의 단계적 구성
- 기반기술 역량을 갖춘 연구원, 대학 및 기업 중심으로 연구단 구성
 - 추진과제별 세부과제는 선택과 집중을 위해 6개 이내로 제한(전문센터 예외)하고, 연구원 및 대학에서 주관 (필요시 기업 주관)
 - 연구단 간 실험적 아이디어 교환, 신기술 발굴, 기술연계 등 연계협력 중점

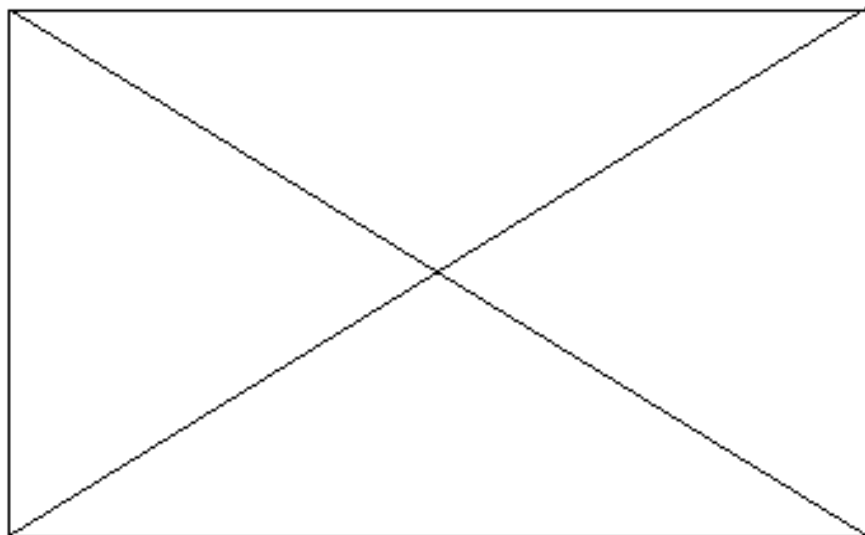
□ 국가연구데이터 플랫폼 및 대소분야 전문센터 구축은 외부환경변화 대응 필요

- 다양한 운용환경에 따른 시험평가가 요구되어 테스트베드와의 긴밀한 협력 체계 구축 필요

<표 8-1> 내역사업 별 주요 특성

구분	사업명	목표	특성	
			기술적	절차적
R&D 사업	국가연구데이터 플랫폼 구축·운영	·6대 분야 요소기술 개발	① 요소기술별 병렬적 기술개발 - 요소기술 내 세부기술간 연계성 존재 ② 외부환경 변화에 따른 변화관리 - 장기투자임을 고려 변화관리 수시화·체계화 필요	① 요소기술 개발 중 유사·중복 기술 도출 및 연계방안 검토를 통해 효율성 극대화 ② 기술개발 후 대소분야 전문센터 등에 기술이전을 통한 성과확산
	대·소분야 전문센터 구축·운영 지원	·11개 분야 전문센터 구축	① 전문센터별 단계적 구축 및 기술개발 ② 전문센터 형태, 임무 등에 따른 기술간 융복합 특성 고려 ③ 변화관리 필수	① 전문센터 공통 기술분야에 대한 전략적인 개발방안 수립 ② 전문센터별 연구데이터의 형태와 성격에 부합하는 특화전략 극대화 ③ 분야별 연구기관의 네트워크 안착 및 강화
非 R&D	데이터 및 컴퓨팅활용 기반 R&D 인재육성	·연구데이터 분석 및 관리 전문인력 양성	① ICT기반 연구데이터 관리 기술교육 반영 ② R&D 응용 시 분석Tool에 대한 기술교육	① 연구인력 역량별 교육과정 ② 분야별 전문성 고려
	테스트 베드	·시험·평가 지원	·시뮬레이션 기반 시험·평가	·다양한 운용환경 연출을 위해 연구단과의 연계성 강화

□ 조직 구성 및 운영체계



<그림 8-1> 주요 조직구성 및 운영체계

- (조직구성) 본 사업 수행을 위한 조직은 크게 관리조직, 지원조직, 연구개발 조직으로 구분
- (관리조직) 한국과학기술정보연구원, 성과관리 위원회, 국가연구데이터 공유·활용 체제 구축 사업단, 연구단장 협의회로 구성되며 주로 연구단 연구프로세스 모니터링, 성과관리, 목표관리, 지원조직 관리, 테스트베드 운영관리 등의 역할 수행
 - (지원조직) 기술이전 전담반, 변화관리 지원단, 기술융합 허브로 구성되며 사업성과 극대화 및 확산을 지원하는 역할 수행
 - (연구개발조직) 국가연구데이터 플랫폼 구축·운영 파트, 대소분야 전문센터 구축·운영 지원 파트, 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 파트로 구성

<표 8-2> 조직별 주요 기능 및 역할

구분	조직명	기능(역할)
관리조직	·한국과학기술정보연구원	·사업 성과 및 목표관리 ·사업단 연구수행 상 애로사항 수집 및 해결방안 모색
	·성과관리 위원회	·연차별 사업 성과검증 ·실패과제 원인분석 및 재활용 가능기술
	·국가연구데이터 공유·활용 체제구축 사업단	·연구단 연구프로세스 모니터링 ·연구단 연구개발 지원 - 기술이전 전담반, 변화관리 지원단 운영 ·연구단장 협의회 기술융합허브 운영 지원 ·테스트베드 연구단과 연계
	·연구단장 협의회 - 연구단장과 총괄사업단장간 협의기구	·기술융합허브 운영 ·연구단별 과제현황 공유 및 연계·협력 방안 모색 - 실패위기 과제관리 - 연구단 애로사항 수렴
지원조직	·기술이전 전담반	·개발기술 DB 구축 ·기술거래 운영지원 ·기술이전 대상 모집 및 기술이전 관리
	·전략지원 전담반	·기술동향(시장·산업동향 분석) ·학계 연구동향 분석 ·주요기업 전략 분석 ·연구단의 논문 및 특허 등록 지원
	·기술융합 허브	·과제 관련 내부세미나, 글로벌 포럼 및 학회 조직 운영 등 연구단장 협의회 내외부 혁신네트워크 조직 및 운영 지원
연구개발 조직	·연구단	·(국가연구데이터 플랫폼 구축 운영단) 요소기술 단위로 연구개발수행 ·(대소분야 전문센터 구축운영 지원단) 전문센터 단위로 연구개발수행 ·(데이터 및 컴퓨팅 활용기반의 R&D인재 육성단) 연구데이터 기반의 R&D 인력 역량 제고를 위한 교육 수행 및 운영 ·(테스트 베드) 국가연구데이터 플랫폼 테스트베드 구축운영

□ 한국과학기술정보연구원

- 한국과학기술정보연구원은 과학기술정보통신부 및 범부처 협의회와 긴밀한 관계를 맺고, 사업기획 역할을 담당하는 주관부처에 사업단 운영현황 등을 보고
- 연구단 및 사업단의 사업 성과와 목표를 관리하며, 산하에 성과관리 위원회를 두어 성과 창출을 극대화하기 위한 전략 수립 및 실행

□ 성과관리 위원회

- (구성) 성과관리 위원회는 사업특성을 고려하여 3대 추진 과제별 전문가와 한국과학기술 정보연구원으로 구성하며, 위원장은 한국과학기술 정보연구원 인사 중 선정
 - － 성과관리 위원회 산하에 ‘3대 추진과제별 책임평가단’ 운영
- (역할) ①사업단 선정 ②연차별 연구 및 지원사업성과를 점검·평가·조정하고 ③실패과제의 경우 원인을 분석하고 개발 프로세스 개선을 요구할 수 있으며, ④실패로 퇴출되는 과제가 발생할 경우 재활용 할 수 있는 연구결과물을 도출하고 수행 중인 과제에 반영이 가능한지 여부를 판단하고 제안할 수 있음

□ 책임평가단

- (구성) 책임평가단은 3개 추진과제 별 기술전문가, 특허전문가, 산업 전문가로 구성하며, 5인 내외로 구성
- (역할) 연구단의 과제수행 과정에 대한 지속적 모니터링, 연구개발 성과의 목표율 달성 여부 분석 및 분석 보고서 작성 → 성과위원회 보고

□ 국가연구데이터 공유·활용 체제구축 사업단(총괄사업단)

- (구성) 사업단은 사무국과 지원반으로 구성, 사무국은 사업단장을 포함하여 5명 내외로 구성하고 사업단 운영기관 인력 활용
- (역할) 사무국은 사업수행에 요구되는 행정업무를 수행하고 지원반은 연구단의 연구개발 과정에서 요구되는 기술이전, 변화관리, 기술융합허브 관리 등을 지원
- (기술이전 전담반) 개발기술의 DB 구축, 기술거래 운영지원, 기술이전 대상 기업 모집 및 기술이전 프로세스 관리 등의 역할 수행
- (변화관리 지원반) 기술동향, 시장·분석, 기업전략 분석, 연구 논문 및 특허 출원 지원

□ 연구단장협의회

- (구성) 연구단 단장과 총괄사업단, 테스트베드 운영사업단 단장으로 구성
- (역할) 연구단 과제 수행 과정에 대한 자체 모니터링, 연구결과 공유 및 연계·협력 방안의 모색, 연구과정에서의 애로사항 수렴, 실패위기 과제에 대한 기술적 위기극복 방안 협의, 기술융합허브 운영
- (기술융합허브) 과제 관련 세미나(수시), 글로벌 포럼 및 학회 조직·운영(정기) 등을 통하여 산·학·연 혁신 네트워크 조직 및 확산 등의 역할 수행

□ 연구단

- (구성) 연구단은 세부과제를 수행하는 연구주체들의 컨소시엄 형태와 테스트베드를 운영하는 조직으로 구성
 - － (R&D 연구단) 연구단장은 추진과제별 연구단 공모시 주관기관으로 참여하는 기관의 연구책임자로 함
 - 연구단장은 상근을 원칙으로 하되, 성과관리위원회의 결정에 따라 프로젝트 관리에 영향이 없는 범위 내에서 최소한의 본업을 유지할 수 있음. 다만, 성과관리위원회의 추천에 따라 기획관리전문가를 연구단장으로 선정할 수 있음
 - － (테스트베드 연구운영단) 사무국과 지원반으로 구성되며, 사무국은 운영단장을 포함하여 10명 내외로 구성하고 운영단 운영기관 인력 활용
 - 사무국은 테스트베드 운영에 요구되는 행정업무를 수행하고 지원반은 테스트베드 운영 및 관리, 동사업 연구개발과 함께 기술융합허브와 연계를 통한 글로벌 활용 수요 발굴 등 테스트베드 활용 확산 등의 역할 수행

8.1.2. 사업관리 방안

□ 사업의 추진절차

추진절차	추진내용	일정
국가전략 프로젝트 심의	사전기획내용 검토 및 기술성 평가(2월) 사업추진여부 확정(과학기술 전략회의, 2월)	'18년 7~8월
↓		
예비타당성 조사	예비타당성조사	'18년 11~
↓		
사업단 구성	국가연구데이터 공유·활용 체제구축 사업단 및 사무국 구성	'20년 02~
↓		
사업기획	과제기획서 및 RFP 작성(사업단, 관계기관 합동)	'20년 03~
↓		
사업심의위원회	사업의 기획, 평가 등 조정(과기부 장관) ※ 관계부처 담당 국장 사업심의위원회 참여	'20년 03~
↓		
사업시행 공고	세부추진계획 및 신규사업 공고	'20년 04~
↓		
신규과제 접수	신규과제 사업계획서 접수(사업단)	'20년 04~
↓		
과제선정·평가	신규과제 평가(사업단) ※ 사전검토-서면평가-발표평가-종합심의 신규과제 확정(과기부 장관)	'20년 04~
↓		
협약체결	주관기관과의 협약 체결(사업단)	'20년 12~
↓		
평가(연차, 단계, 최종)	연차별 실적 및 계획서 평가(사업단) 자문위원회 의견 및 평가위원회 심의	'20년 12~
↓		
사업비 정산	연간 사업비 회계 정산	'21년 02~
↓		
성과활용 보고/ 평가, 사업화	연구활용 성과 조사 및 성과관리	'21년 02~

- (사업단 구성) 본 사업은 '20년~'29년까지 총 10년간 추진하며 '20년 상반기 중에 연구개발을 본격적으로 추진하기 위해 조기에 사업단 및 사무국 구성('20.2월)
- (사업기획) 사업단 및 관계부처, 산학연 전문가로 구성된 과제기획위원회를 구성하고, 이들 합동으로 과제기획서 및 RFP 작성('20.3월)
- (사업심의위원회) 사업계획 수립의 적정성 및 연도별 종합 계획에 관한사항, 신규사업의 기획·평가·관리·예산 등에 사항을 심의·조정('20.3월)
 - － 위원장은 과기부 고위공무원으로 하고 관계부처 고위공무원 및 산업계·학계·연구계의 과학기술 및 인문사회분야 전문가, 경제·산업전문가 중에서 위촉
- (사업시행 공고) 사업별 지원규모, 대상, 자격조건 등 신규사업 공고 추진('20.3월)
- (신규과제 접수) 신규과제 사업계획서 전산 접수('20.4월)
- (과제선정·평가) 전문성, 공정성 확보를 위해 산·학·연 전문가로 구성된 평가위원회에서 심사('20.4월)
 - － 사전검토 → 서면평가 → 발표평가 → 종합심의(과기부) 등을 통해 최종 확정
 - ※ 구체적인 일정은 변동 가능

<표 8-3> 평가단계별 주요사항

구분	주 체	내 용
사전검토	사업단	· 신청자격(신청 제외기준) 및 정책 권장사항 반영 등 과제신청 적합여부를 서류검토를 통해 확인
서면평가	서면평가위원회 (사업단)	· 제출한 사업계획서를 토대로 평가위원회에서 서면 평가
발표평가	발표평가위원회 (사업단)	· 총괄책임자 발표를 통해 평가위원회에서 평가
결과통보 및 이의신청	사업단 ↔ 주관기관	· 종합평가의견(평가점수 제외) 주관기관 통보 및 선정방법, 절차 하자 등 이의신청 접수
종합심의	사업심의위원회 (과기부)	· 최종 심의 후 신규 과제 확정

- (협약체결) 신규과제로 선정된 주관기관과 사업수행에 대한 협약을 체결하고 예산집행 계획에 따른 정부출연금(사업비) 지급('20.5월)
- (평가) 연차, 단계(1단계, 2단계, 3단계), 최종평가로 구분되며 산학연 전문가로 구성된 평가위원회에서 연구결과보고서에 대한 검증·평가를 실시(매년 12월)

- (사업비 정산) 집행잔액 반납 및 불인정 연구비 반납 등 연차별 정산(차기년도 2월)
- (성과활용 보고 및 평가, 사업화) 연구개발 과제의 결과물에 대한 지속적인 성과관리를 위해 매년 성과활용보고서를 작성
 - － 해당년도 성과활용조사를 통해 개선사항을 도출하고 차기년도 성과활용조사에 반영
 - － 연구결과물이 제대로 활용 여부를 자체적으로 검증하고 필요 시 해당 분야 전문가들을 활용하여 검증
- ※ 연구개발사업의 관리 및 평가는 국가연구개발사업의 관리 등의 규정 준용

□ (대·소분야 전문센터) 평가체계 및 추진절차

① 평가개요

- 평가의 기본방향
 - － 과학기술정통부 「국가연구개발 과제평가 표준지침」 준수
 - － 동 지침의 취지 및 내용을 고려하여 평가지침을 마련하고 과제평가를 수행(연구성과평가법 시행령 제9조)
 - ※ 과기정통부가 연구개발 과제평가에 표준지침을 마련하여 제공토록 법에서 규정(연구성과평가법 8조)
 - － 전문센터 구축 및 운영 사업의 성과목표 및 과제 목표와 연계평가
 - － 전략목표에 기반하 성과목표가 설정되고 이에 따른 성과지표를 설정
 - － 사업의 성과목표에 근거하여 과제단위 목표와 이에 따른 성과지표를 설정
- 「국가연구데이터 공유·활용 체제 구축」 사업의 우선분야 고려

① ICT·S/W	⑥ 농림수산·식품	⑪ 재난안전
② 생명·보건의료	⑦ 우주·항공·해양	
③ 에너지·자원	⑧ 건설·교통	
④ 소재·나노	⑨ 환경·기상	
⑤ 기계·제조	⑩ 원자력·핵융합	

※ 우선분야 선정을 위해 정량적 평가지표를 선정하고 AHP분석을 통해 분야별 최종점수 도출 및 우선순위를 기 도출함

- － 전문센터의 해당분야에서의 역할과 범위 평가
- － 전문센터가 속한 해당분야와 분야 내에서의 전문센터의 역할과 연구데이터 확보 범위에 대한 평가
- － 전문센터의 연구데이터 활용도, 활용 시 연구성과에 대한 파급성 평가

② 평가단계

단계	과학기술정보통신부	전문관리기관	주관연구기관
시행계획 수립	연도별 사업시행계획 수립	사업시행계획(안) 마련	전문센터 구축 및 운영 신규사업 공모
	▶ 전문센터 구축 및 운영 신규사업 공모	▶ 전문센터 구축 및 운영 신규사업 공모	
선정평가	선정평가계획(안) 승인	선정평가계획(안) 마련	▼
	◀ 신규 사업 선정결과 확정 (사업추진위 심의)	▶ 선정평가 실시	▶ 사업계획서 작성 및 발표
	▶ 신규 사업 선정결과 통보	▶ 연구개시	
중간평가	중간평가계획(안) 검토 및 승인	중간평가 대상 과제 평가계획(안) 마련	중간보고서 제출
	◀ 중간평가결과 및 조치 확정 (사업추진위 심의)	▶ 중간평가 실시	▶ 구축결과 중간발표
	▶ 중간평가결과 및 조치계획 통보	▶ 구축 및 운영 계속지원 또는 중단	
최종평가	최종평가계획(안) 검토 및 승인	최종평가 대상 과제 평가계획(안) 마련	최종보고서 제출
	◀ 최종평가결과 및 조치사항 확정	▶ 최종평가 실시	▶ 전문센터 구축 및 운영결과 발표
	▶ 평가결과 통보 및 조치사항 수행	▶ 최종평가결과 확인	

- 평가단계를 선정평가 → 중간평가 → 최종평가로 구분 및 정의하여 순차적으로 평가를 진행함
- 평가단계별 구분 및 정의

평가단계	정의
선정평가	구축단계의 시작단계에서 얻어지게 되는 이익을 명확히 하고, 사업계획서 수행의 필요성을 정의하고 목적을 설정하며, 다른 선택 가능한 요소들을 찾아 비교·검토해 보고 비용 대비 이익을 고려하여 최종적인 선택을 하는 일련의 활동
중간평가	당초 설정된 구축계획 대비 진행과정을 점검하고 구축진행 과정에서 장·단점을 체계적으로 점검하여 수행중인 전문센터 구축의 개선을 위한 정보를 제공하는 과정 단위·소규모 구축 및 운영 진행 중에 연차실적 계획서를 검토하는 연차점검과 중·대규모 구축 및 운영 단계 종료 시에 수행하는 단계평가로 구분
최종평가	전문센터 구축 사업이 계획 대비 추진과정과 결과가 성공적으로 수행되었는지를 점검 하는 활동

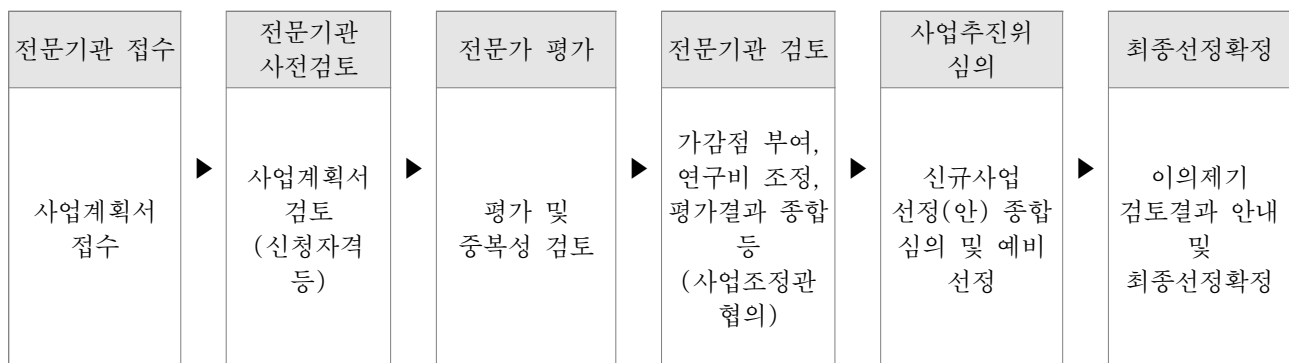
<표 8-4> <전문센터 구축 및 운영지원사업 평가단계 정의>

- (선정평가) 신청 받은 국가연구데이터 전문센터 사업계획서에 대한 검토와 심의 등을 거쳐 신규로 지원할 대상 전문센터를 선정
- (중간평가) 전문센터 구축을 위한 구축기간동안 산출된 구축성과 및 업적에 대한 심층적 평가를 통해 사업 기본목표 달성 확인 및 계속 지원여부를 결정
- (최종평가) 최종 구축 결과물의 목표 달성도 및 구축·운영의 성실수행 점검을 통해 전문센터 구축 성과의 발굴 및 확산

○ 선정평가

- 선정평가 절차는 전문센터구축 사업계획서 접수→사전검토 → 전문가 평가 및 중복성 검토 → 전문기관 검토 → 사업추진위 심의→예비평가 결과 공개→이의제기신청→최종 선정결과 안내 및 평가자정보 공개 순으로 진행
- 선정평가 단계 절차 개요

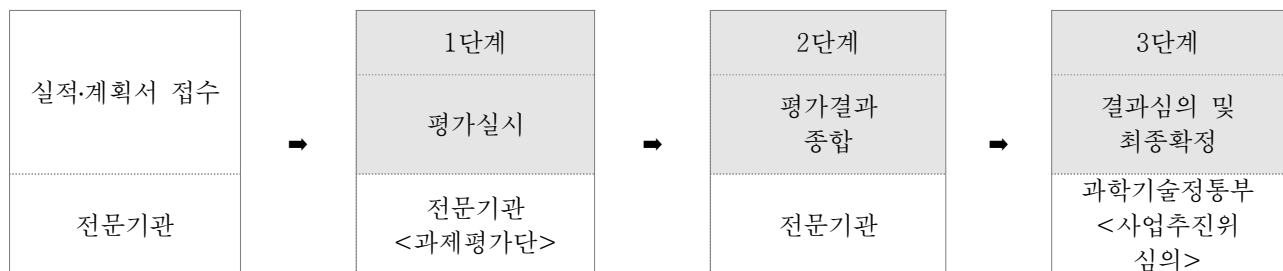
<표 8-5> 선정 평가단계 절차



- (사업공고) 과기정통부가 전문센터 구축 및 운영 신규 사업을 공고
- (전문기관 접수) 전문기관이 전문센터 구축 및 운영사업을 수행하거나 참여하려는 자의 사업계획서를 제출받음
- (사전검토) 전문기관이 신청기관(연구책임자)의 연구수행 상한제도 준수여부, 국가연구개발 사업 참여제한 여부, 암행 위반여부 등 신청자격 등을 검토
- (전문가평가) 전문기관이 사업 공고시 안내된 선정평가기준에 따라 산·학·연 전문가로 구성된 평가위원회가 평가 진행
 - ※ ‘평가’단계에서 서면평가, 발표평가, 토론평가를 하나의 평가방법으로 정의하고 독립적으로 운영이 가능
- (전문기관 검토) 전문기관이 평가결과에 따라 예산규모를 조정하며, 사업 목적에 부합하는 전문센터 선정을 위한 가감점 제도가 운영됨
- (사업추진위원회 심의) 과학기술정통부는 전문기관이 제출한 종합평가서 등을 검토, 평가결과의 타당성 등을 심의하여 지원예산 규모 내에서 전문센터 구축 및 운영 선정(안)을 종합 심의하여 예비 선정함
- (예비평가결과 공개 및 이의신청) 전문기관은 사업별 예비평가결과를 공개하고 특정 기간 이내에 이의제기 신청을 받음
- (최종선정결과 안내 및 평가자 정보 공개) 이의제기 검토결과 안내 및 최종선정을 확정하고, 평가위원 명단을 공개함

○ 중간평가

<표 8-6> 중간 평가단계 절차

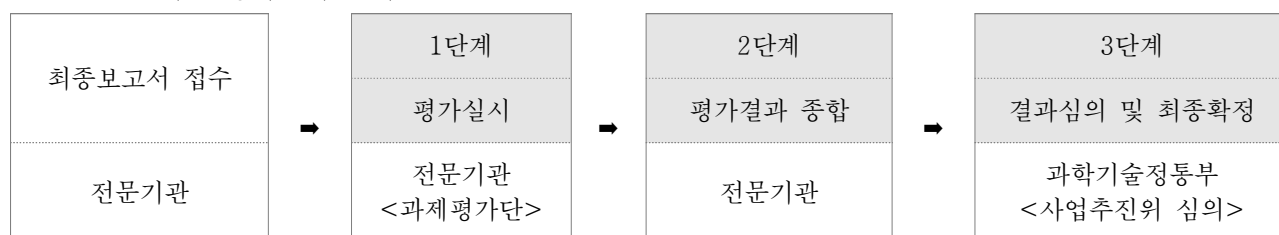


- (기본방향) 전문센터 구축계획 대비 진행과정을 점검하고 구축진행 과정에서 장·단점을 체계적으로 점검하여 수행중인 사업의 개선을 위한 정보를 제공
- 평가 결과를 계속, 성실중단, 불성실중단, 조기완료 등으로 구분하고 중단이 결정된 전문센터 구축에 대해서는 지원이 중단됨
- (평가유형) 과제의 유형 및 평가시기 등을 고려하여 평가유형 구분
- 연차평가는 연차실적·계획서를 제출받고, 이에 기반하여 전년도 실적과 차년도 구축계획을 점검함

- 중간모니터링은 수행진도 및 결과를 점검하고, 부처·전문기관이 필요하다고 판단하여 구축 추진방향 등에 대한 컨설팅을 제공함으로써 연차평가를 대체함
- 단계평가는 구축단계가 종료된 시점에 평가를 통해 단계별 목표 달성도와 차기단계의 구축계획 등을 점검함
- (평가방식) 전문센터의 특성을 고려하여 적합한 평가 방식을 활용
- 연차/단계평가는 평가위원을 통해 실적을 점검하고 향후 계획을 평가하며, 필요시 발표 평가 혹은 현장평가 등의 방식으로 실시함
- 중간모니터링은 성과요약서, 연차실적 및 차년도 계획서 등 간소화된 평가 자료에 기반한 서면검토를 원칙으로 하며, 필요 시 해당분야 전문가를 포함하여 실시하도록 함
- (평가기기) 단계별로 구분되는 연구이기 때문에 단계가 종료되는 연도에 단계평가를 수행하며 단계평가 중간에는 중간모니터링을 수행
- 이때 단계종료 시기는 연구자가 수행일정에 근거하여 직접제시하며, 이 일정에 따라 평가위원 또는 전문기관이 점검 실시
- (결과활용) 중간평가의 결과에 따라 사업의 계속 수행여부를 결정하며, 지원 연구비 증감, 우수 사업 인센티브 등 부가적인 의사결정의 근거로 활용
- 중간모니터링 결과에 따라, 사업 성과나 수행 진도가 부진한 경우 등에 대해 연차평가를 추가적으로 실시할 수 있도록 함
- 구축결과가 불량하여 실패로 판정되거나 중단된 경우에 대해, 참여 제한 및 정부출연금 환수 등의 제재를 가할 수 있도록 함

○ 최종평가

<표 8-7> 최종 평가단계 절차



- (기본방향) 최종목표 달성도, 성과목표 달성도, 사업성과의 질적 우수성 등에 대해 평가함
- 대표 사업성과의 질적 우수성에 대한 평가는 사업에 부합하는 성과
- 사업목표와 사업성과의 비중은 40:60의 수준에서 조정이 가능함
- (평가항목) 사업별 대표 사업결과의 질적 우수성을 중심으로, 기존에 설정된 사업목표에 대한 계획대비 달성도, 수행결과 및 활용실적, 기대효과 등을 평가

- (평가방식) 사업 특성에 따라 질적 우수성을 평가하는 정성평가와 더불어, 사업에 대한 정량평가를 병행할 수 있음
- 사업의 성공 및 실패 여부를 평가위원이 정성평가를 통해 판단하며, 정량평가는 목표달성도 등 평가항목 중 일부에 대해 허용함
- (결과활용) 최종평가의 결과에 따라 사업에 대한 인센티브 및 제재 등의 후속조치에 활용
- 사업착수시점에 설정한 과제목표를 초과달성하거나 도전적으로 설정했던 목표를 달성하는 등의 우수한 성과를 달성한 연구자는 추후의 과제 선정평가에서 가점을 부여하는 등으로 우대
- 당초 제시했던 최종 목표의 달성에 실패하더라도, 대표연구성과의 우수성이 인정될 경우 사업을 실패로 판정하지 않음
- 최종평가 결과 불량하여 실패로 판정된 경우, 참여 제한 및 정부 출연금 환수 등의 제재를 가할 수 있도록 함

③ 평가항목

○ 선정평가

- 우수한 성과활용을 위해 선정평가단계에 성과활용 관련 항목을 통해 원천기술 확보 및 활용계획을 구체적으로 제시하도록 함

<표 8-8> 선정평가 시 평가항목

구분	평가항목	비고
사업 계획	전문센터 구축 및 운영계획의 우수성 및 타당성	
	구축방법(전략)의 합리성 및 명확성	
사업 역량	구축 책임자의 실적 및 수행능력	
	참여 실무자 역량의 우수성	
성과 활용	전문센터 연구데이터 및 원천기술의 확보가능성	전문센터 연구데이터 및 원천기술 확보계획의 구체성과 타당성
	성과창출전략의 적절성	성과목표 및 전략의 타당성

○ 중간평가

- 창출된 성과 및 연구성과 소개서 등을 통해 질적 성과평가를 강화하고 성과평가지표를 적용하여 성과의 질적 우수성을 평가

<표 8-9> 중간평가 시 평가항목

구분	평가항목	비고
사업 단계	단계(연차)목표의 달성도	
	사업성과의 질적 우수성	- 성과분야별 질적 우수성 판단
사업 계획	사업목표의 적절성·구체성·실행가능성	
	수행방법의 적절성	
	성과목표관리의 적절성	- 연차(단계)별 성과목표 및 전략의 타당성

○ 최종평가

- 선정평가 시 제출된 연구계획서에서 제시한 연구목표 및 성과목표 대비 달성도와 성과 요약서, 연구성과 소개서를 바탕으로 대표연구 성과의 질적 우수성을 종합적으로 평가

<표 8-10> 최종평가 시 평가항목

구분	평가항목	비고
연구 성과	연구목표의 달성도	계획 대비 연구목표 달성여부
연구 결과	연구성과의 질적·양적 수준	연구성과의 질적 우수성 성과의 파급효과
연구 계획	연구결과 활용계획의 구체성·가능성	
	성과의 기대효과	

④ 평가위원 구성

○ 평가위원 Pool 구성

- NTIS에서 보유·관리하고 있는 평가위원 후보단 중에서 세부기술별로 적정 규모의 전문가를 확보하여 평가의 전문성을 유지, 필요에 따라 해외전문가 참여
- 한국 연구재단이 보유하고 있는 한국학술진흥재단의 대학 연구자 그룹 DB, 한국과학재단의 출연 연구기관 DB를 활용

○ 평가위원 검증

- 한국연구재단이 구축하고 있는 연구DB에 대해 연초 1회(필요시 2회)의 평가위원 풀 확정 절차를 통해 평가위원에 대한 이력을 검증
- 교수급 또는 최근 5년내 논문 3편 이상의 실적이 있는 연구자 등 전문성을 가진 평가위원 후보는 평가위원 검증 없이 개인이 동의하면 평가위원 후보풀에 등록

－ 특정 사업에 대해서는 별도의 모집 공고를 통해 평가위원을 추가

⑤ 평가위원 추출·선정방식

- (구성절차) 평가위원 추출은 1차추출 --> 평가위원군 구성 --> 평가위원군 우선순위 설정 --> 섭외 절차 순으로 하고 각 절차별로 담당하는 주체를 다르게 설정하고, 평가위원 추출절차를 모두 시스템상에 기록하여 평가위원 선정 시 투명성을 확보
- (구성원칙) 선정기준 상의 평가위원 선정방법 및 원칙, 평가위원 제외대상 등 과기정통부의 선정기준에 따라 구성
 - － 평가위원은 평가위원 후보단(후보 pool) 중에서 선정하는 것을 원칙으로 하며, 평가단·관련부처 평가단 위원, 1명 이상의 기술사업화위원을 포함하여 산·학·연 전문가 7명 내외로 평가위원회 구성
 - － 실용화/성과활용의 중요성이 강조되는 경우, 산업계의 사업화전문가 등을 평가위원으로 포함
- (제척기준) 평가위원 구성 시 피평가자와 상관관계 분석을 통해 상관도 점수에 따라 배제
 - － 현재 소속 대학/단과대학/학부 동일, 소속기관 동일, 사제기간, 최종학위 지도교수 동일 등의 경우 제외
 - － 이해관계자를 연구개발과제 평가단에서 제외하여 평가의 공정성을 유지

<표 8-11> 제척 대상 및 제척기준

구분	제척기준
필수 제척대상	① 평가대상 과제의 참여연구원 ② 불성실·불공정한 평가경력이 있는 전문가 ③ 평가위원 참여자격 제한을 받은 전문가 ④ 법 제11조의2제1항에 따라 국가연구개발사업 참여제한을 받은 전문가
선택적 제척대상	① 중앙 행정기관의 공무원 및 소관 전문기관의 직원 ② 민법 제777조에 따른 친족관계인 자 ③ 상호간 평가자 ④ 평가대상과제의 연구책임자와 사제관계 ⑤ 평가대상과제의 연구책임자와 같은 기관에 소속된 전문가 ⑥ 그 밖에 평가의 공정성을 해할 염려가 있는 전문가

8.1.3. 성과활용 방안

- (사업 성과관리 단계와 주체) 사업단 및 운영단 성과는 1차 자체평가, 2차 성과관리 위원회 평가로 진행
 - (1차) 사업단장 및 운영단장은 사업의 주요 마일스톤 단계에 따라 연차별/사업 분야별로 제시된 성과지표에 근거하여 성과를 측정하여 보고서로 제출
 - － 사업단장(운영단장)의 역할
 - ① 연차별 연구개발 지원 방향, 변화관리 및 기술이전 진행 상황 모니터링 (운영단장은 장비구축 및 운영현황 모니터링)
 - ② 연차별 사업(운영)목표 달성을 평가 및 사업 진행(운영 현황) 보고서 작성(연차별 사업종료 2개월 전) → 성과관리 위원회 제출
 - ③ 성과위원회 조정의견을 수렴하여 사업의 재기획
 - (2차) 사업단장 및 운영단장이 제출한 성과보고서에 기반하여 성과관리 위원회는 사업단과 운영단의 성과를 점검, 실증하고 문제 진단 및 컨설팅
 - － 성과관리 위원회의 역할
 - ① 연차별 사업단 및 운영단의 연구개발 지원활동 관리 및 모니터링
 - ② 사업단 및 운영단에서 제출된 성과 보고서 검토 및 실증
 - ③ 사업단 연차별 성과 확정 및 재계약 여부 검토 (재계약 여부는 연구단 만족도 조사 등에 근거하고, 최초 실행 시기는 사업단 및 연구단의 안정적 사업운영 정착 시기를 고려하여 사업개시 후 3년차부터 시행)
- (기술개발 성과관리 단계와 주체) 성과는 3차에 걸쳐 평가되며, 1차적으로 요소기술별 연구단장 책임 하에 수행 → 2차적으로 책임평가단의 객관적 평가 수행 → 3차적으로 책임평가단의 보고에 기반하여 성과관리 위원회에서 최종 결정
 - (1차) 연구단장은 과제의 주요 마일스톤 단계에 따라 과제별/연차별 연구개발 추진여부를 평가
 - － 연구단장의 역할
 - ① 연차별 과제 수행 방향, 마일스톤에 따른 과제 진행 상황 모니터링
 - ② 연차별 목표 달성을 평가 및 과제 진행 보고서 작성 → 성과관리 위원회 보고
(성과관리 위원회 산하 기술분야별 책임평가 위원에게 제출)
 - ③ Design Review : 과제수행 중 변화관리 지원반과의 협력, 연구단장 협의회 를 통하여 외부환경 변화에 따른 연구개발 방향 재설정 및 재기획 등

- (2차) 연구단 자체 평가 결과를 토대로 책임평가단에서 성과를 점검하고 연구개발 과정상의 문제점 진단 및 컨설팅 → 연차별 성과점검 보고서 작성 → 성과관리 위원회에 제출
 - － 책임평가단의 역할
 - ① 연구단 자체 평가 결과 실증
 - ② 목표 미달성 연구단 애로요인 진단 및 컨설팅
 - ③ 성과관리 위원회 보고
- (3차) 책임평가단의 성과점검 보고서 기반 연구단별 성과 확정 및 인증
 - － 성과관리 위원회의 역할
 - ① 책임평가단 성과점검 보고서 점검
 - ② 목표 미달성 연구단에 대한 책임평가단 컨설팅 조치 내용 검토 및 조치 내용 확정
 - ③ 연차별 연구단 성과 확정 및 총괄 사업단을 통해 결과 보고서 공개

□ (성과 정의) 성과는 사업성과와 연구개발성과로 구분

- (성과창출 시나리오) 성과창출 시나리오는 연구개발 성과에 기반한 사업적 성과 측정을 위한 것으로 투입, 활동, 산출, 성과/영향으로 구분

<표 8-12> 성과창출 시나리오

투입		활동		산출		성과/효과
·연구개발비 ·사업단 운영비 ·장비구축/ 운영비 ·인력 ·장비	⇒	·연구개발 ·변화관리 지원	⇒	·지적재산(공동 원천기술 /차세대 원형플랫폼)		·기술력 제고
		·기술이전 지원	⇒	·대용량 연구데이터 활용 R&D 증가/ 규모 확대 ·대용량 연구데이터 분석 ·활용 기술 이용수 증가		·대용량 연구데이터 활용 산업육성 기반 마련
		·논문/특허 출원·등록 지원	⇒	·전략적 원천기술 관련 논문 /특허		·글로벌 기술경쟁력 인정
		·기술융합허브	⇒	·혁신네트워크 확대		·안정적 혁신체계 구축
		·테스트베드 운영	⇒	·성능 및 안전성 인증 기술 수 증가		·혁신주기 단축

- (사업성과) 본 사업 효과 측정을 위한 성과지표로 사업목표에 따른 대표지표와 세부지표를 구분
 - － (목표지표) 본 사업 목표지표는 기술수준 향상, R&D 인재육성, 연구생산성 제고로 구성
 - － 기술수준은 KISTEP의 기술수준 평가방법론을 준용하여 본 사업 수혜대상에 한정하여

자체 조사평가 실시 후 측정하고,

- R&D 인재육성은 본 사업을 통하여 개설된 빅데이터 및 연구데이터 교육 커리큘럼을 이수 받은 주체들의 수를 계산하여 측정
- 연구생산성 제고는 R&D 인재육성을 통해 배출된 연구원들의 R&D 수행 수, 결과 등을 계산하여 측정
- (사업 특화지표) 본 사업 목표지표는 혁신네트워크 조성, 즉 기술융합허브 운영 관련 사업성과 측정에는 한계가 있어 기술융합허브에 특화된 성과지표 적용
- (연구성과) 연구성과는 요소기술별, 전문센터별 세부과제 단위의 수행실적을 의미하며, 단계별 목표 수행률, 목표 결과물, 연구효율성 등으로 구성

<표 8-13> 연구 성과지표

대표지표	측정방법	세부지표 구성	
		산출지표	결과지표
연구 효율성	·성과관리 위원회의 최종성과 보고서에 근거하여 산출	·과제별 목표 달성율 - 연구개발 마일스톤에 따른 진도 달성의 정도 ·과제별 실제 투입율 - 연구개발 계획에 부합하는 인력(비용)투입 의 정도	·연구효율성 - 산식 : 목표달성율/투입율 (1이상일 경우 효율적, 1일 경우 계획대로 수행, 1미만일 경우 비효율적)

□ 타 연구분야의 연구사업과 연계활용

- R&D 성공과제의 결과를 활용하여 연구하고자 하는 타 연구 분야의 연구사업을 지원 할 수 있도록 연계
- 사용자 등록이 되지 않은 타 연구분야의 연구기관 또는 개별 연구자, 대학 연구 기관 등이 플랫폼이나 전문센터의 연구데이터 또는 분석 툴 등을 활용하고 할 때 실질적인 지원이 가능
- (지원절차) 연계 후보 연구기관 등의 연계(활용) 신청 → 적합성 평가(사업 운영주체) → 적합성 평가시 평가지표 적용 → 최종 심사 후 연계 및 연구사업 지원

□ 대용량 연구데이터 활용 공공조달 사업 참여 인센티브 제공

- 본 사업 기술을 활용하고 연구데이터를 이용하는 연구사업에 참여하는 연구기관이 대용량 연구데이터 활용 공공조달 사업에 참여하고자 할 경우 사업자 선정과정에서 인센티브 제공

- (지원절차) 연계 후보 연구기관 등의 연계(활용) 제안 → 적합성 평가(운영주체) → 적합성 평가시 가산점 적용 → 최종 심사 후 연계 지원

□ 정성적 측면의 성과활용

○ 과학기술적 성과활용

- 데이터에 근거한 분석-예측-최적화를 기반으로 과학기술 분야에 대한 통찰력(Insight)을 제공함으로써, 복잡하고 빠르게 변화하는 과학기술 환경 속에서 경쟁력 높은 과학기술을 확보할 수 있는 과학기술 분석 프레임워크를 제공함.
- 연구 빅데이터 경쟁력 함양을 위한 국가적, 사회적 기반을 정비하고, 산·학·연 협력을 통한 원천기술 개발로 빅데이터 핵심 기반 기술 확보
- 연구데이터 고성능 분석 플랫폼은 데이터 관리, 고급 분석(Advanced Analytics) 및 분야별/유형별 분석 애플리케이션을 통합 프레임워크로 제공함으로써, 과학기술 내 방대한 양의 데이터로부터 가치 있는 정보를 유연하게 제공함.
- 과학기술 연구 빅데이터의 공유·융합 체제는 학제간 연계·융합 연구를 활성화 시키고, 나아가 국가 과학기술 경쟁력을 제고함.

○ 사회·경제적 성과활용

- 국가연구데이터플랫폼 사용으로 세계 유명 연구기관들과의 협업 및 공동 연구를 통해 세계 시장에서 글로벌 기업/연구소보다 경쟁력 높은 기술을 제공함으로써, 날로 치열해지는 환경 속에서 국내 기업들이 새로운 비즈니스 기회를 발굴할 수 있는 패러다임을 제시
- 대용량 연구데이터의 관리, 해석, 정보 추출, 정보 분석, 정보 저장과 관리, 정보 검색과 질의, 가시화 등 빅데이터 관련 핵심 기술의 국산화
- 세계 주요 국가들이 연구데이터의 공유와 활용을 위해 국가적인 정책을 수립하여, 이와 관련하여 기후, 환경, 건강한 삶 유지와 수명 연장 등의 전 지구 및 인류 문제에 대한 공동 대처하고 있으며 우리나라도 이 플랫폼을 활용 가능
- 2011년 1월 인간게놈위원회 회장 리우 소장은 ‘한국이 같은 분야에서 중복 연구를 수행하여 국가 전체적인 낭비를 초래하고 있다’고 지적한 바 있음. 데이터가 체계적으로 관리 된다는 것은 데이터의 생성 주체, 데이터 생성 목적, 데이터 생성환경, 데이터 관리주체 등에 관한 통합 관리가 가능함을 의미함. 따라서 동일한 데이터를 생산하기 위해 예산이 중복으로 투자되는 것을 방지할 수 있으며, 과제 기획단계에서부터 데이터 단위의 유사연구를 조사하는 것으로 중복연구를 피할 수 있음
- 연구의 효율성 및 투명성과 연구결과물의 수준 제고할 수 있음. 즉, 연구데이터의 공동 활용을 통해, 연구자 본인이 직접 수집 및 가공하지 않았더라도 타 연구자들의 데이터

를 활용함으로써 다양한 연구의 시도가 가능

- 다양한 연구자들이 자유롭게 참여해 연구결과의 폭넓은 재연(replication) 과 입증을 통해 연구의 투명성과 연구의 질 제고에 기여할 수 있음
- 온라인 커뮤니케이션과 협력 데이터 분석을 촉진하는 소프트웨어 툴을 활용하면 다양하고 폭넓은 협업 연구의 가능성이 증가할 수 있음
- 데이터 생산, 전송, 재사용에서의 중복성 및 비용 감소연구생산성의 증대와 경제에 기여할 수 있는 지식창출이 가능함

8.2. 과제 · 소요예산 및 재원조달 방안

8.2.1. 전체 과제 및 사업비

□ 세부사업별 투자규모

○ 총 5,751억원 투자(정부 출연금 100%)

－ 본 사업은 국가연구데이터 플랫폼 구축 및 운영, 대·소분야 전문센터 구축 및 운영지원, 데이터 및 컴퓨팅 활용기반의 R&D 인재육성의 3개분야 20개 세부과제로 구성되어 있으며, 각 추진과제별 투자규모는 아래 참조

<표 8-14> 세부 사업별 투자 규모

(단위: 백만 원)

분야별		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계		32,075	41,195	57,887	61,332	69,388	64,176	65,151	63,534	60,598	59,850	575,187
국가 연구 데이터 플랫폼 구축 및 운영	소계	15,595	23,778	22,418	25,315	19,309	15,579	17,283	17,738	17,569	17,581	192,165
	연구데이터수집·연계및공유기술	1,804	2,522	2,405	2,739	2,066	1,671	1,861	1,909	1,898	1,899	20,774
	연구데이터저장·보안기술	3,315	5,324	4,984	5,544	4,186	3,385	3,771	3,869	3,846	3,848	42,072
	연구데이터품질·표준화기술	2,828	3,953	3,769	4,292	3,238	2,618	2,916	2,992	2,975	2,976	32,556
	지능형연구데이터고속 검색 기술	1,634	2,624	2,457	2,733	2,063	1,669	1,859	1,907	1,896	1,897	20,738
	연구데이터융합·분석환경구축기술	2,612	4,029	3,854	4,500	3,566	2,851	3,104	3,190	3,106	3,111	33,923
	연구데이터활용·플랫폼기술	3,402	5,326	4,948	5,508	4,190	3,387	3,772	3,871	3,848	3,850	42,101
대·소 분야 전문 센터 구축 및 운영 지원	소계	9,500	10,900	28,850	29,300	43,350	43,050	42,200	40,200	37,300	36,400	321,050
	바이오분야 전문센터	1,150	1,200	1,200	2,400	2,400	2,400	2,950	2,950	2,950	2,950	22,550
	소재분야 전문센터	2,850	4,200	5,400	5,150	4,700	4,400	4,000	4,000	3,100	2,200	40,000
	인공지능분야 전문센터	5,500	5,500	6,250	5,750	4,250	4,250	3,250	3,250	3,250	3,250	44,500
	농림수산·식품			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	환경·기상			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	기계·제조			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	우주·항공·해양			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	원자력·핵융합					4,000	4,000	4,000	4,000	3,500	3,500	23,000

분야별		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
	에너지·자원					4,000	4,000	4,000	4,000	3,500	3,500	23,000
	건설·교통					4,000	4,000	4,000	4,000	3,500	3,500	23,000
	재난안전					4,000	4,000	4,000	4,000	3,500	3,500	23,000
데이터 및 컴퓨팅 활용 기반의 R&D 인재 육성	소계	6,980	6,517	6,619	6,717	6,729	5,546	5,669	5,596	5,730	5,869	61,971
	ICT기반 기술교육	1,802	1,887	1,976	2,070	2,168	2,271	2,378	2,491	2,609	2,732	22,384
	분야별 전문 활용교육	1,678	1,129	1,143	1,147	1,061	1,076	1,090	906	921	937	11,088
	교육용 데이터뱅크 및 서비스 구축	3,500	3,500	3,500	3,500	3,500	2,200	2,200	2,200	2,200	2,200	28,500

□ 비목별 투자규모

<표 8-15> 비목별 투자 규모

(단위: 백만 원)

연도		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인건비	소계	8,010	9,410	17,409	18,794	25,054	23,806	23,797	22,906	21,614	21,213	192,013
	투입인력(명)	259	295	321	344	432	356	349	339	313	308	3,318
	인건비	31	32	54	55	58	67	68	68	69	69	570
직접비	소계	21,384	28,080	35,128	36,030	36,226	32,745	33,523	33,012	31,749	31,498	319,375
	연구장비/재료비	12,645	17,794	23,334	21,947	20,095	17,838	18,075	17,888	17,062	16,784	183,462
	연구활동비	1,883	2,583	3,225	4,052	4,781	4,354	4,531	4,437	4,249	4,207	38,303
	연구과제추진비	4,255	4,104	4,722	5,172	5,683	5,670	5,855	5,715	5,731	5,840	52,747
	연구수당	1,277	2,029	1,734	2,694	2,734	2,394	2,655	2,661	2,609	2,612	23,400
	위탁연구개발비	1,323	1,569	2,113	2,166	2,934	2,489	2,406	2,311	2,097	2,055	21,463
간접비		2,682	3,705	5,350	6,508	8,108	7,624	7,831	7,617	7,236	7,140	63,800
연구비총액		32,075	41,195	57,887	61,332	69,388	64,176	65,151	63,534	60,598	59,850	575,187

8.2.2. 세부 과제 및 사업비

□ 국가연구데이터 플랫폼 구축 및 운영

○ (총 사업비) 2021년부터 2030년까지 전체 192,165백만원 (단위: 백만 원)

세부과제	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	15,595	23,778	22,418	25,315	19,309	15,579	17,283	17,738	17,569	17,581	192,165
연구데이터 수집·연계 및 공유기술	1,804	2,522	2,405	2,739	2,066	1,671	1,861	1,909	1,898	1,899	20,774
연구데이터 저장·보안기술	3,315	5,324	4,984	5,544	4,186	3,385	3,771	3,869	3,846	3,848	42,072
연구데이터 품질·표준화기술	2,828	3,953	3,769	4,292	3,238	2,618	2,916	2,992	2,975	2,976	32,556
지능형 연구데이터 고속 검색 기술	1,634	2,624	2,457	2,733	2,063	1,669	1,859	1,907	1,896	1,897	20,738
연구데이터 융합·분석환경 구축기술	2,612	4,029	3,854	4,500	3,566	2,851	3,104	3,190	3,106	3,111	33,923
연구데이터 활용·플랫폼기술	3,402	5,326	4,948	5,508	4,190	3,387	3,772	3,871	3,848	3,850	42,101

○ 비목별 사업비 (단위: 백만 원)

비목		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인건비	소계	777	1,554	1,554	2,739	2,739	2,739	3,108	3,108	3,108	3,108	24,533
	투입인력(명)	10	20	20	36	36	36	40	40	40	40	318
	인건비	78	78	78	76	76	76	78	78	78	78	772
직접비	소계	13,345	19,883	18,863	19,468	13,416	10,078	11,111	11,560	11,450	11,460	140,632
	연구장비/재료비	9,713	14,430	14,430	12,904	6,716	4,551	5,051	5,481	5,550	5,550	84,374
	연구활동비	1,070	1,700	1,452	2,257	2,290	2,005	2,224	2,229	2,185	2,188	19,600
	연구과제 추진비	448	711	608	944	958	839	931	932	914	915	8,200
	연구수당	1,277	2,029	1,734	2,694	2,734	2,394	2,655	2,661	2,609	2,612	23,400
	위탁연구개발비	838	1,012	639	669	719	289	250	257	191	195	5,058
간접비		1,474	2,342	2,001	3,109	3,154	2,762	3,064	3,070	3,011	3,014	27,000
연구비총액		15,595	23,778	22,418	25,315	19,309	15,579	17,283	17,738	17,569	17,581	192,165

○ 인건비 소요내역

(단위: 백만 원)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	777	1,554	1,554	2,739	2,739	2,739	3,108	3,108	3,108	3,108	24,533
연구데이터 수집·연계 및 공유기술	84	168	168	296	296	296	336	336	336	336	2,650
연구데이터 저장·보안기술	170	340	340	599	599	599	680	680	680	680	5,370
연구데이터 품질·표준화기술	132	263	263	464	464	464	526	526	526	526	4,153
지능형 연구데이터 고속 검색 기술	84	168	168	295	295	295	335	335	335	335	2,647
연구데이터 융합·분석환경 구축기술	138	275	275	485	485	485	550	550	550	550	4,342
연구데이터 활용·플랫폼기술	170	340	340	600	600	600	681	681	681	681	5,372

○ 인프라 소요내역

(단위: 백만 원)

분야별	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	9,713	14,430	14,430	12,904	6,716	4,551	5,051	5,481	5,550	5,550	84,374
H/W	서버	5,550	8,325	8,325	3,330	4,163	2,220	2,498	2,775	2,775	42,735
	네트워크	1,110	1,665	1,665	1,221	278	444	555	555	555	8,603
	스토리지	1,665	1,665	1,665	2,664	1,166	500	749	833	833	12,571
S/W	DB	555	1,110	1,110	2,276	444	555	500	527	555	8,186
	WEB, WAS, 연계 등	555	1,110	1,110	2,276	444	555	500	527	555	8,186
	보안	278	555	555	1,138	222	278	250	264	278	4,093

□ 대·소분야 전문센터 구축 및 운영 지원

○ (총 사업비) 2021년부터 2030년까지 전체 321,050백만원

(단위: 백만 원)

세부과제	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	9,500	10,900	28,850	29,300	43,350	43,050	42,200	40,200	37,300	36,400	321,050
바이오분야 전문센터	1,150	1,200	1,200	2,400	2,400	2,400	2,950	2,950	2,950	2,950	22,550
소재분야 전문센터	2,850	4,200	5,400	5,150	4,700	4,400	4,000	4,000	3,100	2,200	40,000
인공지능분야 전문센터	5,500	5,500	6,250	5,750	4,250	4,250	3,250	3,250	3,250	3,250	44,500
농림수산·식품 전문센터			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
환경·기상 전문센터			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
기계·제조 전문센터			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
우주·항공·해양 전문센터			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
원자력·핵융합 전문센터					4,000	4,000	4,000	4,000	3,500	3,500	23,000
에너지·자원 전문센터					4,000	4,000	4,000	4,000	3,500	3,500	23,000
건설·교통 전문센터					4,000	4,000	4,000	4,000	3,500	3,500	23,000
재난안전 전문센터					4,000	4,000	4,000	4,000	3,500	3,500	23,000

○ 비목별 사업비

(단위: 백만 원)

비목		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인 건 비	소계	4,233	4,857	12,855	13,056	19,316	19,182	18,804	17,912	16,620	16,219	143,055
	투입인력(명)	60	76	161	163	241	240	235	224	208	203	1,811
	인건비	80	80	80	80	80	80	80	80	80	80	800
직 접 비	소계	4,216	4,837	12,803	13,003	19,238	19,105	18,728	17,840	16,553	16,154	142,478
	연구장비/ 재료비	2,932	3,364	8,904	9,043	13,379	13,287	13,024	12,407	11,512	11,234	99,088
	연구활동비	471	540	1,430	1,452	2,148	2,133	2,091	1,992	1,848	1,804	15,910
	연구과제 추진비	328	376	995	1,011	1,495	1,485	1,456	1,387	1,287	1,256	11,076
	연구수당	0	0	0	0	0	0	0	0	0	0	0
	위탁연구 개발비	485	557	1,474	1,497	2,215	2,200	2,156	2,054	1,906	1,860	16,405
간접비		1,051	1,206	3,192	3,241	4,796	4,763	4,668	4,447	4,126	4,027	35,517
연구비총액		9,500	10,900	28,850	29,300	43,350	43,050	42,200	40,200	37,300	36,400	321,050

□ 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성

○ (총 사업비) 2021년부터 2030년까지 전체 61,971백만원

(단위: 백만 원)

세부과제	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계	6,980	6,517	6,619	6,717	6,729	5,546	5,669	5,596	5,730	5,869	61,971
연구데이터의 R&D 응용을 위한 분야별 전문 활용교육	1,802	1,887	1,976	2,070	2,168	2,271	2,378	2,491	2,609	2,732	22,384
연구데이터 관리를 위한 ICT기반 기술교육	1,678	1,129	1,143	1,147	1,061	1,076	1,090	906	921	937	11,088
교육용 데이터뱅크 및 서비스 구축	3,500	3,500	3,500	3,500	3,500	2,200	2,200	2,200	2,200	2,200	28,500

○ 비목별 사업비

(단위: 백만 원)

비목		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
인 건 비	소계	3,000	3,000	3,000	3,000	3,000	1,885	1,885	1,885	1,885	1,885	24,425
	투입인력(명)	57	40	40	40	41	35	35	35	35	35	393
	인건비	53	75	75	75	73	54	54	54	54	54	620
직 접 비	소계	3,823	3,360	3,462	3,560	3,572	3,562	3,684	3,612	3,745	3,885	36,264
	연구장비/ 재료비	0	0	0	0	0	0	0	0	0	0	0
	연구활동비	343	343	343	343	343	216	216	216	216	216	2,793
	연구과제 추진비	3,480	3,017	3,119	3,217	3,229	3,346	3,469	3,396	3,530	3,669	33,471
	연구수당	0	0	0	0	0	0	0	0	0	0	0
	위탁연구 개발비	0	0	0	0	0	0	0	0	0	0	0
간접비		158	158	158	158	158	99	99	99	99	99	1,283
연구비총액		6,980	6,517	6,619	6,717	6,729	5,546	5,669	5,596	5,730	5,869	61,971

8.2.3. 재원조달방안

- 본 사업은 '21~'30년까지 10년간 총 사업비 5,751억 원(국고 100%)을 조달하여 운영
 - 총 사업비 5,751억 원은 과학기술정통부가 정부출연금 형태로 지원
 - － 정부출연금은 과학기술정통부 연구개발 사업비 회계로부터 조달
- 기존 재원범위 내에서의 수용 가능성
 - 기존 * 분야 연구개발과제는 본 사업으로 포함시켜 계속과제 형태로 진행될 예정
 - ※ KISTI 국가연구데이터 플랫폼 구축 및 운영사업(국가연구데이터 플랫폼 구축 및 운영에 포함), KOBIC 생명연구자원 통합시스템 활용기반 구축사업(대소분야 전문센터 구축 및 운영지원에 포함)

8.2.4. 소요인력

- 투입인력 계획
 - (총 투입 인력) 2021년부터 2030년까지 전체 3,318명 인력 투입

<표 8-16> 추진과제별 인력운영 규모

(단위: 명)

분야별		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
총계		259	295	256	281	288	215	215	214	185	170	2,378
국가 연구 데이터 플랫폼 구축 및 운영	소계	142	179	121	141	149	82	79	80	70	71	1,114
	연구데이터수집·연계및공유기술	34	12	10	16	13	7	8	8	8	8	123
	연구데이터저장·보안기술	16	58	39	31	26	14	15	15	15	15	246
	연구데이터품질·표준화기술	54	19	16	25	20	11	12	12	12	12	193
	지능형연구데이터 고속 검색 기술	8	29	19	15	13	7	8	8	8	8	121
	연구데이터융합·분석환경구축기술	2	4	4	28	50	29	21	22	12	13	184
	연구데이터활용·플랫폼기술	29	58	33	25	27	14	15	15	15	15	248
대·소 분야 전문 센터 구축 및 운영 지원	소계	259	295	321	344	432	356	349	339	313	308	3,318
	바이오분야 전문센터	142	179	121	141	149	82	79	80	70	71	1,114
	소재분야 전문센터	34	12	10	16	13	7	8	8	8	8	123
	인공지능분야 전문센터	16	58	39	31	26	14	15	15	15	15	246
	농림수산·식품	54	19	16	25	20	11	12	12	12	12	193
	환경·기상	8	29	19	15	13	7	8	8	8	8	121
	기계·제조	2	4	4	28	50	29	21	22	12	13	184
	우주·항공·해양	29	58	33	25	27	14	15	15	15	15	248
	원자력·핵융합	60	76	161	163	241	240	235	224	208	203	1,811
	에너지·자원	15	15	15	20	20	20	25	25	25	25	205
	건설·교통	39	55	73	73	71	71	71	69	51	35	608
	재난안전	6	6	7	7	7	7	5	5	4	4	58
데이터 및 컴퓨팅 활용 기반의 R&D 인재 육성	소계	0	0	16	16	18	18	17	16	16	17	134
	ICT기반 기술교육	0	0	16	16	18	18	17	16	16	17	134
	분야별 전문 활용교육	0	0	16	16	18	18	17	16	16	17	134
	교육용 데이터뱅크 및 서비스 구축	0	0	16	16	18	18	17	16	16	17	134

8.2.5. 중장기 재정운영계획과의 연계

- 본 사업은 10년간('21년~'30년) 총사업비 5,751억원을 투입
- 과기부의 '13년~'18년 연구개발 예산의 연평균증가율(CAGR)은 약 3.6%

<표 8-17> 연도별 과학기술정보통신부 연구개발 예산 추이

구분	2013	2014	2015	2016	2017	2018	연평균증가율 (CAGR)
과학기술정보통신부 R&D 예산(억원)	57,008	60,839	64,967	65,571	67,730	68,110	3.6%

- '13년~'18년 과기부 연평균 R&D 예산 증가율을 적용하여 향후('20년~'29년) 과기부 R&D 예산 증가분(전년 대비 증가분)을 추산하면, 연평균 약 3,074억원의 증가분 발생
- － 본 사업 예산은 과기부 R&D 예산 증가액 중 약 12.1%~22.7%를 차지하는 것으로 나타나 예산 증가분으로 충분히 충당 가능할 것으로 판단

<표 8-18> 향후 과기부 R&D 예산 추정액과 본 사업의 비중

(단위 : 억원, %)

구분	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030
과기부 R&D 예산 추정	73,134	75,784	78,529	81,374	84,322	87,377	90,542	93,822	97,221	100,743	104,370
전년대비 R&D 예산 증가액 추정	2,557	2,649	2,745	2,845	2,948	3,055	3,165	3,280	3,399	3,522	3,649
본사업 연도별 투자계획	—	321	412	579	613	694	642	652	635	606	599
R&D 예산 증가액 중 본사업 예산의 비중	—	12.1%	15.0%	20.4%	20.8%	22.7%	20.3%	19.9%	18.7%	17.2%	16.4%

사전타당성 분석

- 9.1. 사전타당성 분석 방법론
- 9.2. 정책적 타당성 분석
- 9.3. 기술적 타당성 분석
- 9.4. 경제적 타당성 분석
- 9.5. 기대효과 및 파급효과

9.1. 사전타당성 분석 방법론

□ 사전타당성 검토 지표

<표 9-1> 사전타당성 검토 지표

2계층 항목	3계층 항목	평가내용(작성항목)	검토결과 (반영여부)	검토근거 (정리 위치)
정책의 일관성 및 추진체제	상위계획과의 부합성	1. 국가정책의 실현을 위해 정책적 일관성을 가지고 추진되었는가?	×	2.2.1. 3.2.2.
	사업 추진체제 및 추진의지	1. 연구개발사업 거버넌스를 실현하였는가?	×	8.1.1
		2. 사업 추진의지와 선호도를 확보하였는가?	○	2.3.2. 2.3.3.
사업 추진상의 위험요인	재원조달 가능성	1. 재원을 조달하는 과정에서 예상되는 위험 요인을 분석했는가?	○	8.2.1.
	법·제도적 위험요인	1. 법·제도와 규정에 부합하여 실행되는가?	×	3.3.1. 8.1.3.
과학기술 개발계획의 적절성	기획과정의 적절성	1. 기획에 참여한 전문가 집단은 적절하게 구성되었는가?	○	1.2.2.
		2. 우선순위 설정과정은 합리적으로 진행되었는가?	○	6.6.
	사업목표의 적절성	1. 해결하여야 할 문제나 이슈가 적절하게 진술되어 있는가?	○	1.1. 3.2.
		2. 사업목표는 달성하고자 하는 효과를 구체적으로 제시하고 있는가?	×	1.2.3.
		3. 사업목표와 해결할 문제와의 연관관계가 존재하는가?	×	1.2.3.
		4. 사업성과에 대한 수혜자의 포적화가 적절히 이루어졌는가?	○	1.2.3.
	구성 및 내용의 적절성	1. 적절한 수준의 세부 활동을 도출하여 제시하고 있는가?	○	5.3.
		2. 세부 활동이 사업목표와 논리적으로 연계되는가?	○	5.3.
		3. 세부 활동의 성과지표를 적절히 제시하고 있는가?	○	5.3.
		4. 세부 활동의 기간 추정과 선후관계는 논리적인가?	○	5.3.
과학기술 개발 성공가능성 기존 사업과의 중복성	-	1. 기술추세 분석	○	3.1.4.
		2. 과학기술 수준 분석	○	3.1.4.
	-	1. 사업의 중복성	×	2.2.2.
		2. 세부 활동의 중복성	×	2.2.2.

9.1.1. 개요

□ 법적근거

- 예비타당성조사는 「국가재정법」 제38조 및 동법 시행령 제13조의 규정에 따라 대규모 신규 사업에 대한 예산을 편성하기 위해 기획재정부장관의 주관으로 실시함

“제38조(예비타당성조사) ①기획재정부장관은 총사업비가 500억 원 이상이고 국가의 재정지원 규모가 300억 원 이상인 신규 사업으로서 다음 각 호의 어느 하나에 해당하는 대규모사업에 대한 예산을 편성하기 위하여 미리 예비타당성조사를 실시하고, 그 결과를 요약하여 국회 소관 상임위원회와 예산결산특별위원회에 제출하여야 한다. 다만, 제4호의 사업은 제28조에 따라 제출된 중기사업계획서에 의한 재정지출이 500억 원 이상 수반되는 신규 사업으로 한다.”

「국가재정법」 제38조

- － 한정된 정부지원재원을 효율적으로 사용하기 위해 제 3자가 전체적인 시각에서 재정사업의 타당성 여부 등을 사전 검토하는 과정임
- － 객관적이고 중립적인 조사를 통해 예산 낭비를 방지하고 한정된 재원을 효율적으로 사용하는 것을 목적으로 함
- － 대형 사업의 실패는 그 피해와 규모가 장기적으로 발전할 가능성이 있으므로, 재정사업의 타당성과 더불어
- － 사업 추진에 있어서의 위험요소 및 대안의 검토 등을 사전에 검토하여 예산 낭비를 방지하고자 함
- 국가연구개발사업에 대한 예비타당성조사는 「국가재정법」 제8조의2, 제38조의2,3의 규정에 따라, 기획재정부 장관이 과학기술정보통신부장관에게 위탁하여 실시함

“제8조의2(재정사업 평가 등에 대한 전문적인 조사·연구기관의 지정 등) ① 기획재정부장관은 주요 재정사업 평가 등을 적절하게 수행하기 위하여 「정부출연연구기관 등의 설립·운영 및 육성에 관한 법률」에 따라 설립된 한국개발연구원 및 한국조세재정연구원과 전문 인력 및 조사·연구 능력 등 대통령령으로 정하는 지정기준을 갖춘 기관을 전문기관으로 지정하여 다음 각 호의 업무 중 전부 또는 일부를 수행하게 할 수 있다.”

「국가재정법」 제8조의2

“제38조의2(예비타당성조사 결과 관련 자료의 공개) 기획재정부 장관은 제38조1항에 따른 예비타당성조사를 제8조의2제1항제2호의 업무를 수행하는 전문기관에 의뢰하여 실시할 수 있으며, ...”

「국가재정법」 제38조의2

“제38조의3(국가연구개발사업 예비타당성조사의 특례) ① 기획재정부장관은 제8조의2, 제 38조 및 제38조의2에 규정된 사항 중 「과학기술기본법」 제11조에 따른 국가연구개발사업에 대한 예비타당성조사에 관해서는 대통령령으로 정하는 바에 따라 과학기술정보통신부장관에게 위탁할 수 있다.”

「국가재정법」 제38조의3

- 또한, 「국가연구개발사업의 관리 등에 관한 규정」 제1조에서 국가연구개발사업의 기획·관리·평가 및 활용을 규정함

“제1조(목적) 이 영은 「과학기술기본법」 제11조, 제11조의2부터 제11조의4까지 및 제16조의2에 따른 국가연구개발사업의 기획·관리·평가 및 활용 등에 필요한 사항을 규정함을 목적으로 한다.”

「국가연구개발사업의 관리 등에 관한 규정」 제1조

□ 방법의 방향 및 내용

- 합리적인 의사결정 지향
 - － 기초자료 분석, 평가항목별 의미 및 분석방법, 분석결과 정리 및 보고, 의사결정 방식의 과정을 통해 객관적인 입장, 평가항목 간 중복 및 항목의 편향성을 방지함
- 법에 기반한 구체적 분석
 - － 「국가연구개발사업 예비타당성조사 운용지침」 및 「국가연구개발사업 예비타당성조사 수행 총괄지침」에 기반 하여 분석의 구체적인 내용을 구성함
 - － 합리적인 의사결정을 지원하기 위해 기초자료 분석, 과학기술적 타당성 분석, 정책적 타당성 분석, 경제적 타당성 분석 등으로 재정사업을 분석

9.1.2. 영역별 타당성 분석 수행 체계

가. 정책적 타당성 분석체계

- 사업의 특수성을 반영하여 정당성을 확보하기 위한 분석
 - － 재정사업의 원활한 진행과 재정사업의 성공을 분석하는 분석과 더불어 사업 추진의 정책적·시기적 정당성을 확보하기 위함
 - － 계량화의 어려움으로 과학기술적·경제적 타당성에서 고려되지 않지만, 사업의 특수성을 반영하여 국가연구개발사업이 가지는 불확실성, 특수성 그리고 복잡성을 고려함
- 일반적으로 ‘정책의 일관성 및 추진체제’와 ‘사업 추진상의 위험 요인’ 등 2가지로 범주화하여 분석을 수행함

□ 정책의 일관성 및 추진체제

① 상위계획과의 부합성

- 국가수준의 파급효과 및 정부차원의 정책적 합의를 평가하기 위한 항목
 - － 국가연구개발사업의 수행, 결과물의 활용, 사업파급 등이 주관부처에서 그치는 것이 아니라 국가전역에 미침
 - － 따라서 정부의 정책 동향뿐만 아니라 부처 간 정책적 합의를 통해 국가수준의 파급효과를 대비할 수 있는 정책적 대응 및 연계가 필요함

② 사업 추진체제 및 추진의지

- 사업 주체 간 적극적인 협조를 파악하고 사업의 거버넌스 체계를 평가하기 위한 항목
 - － 효율적인 국가연구개발사업 거버넌스를 체계적으로 수립하여 사업에 참여하는 주체 간에 협력으로 성공적인 정책목표 달성을 목적으로 함
 - － 사업 참여주체는 법·제도에 준수하여 선정되어야 하며, 참여 예정 주체들의 사업의지 및 이해관계자들의 선호도가 반영되어야 함

□ 사업 추진상의 위험요인

① 재원조달 가능성

- 재원조달에 있어 위험요인과 해당 위험요인에 대한 대응방안의 수립 여부를 검토하는 항목
 - － 타당성조사의 대상이 되는 국가연구개발사업은 대규모의 투자를 수반하는 사업으로, 사

업비 부담주체에 따라, 사업계획에 따라 재원조달의 위험요인을 식별·분석해야 함

- 사업 추진에 있어 법·제도적 위험요인이 예상될 경우 위험을 회피하거나, 피해를 최소화하기 위한 대응계획 마련이 필요함

② 법·제도적 위험요인

- 사업 추진 및 성과물과 관련된 타 법령과의 충돌 가능성을 평가하기 위한 항목
 - 연구개발의 산출물을 전달 및 활용하는데 있어 예상되는 위험요인을 파악하며, 이와 관련되는 국내·외 법·제도적인 위험요인을 식별하고 분석함
 - 사업 추진에 있어 법·제도적 위험요인이 예상될 경우 위험을 회피하거나, 피해를 최소화하기 위한 대응계획의 마련이 필요함

<표 9-2> 정책적 타당성 분석의 평가항목 구성

구분	2계층 항목	3계층 항목	의미
기본 평가 항목	정책의 일관성 및 추진체제	상위계획과의 부합성	정부차원의 정책적 합의
		사업 추진체제 및 추진의지	사업 거버넌스 및 사회적 합의
	사업 추진상의 위험요인	재원조달 가능성	정부·지자체·민간의 재정 여력
		법·제도적 위험요인	법·제도적 합의

자료: 국가연구개발사업 예비타당성조사 수행 세부지침(2018.4)

나. 기술적 타당성 분석체계

- 연구개발사업의 비정형성, 경제성 분석의 한계점을 보완하기 위한 분석
 - 건설사업 중심으로 추진되었던 예비타당성조사의 대상을 연구개발사업, 대량구매사업, 정보화 사업 등으로 확대하는 시기에 도입됨
 - 과학기술분야의 비정형성을 합리적으로 통제하고 국가연구개발사업 예산집행의 결과로써 예상되는 효과를 논리적인 인과관계로 제시하기 위하여 실시함
 - 향후 예상되는 문제를 다양한 관점에서 검토하고 이를 통해 해당 부처에서 제안한 사업이 목표하는 바를 이루기에 적절한 사업인지에 대한 정보를 제공함
- 기술관리 관점에서 과학기술개발의 성공가능성을 분석하는 ‘과학기술개발계획의 적절성’, 기술숙성 관점에서 분석하는 ‘과학기술 개발 성공가능성’ 그리고 ‘기존 사업과의 중복성’ 등 3가지로 범주화하여 분석을 수행함

□ 과학기술개발계획의 적절성

① 기획과정의 적절성

- 사업계획의 배경 및 경위의 적절성 및 합리성을 검토하는 항목
 - 기획단계에서 발생할 수 있는 오류로 인해 향후 사업 추진 시 발생할 수 있는 사업 부실화를 방지하기 위해 실시함
 - 당면한 문제를 정확하게 인식하고 이를 통해 연구개발사업의 필요성이 구체화되면, 예상 활용주체 등의 의견까지 종합하여 사업 수혜자의 요구를 구체화 함
 - 기획과정의 면밀함은 사업의 실패 가능성 및 산출물 무력화 가능성을 낮춤으로 미래의 사업 부실화 방지 가능성을 평가함

② 사업목표의 적절성

- 사업 배경, 문제/이슈 등의 적절성을 검토하는 항목
 - 사업이 추진된 배경, 동 사업을 통해 해결하고자 하는 문제/이슈 등이 사업의 목표와 연계하여 적절하게 수립되었는지를 검토함
 - 적절한 목표의 설정은 사업구조 뿐만 아니라 사업을 평가하는 기준으로 작용하여, 다양한 관점에서 의사결정을 내릴 수 있게 하는 핵심사항임
 - 사업의 목표는 사업의 추진 필요성 및 사업성과/영향과 밀접한 관계를 맺으며 이해관계자의 요구사항을 포함함
 - 또한, 사업목표는 과학기술개발계획의 관리 관점에서의 사업의 성과를 관리하여 사업 성공가능성을 분석하는 요소임

③ 구성 및 내용의 적절성

- 사업이 지향하는 방향 및 내용에 구성 및 내용이 부합하는 지를 검토하는 항목
 - 연구개발과제의 세부내용이 구체적이며 유기적으로 연계되어 제시된 사업의 구성 및 내용이 사업목표에 부합하는 지를 검토함
 - 국가연구개발사업은 연구개발이 가지는 불확실성, 투자회임기간의 장기성 및 사업별로 상이한 성과 등의 특성으로 인해 사업 추진이 어려울 수 있음
 - 적절한 사업의 구성 및 내용은 목표관리, 예산관리 및 일정관리를 통해 사업의 불확실한 요소를 최소화하고, 사업목표를 달성할 수 있도록 함

□ 과학기술 개발 성공가능성

- 기술 관리와 기술속성의 측면에서 연구하는 과학기술의 개발 성공가능성을 분석

- 기술추세 분석을 통해 기술 분야의 동향 및 산출물이 활용되어질 수 있는 경제·사회적 상황 등이 기술 개발에 대규모 투자가 이루어지기에 적절한 시기인지를 분석함
- 과학기술 수준 분석은 주요 연구개발 활동 주체와의 상대적인 기술격차를 분석함으로써 타 국가 대비 우리나라가 가지는 경쟁우위를 파악하는 분석임

□ 기존 사업과의 중복성

- 국가연구개발 포트폴리오의 균형이 적절히 유지되는가를 검토하는 항목
 - 기존 사업과의 중복을 지양하고 사업시행의 효과가 유사성 및 동일성 여부를 분석하여 국가기관의 특정분야에 대한 과잉투자를 방지함
 - 한정된 국가재정을 효과적으로 사용하고 정부의 연구개발투자가 균형을 이루도록 함

<표 9-3> 과학기술적 타당성 분석의 평가항목 구성

구분	2계층 항목	3계층 항목	의미
기본평가항목	과학기술 개발계획의 적절성	기획과정의 적절성	사업이 기획된 배경과 경위의 적절성
		사업목표의 적절성	문제 정의 및 목표 설정의 적절성
		구성 및 내용의 적절성	세부 활동 구성 및 내용의 구체성과 연계성
	과학기술 개발 성공가능성	-	대규모 투자 시점의 적절성 경쟁우위 기준 우선순위
	기존 사업과의 중복성	-	중복 가능성

자료: 국가연구개발사업 예비타당성조사 수행 세부지침(2018.4)

다. 경제적 타당성 분석체계

- 비용-편익의 논리적 인과성을 확보하기 위한 분석
 - 사업 시행과 미시행(with-without)의 관점에서 사업 추진을 위해 소요되는 비용과 사회 전반에 미치는 편익효과를 도출함

□ 비용 추정

① 비용 추정의 기초

- 국가연구개발사업을 수행함에 있어 예산 요건을 충족시키는 현실적인 비용을 추정하는 과정임

- 비용 추정의 정확성, 신뢰성, 시기의 적절성 등에 의해 비용 추정의 질(quality)이 결정됨

② 비용 추정방법론

- 비용 공학적인 관점에서는 일반적으로 유사사례분석, 통계분석, 세부분석, 전문가 의견 등의 방법 활용
 - 유사사례분석(analogy analysis)은 분석하고자 하는 대상과 과거의 자료가 존재하는 유사한 대상을 비교하여 비용을 추정하는 방법으로,
 - 비교 대상과 추정 대상의 유사성이 높을수록, 주관적 조정이 적어질수록, 임의 수정 범위가 한정적일수록 정확한 결과를 도출함
 - 통계분석(statistical analysis)은 과거 수학적 혹은 통계학적 자료를 활용하여 비용을 추정하는 방법으로,
 - 변수값 수정이 용이하고 변화 민감도가 높으며 불확실한 대상에 대해서도 객관적이고 유효한 측정이 가능함
 - 세부분석(detail analysis)은 낮은 위계 혹은 요소별 비용을 계산 한 후, 이를 합산하여 전체 비용을 산출하는 방법으로,
 - 비용 추정방법론 중 가장 이해하기 쉬우며 실제 생산과정에 소요되는 시간과 작업의 형태를 고려할 수 있음
 - 전문가 의견(expert opinion)은 전문가 의견에 근거하여 비용을 추정하는 방법으로, 1대 1인터뷰, 회의 및 델파이 조사 등을 활용함

□ 편익 추정

① 편익 추정의 원칙

- 직접적 편익
 - 재정사업의 목표에 부합하면서 사업 추진의 목적 달성에 따라 얻어지는 편익으로 재정사업의 목표에 따른 수혜자를 구체적으로 식별하여 판단함
 - 사업 전후(before and after)가 아닌 시행 유무(with or without)비교를 통하여 사회후생(social welfare)의 차이를 분석함
- 긍정적 효과의 합산
 - 경제주체를 소비자, 생산자에게 발생하는 효과로 구분하며 수혜자를 뚜렷하게 구분하기 어려운 경우 국가 전체적 관점에서 편익을 평가할 수 있음
 - 소비자에게 발생하는 긍정적 효과는 지불의사액(willingness to pay, WTP)관점에서

평가하고 생산자에게 발생하는 편익은 부가가치(value-added) 관점에서 평가함

② 편익항목의 구분

- 편익항목은 정(positive)의 편익인 가치창출 편익과 부(negative)의 편익인 비용저감 편익으로 구분
 - － 가치창출편익은 국가연구개발사업의 결과로 소비자의 후생을 향상시키거나 새롭게 부가가치를 창출하는 것을 사업의 목적으로 할 때 반영함
 - － 비용저감편익은 국가연구개발사업을 통해 개발된 기술을 적용함으로써 특정 분야에서 기존에 소요되던 비용이 감소할 때 반영함

<표 9-4> 연구개발부문 예비타당성조사의 편익항목 구분

구분	세부 편익항목	발생 경우
가치창출 편익	소비자중심 편익	• 국가연구개발사업의 효과가 소비자에게 영향을 주는 경우 • 후생경제학에 근거 － 사업의 성과물이 국민에게 기존에 없는 새로운 긍정적인 효과를 주면서 그 효과를 화폐가치로 계량화 • 시장가치로 계량화할 수 없는 경우에는 진술선호접근법
	생산자중심 편익	• 국가연구개발사업의 효과가 생산자에게 영향을 주는 경우 • 시장수요접근법 － 사업을 통해 새롭게 개발된 산출물이 상용화되어 시장에서 거래됨으로써 발생하는 부가가치의 증가분을 바탕으로 편익을 추정
비용저감 편익	생산비용저감 편익	• 개발된 기술을 적용함으로써 특정 분야에서 기존에 소요되던 비용이 감소하는 경우 － 예)기존의 생산과정에 존재하는 비효율적인 부분이 제거되거나 투입 요소 대체 등을 통해 생산의 효율성이 증가하여 동일한 양의 산출에 드는 비용이 기존에 비해 낮아지는 경우
	피해비용저감 편익	• 국가연구개발사업을 통해 개발된 기술을 적용하여 기존에 재난·재해, 사고, 질병 등으로 인해 발생하던 피해비용을 낮출 수 있을 경우

자료 : 국가연구개발사업 예비타당성조사 수행 세부지침(2018.4)

□ 경제성 분석

- 예비타당성조사 대상사업인 경우, 국가 전반에 미치는 경제적 파급효과와 투자의 적절성 여부를 판단하기 위해 기본적으로 비용-편익분석(Cost-Benefit Analysis)을 채택하여 분석
- 연구개발 활동의 특성에 따라, 비용편익분석 외에 경제사회적 파급효과를 산출하고 비용효과 분석(Cost-Effectiveness Analysis) 적용 가능

제34조(경제적 타당성 분석) ① 제31조에 의한 경제적 타당성 분석은 국가연구개발사업 예비타당성조사 대상사업의 국민 경제적 파급효과와 투자적합성을 분석하는 핵심적 조사과정으로서 비용-편익분석(Cost-Benefit Analysis)을 기본적인 방법론으로 채택하여 분석한다.

② 비용-편익분석을 위해서 사업 시행에 따른 수요를 추정하여 편익을 산정하고, 총사업비와 해당 사업의 운영에 필요한 모든 경비를 합하여 비용을 산정한다.

(주) 일반적으로 B/C비율이 1보다 클 경우 경제적 타당성이 있음을 의미함

③ 연구개발사업의 유형에 해당하는 사업 중에서 비용-편익분석이 적합하지 않다고 판단되는 사업의 경우에는 경제사회적·과학기술적 파급효과 등을 산출하고 이를 통해 비용-효과분석(Cost-Effectiveness Analysis)을 실시할 수 있다.

④ 기금사업 중 여유자금 등을 활용하여 수입증대를 주요 목적으로 하는 사업은, 경제적 타당성 분석 대신 수익성 분석으로 대체할 수 있다.

「국가연구개발사업 예비타당성조사 운용지침」 제34조

9.2. 정책적 타당성 분석



9.2.1. 상위계획과의 부합성

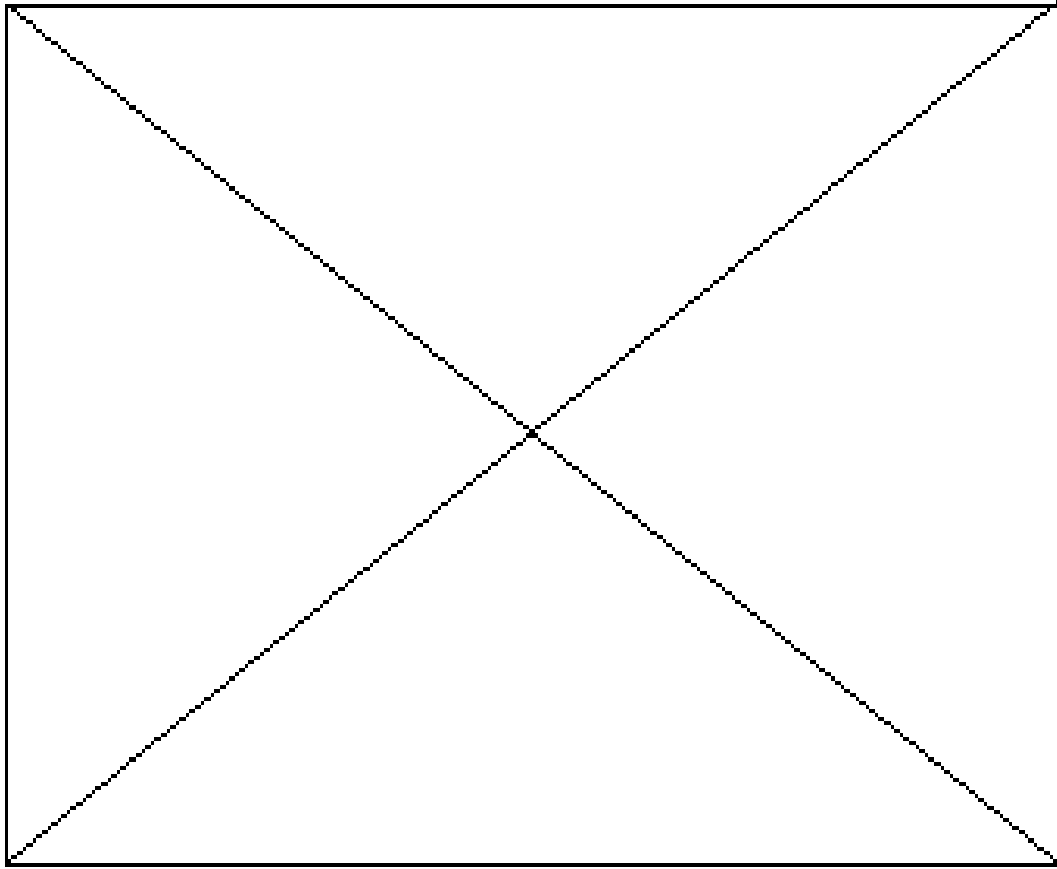
가. 상위계획과의 부합성

(방향) 상위계획과의 부합성은 과학기술분야의 최상위 국가계획인 과학기술기본계획을 기본으로 하여 종합 및 세부계획을 참조하였으며, 그 중 법정계획을 분석대상으로 함

- 국가 연구데이터 공유·활용체제 구축사업은 과학기술분야의 최상위계획인 과학기술기본계획에 부합하는 사업이며,
- 이 외에 문재인 정부 100대 국정과제, 연구성과 관리·활용 기본계획 등에 부합하여 국가정책적 차원에서 상위계획과 부합하는 사업임

□ 문재인 정부 100대 국정과제와의 부합성

- 문재인 정부는 국정운영 계획을 위해 100대 국정과제를 제시하였으며, 이는 향후 5년(‘17~’21)을 위한 국가비전과 국정목표를 달성하기 위한 정부의 업무 과제로,
- 상위계획과의 부합성, 정부지원의 필요성 등 사업의 타당성 평가에 매우 중요한 근거임
- 국가 연구데이터 공유·활용체제 구축사업은 문재의 정부 100대 국정과제 전략 중 ‘과학기술 발전이 선도하는 4차 산업혁명’과 관련됨
 - ‘과학기술 발전이 선도하는 4차 산업혁명’의 주 내용은 초지능·초연결 기술의 확산으로 시대에 부응하는 인재 육성 및 성장 동력을 확보하고 체계적인 컨트롤타워의 설치임
 - 특히, 세계 최초의 5세대 이동통신(5G) 구축과 인공지능(AI) 등의 핵심기술력 확보와 데이터 공유·활용 플랫폼 구축을 초지능·초연결 사회의 기반으로 지정하였음



<그림 9-1> 문재인 정부 국정목표 및 전략

* 자료: 청와대

- 이와 같이, 동 사업은 연구개발 활동 결과물의 일괄적 관리·활용을 통하여 4차 산업혁명 대비와 문재인 정부가 지향하는 생태계 조성에 이바지함으로써 국정과제에 부합하는 사업임

□ 제4차 과학기술기본계획과의 부합성

- 제4차 과학기술기본계획은 장기적인 관점에서 국가의 과학기술 미래비전을 제시하고 이에 따라 수립된 향후 5년 간('18~'22) 기본계획으로
- 대한민국의 과학기술혁신정책의 비전, 목표, 방향을 제시하여 각 부처 과학기술 관련 정책의 수립·추진방향에 있어 최상위에 있는 기본계획임
- 국가 연구데이터 공유·활용체제 구축사업은 제4차 과학기술기본계획의 추진전략 중 '주체 간, 분야 간 협력·융합 활성화'와 관련됨
 - － '주체 간, 분야 간 협력·융합 활성화'의 주 내용은 산·학·연 협력 기반에서 수요 지향적 협력으로 전환하여 연구데이터 수집·공유 플랫폼을 구축하는 것임

- 특히, 융합·공동연구 촉진을 위해 R&D 과정에서 생산되는 연구데이터를 수집·활용하기 위한 체계를 마련하고,
- 연구데이터의 체계적 관리를 위해 국가 연구데이터 센터 및 분야별 전문센터를 운영²⁸⁾ 하는 내용이 포함되어 있음
- 이와 같이, 동 사업은 연구데이터 통합관리체계와 연구데이터 플랫폼 및 전문센터를 운영하고 연구데이터 활용을 지원하는 등 제4차 과학기술기본계획에 부합하는 사업임

<표 9-5> 제4차 과학기술기본계획 추진전략 및 과제

추진전략	중점과제	세부과제
혁신이 일어나는 과학기술 생태계 조성	주체 간, 분야 간 협력·융합 활성화	산·학·연 간 인력교류 활성화
		출연(연) 간 개방·협력 및 중소·중견기업 지원 강화
		민간기업 간 협력 확대 유인
		융합활성화를 위한 기반 구축
		융합·공동연구 촉진을 위한 연구 데이터 수집·공유 플랫폼
	기술혁신형 창업·벤처활성화	대학의 창업 활성화
		공공연구기관의 창업 촉진
		재직자, 일반인 등으로 창업저변 확대
		창업기업의 성장사다리 강화
		창업 투자를 위한 모험투자 강화 및 민간자본의 유입촉진
	경쟁력 있는 지식재산 창출	R&D 전 과정에 지식재산 전략 도입 강화
		4차 산업혁명분야 특허정보 활용 확대
		중소·벤처기업 지식재산 경쟁력 강화
		공공 IP·연구성과의 경제적 활용성 제고
	지역주도적 지역혁신 시스템 확립	지역의 R&D 투자 결정권 강화
		지방정부의 R&D 기획·평가역량 확충
		지역에 대한 중앙정부의 지원체계 개선
		지역R&D 혁신주체 역량강화
		지역 혁신클러스터 고도화
	국민참여 확대 및 컨트롤타워 강화	국민 참여형 ‘열린 국가R&D 시스템’으로 전환
		정부R&D 투자시스템 혁신
		과학기술 분야 중장기 계획연계 강화 및 효율화

자료: 제4차 과학기술기본계획(2018~2022)(2018.2)

28) 바이오, 미래소재, 대형연구장비, 인공지능 등의 데이터 집약형 연구 분야가 기 선정됨

□ 제3차 연구성과 관리·활용 기본계획과의 부합성

- 제3차 연구성과 관리·활용 기본계획은 연구성과의 효과적인 활용·확산을 통해 일자리 창출과 경제 활성화를 도모하는 계획으로
- 연구성과를 효율적으로 활용하기 위한 연구성과의 기획·평가·관리 강화를 목표로 함
- 국가 연구데이터 공유·활용체제 구축사업은 제3차 연구성과 관리·활용기본계획의 추진 전략 중 ‘연구성과와 사업성·시장성간 간극을 최소화’, ‘성과의 수집·관리 체계 개선 및 정부 3.0 제고’ 등에 부합
 - － ‘연구성과와 사업성·시장성간 간극을 최소화’의 주 내용은 성과활용을 위해 지원과 부처간의 협업을 강화하고 연구개발의 성과활용을 증가 시키는 것이며,
 - － ‘성과의 수집·관리 체계 개선 및 정부 3.0 제고’는 연구성과의 수집 체계를 일원화하고 관리 및 유통 체계를 종합적으로 개선하여 연구성과의 질(quality) 제고를 목표로 함
- 이와 같이, 동 사업은 연구데이터의 산업적 활용을 촉진하여 연구성과와 사업성 간극을 최소화하는데 이바지하고,
- 동 사업의 계층적 연구데이터 연계체계 마련은 연구성과의 수집·관리를 원활하게 함으로써, 제3차 연구성과 관리 활용기본계획에 부합하는 사업임

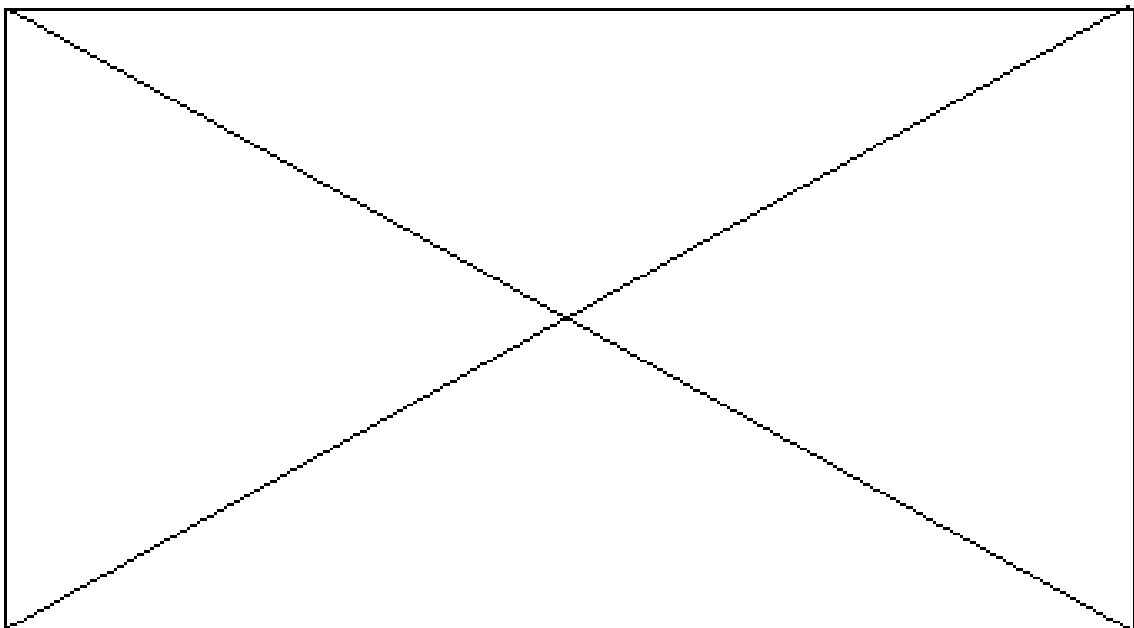
<표 9-6> 제3차 연구성과 관리 활용기본계획 추진전략 및 중점추진과제

추진전략	중점추진과제
성과의 활용성 제고를 위한 연구개발시스템 개선	시장·수요·가치 중심의 R&D기획 강화
	시장친화적 연구관리 개선과 평가결과 예산 반영
연구성과와 사업성·시장성간 간극을 최소화	단절 없는 기술이전·사업화 체제 활성화
	고부가가치 특허창출 역량 강화 및 활용 다각화
	중소·중견기업의 R&D기반 성장지원
	대학·출연(연)의 성과활용 전담조직 정비 및 역량 강화
성과 활용 제도의 경직성 완화	기술료 등 제도 개선 및 탄력적 운영
	제3자 이전 활성화 및 기술이전 기여자에 대한 정당한 보상
성과의 수집·관리 체계 개선 및 정부 3.0 제고	연구성과의 수집·관리 체계 개선
	연구개발 정부 3.0 제고

자료: 제3차 연구성과 관리·활용 기본계획(‘16~’20)(2016.1)

□ I-Korea 4.0 과학기술과 ICT로 열어가는 사람중심의 4차 산업혁명과의 부합성

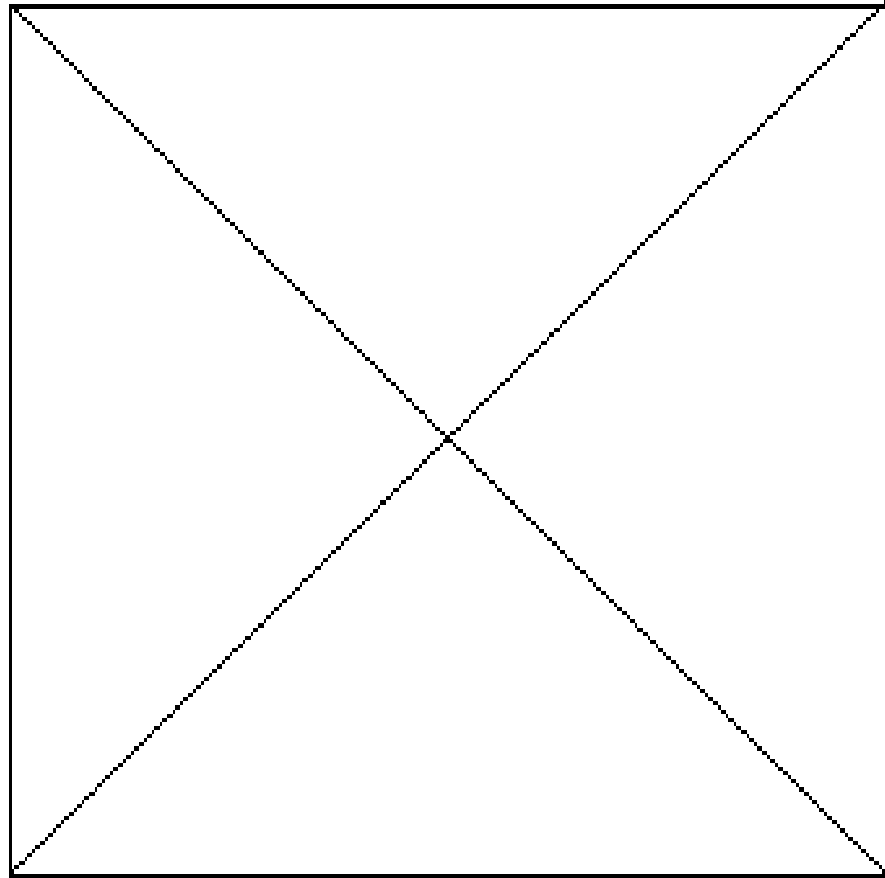
- I-Korea 4.0 과학기술과 ICT로 열어가는 사람중심의 4차 산업혁명은 2018년도 과학기술정보통신부의 업무계획으로 국가 21개 부·처의 향후 5년의 청사진을 제시함
 - － 국민 삶의 질 향상을 위해 ‘초연결 지능화 인프라 구축과 R&D 시스템 혁신으로 기술·산업·사회가 연계된 “사람중심 4차 산업혁명” 구현’을 기본방향으로 하며,
 - － 4차 산업혁명을 최전선에서 대응하며 기술-산업-사회 전 분야에서 유기적 혁신을 통해 연구자-기업인-국민에게 과학기술 혜택 제공을 목적으로 함



<그림 9-2> 과학기술정보통신부 핵심 정책과제

* 자료: 과학기술정보통신부

- 국가 연구데이터 공유·활용체제 구축사업은 I-Korea 4.0 과학기술과 ICT로 열어가는 사람 중심의 4차 산업혁명의 추진 전략 중 ‘국가R&D 시스템 혁신’에 부합함
 - － ‘국가R&D 시스템 혁신’의 주 내용은 국가 전 부·처, 정부출연기관 및 대학의 자율적인 연구 활동을 지원하고 연구개발 활동의 성과를 공유하고 확산하는 것임



<그림 9-3> I-Korea 4.0 과학기술과 ICT로 열어가는
사람중심의 4차 산업혁명 추진전략 및 과제

* 자료: 과학기술정보통신부

- 이와 같이, 국가연구개발사업에서 생산·축적되는 연구데이터를 동 사업은 연구데이터 공유·활용함으로써, 증가하는 연구개발의 성과창출을 지원하는 등 I-Korea 4.0 과학기술과 ICT로 열어가는 사람중심의 4차 산업혁명에 부합하는 사업임

나. 정부 지원의 필요성

□ 사업의 필요성

- 연구 재현성과 투명성 제고 및 연구주체의 신뢰도 향상으로 국가연구개발 경쟁력 향상
 - － 연구개발의 투명성 제고는 연구데이터의 신뢰도를 높이고, 타 연구주체가 국가연구개발의 다양한 데이터를 활용하도록 유도함
 - － 이러한 데이터의 개방과 공유로 연구 재현과 다학제 간 공동·융합 연구 실현이 더 활발히 일어날 수 있는 환경을 조성함

- 개방과 공유로 활발히 활용되는 연구데이터는 국가의 주도로 이루어지는 국가연구개발 사업의 타당성을 입증하여 정부의 공공기금 사용에 정당성을 부여함
- 연구데이터 재사용 및 불필요한 연구데이터 생산 방지로 국가연구개발 효율성 제고
 - 연구데이터 공유·활용을 위한 국가 연구데이터 플랫폼은 정부조직 간 국가연구개발사업에서 발생하는 데이터를 일괄 수집·관리하여 연구데이터의 중복을 방지함
 - 효율적으로 관리되는 연구데이터로 중복 연구를 줄이고 기존 연구데이터의 활용성을 제고 하는 등 한정된 공공자금을 효율적 활용에 기여할 것임
- 타 연구 분야로의 도전 및 융합연구를 통한 일자리 창출과 경제 활성화
 - 공유·활용된 연구데이터는 연구주체가 새로운 연구개발을 하도록 격려하며, 이를 통한 융합연구는 새로운 산업 및 일자리 창출에 기여함
 - 다양한 연구주체는 통합 관리되는 연구데이터에서 영감을 받아 새로운 분야로의 연구개발로 이어질 수 있으며 이를 통해 경제 활성화를 기대할 수 있음

□ 정부지원의 필요성

- 해외 현황으로 본 사업 추진의 시급성
 - 4차 산업혁명의 도래로 초(超)연결 기반의 지능화가 이루어지고 있는 가운데, 대용량의 데이터 수집·활용은 기업 간, 국가 간 경쟁력의 원천으로 자리잡음
 - 해외 주요 선진국은 연구과정 및 성과를 공개·공유하는 오픈 사이언스를 주 정책으로 추진하며 다양한 분야의 융합·공동 연구 시행하고 있음
- 국내 현황으로 본 사업 추진의 적절성
 - 반면, 국내에서는 연구데이터가 연구 성과물로서 인정되지 않을 뿐만 아니라, 연구데이터의 수집·관리를 위한 제도 및 인프라 및 전문 인력이 부족한 실정임
 - 또한, 연구주체의 연구데이터 공유·활용에 대한 인식과 의지가 낮아 연구데이터 공유·활용에 대한 중요성을 환기하고,
 - 각 부처와 중·소규모의 연구기관을 주도하여 추진할 수 있는 강력한 정책마련이 시급하며 이를 통해 정부주도의 사업추진이 적절함

□ 정부주도의 정당성

- 혁신 측면에서의 정당성
 - 정부가 연구데이터 공유·활용을 총체적으로 관리·지원함으로써, 개별적으로 데이터를 관

- 리하는 기관의 연구데이터까지 기관 간 연계 가능하도록 시스템을 구축함
- 자체적으로 연구데이터 관리가 어려운 대학 및 소규모의 기관은 정부의 지원 없이는 연구데이터 공유·활용에 어려움을 겪을 수 있음
- 공동의 안녕을 목표로 국가연구개발사업의 기반을 마련하다는 점에서 수익창출을 주목적으로 하는 민간의 투자를 기대하기가 어려움
- 경제적·형평성 측면에서의 정당성
 - 연구데이터 공유·활용의 시스템이 구축되지 않을 경우, 기존에 수행된 연구 및 기 확보된 연구데이터의 활용이 어려워 중복 사업 시행의 우려가 있음
 - 인프라 및 전문가 부족으로 시스템을 구축하지 못하는 대학연구기관 및 소규모의 연구기관 등은 데이터 공유·활용의 혜택을 받지 못함
 - 연구데이터 공유·활용의 시행 여부로 인해 연구과정 및 성과에 있어 불필요한 낭비 및 효율성 제고에 관한 격차가 심해질 경우 형평성에 어긋남
- 공공성 측면에서의 정당성
 - 공공기금의 투자로 이루어지는 사업의 결과물이기 때문에 그 활용 또한 국민의 삶의 질 향상을 목적으로 함
 - 동 사업으로 정부출연연구소 및 대학은 다양한 융합연구를 시행할 수 있고, 이를 통해 새로운 산업 및 일자리 창출을 도모하는 등 공익성이 매우 높은 사업구조를 가짐

9.2.2. 사업 추진체제 및 추진의지

가. 사업 추진체제

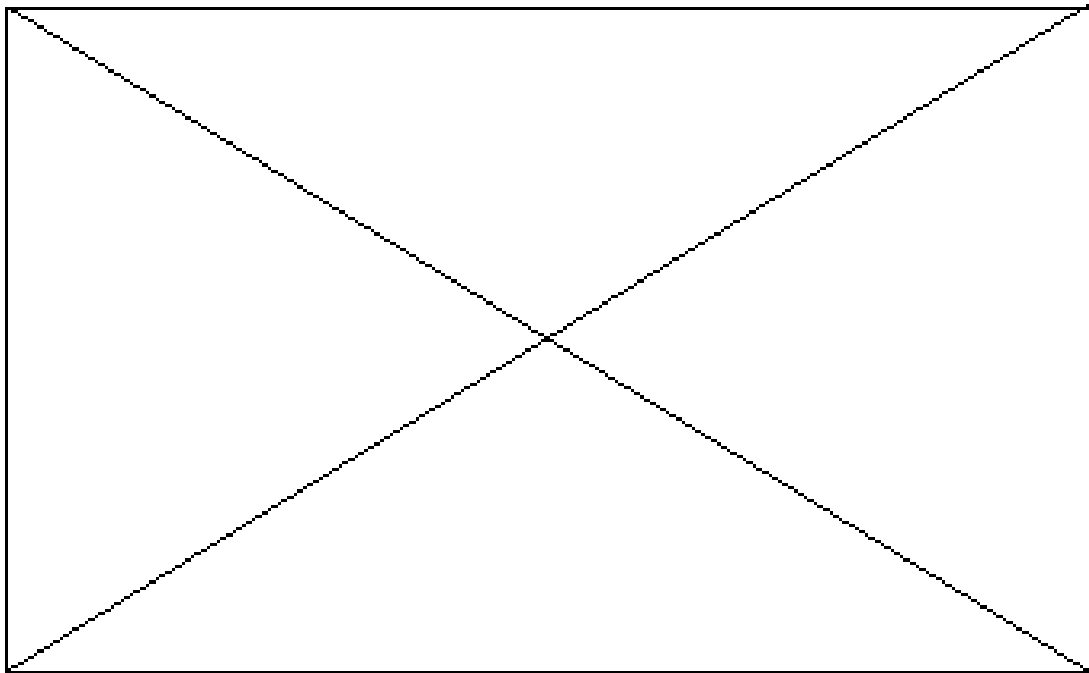
(방향) 사업 추진체제는 사업 추진체제와 관련하여 제출된 자료를 바탕으로 주관부처-수행기관 간 관계와 협조체제, 그 외, 연구 주체들 간 역할분담, 협조체제와 운영방안 등을 분석

- 국가 연구데이터 공유·활용체제 구축사업은 한국과학기술정보연구원을 전문기관으로 하여 사업 추진에 필요한 조직과 협조하였으며,
- 성과관리 위원회, 전담반 및 연구단장 협의회 등 사업목표 달성을 위한 부수적인 조직을 구성하여 동 사업을 전적으로 지원하고 있음으로
- 시행 시, 사업에 참여하는 이해관계자의 연구데이터 공유·활용에 관한 역할분담 및 협조가 기대되는 사업임
- 관리기관은 법·제도에 근거하여 부처로부터 자금이 전달되는 경로에 위치하는데 적합한 기관이며, 주관연구기관과 과제 수행주체가 법·제도적인 결함이 없이 선정됨

□ 협력체제 및 역할분담

- 동 사업은 과학기술정보통신부를 주관부처, 한국과학기술정보연구원을 전담기관으로 하여 구성되며,
- 세부과제별 연구단과, 동 사업에서 선정된 11개의 과학 기술 전문센터별 연구단 그리고 테스트 베드를 위한 연구단이 사업 수행에 있어 체계적인 추진체제를 확보함
 - － 과학기술정보통신부는 주관부처로서 국가 연구데이터 공유·활용체제 구축 기본계획 수립부터 예비타당서 조사 기획까지 동 사업의 주요 의사결정을 수행함
 - － 한국과학기술정보연구원은 과학기술정보통신부의 역할을 대행하여 사업기획을 담당하며 연구단의 사업성과 및 목표를 관리하고 성과창출의 극대화를 위한 전략을 수립함
 - － 연구단은 세부과제를 수행하는 연구주체로서, 세부과제별 연구단과 테스트베드 운영을 위한 연구단이 별도로 구성됨
- 동 사업의 추진체제는 주관부처의 연구개발 관리·감독보다는 연구개발 활동 지원 역할에 초점을 맞추어 연구개발에 필요한 인프라 및 인적자원의 효율적인 활용을 지원하며,
- 연구개발사업 추진에 있어 사업단을 보조하기 위해 성과관리 위원회, 전담반, 연구단장 협의회 등을 추가로 설치하여 사업의 기술적 전문성을 확보하고 효율을 창출함

- 성과관리 위원회는 사업단 선정, 사업성과물 평가·조정, 실패과제 분석, 실패과제의 재 활용 여부 판단 등 사업의 성과 활용도를 제고하기 위한 전반적인 업무를 수행함
- 책임평가단은 성과관리 위원회 산하로, 연구단의 연구 활동 과정을 지속적으로 모니터링 하고 연구개발 성과 목표 달성 여부를 분석하여 성과위원회에 보고함
- 전담반은 사업단은 지원하는 지원조직으로 기술을 개발하고 기술거래 운영지원, 기술이 전 관리 등 기술개발을 지원하고 사후 관리 역할을 수행하며,
- 전략지원 전담반은 시장·산업동향 분석, 연구 활동에서 산출된 논문 및 특허 등록을 지원하여 사업단의 연구개발이 목표한 성과물을 창출하는 과정을 지원함
- 기술융합 허브는 과제 관련 세미나, 포럼 및 학회 조직 운영 등 내·외부의 혁신네트워크를 운영하고 지원하는 역할을 수행함
- 연구단장 협의회는 사업단의 과제 수행을 자체 모니터링 하고 연구 결과물을 공유·연계 하기 위한 협력방안을 모색함



<그림 9-4> 주요 조직구성 및 운영체계

□ 사업 주체의 법·제도적 정당성

- 과학기술정보통신부는 「과학기술기본법」에 따라 중앙행정기관의 사업 기획·관리·평가·활용에 관한 역할을 위임받아 대행함
 - － 과학기술정보통신부는 「과학기술기본법」에 근거하여 정부의 역할을 위임받아, 국가 연구개발사업에 관한 기획·관리·평가 및 활용 등에 관한 업무를 수행할 수 있음

“제11조(국가연구개발사업의 추진) ① 중앙행정기관의 장은 기본계획에 따라 맡은 분야의 국가연구개발사업과 그 시책을 세워 추진하여야 한다.

④ 중앙행정기관의 장은 소관 국가연구개발사업의 효율적 추진을 위하여 필요하다고 인정하는 경우에는 소관법령으로 정하는 기관 또는 단체에 국가연구개발사업의 과제 기획 등에 관한 업무를 대행하게 할 수 있다. 이 경우 중앙행정기관의 장은 기획 등을 대행하는 자(이하 "전문기관"이라 한다)에 대하여 기획 등의 수행에 사용되는 비용의 전부 또는 일부를 지원할 수 있다.”

「과학기술기본법」 제11조

- 한국과학기술정보연구원은 과학기술정보통신부의 유사기관으로 동 사업과 유사한 분야²⁹⁾를 전담하는 연구기관임
 - － 한국과학기술정보연구원은 과학기술 정보 인프라를 바탕으로 지식 인프라 고도화, 과학기술 데이터 생태계 형성 등 과학기술정보 관련 분야를 연구함
 - － 과학기술정보통신부의 유사기관으로 과학기술정보통신부가 정부의 역할을 대행하여 주관하는 사업에 대하여, 지식·정보·데이터 등 관련 사업인 경우에 해당 사업을 전담함
- 한국과학기술정보연구원은 「국가를 당사자로 하는 계약에 관한 법률」에서 정하는 경쟁 입찰의 원칙에 따라 연구기관인 한국과학기술정책플랫폼협동조합과 계약

“(제7조) ① 각 중앙관서의 장 또는 계약담당공무원은 계약을 체결하려면 일반경쟁에 부쳐야 한다. 다만, 계약의 목적, 성질, 규모 등을 고려하여 필요하다고 인정되면 대통령령으로 정하는 바에 따라 참가자의 자격을 제한하거나 참가자를 지명(指名)하여 경쟁에 부치거나 수의계약(隨意契約)을 할 수 있다.”

「국가를 당사자로 하는 계약에 관한 법률」 제7조

29) 연구데이터허브센터, 연구데이터플랫폼센터, NTIS센터, 대용량데이터허브센터, 데이터분석플랫폼센터 등

나. 사업 추진의지

(방향) 사업 추진의지는 주관부처 및 사업 추진체계상에 포함된 참여 주체와 사업 추진의 영향을 받는 이해관계자, 즉 수혜자 및 비용부담자와 같은 관련주체자를 대상으로 사업 추진의지와 선호도를 분석

- 국가 연구데이터 공유·활용체제 구축사업은 동 사업 사전에 결성된 TF와 20여 건의 회의 등으로 보아 주관부처의 사업의지가 확고함
- 동 연구개발사업의 수혜자가 연구개발과정에서 생산되는 연구데이터 활용의 어려움과 그 필요성에 관해 인지하고 있으며,
- 사업 참여예정 주체와 이해관계자의 사업에 관한 의지를 객관화 할 수 있는 공식적인 문서와 구체적인 근거를 확보함

□ 참여예정 주체의 사업 추진의지

- 과학기술정보통신부의 주도로 동 사업의 체계 기획위원회를 구성하기 전 TF를 결성하여 연구데이터 공개 제도 및 수집·관리의 범위 등을 사전에 논의함
 - － 주관부처인 과학기술정보통신부는 ‘모아서 새롭게 TF’를 구성·운영하여 「연구데이터 공유·활용 전략(안)」을 수립하였음
 - － 학·연의 전문가 및 연구관리 전문기관을 포함한 4개의 분과를 구성하고 분과별 화의를 통해 동 사업의 전문성 및 구체성을 제고하는데 기여함
 - － 사업의 기획위원회 구성 이전부터 TF를 결성하고 총 24건의 회의를 거치는 등 동 사업 추진에 있어 강한 의지를 보이고 있음

<표 9-7> 모아서 새롭게 총괄 TF 추진 현황 및 주요 논의 내용

구분	횟수	주요 논의내용
총괄 TF	10	- 연구데이터 공개 의무 제도화, 공유·활용 대상 분야 및 우선순위, 단계별 목표 - 연구데이터 공유·활용 체계 구축 추진전략·로드맵, 분야별 시범사업 후보 선정논리 마련 - 바이오·소재 분야에서 수집 가치가 있는 데이터, 공개 시점, 출연연 역할 등 논의 - 연구데이터 공유·활용 체계 및 커뮤니티 육성방안, 데이터 관리 및 공유·활용 제도화 방안, 공유·활용 인프라 클라우드 구축 방안, 바이오·소재·거대실험장비 분야 추진방안 논의 - 연구데이터 관리 및 공유·활용 법·제도화안 및 대안, 시범분야 추진 방안, 데이터·컴퓨팅 활용 R&D 인력양성 논의 - DMP 제출·이행 및 이행 점검에 있어 일반적인 원칙 및 기준, 데이터 및 컴퓨팅 활용 R&D 인력양성 방안 논의 - TF에서 논의·마련한 연구데이터 관리·활용체계 구축 방안 관련 전문가 의견수렴 - 클라우드 방식의 데이터 정리용 플랫폼을 통한 과거 및 진행 중인 과제 데이터 유실방지 방안, 적용 분야, 제공 필요서비스, 운영 방식·원칙 등 논의 - 총괄 및 분과 TF에서 논의·마련한 연구데이터 관리·활용 방안 집중 토론 - 연구데이터 관리·활용 체계 구축 방안
제도화 분야 TF	4	- 미국, 영국, 호주 등 해외 선진국들의 연구데이터 제도화 사례 및 시사점 분석 - 연구데이터 제도화 시 고려사항 논의(기본방향, 연구현장 의견 등) - 연구데이터 관리 및 공유·활용, 지원을 위한 단계적 입법 추진, 연구데이터를 R&D 성과물로 지정·관리 - 과제 전주기)에 걸쳐 DMP 제도화 적용 여부 및 해외사례(NSF) 분석 등
바이오 분야 TF	4	- 바이오분야 연구데이터 연구내용 현황 - 인공지능 빅데이터를 활용한 신약 개발 방안 - 인공지능 빅데이터를 활용해 신약을 개발하는 벤처 기업 사례 공유 및 빅데이터 활용 신약 개발 방안 1~3차 회의 및 총괄 TF 논의사항 공유 및 향후 운영계획
소재 분야 TF	4	- 소재 분야 빅데이터 필요성·정의, 국내 소재정보은행·참조표준센터 및 미국 NIST 연구데이터 수집 현황 등 - 소재 분야 빅데이터 수집·관리·활용을 위한 전문가 설문 검토 및 분류체계 필요성 논의(소재전문가 설문조사 실시) - 소재 분야 연구데이터 플랫폼 구축 연구자 설문조사 결과 분석, 국내외 연구데이터 수집·활용사례 발표 및 논의 - 소재분야 Tier-1, Tier-2 역할 및 범위 논의, Tier-1,2 예산 및 단계별 계획 정리
대형연구장비 분야 TF	4	- 대형연구장비 데이터 저장, 공유, 활용 현황 및 구조 생물학 분야에서의 대형연구장비 데이터 분석, 활용 현황 등 - 대형연구장비별 데이터 저장, 공유, 활용 현황 등 - IBS 연구단 연구데이터 관리현황 및 향후 공유·활용 방안 - 글로벌 프런티어 연구단별 연구데이터 관리현황 파악, 향후 공유·활용 방안

- 국가 연구데이터 플랫폼 공유·활용 체계 기획위원회를 구성·운영하여, 동 사업의 기능 및 기술체계(안)를 마련함
 - － 플랫폼 도출을 위해 플랫폼, 전문센터 분과위원회를 구분하였고, 연구데이터 활용을 위한 인력교육을 위해 인력양성만을 위한 분과를 설치함
 - － 새롭게 구성된 3개 분과의 산·학·연 전문가의 주기적인 회의와 피드백을 통해 동 사업의 방향성 및 구체성을 확보함
 - － 전문가의 피드백이 동 사업의 기획보고서에 반영이 되어 분과위원회의 전문가와 국가연구개발사업의 연구기관의 상호작용이 긍정적인 효과를 보임

□ 국가 연구데이터의 관리·공유 현황

- 연구개발 전 분야에 걸쳐 산·학·연 이해관계자에게 설문한 결과³⁰⁾ 대부분의 연구데이터가 관리되고 있으며,
- 절반에 가까운 연구개발사업에서 외부와의 연구데이터 공유 의사가 확인되어 국가 연구데이터 관리·공유에 있어 양호한 현황으로 판단됨
 - － 국가연구개발과제 수행 과정에서 생산되는 연구데이터의 현황을 파악하고 산출된 연구데이터의 공유 현황을 파악하기 위해 연구개발 전 분야에 걸쳐 설문조사를 진행함
 - － 현재 진행 중인 과제를 포함하여 국가연구개발사업 종료 후 2년 이내의 과제를 대상으로 하여 진행하였으며, 총 3,205건의 설문이 회수됨

<표 9-8> 연구데이터 관리실태 1차 설문조사 개요

구분	내용
목적	연구과제 수행 시 생산되는 연구데이터 생산 현황 및 외부 공유 의견 파악
기간	2018.04.19. ~ 2018.04.20.
방법	온라인
대상	국가연구개발사업 종료 후 2년 이내 과제 (현재 진행 중인 과제 포함)
회수 설문	3,205건

- － 국가연구개발사업 과정에서 연구데이터를 생산하는 과제는 총 2,081건이며, 이 중 연구실 내부 시스템을 활용하여 연구데이터를 관리하는 과제는 48%를 차지함
- － 서버에서 파일단위로 관리되는 과제는 연구데이터를 생산하는 과제 중 40%를 차지하며, 대부분의 연구데이터는 관리가 이루어지고 있음

30) 본 설문은 2018.04.19.~2018.04.20. 동안 현재 진행 중인 과제를 포함, 국가연구개발사업 종료 후 2년 이내의 과제를 그 대상으로 하여 수행되었으며, 연구과제 수행 시 생산되는 연구데이터 생산 현황 및 외부 공유 의견 파악을 검토함

<표 9-9> 연구데이터 관리실태 1차 설문조사 연구데이터 관리 방법

구분	과제 수(건)	비중(%)
기관 관리시스템	98	5%
연구실 내부 시스템	1005	48%
서버에 파일단위로 관리	824	40%
외부 리포지터리	64	3%
관리 없음	42	2%
기타	48	2%
합계	2,081건	100%

- 국가연구개발사업 과정에서 연구데이터를 생산하는 과제는 총 2,081건이며, 이중 절반 이 넘는 1198건(58%)의 과제는 생산된 연구데이터를 외부와 공유하지 않을 과제임
- 반면, 생산한 연구데이터를 외부와 공유할 의사가 있거나 향후 공유 예정인 과제는 883건으로 전체의 42%를 차지함

<표 9-10> 연구데이터 관리실태 1차 설문조사 연구데이터 공유 의사

구분	과제 수(건)	비중(%)
공유의지가 있는 과제	313	15%
공유 예정인 과제	570	27%
공유하지 않음	1198	58%
합계	2,081	100%

□ 사업 이해관계자의 선호도

- 산·학·연 이해관계자에게 설문한 결과³¹⁾ 응답자의 일부만이 타 분야 연구데이터의 필요성은 인지하지만,
- 대부분이 타 연구 분야의 연구데이터 활용에 있어서 어려움을 겪고 있고, 향후 연구데이터 공유·활용에 긍정적이고 적극적일 것으로 기대됨
- 국가연구개발과제 수행 과정에서 생산되는 연구데이터의 공유·활용 현황을 파악하기 위해 연구개발 전 분야에 걸쳐 설문조사를 진행함
- 현재 진행 중인 과제를 포함하여 국가연구개발사업 종료 후 2년 이내의 과제를 대상으로 하여 진행하였으며, 총 1,731건의 설문이 회수됨

31) 본 설문은 2018.08.14.~2018.09.07. 동안 현재 진행 중인 과제를 포함, 국가연구개발사업 종료 후 2년 이내의 과제를 그 대상으로 하여 수행되었으며, 연구데이터 관리·활용 및 애로사항 조사를 검토함

<표 9-11> 연구데이터 관리실태 2차 설문조사 개요

구분	내용
목적	연구데이터 관리·활용 및 애로사항 조사
기간	2018.08.14. ~ 2018.09.07.
방법	온라인
대상	국가연구개발사업 종료 후 2년 이내 과제 (현재 진행 중인 과제 포함)
회수 설문	1,731건

- 설문조사에 응답한 1,731명 중 국가연구개발사업 과정에서 생산된 연구데이터가 ‘가끔’ 필요하다고 경험한 응답자의 비율은 33%로 가장 많은 비중을 차지함
- 타 연구 분야의 연구데이터가 필요하다(‘매우 자주’와 ‘자주’) 경험한 응답자는 24%를 차지하였으며, 응답자의 21%는 타 연구 분야의 연구데이터는 필요하지 않다고 응답함

<표 9-12> 연구데이터 관리실태 2차 설문조사 연구데이터 공유 필요 경험

구분	빈도(명)	비중(%)
매우 자주	125	7%
자주	300	17%
보통	367	21%
가끔	568	33%
없음	371	21%
합계	1,731	100%

- 타 연구 분야의 연구데이터 활용 시 어려운 이유 중 가장 많이 언급된 이유는 타 분야의 연구데이터가 공개되지 않아서(30%)임
- 연구데이터의 존재 혹은 보유 위치를 몰라서 어려움을 겪은 경우는 24%, 타 분야 연구 데이터로의 접근성 때문인 경우는 19%를 차지함
- 타 연구 분야의 연구데이터 활용에 어려움이 있다고 응답한 응답자 중 타 연구 분야의 연구데이터 활용이 가능할 경우 연구데이터 공유·활용에 참여할 의향이 있다(‘매우 그렇다’, ‘그렇다’)고 응답한 응답자가 49%를 차지함

<표 9-13> 연구데이터 관리실태 2차 설문조사 타 연구데이터 활용의 어려움



- 클라우드 기반의 연구 환경이 제공 될 때, 필요한 연구데이터 활용 기능으로 언급된 기능은 연구데이터 저장·공유 공간으로 43%를 차지함
- 연구데이터 장기 보존을 위한 공간에 대한 요구도 30%를 차지하며 이외 데이터 처리 및 분석을 위한 컴퓨팅 리소스가 필요하다고 응답한 응답자가 22%임
- 연구데이터 공유·활용체제의 구축을 통해 연구 생산성이 증가할 것('매우 그렇다', '그렇다')이라고 응답한 응답자는 57%를 차지함

<표 9-14> 연구데이터 관리실태 2차 설문조사 필요한 연구데이터 활용 기능



9.2.3. 사업 추진 상의 위험요인 및 대응방안

가. 재원조달 가능성

(방향) 재원조달 가능성은 국비, 지방비, 민자로 구분되는 연구개발사업의 전체 재원을 통합분석하고 재원을 재원부담주체별, 연차별로 구체적으로 분석하여 위험요소 식별

- 국가 연구데이터 공유·활용체제 구축사업은 총사업비 5,752억 원이 100% 정부출연금으로 추진되는 사업으로,
- 정부의 재정지원 능력 및 사업 추진 의사가 확보되어 사업 추진 중 예상되는 재정적 위험수준이 양호한 사업임

□ 재원부담주체·연차별 예산투입규모

- 국가 연구데이터공유·활용체제구축 사업은 사업 추진을 위해 2021년부터 2030년까지 10년 동안 총 5,752억 원의 정부출연 사업비로 계획됨
 - － 사업비 부담주체는 정부이며, 사업비 5,752억 원 중 정부출연금 5,752억 원으로 동 사업은 100% 정부출연금으로 추진되는 사업임
 - － 동 사업은 ‘국가 연구데이터 플랫폼 구축’, ‘대·소 분야 전문센터 플랫폼 구축 및 운영지원’, ‘데이터 및 컴퓨팅 활용기반의 R&D인재육성’의 3가지 추진과제로 구성됨
 - － 국가 연구데이터 플랫폼 구축 및 운영은 6가지의 세부 기술개발을 진행하며, 사업기간 동안 192,165백만 원의 사업비가 소요됨
 - － 대·소 분야 전문센터 플랫폼 구축 및 운영 지원은 바이오 분야, 소재 분야, 인공지능 분야와 기타 8개 분야의 전문센터 구축 및 운영지원을 진행하며, 사업기간 동안 321,050백만 원의 사업비가 소요됨
 - － 데이터 및 컴퓨팅 활용기반의 R&D인재육성은 위탁, 심화, 양성교육 등을 진행하며, 사업기간 동안 61,971백만 원의 사업비가 소요됨

<표 9-15> 세부과제 별 연차별·재원부담별 재원조달 계획

(단위 : 백만원)

분야별		1단계			2단계				3단계			계
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	
총계		32,075	41,195	57,887	61,332	69,388	64,176	65,151	63,534	60,598	59,850	575,187
국가 연구데이터 플랫폼구축 및 운영	소계	15,595	23,778	22,418	25,315	19,309	15,579	17,283	17,738	17,569	17,581	192,165
	연구데이터수집· 연계및공유기술	1,804	2,522	2,405	2,739	2,066	1,671	1,861	1,909	1,898	1,899	20,774
	연구데이터저장· 보안기술	3,315	5,324	4,984	5,544	4,186	3,385	3,771	3,869	3,846	3,848	42,072
	연구데이터품질· 표준화기술	2,828	3,953	3,769	4,292	3,238	2,618	2,916	2,992	2,975	2,976	32,556
	지능형연구데이터 고속 검색 기술	1,634	2,624	2,457	2,733	2,063	1,669	1,859	1,907	1,896	1,897	20,738
	연구데이터융합· 분석환경구축기술	2,612	4,029	3,854	4,500	3,566	2,851	3,104	3,190	3,106	3,111	33,923
	연구데이터활용· 플랫폼기술	3,402	5,326	4,948	5,508	4,190	3,387	3,772	3,871	3,848	3,850	42,101
대·소분야 전문센터 구축 및 운영지원	소계	9,500	10,900	28,850	29,300	43,350	43,050	42,200	40,200	37,300	36,400	321,050
	바이오분야 전문센터	1,150	1,200	1,200	2,400	2,400	2,400	2,950	2,950	2,950	2,950	22,550
	소재분야 전문센터	2,850	4,200	5,400	5,150	4,700	4,400	4,000	4,000	3,100	2,200	40,000
	인공지능분야 전문센터	5,500	5,500	6,250	5,750	4,250	4,250	3,250	3,250	3,250	3,250	44,500
	농림수산·식품			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	환경·기상			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	기계·제조			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	우주·항공·해양			4,000	4,000	4,000	4,000	4,000	3,500	3,500	3,500	30,500
	원자력·핵융합					4,000	4,000	4,000	4,000	3,500	3,500	23,000
	에너지·자원					4,000	4,000	4,000	4,000	3,500	3,500	23,000
	건설·교통					4,000	4,000	4,000	4,000	3,500	3,500	23,000
	재난안전					4,000	4,000	4,000	4,000	3,500	3,500	23,000
데이터 및 컴퓨팅 활용 기반의 R&D 인재육성	소계	6,980	6,517	6,619	6,717	6,729	5,546	5,669	5,596	5,730	5,869	61,971
	ICT기반 기술교육	1,802	1,887	1,976	2,070	2,168	2,271	2,378	2,491	2,609	2,732	22,384
	분야별 전문 활용교육	1,678	1,129	1,143	1,147	1,061	1,076	1,090	906	921	937	11,088
	교육용 데이터뱅크 및 서비스 구축	3,500	3,500	3,500	3,500	3,500	2,200	2,200	2,200	2,200	2,200	28,500

□ 재원조달 가능성

- 사업 추진에 필요한 재원을 충분히 제공할 능력이 있고 동 사업 추진에 확고한 지원 의사가 있으므로 재원조달 위험수준이 양호함
 - － 정부재원 측면에서 현행정부의 국정과제, 정부 정책 및 상위계획과의 부합성 및 재원조달 등 사업에 관한 정부의 의지가 확고한 바, 재원조달 가능성에 있어 위험요인은 대체적으로 크지 않음
 - － 민간재원 측면에서는 동 사업의 재원이 100% 정부출연금으로 조달되기 때문에 민간재원을 조달하기 위한 위험요인이 없음
- 기획재정부의 2018년도 이후 R&D 예산 계획의 증가 수준을 고려해 볼 때, 연간 320~690억 원 수준의 신규 예산은 조달 가능한 규모로 보임
 - － 2018년도 기획재정부의 총 지출은 432.9조원이며 이 중 R&D 분야 예산으로 19.7조원이 투입되었음
 - － 기획재정부의 향후 5년간 재정운용계획에서 R&D 분야의 예산 증가율이 약 5.2%의 수준으로 높지 않은 편이나, 향후 5년 동안 예산의 비율이 증가하며,
 - － R&D분야에 배분된 투자금액 타 투자금액에 비해 크고, 동 사업의 총사업비는 R&D에 배분된 사업비의 약 2%만을 차지함으로 재원조달 위험요인이 없는 것으로 보임

<표 9-16> 2018~2022년 분야별 재원배분 계획

(단위: 조 원, %)

구분	'18	'19	'20	'21	'22	'18~'22 연평균
총지출	428.8 (7.1)	470.5 (9.7)	504.6 (7.3)	535.9 (6.2)	567.6 (5.9)	7.3
1. 보건·복지·고용	144.6 (11.7)	162.2 (12.1)	179.0 (10.4)	196.4 (9.7)	214.3 (9.1)	10.3
2. 교육	64.2 (11.8)	70.9 (10.5)	76.0 (7.2)	80.1 (5.3)	84.0 (4.9)	7.0
3. 문화·체육·관광	6.5 (△6.3)	7.1 (10.1)	7.4 (4.7)	7.8 (4.3)	8.0 (3.0)	5.5
4. R&D	19.7 (1.1)	20.4 (3.7)	21.4 (5.1)	22.6 (5.5)	24.0 (6.3)	5.2
5. 산업·중소기업·에너지	16.3 (1.8)	18.6 (14.3)	19.4 (3.9)	19.9 (2.8)	20.2 (1.3)	5.5
6. SOC	19.0 (△14.2)	18.5 (△2.3)	18.0 (△2.7)	17.7 (△1.8)	17.5 (△1.1)	△2.0
7. 농림·수산·식품	19.7 (0.5)	19.9 (1.1)	19.8 (△0.5)	19.7 (△0.5)	19.6 (△0.5)	△0.1
8. 환경	6.9 (△0.3)	7.1 (3.6)	7.0 (△2.0)	6.9 (△1.9)	6.7 (△1.8)	△0.5
9. 국방	43.2 (7.0)	46.7 (8.2)	49.9 (6.8)	52.8 (5.9)	55.5 (5.0)	6.5
10. 외교·통일	4.7 (3.5)	5.1 (7.5)	5.4 (6.8)	5.7 (5.7)	6.0 (5.1)	6.3
11. 공공질서·안전	19.1 (5.2)	20.0 (4.9)	20.9 (4.3)	21.7 (4.1)	22.6 (3.8)	4.3
12. 일반·지방행정	69.0 (9.0)	77.9 (12.9)	84.1 (7.9)	89.2 (6.1)	94.0 (5.4)	8.0

자료: 기획재정부. 2018.8

나. 법·제도적 위험요인

(방향) 국가연구개발사업은 출연 및 공공기금으로 추진되는 사업으로 규정된 절차³²⁾의 수행 여부에 따라 법·제도적인 쟁점사항을 분석하고 연구개발 활동의 수행 및 연구개발사업 성과물의 활용 등에 따라 국내·외에 저촉되는 법률 검토를 통해 위험요인을 분석

- 국가 연구데이터 공유·활용체제 구축사업은 「과학기술기본법」과 「정보통신 진흥 및 융합 활성화 등에 관한 특별법」 등 적절한 법적 근거를 기반으로 추진된 사업이며,
- 동 사업 수행 중 연구데이터 공유·활용과 마찰을 일으키는 법·제도적 제약이 없고, 동 사업의 산출물이 외국과의 교역에 있어 대립될 위험요인이 없는 사업임

□ 사업 추진근거의 적절성

- 국가 연구데이터 공유·활용체제 구축 사업의 주관부처인 과학기술정보통신부는 중앙행정기관으로 「과학기술기본법」에 따라 국가연구개발사업 추진에 필요한 정책을 추진함

“제11조(국가연구개발사업의 추진) ① 중앙행정기관의 장은 기본계획에 따라 맡은 분야의 국가연구개발사업과 그 시책을 세워 추진하여야 한다.” 「과학기술기본법」 제11조

- 국가 연구데이터 공유·활용체제 구축 사업은 「정보통신 진흥 및 융합 활성화 등에 관한 특별법」을 근거로 하여 추진함
 - － 「정보통신 진흥 및 융합 활성화 등에 관한 특별법」 제1조(목적)에서는 융합의 활성화를 목적으로 하는 정책, 규제, 인력 양성 등을 규정하여 국민 삶의 질 향상에 이바지함을 명시하고 있음
 - － 또한, 제11조에서는 정보통신 분야에서 전문 분야 기술 및 지식을 가진 인재양성을 위한 특화교육의 활성화에 관하여 명시함
 - － 그 외, 제16, 17조에서는 정보통신 진흥 및 융합 활성화에 기여하기 위한 정보통신 기술 및 서비스의 표준화 및 품질인증에 관해

32) 대체적으로 「국가연구개발사업의 관리 등에 관한 규정」에 근거하여 진행

“제1조(목적) 이 법은 정보통신을 진흥하고 정보통신을 기반으로 한 융합의 활성화를 위한 정책 추진 체계, 규제 합리화와 인력 양성, 벤처육성 및 연구개발 지원 등을 규정함으로써 정보통신의 국제 경쟁력을 제고하고 국민경제의 지속적인 발전을 도모하여 국민의 삶의 질 향상에 이바지함을 목적으로 한다.”

「정보통신 진흥 및 융합 활성화 등에 관한 특별법」 제1조

“제11조(국내 전문인력의 양성) ① 과학기술정보통신부장관은 정보통신 분야의 전문적인 기술, 지식 등을 가진 인력(이하 "전문인력"이라 한다)의 육성에 관한 시책을 수립·추진하여야 하며, 특히 소프트웨어 교육의 저변 확대 및 지역산업의 발전을 위한 소프트웨어 특화교육 활성화를 위하여 노력하여야 한다.”

「정보통신 진흥 및 융합 활성화 등에 관한 특별법」 제11조

“제16조(기술·서비스 등의 표준화) ① 과학기술정보통신부장관은 정보통신 진흥 및 융합 활성화를 위하여 정보통신융합등 기술·서비스 등의 표준화에 관한 다음 각 호의 사업을 추진할 수 있다.”

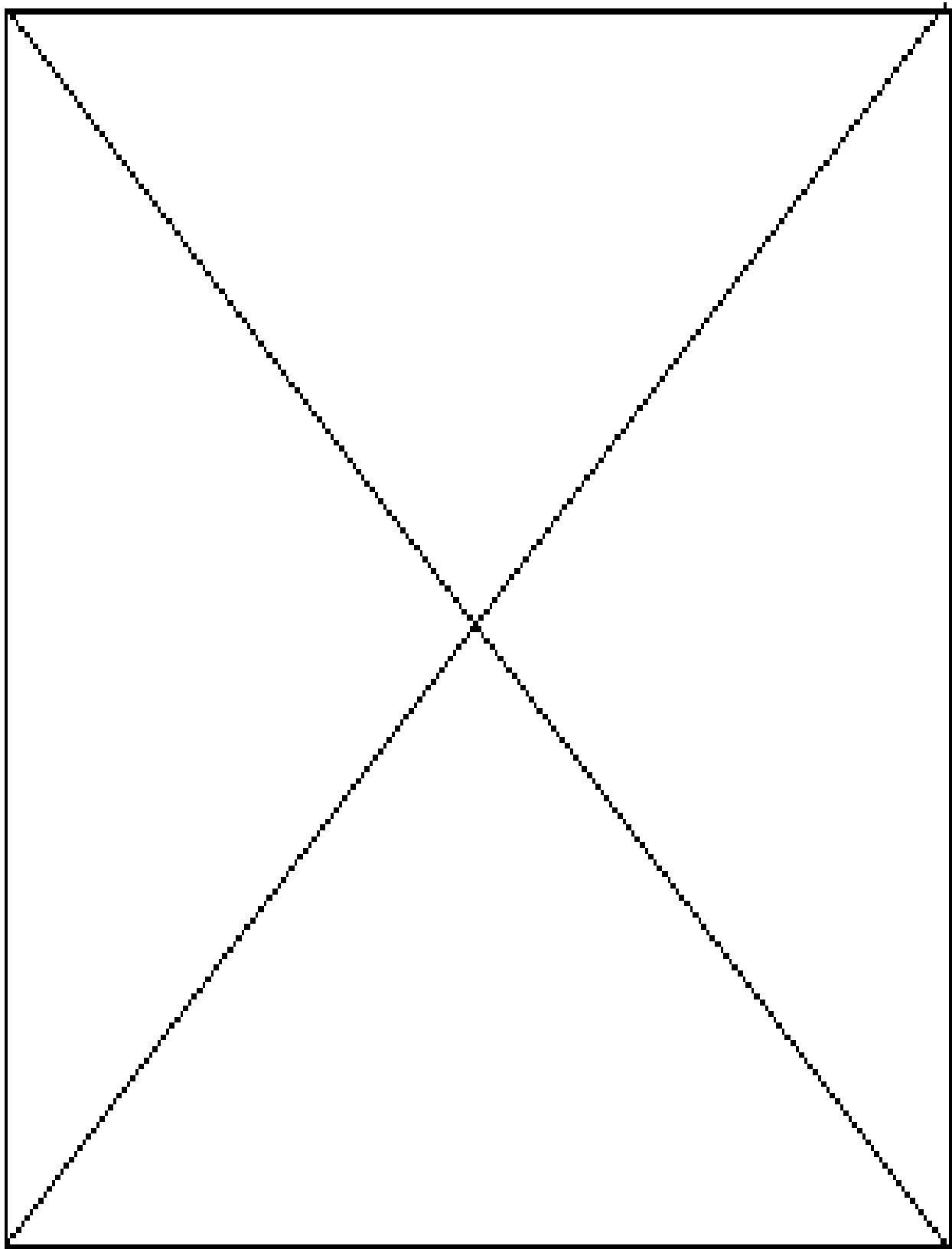
「정보통신 진흥 및 융합 활성화 등에 관한 특별법」 제16조

“제17조(기술·서비스 등의 품질인증) ① 과학기술정보통신부장관은 정보통신융합 등 기술·서비스 등의 편의성·안정성·신뢰성·확장성 등에 관한 인증기준(이하 "품질기준"이라 한다)을 정하여 고시할 수 있다.”

「정보통신 진흥 및 융합 활성화 등에 관한 특별법」 제17조

□ 사업 추진 절차의 적절성

- 동 사업은 국가연구개발사업의 일반적인 절차에서 크게 벗어나지 않으며, 절차와 더불어 사업 추진일정을 함께 제시하여 사업 수행에 불확실한 사항이 없는 것으로 보임
- － 동 사업을 위해 '18년 2월에 사전기획 및 기술성을 평가를 진행하는 등 국가전략 프로젝트를 심의하였으며, 과학기술 전략회의를 통해 사업추진을 확정함
- － 동 사업은 '20년~'29년까지 10년간 추진되는 사업으로, '20년 상반기에 동 사업을 위한 사업단 및 사무국을 구성하고,
- － 사업단 및 관계부처, 산·학·연 전문가로 위원회를 구성하여 과제기획서 및 RFP를 작성할 것으로 명시함
- － '20년 상반기에 신규 사업을 위한 공고를 추진하고, 접수된 과제를 검토·평가·심의함
- － '20년 말에는 주관기관과 협약 체결하고 연차별 실적 및 계획서를 평가하고 '21년 말에 연간 사업비 정산 및 연구활용 성과 조사 및 성과관리를 수행함



<그림 9-5> 사업 추진 절차 및 일정

□ 법·제도적 마찰 가능성

○ 연구데이터 공유·활용에 부정적 입장의 집단

- 국가연구개발과제 수행 과정에서 생산되는 연구데이터의 공유·활용 현황을 파악하기 위해 진행한 설문조사³³⁾ 응답 중
- 타 과학기술 분야의 연구데이터 활용의 어려움이 해결될 경우 연구데이터 공유·활용 참여에 부정적인 입장('그렇지 않다', '매우 그렇지 않다')으로 응답한 응답자가 298명 (17.21%)으로 확인 됨
- 따라서 연구데이터 공유·활용을 지지하지 않는 집단에 대해 어떻게 대응할 것인지에 대한 대응책을 마련할 필요가 있음
- 그러나 동 사업은 본인의 연구데이터가 공유·활용되는 것에 동의한 연구자를 대상으로 진행될 것이므로, 동 사업의 추진에 있어서의 위험요인은 없는 것으로 보임

○ 데이터에 대한 연구자의 권리

- 동 사업 추진 과정 중 연구데이터를 공유·활용하는데 있어 직접적으로 영향을 미치는 제약사항이나 충돌할 타 법령은 없으나,
- 국가연구개발사업 과정에서 산출되는 연구데이터에 대한 연구자의 권리와 이해관계에 대한 합의에 관한 제도적인 선결요건이 필요함
- 동 사업은 엠바고와 연구데이터 공개에 관한 사항은 연구자가 직접 설정할 수 있도록 하여 연구데이터에 대한 연구자의 선택을 반영함

○ WTO 보조금협정

- 연구데이터 공유·활용체제 구축사업의 산출물이 원천특허, 직접고용 및 전문분야센터의 설립 등으로 국제 교역재에 적용되어 외국과의 관계에 마찰을 일으킬 가능성을 검토함
- WTO의 보조금협정에서는 정부 또는 공적기관에 의한 재정적 기여로 인해 수혜자에게 혜택이 발생하는 경우 이를 보조금으로 인정함

“제1조 보조금의 정의

1.1 이 협정의 목적상 아래의 경우 보조금이 존재하는 것으로 간주된다.

가. (1) 회원국의 영토 내에서 정부 또는 공공기관(이 협정에서는 "정부"라 한다)의 재정적인 기여가 있는 경우, 즉

(가) 정부의 관행이 자금의 직접이전(예를 들어, 무상지원, 대출 및 지분참여), 잠재적인 자금 또는 채무부담의 직접이전(예를 들어, 대출보증)을 수반하는 경우,

(나) 정부가 받아야 할 세입을 포기하거나 징수하지 아니하는 경우(예를 들어, 세액공제와 같은 재정적 유인(Re.1)).”

「보조금 및 상계조치에 관한 협정」 제1조

33) 본 설문은 2018.08.14.~2018.09.07. 동안 현재 진행 중인 과제를 포함, 국가연구개발사업 종료 후 2년 이내의 과제를 그 대상으로 하여 수행되었으며, 연구데이터 관리·활용 및 애로사항 조사를 검토함

- 그 중에서도 ‘특정성’이 있는 보조금만을 그 규제 대상으로 하는데, 동 사업의 산출물은 정부출연연구원 및 대학에 적용되어 사용되어 짐으로,
- 보조금 협정에1서 명시하는 특정성의 요건을 성립하지 못하므로 국제통상 관련 위험요인의 분석범위에서 제외됨

“제2조 특정성

2.1 제1조제1항에 정의된 보조금이 공여당국의 관할내에 있는 특정 기업이나 산업 또는 기업군이나 산업군(이 협정에서는 "특정 기업"이라 한다)에 대해 특정적인지 여부를 결정함에 있어서 아래의 원칙이 적용된다.

가. 공여당국 또는 공여당국이 그에 따라 활동하는 법률이 보조금에 대한 접근을 특정 기업으로 명백하게 한정하는 경우 이러한 보조금은 특정성이 있다.

나. 공여당국 또는 공여당국이 그에 따라 활동하는 법률이 보조금의 수혜조건 및 금액을 규율하는 객관적인 기준 또는 조건 (Re.2) 을 설정하고, 수혜조건이 자동적이며 이러한 기준과 조건이 엄격히 준수되는 경우, 특정성은 존재하지 아니한다. 이러한 기준 및 조건은 검증이 가능하도록 법률, 규정 또는 그 밖의 공식문서에 명백하게 규정되어야 한다.

(Remark 2) 여기에서 사용된 객관적인 기준 또는 조건이라 함은 종업원 수 또는 기업의 규모와 같이 중립적이고, 특정 기업에 대하여 다른 기업보다 특혜를 주지 않으며, 성격상 경제적이며 적용시 수평적인 기준 또는 조건을 의미한다.

다. 가호 및 나호에 규정된 원칙의 적용결과 외견상 특정성이 없음에도 불구하고 보조금이 사실상 특정적일 수 있다고 믿을 만한 이유가 있는 경우에는 다른 요소들이 고려될 수 있다. 이러한 요소는 제한된 숫자의 특정 기업에 의한 보조금 계획의 사용, 특정 기업에 의한 압도적인 사용, 특정 기업에 대해 불균형적으로 많은 금액의 보조금 지급 및 보조금 지급 결정에 있어서 공여기관의 재량권 행사방식(Re.3)과 같은 것이다. 이 호를 적용함에 있어서 보조금계획이 집행되는 기간 뿐 아니라 공여기관의 관할하에 있는 경제활동의 다양화의 정도가 고려된다.

(Remark 3) 이와 관련하여, 특히 보조금 신청이 거부 또는 승인된 빈도에 대한 정보와 그러한 결정에 대한 이유가 고려된다.

2.2 공여기관의 관할지역중 지정된 지역내에 위치라는 특정 기업에 한정된 보조금은 특정적이다. 이 협정의 목적상 권한이 있는 각급 정부에 의해 일반적으로 적용될 수 있는 세율의 설정 또는 변경은 특정적인 것으로 간주되지 아니한다고 양해된다.

2.3 제3조의 규정에 해당되는 모든 보조금은 특정적인 것으로 간주된다.

2.4 이 조의 규정에 따른 특정성에 대한 판정은 명확한 증거에 기초하여 명백히 입증되어야 한다”
「보조금 및 상계조치에 관한 협정」 제2조

9.3. 기술적 타당성 분석



9.3.1. 사업 기획의 적절성

(방향) 사업 기획의 적절성은 예산투입 전 검토단계에서 사업 기획의 충실성 및 과학기술 개발 성공가능성에 영향을 주는 요인을 분석

(구성) 사업기획의 적절성은 사업의 배경 및 과정을 분석하는 ‘기획과정의 적절성’, 사업의 작동 논리를 분석하는 ‘사업목표의 적절성’, 세부 활동의 구체성 및 연계성을 분석하는 ‘구성 및 내용의 적절성’으로 구분

- 국가연구데이터 공유·활용체제 구축사업에 참여한 전문가 집단은 과학기술별 전문성과 전문가가 속한 기관 간 균형성을 확보하였으며,
- 우선순위 선정을 위한 적절한 평가기준의 적용 및 가중치 선정 등 공정하고 객관적인 합리적인 기획과정을 거쳤으며,
- 동 사업이 직면한 이슈/문제를 해결하기 위한 사업목표의 설정 그리고 사업목표에 유기적으로 연계되며 구체적인 세부과제의 내용이 적절하게 제시된 사업임

가. 기획과정의 적절성

(방향) 기획과정의 적절성은 과학기술 기획에 참여한 전문가 집단의 전문성·포괄성·균형성 및 관련분야의 연구개발수요 등과 대상사업의 선정과정 중 우선순위 설정과정의 합리성을 분석

- 국가 연구데이터 공유·활용체제 구축사업은 사업 규모에 적절하게 전문가 집단 및 소속의 다양성을 확보하였으며,
- 중앙부처의 중기 과학기술기본계획 등에 근거하여 사업 추진의 시급함이 증명된 사업임
- 또한, 동 사업을 위한 전문센터에 우선순위 선정과정에서 논리적이고 객관적인 절차 및 전문가 집단의 선정, 사업의 목표에 부합하는 세부 평가지표를 선정하여,

- 우선순위 선정의 합리성 및 적절성을 확보한 사업으로 사업의 목표 달성을 위한 적합한 과정을 제시한 사업임

□ 전문가 집단의 적절성

- 국가 연구데이터 공유·활용체제 구축사업의 규모 및 범위 등을 고려했을 때, 기술 분야별 다양한 전문가 집단이 참여하여 전문분야별 균형성을 확보함
 - 사업을 구성하는 개별 전략사업의 규모에 비교해 보았을 때, 국가 연구데이터 플랫폼 사업과 대·소 분야 전문센터 지원사업의 경우 20명에 가까운 전문가로 구성됨
 - 연구 분야의 전문가가 43%, 학문 분야의 전문가가 20%를 차지하는 등 분야별 사업에 참여한 산·학·연의 전문가가 불균형한 것처럼 보이나,
 - 이는 사업의 직접적인 수혜자가 연구원인 것을 감안하면 전문가 균형성에 문제가 되지 않을 것으로 판단됨
- 세부 사업의 특성에 맞게 해당 분야의 현황과약, 요구사항을 적절하게 대표할 수 있도록 전문가를 구성하여 균형성을 확보하고 있음
 - 국가 연구데이터 플랫폼 사업의 경우, 연구데이터 공유·활용의 직접적인 이해관계자인 연구 분야의 전문가 및 산업 분야와 학문 분야의 전문가를 균형 있게 선정하여 구성함
 - 전문센터 지원사업의 경우, 과학기술, 화학, 지질, 항공·우주, 천문 등 다양한 과학 기술 분야의 전문가를 포함하였으므로 전문가 집단의 균형성을 확보함
 - 인력양성 사업의 경우, 학문 분야 전문가의 비율을 증가하여, 해당 분야의 현황, 수요 및 요구 등을 적절하게 반영할 수 있도록 구성함
- 또한, 전문가 소속 측면에서도 특정 연구기관의 주도가 아닌 참여기관의 다양성을 확보하여 전문가의 의견이 한 쪽에 치우치지 않고 균형 있게 반영되기에 유리한 구성임

<표 9-17> 세부 사업별 참여 전문가

구분		참여자				
		산	학	연	기타	합계
플랫폼	인원수	5	3	6	4	18
	소속기관	충북대학교(1) 충남대학교(1) 한국과학기술원(1)	SK텔레콤(1) NVIDIA(1) 오라클(1) 아이티센(1) 솔리데오(1)	국가보안연구소(1) 한국전자통신연구원(1) 한국한의학연구원(1) 한국화학연구원(1) 국가핵융합연구소(1) 한국과학기술정보연구원(1)	국가생명연구자원정보센터(1) 정보화진흥원(1) 데이터베이스진흥원(1) 한국정보통신기술협회(1)	
전문센터	인원수	2	2	11	2	17
	소속기관	코아아이티(1) 버추얼랩(1)	충북대학교(1) 서울대학교(1)	한국과학기술정보연구원(4) 한국화학연구원(1) 한국지질자원연구원(1) 기초과학연구원(1) 재료연구소(1) 한국과학기술연구원(1) 한국항공우주연구원(1) 한국천문연구원(1)	국가생명연구자원정보센터(1) 한국정보화진흥원(1)	
인력양성	인원수	1	4	2	2	9
	소속기관	인텔코리아(1)	강원대학교(1) 전주교육대학교(1) 건국대학교(1) 과학기술연합대학원대학교(1)	한국과학기술정보연구원(2)	한경아카데미(1) 한국데이터진흥원(1)	
합계(명)		7	9	19	9	44
비중		16%	20%	43%	20%	100%

□ 연구개발수요의 적절성

- 국가연구데이터 공유·활용체제 구축사업은 국가전략적 시급함을 확보한 사업으로 연구개발수요를 위한 기술수요조사가 불필요함
- － 「국가연구개발사업의 관리 등에 관한 규정」에서는 사업의 특성에 따라 시급하거나 전략적으로 수행할 필요가 있는 사업의 경우에는 기술수요조사가 불필요함을 언급함

“제5조(기술수요조사) ① 중앙행정기관의 장은 정기적으로 기술수요조사를 하고, 그 결과를 반영하여 연구개발과제를 발굴하여야 한다. 다만, 연구에 참여하려는 연구자가 직접 연구기획 결과를 제출하도록 하는 연구개발과제 또는 시급하거나 전략적으로 반드시 수행할 필요가 있는 연구개발과제의 경우에는 기술수요조사 결과를 반영하지 아니할 수 있다.”

「국가연구개발사업의 관리 등에 관한 규정」 제5조

- 동 사업은 과학기술정보통신부에서 발표하는 과학기술기본계획과 현행정부의 국정과제, 그 외 중앙부처의 중·장기계획에 근거한 사업이며,
- 사업의 혜택이 연구소, 기업 나아가 전 국민을 대상으로 한다는 점에서 공공성이 매우 높은 사업임으로 국가전략적으로 시급히 시행할 필요가 있음

□ 우선순위 선정과정의 적절성

- 국가 전략적 중요성과 연구데이터 활용성 그리고 파급효과를 평가지표로 사업 특성에 적합한 항목 및 기준을 적용하여 과학기술 분야의 공공성, 기술성 및 시장성을 고려함
 - 정부 정책에의 부합성 및 해외 연구데이터 공유·활용 분야를 고려하여 국가 전략적으로 시급한 사업을 우선 선정함
 - 연구데이터의 생산량, 연구자의 연구데이터 활용의사 및 타 분야와의 연계 가능성을 고려하여 개별 과학 기술 분야가 가진 연구데이터 공유·활용의 가능성을 평가함
 - 개별 과학 기술 분야가 가진 시장성을 판단하기 위해 연구개발사업의 투자효율, 시장 파급성을 고려하였고, 기술 개발이 사회에 미치는 효과를 고용효과로 적용함
- 전문센터 우선순위 선정을 위한 절차는 객관적인 자료를 근거로 진행되었으며, 가중치 선정을 위해 평가에 참여한 전문가 집단의 객관성과 공정성을 확보함
 - 우선분야를 선정하는 절차는 전체 과학 기술 분야 중 대상 분야를 선정, 평가지표 선정 및 평가척도 산정, 정량적 평가 및 평가지표 중요도 산정 등을 거쳐 진행되어 최소한의 필요 과정을 공식적인 자료를 근거로 하여 진행함
 - AHP 분석을 통해 8개의 전략적 지표에 상대적 중요도를 배분하고, 중요도 산정을 위한 설문조사 대상으로 세부 과제별 분과위원회 의원 14명을 고루 선정하여 실시함
 - 국가연구데이터 공유·활용 실현을 위해 세부 과제에서 객관적이고 공정하게 전문가를 선정하여 평가지표별 가중치를 도출함
- 전문센터 우선분야 선정을 위한 평가지표는 국가연구개발사업이 가지는 정책적 정당성 및 수혜자의 선호도 그리고 사회에 미치는 파급성을 고루 고려하여 국가연구데이터 공유·활용을 활성화 한다는 사업의 목표에 부합하며 합리적임

나. 사업목표의 적절성

(방향) 사업목표의 적절성은 당면한 문제를 해결하기 위해 사업수요가 사업목표에 구체적이고 논리적으로 연결되었는지를 분석

- 국가 연구데이터 공유·활용체제 구축사업은 사업의 문제 및 이슈를 해외 사례 및 정책적·경제적·사회적·기술적 환경 등 다양한 관점에서 분석하였으며,
- 이를 통해 사업이 직면한 문제 및 이슈를 해결할 수 있는 구체적이고 타당한 근거를 가진 사업목표를 단계별로 제시하였음
- 사업목표와 함께 제시된 성과지표는 사업을 관리함에 있어 체계적이고 객관적임으로, 동 사업의 목표는 동 사업이 직면한 문제와 논리적으로 연결된 있는 사업임

□ 문제 및 이슈 인지

- 동 사업은 연구데이터 과정에서 생산되는 연구데이터가 연구 성과물로서 활용되어지지 못하는 국내 법·제도적인 한계를 인지하고,
- 연구데이터의 지식자산 활용과 빅데이터화를 비전으로 하여, 수요에 비해 부족한 제도적·물리적 열악함을 지원하고 개선하는 방향으로 사업의 논리와 목표를 제시함
 - － 연구데이터의 관리·활용이 중요해지고 있는 시점에서, 국내 및 해외 동향에 맞추어 국가 연구데이터 공유·활성화를 위한 제도적·물리적(인프라) 기반을 마련함
 - － 인프라 부족의 문제뿐만 아니라, 연구데이터 공유·활용의 직접적인 수혜자인 연구자 인식개선 및 의지 제고를 위한 방안을 도출함
- 해외의 선진사례, 국내 정책 및 경제·사회·기술적 관점 등 연구데이터의 공유·활용과 관련한 다양한 특성 및 현황을 광범위하게 분석함
 - － 해외 선진국의 과학데이터 관련 법·제도, 추진·지원을 갖춘 조직 및 인프라 등을 분석하고 이를 통해 국가 주도의 연구데이터 공유·활용체제 구축사업을 위한 시사점을 도출함
 - － 세계적으로 과학기술 연구성과 및 연구과정을 개방화하는 오픈사이언스 정책이 확산하고 있는 추세임에도 불구하고, 국내 법·제도는 이러한 특징을 반영하지 못함
 - － PEST 분석을 통해 정책적·경제적·사회적·기술적 환경을 분석하였으며, SWOT 분석으로 내·외부 환경 분석과 더불어 성공적인 연구데이터 공유·활용을 위한 대응전략을 제시함
- 현재 상황과 목표로 하는 미래 상황의 차이에 의해 발생할 수 있는 장애용인 규명, 문제 해결에 있어 한계점을 인지하고, 이를 사업목표를 수립할 때 중요한 요소로 고려

- 연구데이터를 개인의 성과물로 인지하여 연구데이터 공유·활용에 부정적인 연구자가 있더라도, 연구데이터 공유·활용에 동의한 연구자만을 대상으로 선결진행 하며,
- 연구데이터 엠바고, 공유·활용함에 있어 동 사업의 목표와 연구자 간에 발생하는 이해관계를 향후 협의·조정하는 방향으로 사업을 진행함

□ 사업목표의 구체성 및 직면한 문제와의 연계성

- 동 사업은 연구데이터 지식자산化·빅데이터化로 국가연구개발의 효율성과 경쟁력을 강화하는 비전을 달성하기 위해 장기적이고 서로 보완적인 전략적 목표 4가지를 설정함
 - 동 사업의 목표는 연구데이터 공유·활용의 인식제고, 다양한 분야들을 결합한 융합 연구 활성화, 연구의 효율성 및 투명성과 연구결과물의 수준 제고, 연구데이터 공유·활용을 위한 전문인력 양성임
 - 효과적인 목표 달성을 위해 연구개발·변화관리, 기술이전, 논문/특허 출원·등록 등을 지원하고 기술융합허브 및 테스트베드 운영을 통해 사업 실현화 및 응용가능성을 제고함
 - 동 사업에서 산출되는 결과물은 공통 원천기술 및 차세대 원형플랫폼 등의 지적재산과 혁신네트워크 등이 있으며,
 - 이 외에 연구데이터를 활용한 연구개발의 규모가 확대되고 기술을 활용하는 이용자 수가 증가하며 기술성능·안전성 인증 기술의 증가 등을 기대할 수 있음

<그림 9-6> 사업 비전 및 목표

- 연구데이터 지식자산化·빅데이터化를 통한 국가 R&D성과 혁신 달성을 위해 동 사업이 제시한 목표는 동 사업이 직면한 이슈/문제를 해결하기에 구체적이고 체계적이며,
- 사업목표에 따른 성과관리를 위해 객관적인 성과지표를 함께 제시하여 사업목표가 달성하고자 하는 효과를 구체적으로 제시함
 - － (연구데이터 공유·활용의 인식제고) 해당 사업목표는 연구개발 분야에서 연구데이터가 중요한 요소로 작용하고 이를 공유·활용하는 전반 적인 분위기를 창출하는 것으로,
 - － 연구데이터가 연구 성과물로 인정되지 않는 시점에서 연구데이터를 관리하고 공유·활용 까지 지원하여 동 사업이 직면한 근본적인 문제, 연구데이터의 낭비를 방지하고 연구데이터가 가진 잠재력을 활용하여 이슈/문제를 해결하기에 적절한 목표임
 - － 해당 목표를 관리하기 위해 과학적, 기술적, 사회적 성과 및 인프라적 성과지표를 선정 하였으며 사업 기간 동안 달성할 객관적인 성과목표를 함께 설정함
 - － (다양한 분야들을 결합한 융합 연구 활성화) 해당 사업목표는 전문분야별 연구데이터의 집약과정을 지원하여 공통의 연구데이터를 활용한 공동연구 기반을 확보하는 것으로,
 - － 과학기술 전문센터에서 수집·관리되는 연구데이터는 국가 연구데이터 플랫폼에서 활발한 연구데이터 공유·활용에 사용되어 데이터기반의 연구개발을 정착하는데 기반이 됨
 - － 해당 목표는 전문분야별 연구데이터 관리를 위한 인프라를 지원함으로써 연구데이터 기반의 연구 환경조성과 더불어 연구데이터 공유·활용 생태계를 조성하는데 이바지함
 - － 해당 목표를 관리하기 위해 과학적, 기술적, 경제적, 사회적 성과 및 인프라적 성과지표를 선정하였으며 사업 기간 동안 달성할 객관적인 성과목표를 함께 설정함
 - － (연구의 효율성 및 투명성과 연구결과물의 수준제고) 해당 사업목표는 연구데이터의 공유 및 재사용 즉 활용을 통해 달성할 수 있는 목표로,
 - － 연구데이터 공유·활용의 인식제고와 다양한 분야의 융합 연구 활성화 과정을 통해 연구데이터의 투명성을 확보할 수 있고, 데이터를 위한 새로운 연구의 방지 등 비용 및 연구과정 절감을 통한 효율성을 제고할 수 있음
 - － (연구데이터 공유·활용을 위한 전문인력 양성) 해당 사업목표는 향후 연구데이터 공유·활용 생태계를 이끌어 나갈 수 있는 전문가 및 인재를 육성하는 것으로,
 - － 연구데이터 관리를 위한 기술교육 외에도 연구데이터를 연구개발에 응용할 수 있도록 전문분야별 활용교육을 실시하여 전문분야별 연구데이터 활용 및 공동연구 실현에 기여하고 나아가 장기적으로는 국가 연구데이터 공유·활용 생태계 구축에도 이바지함
 - － 해당 목표를 관리하기 위해 기술적, 사회적 성과 및 인프라적 성과지표를 선정하였으며 사업 기간 동안 달성할 객관적인 성과목표를 함께 설정함

- 동 사업은 데이터기반 연구개발 정착을 달성하기 위해 다수의 목표를 제시하여 사업추진의 효율성을 제고하고 이슈/문제를 전략적으로 해결하고자 함
 - － 동 사업이 직면한 이슈/문제는 법·제도, 인프라, 인력 등 여러 가지 문제가 얹혀 있기 때문에 하나의 일관된 목표로 동 사업의 문제를 해결하는 것 보다는, 사업의 목적 달성을 위해 서로 유기적으로 연계된 다수의 사업목표 설정이 더 적합함
 - － 따라서 국가 전체적으로는 연구데이터 공유·활용 인식을 제고하여 연구데이터 활용이 활발해 질 수 있는 연구 환경 구축 및 신뢰높은 연구결과물을 산출할 수 있고,
 - － 부분적으로는 연구데이터 공유·활용이 보편화된 분위기 속에서 다양한 분야의 융합 연구가 가능해지며, 이 과정에서 기존 연구데이터의 사용을 통해 데이터 활용도를 제고함
 - － 전문분야별 연구데이터의 집약과 공동연구 기반을 확보함으로써 연구데이터 활용에 있어 존재하는 법·제도적 한계와 열악한 인프라를 지원함
 - － 또한, 데이터를 활용하는 전문인력 성장 지원 사업을 통해 일시적인 연구데이터 공유·활용이 아니라, 향후 지속적인 데이터 관리와 공유·활용이 유지되도록 지원함
- 제시된 성과지표는 시간의 관점에서 사업이 직면한 이슈/문제를 해결위해 사업의 수준 및 시기를 고려한 단계별 성과목표를 구성하여 체계적이고 일관적이며,
- 「국가연구개발사업 표준 성과지표」³⁴⁾를 근거로 한 성과지표를 활용하였으므로 사업의 수행성과를 측정·관리하기 위한 기본단위로서의 타당성과 구체적인 근거를 확보함
- 또한, 설정된 사업지표는 사업을 통해 발생하는 직접적인 결과물을 목표로 제시하여, 성과지표를 측정하기 때문에 객관적으로 사업의 목표를 관리할 수 있음
 - － 동 사업은 사업의 목표를 총 3단계에 걸쳐 수행함으로써 시간적 관점에서 동 사업이 직면한 이슈/문제를 관리할 수 있으며,
 - － 미래창조과학부(現 과학기술정보통신부)에서 제시한 국가연구개발사업을 위한 표준 성과지표 10대 사업유형, 5대 성과지표를 기준으로 성과지표를 선정하여, 성과지표가 사업목표 달성에 있어 대표성을 확보하며, 사업목표를 관리하는 기본단위로서 타당함
 - － 또한, 제시된 성과지표는 사업을 수행하는 과정에서 생산되는 성과물³⁵⁾을 체계화·구체화 하여 객관적인 사업목표 관리를 실현함

34) 「국가연구개발사업 표준 성과지표(4차)」, 2014.12

35) 논문, 특허, 표준모델, 교육을 통해 양성된 인력의 수, 사업을 통해 새롭게 창출된 일사리의 수 등 사업 수행과정에서 발생할 것으로 기대되는 성과물

<표 9-18> 사업목표에 따른 성과지표 및 단계별 성과목표

사업목표	성과지표		성과목표			
	구분		1단계 (‘21~’23)	2단계 (‘24~’27)	3단계 (‘28~’30)	계
국가 연구데이터 의 공유·활용 생태계 조성	과학 적 성과	논문(SCI)	64	88	60	212건
		논문(KCI)	145	189	130	464건
	기술 적 성과	특허	출원(70)/ 등록(28)	출원(90)/ 등록(35)	출원(63)/ 등록(24)	출원(223)/ 등록(87)
		기술혁신 (표준모델)	20	26	12	58개
		콘텐츠S/W	10	3	0	13건
전문분야별 연구데이터 의 집약과 공동연구 기반확보	경제 적 성과	기술 활용 효과		5조	5조	10조
		중소기업 지원		기업양성(1)	기업양성(1)	2개
		연구개발 서비스	4	8	6	18개
	사회 적 성과	인력양성	8,195	12,523	10,712	31,430명
		일자리창출	1,319	2,046	1,783	5,148 명
정책효과		3	4	3	10건	
홍보확산		3	4	3	10건	
빅데이터, AI기술 등 데이터를 활용하는 R&D 확산		국제협력	1	1	1	3건
	인프 라 성과	공동 활용 시설장비	데이터set (1000)	데이터set (2500)	데이터set (2500)	6,000개
			교육용DB 구축(3)	교육용DB 구축(3)	교육용DB 구축(4)	11개
			서버장비(15)	서버장비(3)		18개
	전산시스템	39Core /306TB	12Core /408TB	9Core /306TB	30Core /1PB	
		지원시스템(1) 세부별(5)	세부별(1)		7개	

□ 수혜자의 표적화

- 동 사업은 데이터기반의 연구개발을 정착을 목표로 하며 사업목표 달성을 통해 발생하는 편익의 직접적 수혜자는 연구자임
 - 국가 연구데이터 공유·활용을 통해 연구의 중복을 방지하고, 타 과학기술 분야와의 융합 연구 및 공동 연구를 통해 연구자의 연구개발 활동 범위를 확대함
 - 국가 연구데이터 공유·활용은 연구자로 하여금 기존의 연구데이터 관리를 통해 새로운 연구를 위한 예산을 낭비하지 않도록 하며, 이를 통해 사업의 효율성을 제고함
 - 또한, 연구자는 기존에 쉽게 접근하지 못했던 타 분야 연구데이터로의 접근을 통해 데이터 확보를 위한 불필요한 과정을 줄이고 연구데이터 융·복합 활용을 실현할 수 있음
- 국가 연구데이터 공유·활용체제 구축사업의 기술력 제고 및 전문센터 설립의 경우 편익의 수혜자와 동 사업이 산출하는 성과물의 전달 과정이 명확함
 - 전문분야별 연구데이터 집약 및 공동연구의 기반마련을 통해 연구자가 연구데이터를 관리·공유 할 수 있는 인프라 환경을 조성하고,
 - 국가 연구데이터의 공유·활용이 활발하게 이루어지는 연구 환경 조성은 연구자 간 융합 연구 및 공동연구를 적극적으로 지원하는 수단이 되어 과학 기술력을 제고함
 - 동 사업이 목표한 성과물은 과학·기술적 역량을 제고하는데 이바지하며, 연구자는 증가한 과학·기술적 역량을 통해 더욱 활발한 연구개발 활동을 수행하는 등 연구데이터 공유·활용의 선순환 구조를 형성함

다. 구성 및 내용의 적절성

(방향) 구성 및 내용의 적절성은 연구개발사업의 목표달성을 위해 연구개발사업 및 세부과제가 효과적, 유기적, 효율적으로 연계되었는지를 분석

- 국가연구데이터 공유·활용체제 구축사업에서 제안된 세부과제는 사업목표와 유기적으로 연계되어 세부과제의 수행을 통해 사업의 목표를 달성할 수 있으며,
- 사업목표 달성을 위해 세부과제 간 서로 다른 세부과제를 보완하여 사업목표를 효과적으로 달성함
- 세부과제를 관리하기 위한 성과지표는 성과목표를 연차별로 계획하여 세부사업 간 선·후 관계에 있어 동 사업의 목표 달성에 있어 효율성을 제고하는 사업임

□ 국가연구데이터 플랫폼 구축·운영

- 연구데이터 수집·연계 및 공유 기술
 - － 연구데이터 수집·연계 및 공유 기술은 국내·외 연구데이터/메타데이터 통합 및 공유를 위한 수집·전송 시스템 개발을 목표로 하는 기술 개발임
 - － 연구데이터/메타데이터 연계는 대·소 분야 전문센터와 국내·외 연구기관이 생산한 연구데이터 및 메타데이터를 수집하고 수집된 데이터를 연계·활용하는 기술이며,
 - － 연구데이터 연계기술 개발은 국내·외 연구기관이 생산된 연구데이터를 연계하고 공유하여 기술개발에 활용하는 것을 지원하는 기술임
- 연구데이터 저장·보안 기술
 - － 연구데이터 저장·보안 기술은 메타데이터 및 DMP정보 통합 관리와 대용량 연구데이터 분석을 위한 연구데이터 저장, 보안 시스템 개발을 목표로 하는 기술 개발임
 - － 클라우드 기반 분산, 병렬 저장 기술 개발은 연구데이터/메타데이터를 저장하기 위한 기술로, 대용량의 이질적인 데이터의 저장·분석을 처리하는 하둡을 활용하여 연구데이터를 저장하고 관리하는 기술임
 - － 데이터 저장 및 연계 보안기술 개발은 메타데이터에 포함되어있는 개인정보를 보호하기 위한 기술로, 데이터로의 접근 및 공유를 위해 개인인증체계를 적용하는 기술임
- 연구데이터 품질·표준화 기술
 - － 연구데이터 품질·표준화 기술은 연구데이터/메타데이터의 품질제고를 위한 연구데이터

품질관리 시스템 개발을 목표로 하는 기술임

- 데이터 표준 수립 및 품질관리 기술 개발은 수집·연계되는 연구데이터/메타데이터에 대해 조회, 표준 체크 등의 관리를 통해 데이터의 활용성과 품질을 제고하는 기술임
- 연구데이터 저작권 관리 기술 개발은 수집·공유된 연구데이터의 출처 및 소유권 정보를 관리하여 저작권을 보호하고 관리하는 기술임

○ 지능형 연구데이터 검색 기술

- 지능형 연구데이터 검색 기술 개발은 통합 검색 서비스 제공 및 연구플랫폼 서비스 활성화를 위한 검색 기술 개발을 목표로 하는 기술 개발임
- 연구데이터 분석 정보를 실시간으로 수집하여 색인관리, 검색관리 그리고 상세검색 영역으로 구분하여 검색기술의 기능을 구현함

○ 연구데이터 융합·분석 환경 구축 기술

- 연구데이터 융합·분석 환경 구축 기술 개발은 손쉬운 연구데이터 분석 환경을 제공하기 위한 융합·분석 환경 개발을 목표로 하는 기술 개발임
- 연구데이터의 관리·공유뿐만 아니라 연구데이터 활용을 활성화하기 위해 클라우드를 기반으로 한 빅데이터 분석 환경을 제공하는 기술로,
- 연구플랫폼에 BI, AI, ML, DL 솔루션을 지원하는 기술과 분석 솔루션을 활용하기 위한 workflow 서비스 기술을 지원하는 개발임

○ 연구데이터 활용·플랫폼관리 기술

- 연구데이터 활용·플랫폼관리 기술 개발은 24·365 무중단 연구데이터 활용을 위한 클라우드 서비스 기술 개발을 목표로 하는 기술 개발임
- 협업 서비스 및 관리 기술 및 커뮤니티 생성 및 관리 기술 개발은 연구자 간 협업 및 커뮤니티 형성을 지원하는 기술로,
- 연구플랫폼 클라우드의 안정적인 서비스 제공을 위해 클라우드 고가용성 기술, 클라우드 어드민/아키텍처 관리 기술, 클라우드 보안 관리 기술 및 클라우드 가상머신 관리 기술을 도출함

<표 9-19> 국가 연구데이터 플랫폼 구축 및 운영 과제 연도별·단계별 개발 활동

세부과제	핵심 기술	세세부 기술	기술 수준	1단계			2단계				3단계		
				'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
연구데이터 수집·연계 및 공유 기술	연계 프레임워크	연구데이터/메타데이터 연계	TRL 3 → TRL 6	설계 및 기술 개발	최적화 및 안정화		기능 고도화		정상운영				
		연구데이터 연계기술 개발											
연구데이터 저장·보안 기술	저장·보안 프레임워크	클라우드 기반 분산·병렬 저장 기술 개발		-	설계 및 기술 개발	최적화 및 안정화	기능 고도화		정상운영				
		데이터 저장 및 연계 보안기술 개발											
연구데이터 품질·표준화 기술	품질·표준 프레임워크	데이터 표준 수립 및 품질관리 기술 개발		설계 및 기술 개발	최적화 및 안정화		기능 고도화		정상운영				
		연구데이터 저작권 관리 기술 개발											
지능형 연구데이터 검색 기술 개발	검색 프레임워크	연구데이터 검색 기술 개발		-	설계 및 기술 개발	최적화 및 안정화	기능 고도화		정상운영				
연구데이터 융합·분석 환경 구축 기술	융합·분석 프레임워크	Workflow 서비스 기술 개발		-			설계	기술 개발	최적화 및 안정화	기능 고도화		정상운영	
		Batch&Realtime 분석 기술 개발											
		BI 서비스 기반기술 개발											
		AI 서비스 기반기술 개발											
		ML 서비스 기반기술 개발											
		DL 서비스 기반기술 개발											
		가시화 기반 기술 개발											
연구데이터 활용·플랫폼관리 기술	품질·표준 프레임워크	협업 서비스 및 관리 기술 개발		설계	기술 개발	최적화 및 안정화	기능 고도화		정상운영				
		커뮤니티 생성 및 관리 기술 개발											
		클라우드 고가용성 기술 개발											
		클라우드 어드민/아키텍처 관리 기술 개발											
		클라우드 보안 관리 기술 개발											
		클라우드 가상머신 관리 기술 개발											

- 국가 연구데이터 플랫폼에서 연구데이터 공유·활용을 실현하는 융합연구 및 공동연구를 지원하기 위한 과학기술 분야 전문센터 11개를 선정하고,
- 그 중 국가전략적, 연구데이터 공유·활용 효과성 및 파급효과를 기준으로 우선적으로 사업을 시행할 3개 과학기술분야를 선정함
- 바이오 전문센터 구축
 - 바이오 분야는 사업 초기 년도에 생명연구자원 및 연구 성과물을 확보하고 정제하는 작업을 지속적으로 진행하며, 연구데이터 연계체계를 구축함
 - 사업 중반 부에 들어 연구데이터 전문센터를 운영하고 정보기반의 멀티오믹스 분석시스템을 구축·운영함
- 소재 전문센터 구축
 - 소재 분야는 사업 전반에 걸쳐 소재 연구데이터 수집·관리·공유체계를 구축하며 에너지·환경소재 데이터, 스마트/IT 소재 데이터 그리고 안전(구조) 소재 데이터로 구분하여 사업을 추진함
- AI 전문센터 구축
 - AI 분야는 사업 전반에 걸쳐 AI 데이터 서비스를 통한 연구 데이터 허브를 마련하기 위한 전처리 기술과 빅데이터 네트워크를 개발함
 - 사업 초기 년도에 AI 연구데이터 관리체계와 보안체계를 구성하여 AI 데이터를 지원할 수 있는 인프라 구조를 마련하고,
 - AI 연구데이터 거버넌스를 정립 및 AI 연구데이터 전문 인력 양성을 통하여 AI 연구생태계를 형성함
 - 또한 AI 연구데이터 챌린지 센터를 건립하여 AI 기반 연구의 혁신성과, 응용성, 표준화 겸비한 소분야 발굴을 계획함

<표 9-20> 대·소 분야 전문센터 구축 및 운영지원 과제 연도별·단계별 개발 활동

전문센터 분야	추진계획	1단계			2단계				3단계		
		'21	'22	'23	'24	'25	'26	'27	'28	'29	'30
바이오 전문센터 구축	바이오 연구데이터 연계체계 구축										
	바이오 연구데이터 수집										
	바이오 연구데이터 활용시스템 구축 및 지원을 위한 인프라 구축										
소재 전문센터 구축	소재 연구데이터 수집·관리·공유 체계 구축										
	에너지·환경소재 데이터 수집·활용 체계 구축										
	다공성 소재 연구데이터 수집·활용 연구										
	금속화합물 소재 연구데이터 수집·활용 연구										
	이차원 소재 연구데이터 수집·활용 연구										
	스마트/IT 소재 데이터 수집·활용 체계 구축										
	차세대 반도체 소재 연구데이터 수집·활용 연구										
	센서 소재 연구데이터 수집·활용 연구										
	바이오/의료 소재 연구데이터 수집·활용 연구										
	안전(구조)소재 데이터 수집·활용 체계 구축										
	극한환경 소재 연구데이터 수집·활용 연구										
	경량 소재 연구데이터 수집·활용 연구										
	수소인프라 소재 연구데이터 수집·활용 연구										
인공지능 전문센터 구축	AI 데이터서비스										
	AI데이터기반										
	AI연구생태계 구축										
	소분야 AI연구데이터 챌린지 센터										

□ 데이터 및 컴퓨팅 활용 R&D 인재 성장 지원

- 연구데이터 관리를 위한 ICT 기반기술 교육
 - － 데이터 관리자와 데이터 운영자로 교육내용을 구분하여 직무별 전문성을 제고함
- 연구데이터의 R&D 응용을 위한 분야별 전문 활용 교육
 - － 선정된 과학기술 전문센터의 데이터 분석가를 대상으로 하여 다양한 데이터 분석 프로

그림과 평가 시스템 등을 공통으로 교육하고

－ 분야별로 해당 전문센터 분야의 알고리즘과 수치해석, 가시화 기법 등을 교육내용으로 함

○ 연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축

－ 연구데이터를 활용한 교육을 위해 교육 인프라 및 서비스를 구축하는 연구개발 활동으로 전문분야 데이터셋을 구축하고 CX-Learning 도구개발 및 교재를 개발함

<표 9-21> 데이터 및 컴퓨팅 활용 기반의 R&D 인재육성 과제 연도별·단계별 개발 활동

추진계획	운영내용		1단계			2단계			3단계		
			'21	'22	'23	'24	'25	'26	'27	'28	'29
연구데이터 관리를 위한 ICT 기반기술 교육	데이터 관리자										
	데이터 운영자										
연구데이터의 R&D 응용을 위한 분야별 전문활용 교육	공통										
		ICT·SW									
		생명·보건									
		소재·나노									
		농림수산·식품									
		환경·기상									
		기계·제조									
		우주·항공·해양									
		원자력·핵융합									
		에너지·자원									
		건설·교통									
		재난안전									
연구데이터 활용 인재육성을 위한 교육용 Databank 및 서비스 구축	전문분야 데이터셋 구축 및 데이터 기반 교육 콘텍스트 개발										
		데이터 기반 연구 콘텍스트 수립									
		데이터셋 수집 및 생산									
	CX-Learning 도구개발 및 콘텐츠 교재제작										
		학습/제작 도구 개발 및 사용 테스트									
		강의형태 교재 제작									
	데이터교육 지원센터 지원시스템 구축										

□ 세부과제의 적절성

- 국가 연구데이터 공유·활용체제 구축을 위한 세부과제는 연도별·단계별 개발 내용이 명확하게 제시되었으며,
- 개별 기술은 기술 개발의 명확한 목표를 가지고 있어, 목표 달성을 위한 구체적인 산출 기술을 명시함
 - － 국가 연구데이터 공유·활용체제 구축 기술은 연구데이터/메타데이터를 수집·전송 및 저장(보완기술 포함)하고, 수집·공유된 데이터의 표준화를 통해 품질을 관리하는 기술로,
 - － 연구데이터의 공유·활용을 위해 기본적으로 수행되어야 하는 수집·관리·저장 기술을 목표로 함으로 전체 사업목표인 국가 연구데이터 공유·활용의 활성화에 기여함
 - － 또한, 통합 검색 기술과 데이터 분석 기술 개발을 통해 연구플랫폼 서비스를 활성화하고 언급된 서비스를 24시간, 365일 연중 안정적으로 활용 가능하도록 개발하여,
 - － 단순한 연구데이터의 공유·활용의 플랫폼이 아니라 수집된 데이터의 표준화, 공유된 데이터의 실시간 분석 등 대·소 분야의 전문센터에서 연구데이터 공유·활용을 실현할 수 있도록 지원하는 사업임

□ 사업목표와 세부과제의 논리적 연계성

- 사업목표로의 종속성
 - － 동 사업의 세부과제는 연구데이터 공유·활용을 위한 기반 기술과 지능형 검색기술, 분석 서비스를 함께 제공하기 위한 기술 개발과
 - － 개발된 기술로 구축하는 다양한 과학기술 연구 분야의 연구자가 연구데이터 공유·활용을 기반의 연구개발을 실현하도록 과학기술 분야 전문센터를 구축하고 운영하는 과제,
 - － 그리고 미래의 연구데이터 공유·활용을 위한 전문가를 양성하고 산업적 활용 촉진을 통한 일자리 창출 등 사업목표 달성을 위한 일관성 있고 상호보완적인 구조적 연계성을 가짐
- 국가연구데이터 플랫폼 구축·운영
 - － 연구데이터 관리와 더불어 지능형 검색 및 데이터 융합·분석 서비스를 함께 제공함으로써, 연구자가 연구데이터를 기반의 연구개발을 수행할 수 있는 환경을 제공하고
 - － 연구데이터의 공유·활용을 실현하는 인프라를 설치하여 국가 연구데이터의 공유·활용 생태계 조성에 이바지함
 - － 국가전략적, 연구데이터의 공유·활용 가능성 및 파급효과를 기준으로 전문센터별 동 사업을 수행하는 우선순위를 선정하여 사업의 효율성을 제고하고

- 전문 분야별로 연구데이터를 수집·관리할 수 있는 인프라를 지원하여 전문분야별 연구데이터의 집약과 공동연구 기반을 확보하는 데 이바지함
- 데이터 및 컴퓨팅 활용 R&D 인재 성장 지원
 - 연구데이터 관리를 위한 기술교육과 과학기술 분야별 특성에 맞는 전문 활용 교육을 제공하여,
 - 미래에 양성된 전문가를 통해 빅데이터, AI 기술 등 데이터를 활용하는 R&D가 확산될 수 있는 기반을 마련하는 과제임
- 연구데이터 관리·공유·활용 관련 법제도 마련
 - 연구데이터의 관리·공유·활용에 관련한 법적 근거를 마련하는 것은 연구데이터 공유·활용의 인식 제고 및 장기적이고 실질적인 활용을 위한 기반이 되는 사업임
 - 법적 근거 마련을 통해 연구데이터 공유·활용에 관한 인식을 제고할 수 있고, 공식적인 가이드 라인을 통해 공유·활용에 있어 혼란을 최소화 할 수 있음

□ 세부과제 성과지표의 적절성

- 동 사업의 세부과제 성과목표는 세부과제 목표 및 사업목표를 달성을 위해 구체적이고 측정가능하며 사업 기간 내 단계별 성과관리가 가능하도록 연차별 성과목표를 제시함
 - 동 사업 세부과제 별 성과지표는 미래창조과학부(現 과학기술정보통신부)의 10대 사업 유형, 5대 성과지표³⁶⁾를 기준으로 하여 선정됨
 - 동 사업의 유형은 연구개발을 위한 기초연구, 중·장기산업 기술개발과 연구기반을 조성하는 인력양성 및 시설장비구축으로 분류되며, 해당하는 사업유형에 따라 5대 성과분야가 모두 포함되어, 총 15가지의 성과지표가 선정됨
 - 동 사업의 산출물이 과학 분야에 기여하는 학문적 수준을 평가하기 위해 논문 게재 수, 학술지 수준 등을 성과목표로 선정하여 세부과제를 관리함
 - 연구데이터의 수집·관리를 위해 빅데이터화, 클라우드 기반 환경 등 기술적 혁신이 예상되며, 이러한 혁신들의 성과물을 관리하기 위해 개발된 기술의 특허, 표준획득 및 S/W 등록 등을 성과목표로 산정함
 - 플랫폼 및 전문센터에서 제공하는 공유·활용된 연구데이터, 분석 인프라 등으로 간접적 성과가 발생하며, 분석환경 및 연구개발서비스 등으로 경제적 성과를 관리함
 - 연구데이터 공유·활용을 위한 전문가 양성과 기술의 개발 및 융합연구 등의 발전으로 열린 신산업 창출, 일자리 창출 등 다양한 사회적 성과로 세부과제를 평가함

36) 「국가연구개발사업 표준 성과지표(4차)」, 2014.12

□ 세부과제 간 연계성 및 시간적 선·후행 관계 적절성

- 동 사업의 세부과제는 연구데이터의 공유·활용을 위한 국가 연구데이터 플랫폼을 구축하고 선정된 전문센터에서 수집된 연구데이터를 공유·활용 할 수 있는 환경을 제공하여
- 사업업의 비전을 실현함에 있어 일관성이 있으며, 과제 간 특성을 고려하여 일부 동시에 진행하는 일정을 계획하여 사업의 효율성을 제고하기도 함
 - － 국가 연구데이터 플랫폼이 제공하는 연구데이터의 수집·관리·표준화와 데이터 품질 제고 등은 전문센터 구축 사업과 긴밀하게 연계됨으로 단계별 일정을 함께 계획함
 - － 국가 연구데이터 플랫폼 구축 및 운영 사업에서 개발된 기술을 대소·분야 전문센터 구축·운영 지원 사업과 연계하여 전문센터가 데이터 생성·저장·유통의 역할을 수행하며,
 - － 다양하고 표준화된 데이터를 제공하여 연구자 간 공동 연구의 기회를 제공하고 연구자 커뮤니티 형성 및 협업을 지원하는 등 기술개발을 통한 사업간 전략적 연계가 가능함
- 미래 연구데이터 전문가를 양성하는 교육 사업을 통해 실질적인 연구데이터 공유·활용과 더불어 연구데이터의 산업적 활용 촉진 및 일자리 창출에 기여함
 - － 데이터 활용 기반의 연구방법이 확산되는 분위기 조성을 위해 사업 초기부터 지속적인 인재육성 사업을 구축하고,
 - － 전문 연구기관의 교육 프로그램 및 분석 플랫폼을 활용한 심화 교육과정은 개별 전문센터가 구축되어 최적화 및 안정화 단계와 연계하여 과제의 효율성을 제고함

9.3.2. 기술개발 성공가능성

(방향) 기술개발 성공가능성은 기술 속성의 관점에서 동 사업과 관련된 쟁점사항을 통해 기술개발 성공가능성에 영향을 주는 요인을 검토하고 이를 통해 일정지연(schedule delay) 및 비용증가(cost overrun)를 분석

- 국가 연구데이터 공유·활용체제 구축사업은 기술추세 측면에서 세계 주요 국가가 성장기에 해당하여 대규모의 연구개발 투자를 위해 합리적인 시기임
- 동 사업에서 제시된 기술이 해외 선진국에 비해 우위에 있는 기술은 아니나 분명, 세계 시장을 목표로 한 기술개발에 가능성이 있으며,
- 상대적으로 저조한 질적 수준을 제고하기 위하여 사업목표 달성을 위한 적절한 기술 수준을 확보하기 위하여 대규모의 재원투입 정당성이 확보되는 사업임

가. 기술추세 분석

(방향) 기술추세 분석은 사업에서 제시된 분야 및 사업기간을 중심으로, 연구개발사업의 착수, 기술개발, 결과물의 활동 등을 종합적으로 분석

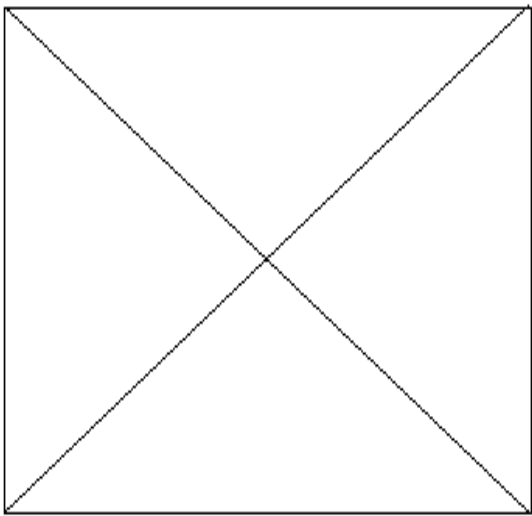
- 국가 연구데이터 공유·활용체제 구축사업은 연구데이터 공유·활용 관련 기술이 세계시장에서 주요 국가에서 기술 성장기에 속함으로
- 연구개발 투자를 위해 기술적 합리적 추세를 확보하고 대형 연구개발 사업으로 추진되기 위해 적합한 시점에 있는 사업임

□ 기술성장 단계 추정; 연구자-연구문헌 포트폴리오 분석

- 기술추세를 분석하기 위한 기술성장 단계 추정
 - 기술성장 단계는 기술혁신의 확산모형에 따르는데, 과학기술과 관련된 새로운 개념 혹은 주제가 도입되는 시기(태동기)에 소수의 연구개발 결과물이 발표됨
 - 발표된 연구개발 결과물을 다수가 받아들이기 시작하며 기술은 빠른 성장(성장기)에 들어서게 되고 이에 따라 해당 과학기술과 관련된 연구개발이 활성화됨
 - 이후, 과학기술이 더 많은 연구자에게 확산되어 연구개발은 꾸준히 증가하지만 속도가 둔화되기 시작하고(성숙기), 쇠퇴기로 넘어가며 일부 기술 개념이 활용되기 시작하지

만 일반적인 다수에게 진부화 되는 시기까지 이르게 됨

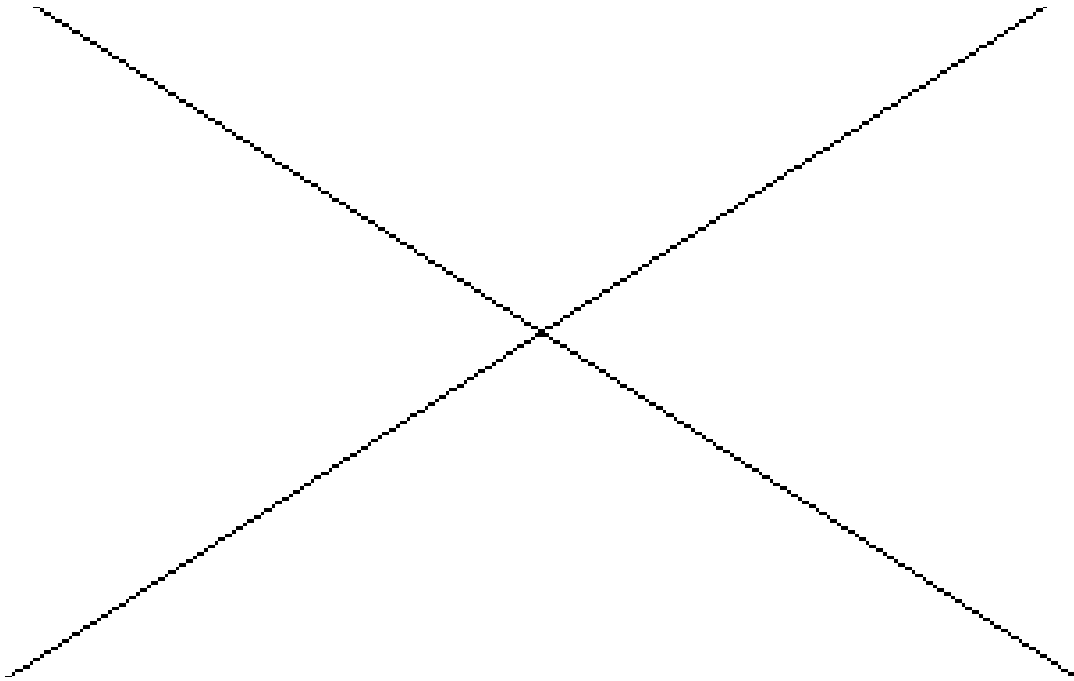
<표 9-22> 연구자-연구문헌 포트폴리오

연구자-연구문헌 포트폴리오	단계	특징
	I. 태동기	- 신기술의 출현 - 연구문헌 수와 연구자 수가 천천히 증가
	II. 성장기	- 연구개발 활발, 경쟁의 격화 - 연구문헌 수와 연구자 수가 빠르게 증가
	III. 성숙기	- 지속적인 연구개발, 일부 도태 - 연구문헌 수 정체 - 연구자 수 정체 혹은 감소
	IV. 쇠퇴기	- 대체기술의 출현, 기술발전 불연속 - 연구문헌 수 정체 혹은 감소 - 연구자 수 정체 혹은 감소
	V. 회복기	- 대체기술 쇠퇴 혹은 유용성 재발견 - 연구문헌 수와 연구자 수가 증가추세로 전환

자료: 국가연구개발사업 예비타당성조사 수행 세부지침, 2018.4

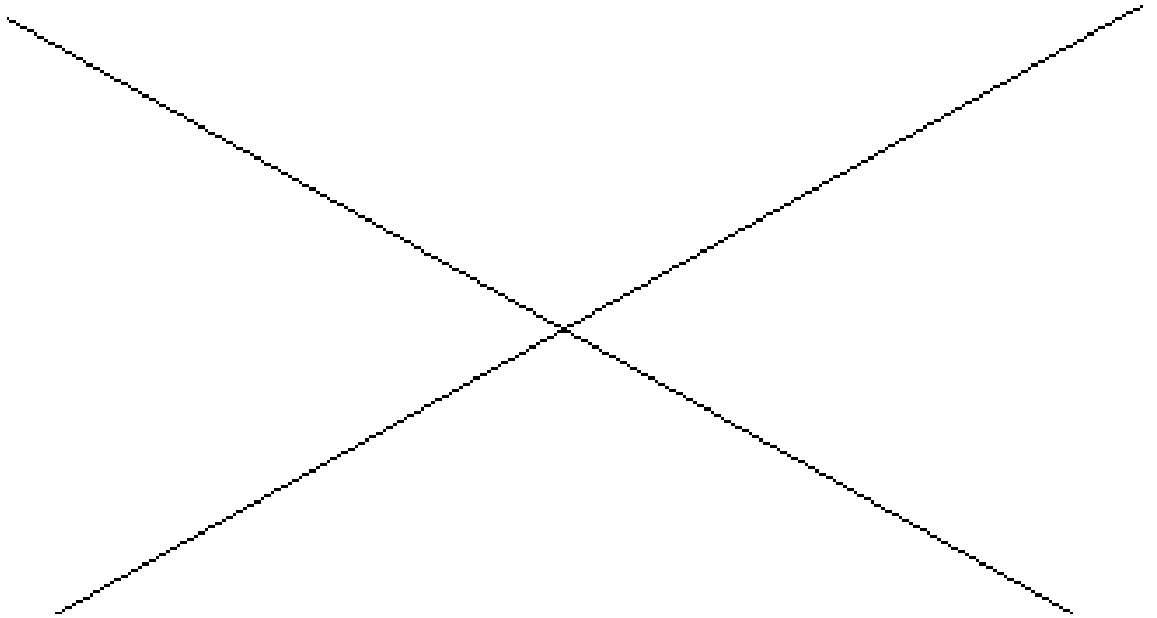
□ 전체 기술 분야의 사업추진 시기 적절성

- 연구데이터 공유·활용 분야의 특허기술은 원천특허에서부터 응용특허까지 세계 주요 시장에서 특허권 선점을 위한 활동이 매우 활발함으로 대규모 투자에 적합한 기술임
 - 1985년부터 미공개문헌이 존재하는 최근(2017~2018년) 출원된 특허를 제외하면, 2013년부터 2016년까지 출원된 특허가 전체의 56.1%를 차지함
 - 1980년대에는 연구데이터 공유·활용과 관련한 특허출원이 미미하다가, 1990년도에 이르러 조금씩 증가하는 추세를 보였으며, 2000년도에 들어서 눈에 띄는 활동을 보임
 - 2014년, 2015년도에는 200건이 넘는 특허가 출원되었으며, 2016년도에 출원된 특허 수는 500건에 다다를 정도로 특허권 선점을 위한 활동이 급격히 증가하는 추세임



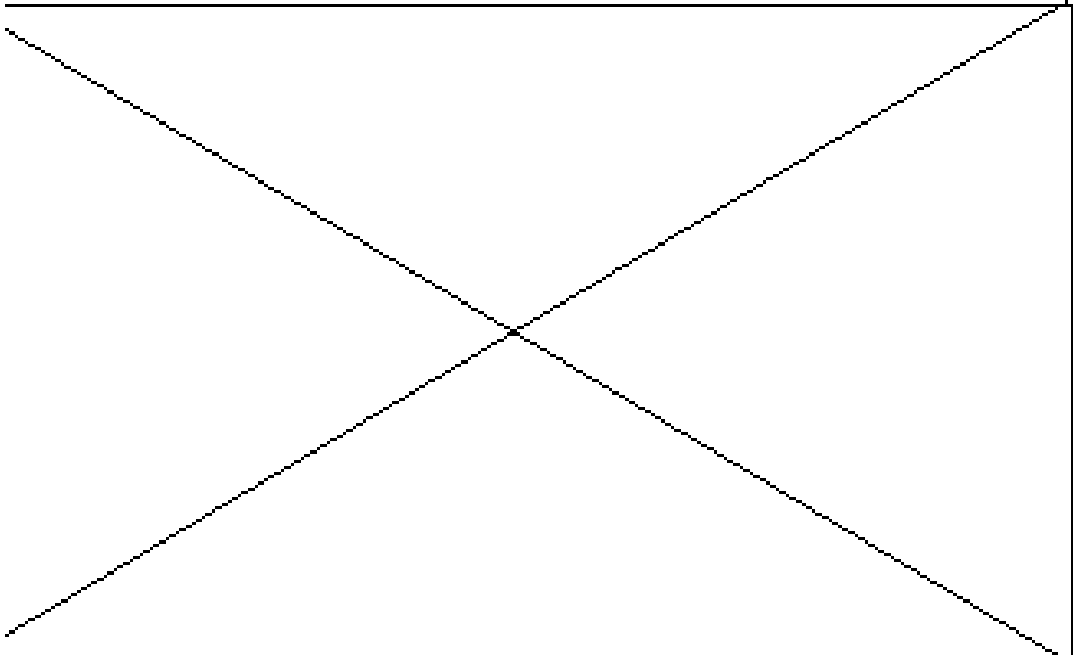
<그림 9-7> 국가 연구데이터 공유·활용 체계 구축 관련 특허출원 동향

- 세계시장의 주요 국가가 연구데이터 공유·활용 관련 분야 특허에 있어 기술의 발전이 기대되는 성장기 기술성장단계에 해당하며, 적절한 대규모 투자를 통한 개발이 적합한 기술임
 - 동 사업에서 개발하고 있는 국가 연구데이터 공유·활용체계 구축 기술은 특허 출원 건수 및 출원인 수가 모두 가파르게 증가는 성장기 단계인 것으로 파악됨
 - 일본을 제외한 대부분의 국가에서는 특허와 특허 출원 인이 빠르게 증가하는 성장기 단계인 것으로 보이며,
 - 이 단계는 해당 기술 분야에 관해서 연구개발이 증가하고 경쟁은 격화하는 등 대규모 투자가 이루어지기에 적절한 시기임



<그림 9-8> 국가 연구데이터 공유·활용 체제 구축 관련 특허 기술성장단계

- 2010년 이후의 세계시장 전체 연평균은 출원성장을 32%이며, 한국은 17%에 머문 반면 중국은 평균 이상인 82% 달성
 - 특허증가율과 점유율을 기준으로 중국에서 출원되는 특허는 뒤늦은 참여에도 불구하고, 기술개발의 속도가 타 국가에 비해 활발하여, 빠르게 성장할 것으로 판단됨
 - 미국 및 독일에서 활동하는 특허 출원의 경우, 자국 내 출원 뿐만 아니라, 세계 주요 시장에서 선점하기 위한 활동으로 활발히 진행되고 있음



<그림 9-9> 구간별 내·외국인 특허출원 및 외국인 국적 현황

- 연구데이터 공유·활용을 위한 구축 기술은 세계시장 주요 국가에서 대부분 성장기에 해당하므로 대규모의 투자가 이루어지기에 적절하며,
- 분석 대상 기술은 국가 연구데이터 공유·활용을 활성화하는 동 사업의 목표 달성에 이바지함으로 투자 규모 및 시기에 있어 적절한 것으로 판단됨
 - － 연구데이터 공유·활용과 관련된 특허 출원은 출원되는 특허의 수가 급격히 증가하는 등 최근들어 더욱 활발히 진행되고 있으며,
 - － 특허 기술성장단계 또한 연구개발이 증가하고 경쟁이 심화되는 단계에 해당하는 등 동 사업과 같은 대규모 사업 추진에 있어 정당성을 확보한 것으로 판단됨

나. 과학기술 수준 분석

(방향) 과학기술 수준 분석은 기술 뿐 아니라 지식까지 포함하여, 주요 연구개발 활동 주체와의 상대적 기술격차를 분석

- 국가 연구데이터 공유·활용체제 구축사업은 세계시장에서 상대적인 우위를 차지하는 경쟁력은 없을 뿐 아니라 최근 10년 저조한 활동을 보임으로 세계시장에서 주요 국가의 기술 경쟁력과의 격차를 줄이기 위한 투자가 시급함
- 국가 연구데이터 공유·활용체제 구축 기술의 6대 핵심기술 개발은 특허 활동도, 기술 경쟁력과 질적 수준에 있어 서로 다른 우위수준을 확보하기 때문에 향후 기술개발 보강을 통해 목표달성을 이루기 위한 전략적·정책적 전략을 취할 필요가 있는 사업임

□ 과학문헌계량분석

- 기술 수준을 분석하기 위해 과학문헌계량분석을 진행
 - － 과학문헌계량분석은 과학기술의 발전과 경제발전과의 상호의존관계를 대리지표 즉 논문이나 특허를 활용하여 측정하고자 하는 분석 방법임
 - － 동 사업의 기술 수준 분석을 위해서는 기술 경쟁력을 판단하기 위한 현재영향력 지수(CII)지수, 시장확보(PFS)지수와 기술의 질적 수준을 판단하기 위한 특허 인용도(CPP)지수 그리고 특허영향(PII)지수를 활용함

<표 9-23> 과학기술 수준 분석을 위한 지수

기호	수식	특징
현재영향력 지수 (CII: Current Impact Index)	$CII = \frac{\sum_{t=y-4}^y \left(\frac{c_t/n_t}{c_t/N_t} \times n_t \right)}{\sum_{t=y-4}^y n_t}$ t: 분석대상연도, y: 현재 연도 c_t: 특정주체의 t연도 연구출판물이 현재 연도에 인용된 전체 횟수 C_t: t연도의 전체 연구출판물이 현재 연도에 인용된 전체 횟수 n_t: 특정주체의 t연도 연구출판물 수 N_t: t연도의 전체 연구출판물 건수	- 출판물당 피인용 수가 가지는 최근 기술동향 반영의 어려움을 극복하기 위하여 조사하는 지표 - 보편적으로 현재 시점 기준, 이전 5년 동안의 값을 활용하여 특정주체의 연구출판물이 얼마나 인용되었는가를 상대적으로 계산
시장 확보 지수 (PFS; Patent Family Size)	$PFS^a = \frac{FS^a}{FS^{av}}$ FS^a: 출원인 a의 특허 패밀리 규모 FS^{av}: 해당분야 전체 특허 패밀리 규모	해당 특허의 보호범위가 지역적으로 어느 정도까지 미치고 있는가를 분석하는 지표
출판물 당 피인용 수 (CPP; Cites per Publications)	$CPP = \frac{\sum_{i=1}^t C_i}{N_t}$ N_t: 전체 출판물 수 C_i: 연도의 피인용 수 t: 전체 분석기간년)	- 연구결과에 대한 출판물 인용정보를 분석하는 지표 - 분석기간 동안 출판물당 피인용 수의 평균을 계산
출판영향력 지수 (PII; Publication Impact Index)	$PII_a = \frac{CPP_a}{CPP_t} = \frac{C_a/N_a}{C_t/N_t}$ C_a: a의 연구출판물 피인용 수 N_a: a의 연구출판물 수 C_t: 전체 연구출판물의 피인용 수 N_t: 전체 연구출판물 수	- 특정 주체의 기술혁신 성과의 질적 수준을 평가하는 지표 - 분석대상 기술 분야의 연구문헌의 평균 피인용 건수에 대한 특정 주체의 연구문헌의 피인용 건수의 상대적인 비율을 계산

□ 전체 기술 분야의 주요 국가별 기술수준

- 한국의 전체 기술에 관한 기술수준은 특허활동성, 기술 경쟁력 및 질적 수준에 이르기까지 전체적으로 타 주요 국가에 비해 뒤처지는 추세임
- 특허 점유율의 경우 18%를 차지하는 등 적지 않은 비중을 차지하나, 타국에 비해 시장 확보력이 뒤처지며, 인용 및 피인용의 상대 값에 있어 경쟁력 미확보
- 따라서 세계 시장을 겨냥한 기술개발과 더불어 선진국과의 경쟁력 차이를 줄이기 위해

기술 개발을 위한 동 사업에 투자 시급함

- 미국은 특허 활동도에 있어서는 중국 및 네덜란드, 아일랜드 등에 뒤처져 있지만, 기술 경쟁력 및 질적 수준에 있어서는 고르게 우수한 수준을 보이고 있음
- 중국과 네덜란드는 특허 활동도가 세계 주요 국가에 비해 월등하나, 중국의 경우 급증한 활동에 비해 기술 경쟁력 및 질적 수준이 확보되지 못한 반면, 네덜란드의 경우 시장확보력을 시작으로 지속적인 관찰을 통한 동향 파악이 필요함
- 일본의 경우 최근 10년 동안 특허의 증가율이 마이너스(-)를 유지하고 있으나, 상대적으로 우수한 기술 경쟁력과 질적 수준을 확보함

<표 9-24> 주요 국가별 특허지수 현황

국적(코드)	특허 활동도		기술 경쟁력		질적 수준	
	특허증가율 (최근10년)	특허점유율 (PS)	현재영향지수 (CII)	시장확보력 (PFS)	인용도 (CPP)	특허영향지수 (PII)
중국(CN)	2,270%	37%	0.13	0.62	1.39	0.13
미국(US)	134%	29%	1.78	1.52	20.18	1.85
한국(KR)	93%	18%	0.22	0.75	2.27	0.21
일본(JP)	-14%	8%	1.18	0.98	13.75	1.26
독일(DE)	264%	2%	0.81	0.85	9.46	0.87
아일랜드(IE)	300%	0.5%	0.33	1.54	3.50	0.32
인도(IN)	0%	0.5%	0.30	1.46	3.50	0.32
대만(TW)	67%	0.4%	1.85	1.90	18.43	1.69
캐나다(CA)	-33%	0.4%	4.84	2.52	56.33	5.16
스위스(CH)	100%	0.4%	1.78	2.88	21.00	1.92
네덜란드(NL)	500%	0.3%	0.37	1.63	4.25	0.39
러시아(RU)	0%	0.3%	0.09	1.36	1.00	0.09
프랑스(FR)	200%	0.3%	2.24	2.17	26.00	2.38
호주(AU)	-50%	0.3%	-	2.71	-	-
핀란드(FI)	200%	0.3%	9.67	2.17	112.50	10.30
영국(GB)	-33%	0.2%	1.50	1.72	17.50	1.60

□ 국가 연구데이터 공유·활용체제 구축 핵심6대 기술의 상대적 기술수준

- 국가 연구데이터 공유·활용을 위한 체제 구축 6대 핵심 기술 중 연구데이터 융합·분석 환경구축 기술이 특허 활동도가 가장 활발하며,
- 기술 경쟁력 부문에서는 연구데이터 활용·플랫폼 기술이, 질적 수준에 있어서는 지능형 연구데이터 고속 검색 기술이 전반적으로 높은 수준 유지
- 연구데이터 수집·연계 및 공유 기술과 연구데이터 품질·표준화 기술은 인용도가 평균이

- 상으로 세계시장 동향에 따라 기술 선점이 가능한 기술개발 확보를 위해 보강이 필요함
- 연구데이터 저장·보안 기술은 다른 핵심기술에 비해 평균 인용도를 확보했으며, 세계시장 동향을 반영하여 향후 선점이 가능한 기술 개발이 필요하며, 높은 특허증가율을 보여 평균 정도의 기술 경쟁력 및 질적 수준을 확보함
 - 지능형 연구데이터 고속 검색 기술은 빅데이터의 효율적 활용에 대한 시장요구에 따라 시장확보력을 포함한 기술 경쟁력과 질적 수준이 모두 높은 수준으로 나타남
 - 연구데이터 융합·분석 환경구축 기술은 손쉬운 연구데이터 활용을 위해 다른 핵심 기술과 독립하여 개발하는 기술로 기술 경쟁력과 질적 수준 모두 평균 이상으로 확보함
 - 연구데이터 활용·플랫폼 기술은 최근 10년의 특허증가율과 현재영향지수가 높은 수준을 확보하였으며, 이를 통해 물리적 자원과 가상 자원을 효율적이고 안정적으로 관리·제공할 것으로 보임

<표 9-25> 6대 핵심기술의 별 특허지수 현황

기술 분류	특허 활동도		기술 경쟁력		질적 수준	
	특허증가율 (최근10년)	특허점유율 (PS)	현재영향지수 (CII)	시장확보력 (PFS)	인용도 (CPP)	특허영향지수 (PII)
연구데이터 수집·연계 및 공유 기술	104%	9%	1.64	0.90	12.78	1.17
연구데이터 저장·보안 기술	271%	15%	1.41	0.97	10.96	1.00
연구데이터 품질·표준화 기술	246%	5%	1.43	0.75	11.10	1.02
지능형 연구데이터 고속 검색 기술	144%	18%	2.11	1.10	16.38	1.50
연구데이터 융합·분석 환경구축 기술	252%	52%	1.23	1.02	9.53	0.87
연구데이터 활용·플랫폼 기술	298%	18%	7.77	0.94	7.77	0.71

- 연구데이터 공유·활용체제 구축을 위한 핵심6대 기술은 국가 연구데이터 공유·활용을 활성화하는 것을 목표로 하는 동 사업의 사업목표 달성에 이바지 하며,
- 대부분의 기술이 특정 분석 분야에서 우수한 수준을 확보하여 기술 추세에 있어 뒤떨어져있다고 판단하기 어려움
- 6대 핵심 기술에 투자하여 세계시장을 겨냥한 기술개발과 더불어 핵심·원천기술을 확보하고 선진국과의 경쟁력 차이를 줄이기 위한 전략적·정책적 전략적 투자 사업이 시급함

9.4. 경제적 타당성 분석

9.4.1. 비용 추정

- 국가연구데이터 공유·활용체제 구축사업은 3개의 세부 사업으로 총사업비는 5,752억 원이며 동 사업의 경제성 분석을 위한 총비용은 부가가치세를 제외³⁷⁾한 5,404억 원으로 산정됨
- 현재가치는 3,847억 원으로 추정되며, 「국가연구개발사업 예비타당성조사 수행 제부 지침」에 따라, 사업 의뢰 전연도인 2018년을 기준연도로 하여 산정

<표 9-26> 국가 연구데이터 공유·활용 체제 구축 사업 경제성분석용 비용

(단위: 억 원)

구분	세부사업별 비용			합계	
	국가 연구데이터 플랫폼 구축 및 운영	대·소 분야 전문센터 구축 및 운영지원	데이터 및 컴퓨팅 활용 기반의 R&D 인재 육성	경상가치	현재가치
2021	142	90	66	299	262
2022	218	104	62	383	321
2023	205	274	63	542	435
2024	233	278	64	575	441
2025	178	412	64	654	480
2026	144	409	52	605	425
2027	160	401	53	614	413
2028	164	382	53	598	385
2029	163	354	54	571	352
2030	163	346	55	563	332
합계	1,769	3,049	586	5,404	3,847

37) 부가가치세는 인건비를 제외한 직접비용에서만 제외함

9.4.2. 편익 추정

가. 연구개발사업의 편익

- 연구개발사업의 경우 연구개발 활동을 통해 기대되는 파급효과의 종류에 따라 지식파급, 시장파급, 네트워크파급으로 구분되며 유형에 따라 각각 다른 경제성 분석방법 적용
- － 타당성조사의 편익 추정에서 중요한 기준은 객관적인 계량화의 용이성이므로 시장파급은 비용편익 분석, 지식파급은 비용효과 분석으로 경제성 분석에 반영하는 것을 원칙으로 함
- － 사업기간 중의 비용과 편익이 아닌 사업 수행으로 소요되는 총비용과 발생하는 실질적인 순편익만을 반영함

<표 9-15> 연구개발 활동 파급 유형별 경제성 분석 방법

구분	정의 및 발생요인	예비타당성조사 반영 여부
시장파급 (Market spillovers)	- 시장 기능에 의해 여타 주체에게 제품이나 공정상의 편익을 전달하는 경우 - 추가기능의 구비, 가격의 인하, 저렴한 제품 및 서비스 제공 등의 형태로 발생	- 시장정보를 통해 계측 가능 - 화폐단위로 표현 - 비용편익 분석을 통해 결과 도출
지식파급 (Knowledge spillovers)	- 지식의 창출자와 사용자가 다른 경우 - 역설계, 발간, 특허공개, 연구자 이동 등의 방식으로 발생	- 계측이 일부 가능 - 화폐단위로 표현하기 어려움
네트워크파급 (Network spillovers)	- 신기술이 연관된 기술들의 집합적 기능을 제고함으로써 가치를 갖게 되는 경우 - 단일 기술이 다수 기술들의 조합을 통해 향상된 기능이 제공될 경우에 발생	계측이 불가능

자료: 국가연구개발사업 예비타당성조사 수행 세부지침

- 타당성조사 비용-편익 분석 시 가치창출 및 증대, 기술거래로 인한 정(+)의 사업효과와 생산·피해비용저감으로 인한 부(-)의 사업효과를 편익 항목으로 활용
- － 연구개발 활동을 통해 발생할 수 있는 경제적 편익을 직접적 경제적 편익, 간접적 경제적 편익 그리고 산업연관 유발효과로 구분하며 직접적·경제적 편익만을 사업의 편익항목으로 반영함
- － 최종 산출물을 직접적인 생산요소로서 활용하는 제품이나 서비스가 최종소비재로서 시장에서 거래되는 경우, 산출물의 시장가치 혹은 기여분을 비용편익 분석에 반영할 수 있으나, 네트워크파급으로 분류되는 유발효과는 비용편익 분석에 반영하지 않음

<표 9-27> 연구개발부문 예비타당성조사의 편익 반영/미반영 구분

구분	편익 항목 반영	편익 항목 미반영	
정(+)의 가치 증가	■ 부가가치 창출·증대 - 신기술 적용을 통한 생산량 증가 - 신기술 개발로 인한 가치창출 ■ 기술거래 - 기술이전에 의한 로열티 수입	■ 과학기술 지식* (논문, 특허 등) ■ 과학기술자 교육훈련 ■ 지역개발효과 ■ 지역산업구조 개편 ■ 생산 유발효과	■ 부가가치유발효과 ■ 고용유발효과 ■ 수입유발효과 ■ 수출유발효과 ■ 소득분배효과 ■ 취업유발효과
부(-)의 가치 감소	■ 생산비용 저감 - 생산투입 자원·시간의 저감 - 연구기간, 출장횟수 등의 연구수행 비용 저감 ■ 물류비용 저감 ■ 피해비용 저감 - 재난·재해·사고 피해 감소 ■ 질병비용저감* ■ 환경비용저감*		

자료: 국가연구개발사업 예비타당성조사 수행 세부지침

나. 편익산정의 기본가정

□ 기초자료의 선정

- 국가연구개발사업의 편익 분석에는 국가통계자료를 기본 자료로 활용
 - 출처가 다양한 자료를 활용할 때 발생할 수 있는 일관성 저해 및 자료의 불확실성을 최소화함
 - 국가연구개발과 관련된 논문 및 특허에 대한 사항은 국가과학기술지식정보서비스 자료를 원칙으로 함

□ 분석 범위 설정

- 시간적 범위
 - 「국가연구개발사업 예비타당성조사 수행 총괄지침」에 의거 할인을 적용하는 분석의 기준일은 사업의 분석이 착수된 전년도 말일로 정함
 - 동 사업은 사업이 추진되는 2020년을 시작으로 최종년도 2029년까지 총 10년을 사업 수행기간으로 정함
- 공간적 범위
 - 사업 추진을 통해 영향을 받는 범위를 의미함으로 연구개발투자의 효과 비배제성을 반영하여 국내 전체로 정함

□ 분석 기간

○ 편익 회임기간

- 편익이 발생하기 시작하는 시점으로 투자가 이루어진 후, 경제적인 편익 또는 효과가 발생하기 전까지의 시간적 지연이 발생하는 기간을 의미하며,
- 연구개발사업 종료시점의 기술성숙도(TRL; Technology Readiness Level)를 고려하여 편익 회임기간을 다르게 산정함
- 동 사업은 TRL 6단계 기술개발을 목표로 하고 있으며, 「국가연구개발사업 예비타당성 조사 수행 세부지침」에서는 개발·응용연구의 회임기간을 3년으로 제시하고 있으나 SW 개발연구의 특성을 고려하여 2년으로 단축하여 적용함

○ 편익발생기간

- 편익이 발생할 것으로 기대할 수 있는 기간으로 해당 기술의 특성을 반영하여 실제 사업의 유효한 편익과 기술의 수명을 고려하여 설정함
- 그러나, 적절한 방법론이 없을 경우 기술수명기간(TCT; Technology Cycle Time)을 도입하여 편익기간으로 설정함

○ 기술수명기간(TCT)

- TCT는 서지정보 중 특허관련 정보를 정량적으로 활용하여 산출하는 지표로써, 인용된 특허의 발행연도와 차이 값의 중앙값(median)(인용-피인용 특허 시차 중앙값)을 말하며, 특허인용수명 또는 기술순환주기라고도 명명함
- 산업통상자원부는 「기술가치평가 실무가이드」(2014.12)에서 생활필수품, 처리조작; 운수 등을 국제특허분류(IPC; International Patent Classification) 8개 섹션으로 분류하고 기술가치평가를 위한 4단위로 주요 통계값을 제시함

<표 9-28> IPC 분류

A섹션	B섹션	C섹션	D섹션	E섹션	F섹션	G섹션	H섹션
생활필수품	처리조작; 운수	화학; 야금	섬유; 지류	건조물	기계공학; 조명; 가열; 무기; 폭발	물리학	전기

산업통상자원부, 2014.12

- 동 사업의 빅데이터 기술은 IPC 분류에서 G섹션(G06F, 전기에 의한 디지털 데이터처리)에 해당하며, 이하 유관 기술의 특허인용수명을 참고하여 기술수명기간을 8년(G06F : 7.31년)으로 산정함

<표 9-29> 빅데이터 관련 기술수명기간

IPC 분류	IPC 코드	내용	평균(년)
G섹션 (물리학)	G06F	전기에 의한 디지털 데이터처리	7.31

산업통상자원부, 2014.12

- 따라서 국가 연구데이터 공유·활용체제 구축사업의 사업기간은 2021년부터 2030년까지 (10년)이며, 편익은 사업종료 후 편익회임기간 2년이 지난 2033년부터 8년 동안 발생

<표 9-30> 사업 추진기간 및 편익발생기간

구분	1단계			2단계				3단계			편익회 임기간		편익발생기간							
	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	'31	'32	'33	'34	'35	'36	'37	'38	'39	'40
국가 연구데이터 플랫폼 구축 및 운영																				
대·소 분야 전문센터 구축·운영 지원																				

 : 사업기간 : 편익회임기간 : 편익발생기간

□ 할인율

- 미래의 비용과 편익을 현재가치로 바꾸기 위한 할인율(discount rate) 4.5% 적용
 - － 연구개발사업의 비용과 편익은 미래의 시점에서 발생하는 반면, 경제성 분석을 시행하는 시점은 현재 시점이므로 현재가치로 환산하여 비교하여 분석해야함
 - － 타당성조사의 할인율은 사회적 할인율의 적용³⁸⁾을 원칙으로 하며, 「예비타당성조사 수행 총괄지침」 제50조에 따라, 그 비율은 4.5%를 적용함

“제50조(사회적 할인율) ①예비타당성조사에서 사회적 할인율은 4.5%를 적용한다. 다만 분석기간이 30년 이상인 철도와 수자원 사업은 운영 30년 동안은 4.5%를 적용하고 이후는 3.5%의 할인율을 적용한다.

「예비타당성조사 수행 총괄지침」 제50조

다. 편익 추정

① 편익추정 대상

38) 「국가연구개발사업 예비타당성조사 수행 세부지침」

- 동 사업의 세부사업은 1. 국가 연구데이터 플랫폼 구축 2. 대·소 분야 전문센터 플랫폼 구축·운영 지원 3. 데이터 및 컴퓨팅 활용기반의 연구개발 인재육성 등 3개로 구성
- 세부사업 중 국가 연구데이터 플랫폼 구축과 대·소 분야 전문센터 플랫폼 구축·운영 사업은 시장수요접근법(가치창출편익)인 부가가치 창출 편익을 적용하고, 데이터 및 컴퓨팅 활용기반의 연구개발 인재육성 사업은 수요자 교육·효과 편익³⁹⁾(교육으로 인한 업무 처리능력 향상, 임금상승 효과)을 적용함
- 국가 연구데이터 플랫폼 구축 및 대·소 분야 전문센터 플랫폼 구축·운영은 빅데이터라는 대상 시장이 존재하고 사업 산출물의 구체적인 활용처가 명확함으로 편익 추정 대상으로 선정함
 - － 국가 연구데이터 플랫폼 구축 및 대·소 분야 전문센터 플랫폼 구축·운영은 연구데이터의 수집·관리 지원을 시작으로 공유·활용 활성화를 목표로 하는 사업임
 - － 동 세부과제의 핵심기술은 방대한 양의 연구데이터를 수집·관리 가능한 플랫폼 및 인프라 구축으로 대상 시장을 데이터를 수집·관리하며 활용하는 빅데이터 시장으로 선정하여 편익을 추정함
- 연구개발 인재육성 관련 사업은 빅데이터 분석과 컴퓨터 활용 분야의 인재를 양성하는 사업으로 교육훈련에 의한 업무처리능력 향상에 따른 효과를 편익추정 대상으로 선정함
 - － 인력양성 사업은 데이터 및 컴퓨팅 활용에 관한 교육·실습 등의 환경을 제공하여 연구데이터 공유·활용을 위한 미래의 연구자 생산효과가 기대되는 사업임
 - － 국가 연구데이터 공유·활용 활성화를 위한 기반 마련의 기능으로 이에 따른 편익의 화폐가치화 및 정량적 계량화가 어려우나 교육훈련을 포함하는 예비타당성조사 사례⁴⁰⁾를 적용하여 교육훈련에 의한 편익을 추정함

<표 9-31> 세부 사업별 편익검토

사업구분		편익검토
기술개발부문	국가 연구데이터 플랫폼 구축	부가가치 창출 편익
	대·소 분야 전문센터 플랫폼 구축·운영	
인프라 구축부문	데이터 및 컴퓨팅 활용기반의 연구개발 인재육성	교육효과 편익

② 편익 추정방법

39) 국가연구개발사업 예비타당성조사 수행 세부지침에서는 과학기술자 교육훈련에 대한 편익을 반영하지 않는 것으로 되어 있으나, 과학기술자가 아닌 일반 수요자에 대한 교육은 시장수요측면에서 편익 항목으로 반영할 필요가 있을 것으로 판단

40) 한국개발연구원, 식생활교육문화연구센터 건립사업, 2017

□ 부가가치 창출 편익

- ‘사업 분석-편익항목 도출-자료의 가용성 확인-추정방법 선택’의 과정에 따라 연구데이터 공유·활용체제 구축사업의 산출물이 시장에서 거래됨으로써 발생하는 편익을 가치 창출 편익으로 제시
 - 기술개발 부문을 구성하고 있는 국가 연구데이터 플랫폼 운영, 대·소 분야 전문센터 인프라 구축·운영은 각 부문에 대한 시장 구분 및 각 부문에 대한 기술적 기여도 구분이 어려움
 - 국가 연구데이터 플랫폼 운영과 대·소 분야 전문센터 인프라 구축·운영사업을 전체 기술개발과제로 간주하여 동 사업을 통해 빅데이터 시장에서 발생하는 편익을 산정함
- 「국가연구개발사업 예비타당성조사 수행 세부지침」에 의거 새롭게 생산된 산출물의 상용화로 인해 부가가치의 증가분을 바탕으로 편익을 추정하는 시장수요접근법으로 추정
 - 시장수요접근법은 가치창출편익방법 중 연구개발사업의 최종 산출물이 시장가치 창출을 목적으로 하는 경우 주로 활용됨
 - 해당 연구개발사업의 미래 시장규모 증가분에 연구개발사업으로 창출된 미래의 직접적 편익을 한정하여 산정하는 접근법임
 - 시장수요접근법을 활용한 편익은 동 사업과 관련한 전체(국내·외) 미래 시장규모에 예상 시장점유율, R&D기여율, 사업화 성공률과 부가가치율을 곱한 값으로 산정함

<표 9-32> 시장수요접근법 편익 산정식

사업 편익	=	① 미래 시장규모	×	② 사업 기여율	×	③ R&D 기여율	×	④ R&D 사업화 성공률	×	⑤ 부가 가치율

자료: 국가연구개발사업 예비타당성조사 수행 세부지침(2018.4)

□ 교육효과 편익

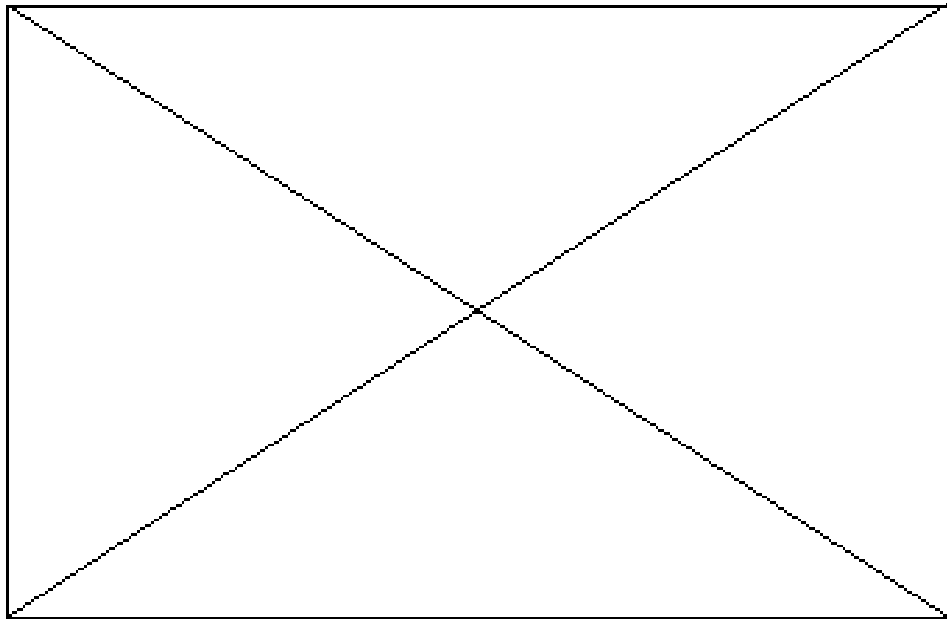
- 공공교육시설 등에서 수행하는 직무관련 교육의 성과는 특정 기간 동안 피교육생의 업무효율을 높이고, 그 높아진 효율의 가치는 교육을 받지 못한 동료 대비 임금상승률에 반영된다고 볼 수 있음
- 이는 직업관련 공공재의 특성이 임금수준에 반영된다는 특성임금법(hedonic wage method, Bockstael and McConnell, 2007⁴¹⁾)의 사례로 볼 수 있으며, 한국개발연구원

41) Bockstael and K .E. McConnell, Environmental and Resource Valuation with Reveal Preferences: A Theoretical Guide to Empirical Models, Berlin, Springer. 2007

의 2017년 예비타당성조사 결과보고서⁴²⁾를 참고하여 교육에 따른 임금 상승효과를 연구개발 인재육성(교육효과) 편익으로 제시함

<표 9-33> 교육효과 편익 산정식

사업 편익	=	① 교육수요	×	② 12(월)	×	③ 월 평균임금	×	④ 교육에 의한 임금상승률



<그림 9-10> 연구개발부문 예비타당성조사의 편익 추정방법 선택과정
* 자료: 한국과학기술기획평가원(2018.4)

42) 한국개발연구원, 식생활교육문화연구센터 건립사업, 2017

라. 편익산정

① 부가가치 창출 편익

□ 미래 시장규모

미래 시장규모란 연구개발의 직접적인 효과를 받는 국내 산업의 미래 총생산액 또는 매출액을 의미하며 산정식은 다음과 같음

$$\text{미래 시장규모} = \text{세계 빅데이터 시장} \times \text{시장점유율}^{43)}$$

국가 연구데이터 공유·활용체제 구축사업의 미래 시장규모는 세계 빅데이터 시장의 2039년도 미래 시장규모는 920억 달러이며, 시장점유율은 1%로 산정

- 미래 시장규모란 연구개발의 직접적인 효과를 받는 국내 산업의 미래 총생산액 또는 매출액을 의미하며, 일반적으로 미래시장에서 국내 수요 중 국산품이 차지하는 규모와 해외에 수출되는 규모를 포함
 - － 편익 산정과정에서 활용하는 미래 시장규모는 특정 산업 전체의 시장이 아니라, 조사 대상사업과 직접적으로 관련된 빅데이터 관련 시장규모를 의미
- 빅데이터 미래 시장규모 산정을 위해 사용한 자료는 IDC(International Data Corporation)의 예측자료를 사용함
- IDC는 IT, 전기통신 등에 관해 시황 및 시장 자문 서비스를 제공해주는 전문기관으로 다수의 공공기관에서 IDC의 자료를 정책수립 등에 활용하고 있어, 빅데이터의 미래 시장 규모 전망에 적합한 자료로 볼 수 있음
 - － IDC는 기술·산업 기회 등에 관해 110개국에 넘는 국가의 동향을 살피고, 글로벌 기업 및 중견 기업까지 아우르는 시장을 조사하며 광범위한 분석 범위를 가진 기관
 - － 과하기술통신부, 소프트웨어정책연구소, 한국데이터진흥원 등 다양한 공공기관이 빅데이터 및 소프트웨어 시장 전망을 위한 현황자료로 인용함
 - － 또한, 미래창조과학부 및 정보통신기술진흥센터는 2018년 ‘ICT R&D 중장기 기술로드맵 2022’ 제시를 위해 IDC의 보고서를 전반적으로 참고하는 등 빅데이터 미래 시장 규모를 전망하기에 적절한 공신력 있는 기관으로 적절함

43) 국내 빅데이터 시장(연구개발 빅데이터)의 점유율을 직접적으로 산정이 어려워, 빅데이터가 소프트웨어 분야의 한 분류임을 참고하여 소프트웨어 분야의 시장점유율을 적용함

- IDC의 ‘Worldwide Big Data and Analytics Software Forecast, 2018–2022’는 세부 분류를 business intelligence and analytics tools, Analytic data management and integration platforms, analytic and performance management applications 등으로 구분하고 있어, 국가 연구데이터 공유·활용을 위한 동 사업의 대상 시장과 유사한 것으로 볼 수 있음

<표 9-34> IDC 빅데이터 및 분석 소프트웨어 범위

Business intelligence and analytics tools (비즈니스 인텔리전스 및 분석 툴)	End-user query, reporting, and analysis
	Advanced and predictive analytics
	Cognitive/AI software platforms
	Content analytics tools
	Search systems
	Spatial and location analytics
Analytic data management and integration platforms (분석적 데이터 관리 및 통합 플랫폼)	Analytic data integration and integrity
	Continuous analytics
	Nonrelational analytic data stores
	Relational data warehouse management
Analytic and performance management applications (분석 및 성능 관리 애플리케이션)	Customer relationship analytic applications
	Enterprise performance management applications
	Supply chain and product analytic applications
	Services operations analytic applications
	Production planning applications
	Workforce analytic applications

자료: IDC's Worldwide Semiannual Big Data and Analytics Software Tracker, May 2018

- 또한 IDC의 Worldwide Big Data and Analytics Software Revenue by Deployment Type(배치유형)을 보면 On-premises/other software(사내용/기타 소프트웨어)와 Public cloud services(공공 클라우드 서비스)로 구분하고 있어 클라우드 서비스 시장도 포함하고 있는 것 나타남
- IDC는 발표한 보고서에서 세계 빅데이터 시장이 2013년 409억 달러에서 연평균 11.2%로 성장하며, 2022년에는 그 규모가 921억 달러에 이를 것으로 예측
- IDC는 ‘Worldwide Big Data and Analytics Software Revenue by Segment(2013–2022)’ 보고서에서 세계 소프트웨어 시장의 규모 및 추이를 제시함

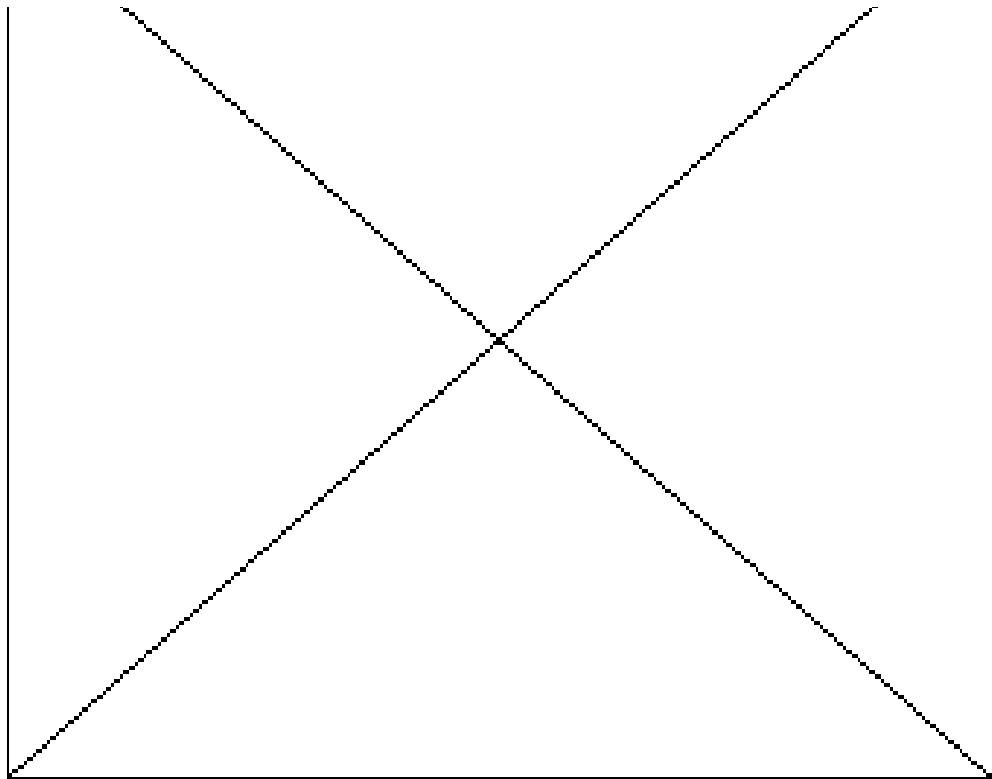
- IDC자료에서는 빅데이터 및 분석 소프트웨어 시장을 Business intelligence and analytics tools, analytic data management and intergration platforms, analytic and performance management applications 등의 분야로 구분하고 2022년 각각 370억 달러, 308억 달러, 244억 달러의 규모로 제시하고 있음

<표 9-35> 세계 빅데이터 및 분석 소프트웨어 시장 규모 및 추이(by segment)

(단위: 10억 원, %)

구분	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2017-2022 CAGR
Business intelligence and analytics tools	16.5	17.7	18.3	19.9	21.6	23.9	26.3	29.1	32.6	37	11.3
Analytic data management and integration platforms	12.2	13.1	13.7	15	17	19.3	21.7	24.5	27.5	30.8	12.6
Analytic and performance management applications	12.2	13	13.3	14.2	15.5	17	18.6	20.3	22.2	24.4	9.5
합계	40.9	43.8	45.2	49.1	54.1	60.2	66.5	73.9	82.3	92.1	11.2

자료: IDC's Worldwide Semiannual Big Data and Analytics Software Tracker, May 2018



<그림 9-11> 세계 빅데이터 및 분석 소프트웨어 시장 규모 및 성장추이(by deployment)

- 2021년 이후의 미래 시장규모는 연평균 성장률(CAGR)을 적용하여 추정하며, 2040년

도에는 6,847,602억 원까지 성장

- IDC에서 제시하지 않은 2021년 이후의 빅데이터 미래 시장규모는 IDC에서 제시한 연평균 성장률 11.2%를 적용하며, 1달러당 환율은 기준년도인 2018년 기준인 1,100원으로 가정함

<표 9-36> 세계 빅데이터 미래시장규모 및 성장률 추정

구분	세계시장 규모	
	(10억 달러)	(억 원)
2020	73.90	812,900
2021	82.30	905,300
2022	92.10	1,013,100
2023	102.42	1,126,567
2024	113.89	1,252,743
2025	126.64	1,393,050
2026	140.82	1,549,072
2027	156.60	1,722,568
2028	174.14	1,915,495
2029	193.64	2,130,031
2030	215.33	2,368,594
2031	239.44	2,633,876
2032	266.26	2,928,871
2033	296.08	3,256,904
2034	329.24	3,621,677
2035	366.12	4,027,305
2036	407.12	4,478,363
2037	452.72	4,979,940
2038	503.43	5,537,693
2039	559.81	6,157,915
2040	622.51	6,847,602

- 국내 빅데이터 시장의 점유율 산정을 위해 국내 소프트웨어 시장의 점유율을 활용하며, 2040년도의 한국 소프트웨어 시장 규모는 전체 소프트웨어 시장에서 약 1%를 차지할 것으로 예상

- 시장점유율을 파악하기 위해 국내 빅데이터 시장의 규모를 적용해야하나, 국내 빅데이터 시장은 아직 완전히 정착하지 않는 신생 산업이기 때문에, 이를 적용하여 시장점유율을 산정하는 것은 적절하지 않음
- 따라서 빅데이터를 상위분야로서 포함하는 소프트웨어산업의 시장점유율을 근거로 하여 국내 빅데이터 시장의 점유율을 산출함
- IDC에서 제시한 국내 SW 시장의 연평균 증가율은 2.5%로 2014년 국내 SW 시장의 규모는 115,190억 원이며 2021년에는 136,527억 원까지 성장함

<표 9-37> 세계 및 국내 SW시장 규모 및 추이

(단위: %)

구분		2014	2015	2016	2017(E)	2018(E)	2019(E)	2020(E)	2021(E)	CAGR (’14~’21)
세계 시장 규모 (억 달러)	패키지SW (증감률)	3,889 (6.1)	3,908 (0.5)	4,155 (6.3)	4,486 (0.8)	4,670 (4.1)	4,999 (7.0)	5,353 (7.1)	5,730 (7.0)	5.7
	IT서비스 (증감률)	6,707 (3.1)	6,382 (-4.8)	6,506 (1.9)	6,728 (3.4)	6,870 (2.1)	7,072 (2.9)	7,282 (3.0)	7,496 (2.9)	1.6
	합계 (증감률)	10,596 (4.3)	10,290 (-2.9)	10,661 (3.6)	11,215 (5.2)	11,540 (2.9)	12,071 (4.6)	12,635 (4.7)	13,226 (4.7)	3.2
국내 시장 규모 (억 원)	패키지SW (증감률)	36,843 (2.8)	39,045 (6.0)	40,379 (3.9)	41,940 (3.9)	43,410 (3.5)	44,773 (3.1)	46,064 (2.9)	47,302 (2.7)	3.6
	IT서비스 (증감률)	78,349 (1.5)	79,475 (1.4)	80,149 (0.8)	83,545 (4.2)	85,494 (2.3)	87,084 (1.9)	88,296 (1.4)	89,225 (1.1)	1.9
	합계 (증감률)	115,190 (1.9)	118,520 (2.9)	120,528 (1.7)	125,485 (4.1)	128,904 (2.7)	131,857 (2.3)	134,360 (1.9)	136,527 (1.6)	2.5

자료: IDC, 2017.10

□ 사업기여율

사업기여율이란 미래 시점의 연구개발 활동 중 연구개발 결과물이 대상사업의 시장규모에서 차지하는 비율을 의미하며, 산정식은 다음과 같음

$$\begin{aligned} \text{사업기여율} &= \frac{(\text{동 사업투자규모})}{(\text{동 사업분야국가 RD투자규모}) + (\text{동 사업분야민간 RD투자규모}) + (\text{동 사업투자규모})} \\ &= 2,592 / (543 + 1,628 + 2,592) = 54.42\% \end{aligned}$$

국가 연구데이터 공유·활용체제 구축사업의 미래 사업기여율은 54.42%로 산정

- 사업기여율이란 미래 시점의 연구개발 활동 중 연구개발 결과물이 대상사업의 시장규모

에서 차지하는 비율

- 연구개발사업의 결과물이 미래 연구개발업 시장에 기여할 때, 오직 동 사업의 연구개발 사업으로 인한 시장점유효과가 발생한다고 단언할 수 없음
- 따라서, 미래의 어느 시점에서의 전체 시장 규모 중에서 어느 부분까지를 동 사업의 기여분을 산정하는 과정이 필요함
- 「국가연구개발사업 예비타당성조사 수행 세부지침」에 의거 정부와 민간의 연구개발투자 비율을 활용
 - 사업기여율은 미래 편익발생 시점에서 진행되는 연구개발 활동 중 대상사업이 차지하는 비중만을 적용하는 개념임
 - 동 산업분야의 미래 연구개발투자 규모 예측자료, 동 산업에서 유사한 성격의 정부 및 민간 투자 규모 자료를 확보했을 경우, 해당 자료를 최대한 활용하여 산정하는 것이 바람직함
- 과학기술정보통신부에서 발표한 ‘국가R&D사업 조사분석 보고서’에서 발표한 빅데이터 및 클라우드 사업의 정부·민간 R&D 투자규모를 통해 사업기여율 산출
 - 동 사업과 유사한 사업의 연구개발로 빅데이터 및 클라우드 분야를 선정하여 해당 사업의 정부·민간의 투자금액을 활용하여 사업기여율을 계산함
 - 빅데이터 및 클라우드분야에 정부가 투자한 금액은 543억 원, 민간의 투자금액은 1,628억 원이며, 확보된 투자금액으로 동 사업의 사업기여율 54.42%를 산출함

$$= 2,592 / (543 + 1,628 + 2,592) = 54.42\%$$

<표 9-38> 정부 및 민간 R&D 투자규모 비교('16)

(단위: 억 원)

구분	반도체 및 디스플레이	방송·통신 및 네트워크	소프트웨어 및 콘텐츠	사물인터넷(IoT)	빅데이터 및 클라우드	정보보안
정부투자(A)	7,364	4,002	5,761	748	543	1,608
민간투자(B)	199,144	15,837	13,038	3,769	1,628	2,689
B/A	26.1	4.1	2.5	4.8	2.5	2.2

자료: 과학기술정보통신부 외, 2016

□ R&D기여율

R&D기여율이란 연구개발의 상업화를 통해 부가가치가 창출되었을 때, 해당 연구개발이 부가가치 창출에 기여하는 비율을 의미

국가 연구데이터 공유·활용체제 구축사업의 R&D기여율은 40.0%로 적용

- R&D기여율이란 연구개발의 상업화를 통해 부가가치가 창출되었을 때, 해당 연구개발이 부가가치 창출에 기여하는 비율로 총 요소생산성에 대한 연구개발투자의 탄력성 추정 값
- 한국과학기술기획평가원의 「국가연구개발사업 예비타당성조사 수행 세부지침」(2018.4)에서 제시한 방향을 고려하여 40.0%를 R&D 경제성장 기여율로 적용
 - － 한국과학기술기획평가원에 따르면 R&D기여율의 정확성을 확보하는 것은 어렵기 때문에 공신력 있는 수치를 활용하는 것이 적절함

“정확한 R&D기여율을 도출하기 위해서는 대상 기술을 상용화하기 위해 필요한 비R&D비용을 추정하는 것이 적절한 것...하지만 이러한 자료는 획득이 어려우며 근거가 불충분한 경우가 많이 때문에 거시적으로 추정된 R&D 기여율 수치를 적용할 수밖에 없는 경우가 대부분...조사 착수 시점을 기준으로 가장 최근에 발표된 공신력 있는 수치를 적용...

「국가연구개발사업 예비타당성조사 수행 세부지침」, 2018.04

- － ‘국가연구개발사업 예비타당성조사 수행 세부지침(2018.04)’은 「제3차 과학기술기본계획」에서 최근 데이터를 적용하여 새롭게 구한 수치를 활용할 것을 권고함
- － 「제3차 과학기술기본계획」에서 제시한 R&D 경제성장 기여율은 '81~'10년 에는 35.4%, '13~'17년 에는 40.0%이므로 동 사업이 진행되는 시점을 기준으로 가장 최근에 발표된 「제4차 과학기술기본계획」을 참고할 필요가 있으나,
- － 「제4차 과학기술기본계획」에서는 새로운 R&D경제성장 기여율을 제시하지는 않고, 「제3차 과학기술기본계획」을 보완하는 방향으로 기본계획 전략 및 과제를 도출하였음
- － 따라서 기존의 「제3차 과학기술기본계획」에서 제시한 가장 최근의 R&D경제성장 기여율 40.0%를 적용함

<표 9-39> 박근혜 정부 과학기술기본계획 성과목표 (안)

지표명		2012	2017	참고
투자확대	정부R&D예산	16.0조 원	'13년 대비 8.1조 원 추가 투입	5년 간 총 92.4조 원 투자
	기초연구 비중 (정부 R&D예산 대비)	35.2%	40.0%	
	중소·중견기업 투자비중 (정부 R&D예산 대비)	12.0%('10)	18.0%	
국가전략 기술개발	국가전략기술경쟁력 제고 (세계 최고 수준 대비)	77.8%	85.0%	기술수준평가 결과
중장기 창의역량 강화	과학기술혁신역량 순위	9위	7위	COSTⅡ 기준
	상위 1%논문 순위	15위('06~'10)	10위('13~'17)	
	인구 중 이공계 박사비율	0.40%	0.60%	0.64%(OECD평균)
	연구원 천 명 당 국제공동특허 수	0.39건(2011)	0.50건	0.51건(OECD평균)
	중소기업 기술경쟁력	74.8%(2011)	85.0%	세계 최고 수준 대비
일자리 창출	창업활동지수(TEA)	7.8%	10%	1위 슬로바키아(14.2%) 2위 미국(12.3%)
	과학기술인력일자리 창출	605만 명	669만 명	신규64만 개
경제부흥	R&D 경제성장 기여율	35.4%('81~'10)	40.0%('13~'17)	
	국민 1인당 산업부가가치	1.9만 달러	2.5만 달러	1위 룩셈부르크(6.1만 달러) 2위 노르웨이(3.7만 달러)
	기술 수출액	4,032(백만 달러)	8,000(백만 달러)	14,812(백만 달러)(OECD평균)
삶의 질 기여	삶의 질 투자 확대 (정부 R&D예산 대비)	15.0%	20.0%	보건 및 환경 분야 정부 R&D 투자 비중 확대

자료: 제3차 과학기술기본계획(2013~2017), 2013.7

<표 9-40> 4차 기본계획 전략 및 과제

3차 기본계획			4차 기본계획 보완방향
전략	성과 및 한계		
연구개발 투자확대	• (성과) R&D투자 확대 • (한계) 투자시스템 선진화 필요	➡	• ‘정부R&D 투자시스템 혁신’과제 신설(과제 10-2) - R&D 예비타당성조사 제도 개선, 패키지형 R&D 투자플랫폼 도입 등
국가 전략기술 개발	• (성과) 120개 국가전략기술 확보노력 • (한계) 30개 중점기술의 실효성 저조	➡	• 국가전략기술을 ‘기본계획 실현과 관계된 중점과학기술’로 변경하고 30개의 중점기술 미선정 • 기술개발방향과 함께 인력양성 및 기초과학과의 연계방안 함께 제시
중장기 창의역량 강화	• (성과) 연구몰입환경 조성 및 출연(연) 역할 강화 • (한계) R&D의 창의·도전성 강화 등 혁신지속 필요	➡	• R&D 창의·도전성 강화 및 연구 관리제도 혁신을 중점과제로 추진 - (과제1) 과학적 지식탐구 및 창의·도전적 연구 진흥 - (과제2) 연구자 중심의 연구몰입환경 조성
신산업 창출 지원	• (성과) 중소·벤처 중심의 R&D 지원체계 구축 • (한계) 신기술·성장동력 확보 및 중소기업 지원 실효성 제고 필요	➡	• 4차산업혁명에 대응한 신기술·성장동력 확보 중점 추진 - (과제11) 4차 산업혁명 대응기반 강화 - (과제12) 국민이 체감하는 혁신 성장동력 육성 • 주소기업 R&D 투자 확대 및 효율화 추진(과제 14-1)
일자리 창출	• (성과) 벤처투자 및 창업 활성화 • (한계) 일자리 창출 강화 및 미래 일자리 변화 대응 필요	➡	• (과제 15) 과학기술기반 일자리 창출 강화 과제 추진 - 과학기술기반 이 자리 26만개 창출목표 및 미래사회 일자리 변화 대응강화 보완

자료: 「제4차 과학기술기본계획」, 2018.02

□ R&D 사업화 성공률

- R&D 사업화 성공률이란 연구개발사업의 산출물이 시장에서 부가가치를 창출하는 과정에서 발생하는 불확실성을 반영한 값

R&D 사업화 성공률이란 연구개발사업 산출물이 시장에서 부가가치를 창출하는 과정에서 사업화에 성공할 비율을 의미

국가 연구데이터 공유·활용체제 구축사업의 R&D 사업화 성공률은 24.7%로 적용

- 기술개발의 실패할 경우 혹은 사업화의 실패할 경우 사업화 성공률이 '0'이 되어 연구개발사업의 편익이 기대되기 어려움
- 따라서 기술개발이 시장에서 실제적인 경제적 효과 창출로 이어지기 위해서는 해당 기술의 실증 및 상용화 과정이 전제되어야 하며, 그 산정식은 다음과 같음

$$\text{R\&D 사업화 성공률} = \frac{\text{기술료징수대상중총사업화성공과제수}}{\text{기술료징수대상과제수}} \times 100$$

- '2016년 국가연구개발사업 성과분석 보고서'에서 제시한 정보통신분야 사업화성공률을 24.7%를 적용
- 정확한 편익 추정을 위해서는 연구개발사업의 특성에 맞는 사업화성공률을 적용해야 하는데, 특정 기관에서 발간하는 성과분석보고서를 일괄적으로 적용하는 것은 이러한 특징을 반영하기 어려움

R&D사업화성공률은 특정 기관에서 발간하는 성과분석보고서를 단일화하여 적용하는 것은 어려운 실정...따라서 충분한 근거가 있다면 주관부처에서 제출하는 구개발성과의 활용에 대한 조사·분석보고서를 우선 활용하고, 필요에 따라 기타 자료를 활용...

「국가연구개발사업 예비타당성조사 수행 세부지침」, 2018.04

- 따라서 '국가연구개발사업 예비타당성조사 수행 세부지침(2018.04)'에 근거 주관부처에서 제출하는 성과분석보고서를 우선 적용하며, 필요에 따라 기타 자료를 활용함
- 과학기술정보통신부와 한국과학기술기획평가원이 발표한 '2016년 국가연구개발사업 성과분석 보고서'에 따르면 2016년 사업화 건수는 28,031건 이며 전년대비 39.5% 증가함

<표 9-41> 국가연구개발사업 성과 총괄현황

구분	과학적 성과	기술적 성과		경제적 성과	
	SCI(E)논문	국내등록 특허	해외등록 특허	기술료	사업화
'16년 성과	37,385건	16,670건	2,121건	8,849건 (2,661억 원)	28,031건
전년대비 증감률	4.3%	11.3%	12.2%	20.0% (△16.0%)	39.5%
연평균 증가율 ('12년~'16년)	6.9%	10.7%	20.7%	12.4% (△1.9%)	18.0%

자료: 2016년도 국가연구개발사업 성과분석 보고서

- 미래유망기술(6T)분야별 사업화 성과의 경우 IT분야가 6,920건으로 가장 높은 건수를 달성하였으며, 24.7%의 비중을 차지함

<표 9-42> 2016년도 국가연구개발사업 성과 총괄표

구분		SCI(E)논문		국내 특허				해외 특허				기술료		사업화	
				출원		등록		출원		등록					
		건수	비 중	건수	비중	건수	비중	건수	비중	건수	비중	건수	비중	건수	비중
6T	IT	5,162	13.8	9,764	31.7	4,752	28.5	1,660	33.7	788	37.1	2,724	30.9	6,920	24.7
	BT	14,375	38.5	6,864	22.3	3,761	22.6	1,287	26.1	541	25.5	2,181	24.7	7,800	27.9
	NT	5,350	14.3	3,010	9.8	1,541	9.2	514	10.4	207	9.8	438	5.0	1,463	5.2
	ST	528	1.4	463	1.5	289	1.7	76	1.5	45	2.1	64	0.7	135	0.5
	ET	5,002	13.4	5,037	16.4	3,252	19.5	772	15.7	281	13.3	1,360	15.4	4,394	15.7
	CT	136	0.4	608	2.0	193	1.2	94	1.9	22	1.0	341	3.9	793	2.8
	기 타	6,752	18.1	5,048	16.4	2,874	17.2	520	10.6	237	11.2	1,715	19.4	6,495	23.2

자료: 2016년도 국가연구개발사업 성과분석 보고서

□ 부가가치율

부가가치율이란 연매출액 중 새롭게 창출된 경제적 효과(편익)가 차지하는 비율을 의미

국가 연구데이터 공유·활용체제 구축사업의 부가가치율은 한국은행 산업연관표를 근거로 하여 53.94%로 적용

- 부가가치율이란 매출액 중 새롭게 창출된 경제적 효과(편익)가 차지하는 비중을 나타내는 값임
 - 편익이란 사업의 시행으로 인해 창출된 매출액 전체가 아닌 부가가치 기준으로 추정하는 것으로 부가가치율을 적용함
 - 통계자료에서 사용하는 부가가치율은 기관에 따라 개념적으로 차이가 있으나, 연구개발 부문 예비타당성조사를 수행함에 있어서는 가장 최근에 한국은행에서 발간한 산업연관표를 적용함

부가가치율에 대한 개념과 수치는 출처에 따라 상이하며 예비타당성조사에 어떤 출처의 부가가치율을 적용해야 하는지에 대한 논의는 지속적으로 제기...기존 「연구개발부문 사업의 예비타당성조사 표준지침 연구(제1판)382)」에 따르면 부가가치율에 대한 출처는 한국은행에서 가장 최근에 발간한 산업연관표를 적용할 것으로 명시...

「국가연구개발사업 예비타당성조사 수행 세부지침」, 2018.04

- 한국은행 산업연관표의 산업 분류 중 소프트웨어 개발 및 공급의 부가가치율(53.94%)⁴⁴⁾을 활용
- － 한국은행 산업연관표의 산업 분류 중 빅데이터와 가장 유사한 소프트웨어 개발 및 공급업의 부가가치율인 53.94%를 적용함

□ 연도별 편익 추정과 현재가치 계산

- 앞에서 제시한 미래시장규모 및 지표를 반영한 8년간의 총 편익은 경상가치로 9,355억원, 현재가치로 4,071억 원이 발생하는 것으로 산정됨

<표 9-43> 연도별 부가가치 창출 편익

(단위 : 억원, %)

구분	미래시장규모 (억원)	시장점유율	R&D기여율	사업기여율	사업화성공률	부가가치율	편익	
							경상가치	현재가치
2033	3,256,904	1%	40%	54.4%	24.7%	53.94%	805	416
2034	3,621,677	1%					889	440
2035	4,027,305	1%					982	465
2036	4,478,363	1%					1,084	491
2037	4,979,940	1%					1,198	519
2038	5,537,693	1%					1,323	548
2039	6,157,915	1%					1,461	580
2040	6,847,602	1%					1,614	613
합계	38,907,401						9,355	4,071

② 교육효과 편익

- 연구개발 인재육성 성과는 특정 기간 동안 피교육생의 업무효율을 높이고, 그 높아진 효율의 가치는 교육을 받지 못한 동료 대비 임금상승률에 반영된다고 볼 수 있음

44) [연구개발부문 사업의 예비타당성조사 표준지침 연구]

- 이러한 교육효과 편익은 다음과 같이 추정함

$$\text{교육효과} = \text{교육수요} \times 12 \times \text{월 평균임금} \times \text{교육에 의한 임금상승률}$$

□ 교육수요

- 동 사업에서 교육수요는 앞에서 산정한 교육대상 인원 22,384명을 적용함

<표 9-44> 연도별 교육대상 인원

(단위 : 명)

구분	'21	'22	'23	'24	'25	'26	'27	'28	'29	'30	계
교육대상	1,802	1,887	1,976	2,070	2,168	2,271	2,378	2,491	2,609	2,732	22,384

□ 월 평균임금

- 월 평균임금은 해당분야의 개별임금을 조사하여 파악할 수 있으나, 각 직종의 종사자 수가 파악되지 않아 가중평균을 취할 수 없다는 문제가 있음
- 따라서, 분석기준년도인 2018년 「고용형태별근로실태조사」 컴퓨터프로그래밍·시스템통합 및 관리업과 정보서비스업의 평균임금을 이용하여 대상 분야의 임금을 산정함

□ 임금상승률

- 동 사업의 인력양성에 의한 편익은 교육훈련 이수 후의 임금 상승효과를 반영하여 도출할 수 있으며, 이는 직업관련 공공재의 특성이 임금수준에 반영된다는 특성임금법(hedonic wage method, Bockstael and McConnell, 2007⁴⁵)의 사례로 볼 수 있음
- 이러한 추정방법에서 추가로 파악되어야 할 부분이 교육에 의한 임금상승률 추정치이며, 이에 관련한 선행연구로 유경준·강창희(2010)가 제시한 고정효과모형과 성향점수 매칭법이 있음⁴⁶)
 - 우리나라에서 교육훈련(직업훈련)을 이수하는 경우 고정효과모형은 1년간 월평균 2.6%~4.7%, 성향점수매칭법은 7.5%~9.8%의 임금상승효과가 있다고 제시하였으나, 유사한 해외 선행연구 검토 결과 그들의 추정치보다 낮게 추정되거나 통계적 유의성이

45) Bockstael and K .E. McConnell, Environmental and Resource Valuation with Reveal Preferences: A Theoretical Guide to Empirical Models, Berlin, Springer. 2007

46) 유경준·강창희, “직업훈련의 임금효과 분석: 「경제활동인구조사」를 중심으로”, 한국개발연구 32(2), 2010, pp.29-53.

발견되지 않는 경우가 있음을 지적하였음

- 한국개발연구원의 2017년 예비타당성조사 결과보고서에서는 이러한 불확실성을 감안하여 고정효과 분석으로 추정한 임금상승효과를 적용함⁴⁷⁾
- 따라서 이러한 점을 감안하여 동 사업에서는 보수적인 값인 고정효과모형의 분석치(평균 3.65%)를 사용하며, 월평균임금은 통계자료를 사용하여 교육효과 편익을 추정함

□ 추정결과

- 앞에서 제시한 지표를 반영한 교육효과 편익은 사업기간인 10년간 발생하며 경상가치로 356.3억 원, 현재가치로 253.9억 원이 발생하는 것으로 산정됨

<표 9-45> 연도별 교육훈련 편익

(단위 : 억원, %)

구분	교육대상인원 (명)	월 평균임금(원)	직무교육에 의한 임금상승률	편익	
				경상가치	현재가치
2021	1,802	3,634,500	3.65%	28.7	25.1
2022	1,887			30.0	25.2
2023	1,976			31.5	25.2
2024	2,070			33.0	25.3
2025	2,168			34.5	25.4
2026	2,271			36.2	25.4
2027	2,378			37.9	25.5
2028	2,491			39.7	25.5
2029	2,609			41.5	25.6
2030	2,732			43.5	25.6
합계	22,384			356.3	253.9

47) 한국개발연구원, 식생활교육문화연구센터 건립사업, 2017

9.4.3. 경제성 분석

가. 경제성 분석 개요

□ 경제성(비용편익) 분석 검토지표 정의

- 비용편익 분석은 비용편익 비율(benefit-cost ratio, BCR), 순현재가치(net present value, NPV), 내부수익률(internal rate of return, IRR) 등을 결과지표로 활용
- 비용편익 비율(benefit-cost ratio, BCR)
 - 연구개발부문 타당성조사에서 기본 지표로써 사용되는 비율로, 사업추진으로 인한 총 편익과 총비용을 현재가치로 환산한 후 총 편익의 현재가치를 총비용의 현재가치로 나눈 값
 - 산출된 값이 1보다 크거나 같다($BCR \geq 1$)로 산출되면 대상 사업이 경제적 타당성이 있는 것으로 평가

$$B/C \text{비율} = \sum_{t=0}^n \frac{B_t}{(1+r)^t} / \sum_{t=0}^n \frac{C_t}{(1+r)^t}$$

Bt: t시점의 편익, Ct: t시점의 비용, r: 할인율, n: 분석기간

- 순현재가치(net present value, NPV)
 - 예산이 투자되는 전 기간 동안 서로 다른 시간에 발생하는 편익과 비용을 기준년도의 현재가치로 환산하여 총 편익에서 총 비용을 뺀 차이 값
 - 산출된 값이 양수($NPV > 0$)로 산출되면 대상 사업이 경제적 타당성이 있는 것으로 평가

$$\text{순현재가치}(NPV) = \sum_{t=0}^n \frac{B_t}{(1+r)^t} - \sum_{t=0}^n \frac{C_t}{(1+r)^t}$$

- 내부수익률(internal rate of return, IRR)
 - 국가연구개발사업이 원만하게 수행된다는 전제하에 기대되는 수익률로써, 예산이 투자되는 전 기간에 발생하는 편익의 현재가치와 비용의 현재가치를 동일하게 만드는 할인율
 - 산출되는 값이 사회적 할인율보다 클 경우 대상 사업이 경제적 타당성이 있는 것으로 평가

$$\text{내부수익률}(IRR) = R$$

$$s.t. \sum_{t=0}^n \frac{B_t}{(1+R)^t} = \sum_{t=0}^n \frac{C_t}{(1+R)^t}$$

<표 9-46> 경제성 분석기준의 장·단점 비교

지표	판단	장점	단점
비용편익 비율 (B/C)	$B/C \geq 1$	• 이해용이, 사업규모 고려 가능 • 투자규모 대비 편익의 비율을 직접적으로 표시	• 상호배타적 대안선택의 오류 발생 가능 • 투자액이 작은 사업의 수익성이 과장되기 쉬움 • 비용과 편익의 구분이 명확하지 않은 경우 B/C비율이 적절한 수치가 되지 못할 수 있음
순현재 가치 (NPV)	$NPV \geq 0$	• 투자 순가치 직접적으로 표시 • 대안선택 시 명확한 기준 제시 • 장래 발생 편익의 현재가치 제시 • 한계 순현재가치 고려 • 타 분석에 이용 가능	• 이해의 어려움 • 대안 우선순위 결정시 오류발생 가능 • 투자사업 규모가 클수록 크게 나타나지만 자본투자의 효율성이 드러나지 않음
내부 수익률 (IRR)	$IRR \geq r$	• 사업의 수익성 측정 가능 • 타 대안과 비교가 용이 • 평가과정과 결과 이해 용이 • NPV나 B/C 적용 시 할인율이 분명할 경우 이용 가능함	• 사업의 절대적 규모 미고려로 투자액이 작은 사업의 수익성이 과장될 가능성이 있음 • 복수의 내부수익률이 동시에 도출될 가능성 내재

자료: 국가연구개발사업 예비타당성조사 수행 세부지침

나. 경제성 분석

□ 경제성 분석

- 국가 연구데이터 공유·활용 체제 구축 사업의 경제성 분석을 위한 총비용과 총편익의 흐름을 정리하면 아래와 같음

<표 9-47> 국가 연구데이터 공유·활용 체제 구축 사업의 비용 및 편익 추정 결과

(단위: 억 원)

구분		비용					편익			
		플랫폼 구축 및 운영	대·소 분야전문 센터구축 및운영지 원	R&D 인재육성	합계		부가가치 편익	교육효과 편익	합계	
					경상가치	현재가치			경상가치	현재가치
사업기간	2021	142	90	66	299	262		29	29	25
	2022	218	104	62	383	321		30	30	25
	2023	205	274	63	542	435		31	31	25
	2024	233	278	64	575	441		33	33	25
	2025	178	412	64	654	480		35	35	25
	2026	144	409	52	605	425		36	36	25
	2027	160	401	53	614	413		38	38	25
	2028	164	382	53	598	385		40	40	26
	2029	163	354	54	571	352		42	42	26
	2030	163	346	55	563	332		43	43	26
회임기간	2031				0	0			0	0
	2032				0	0			0	0
편익발생 기간	2033				0	0	805		805	416
	2034				0	0	889		889	440
	2035				0	0	982		982	465
	2036				0	0	1,084		1,084	491
	2037				0	0	1,198		1,198	519
	2038				0	0	1,323		1,323	548
	2039				0	0	1,461		1,461	580
	2040				0	0	1,614		1,614	613
합계		1,769	3,049	586	5,404	3,847	9,355	356	9,711	4,325

- 비용과 편익을 종합하여 경제적 타당성을 분석한 결과, 비용편익 비율(B/C Ratio)은 1.12, 순현재가치(NPV)는 477억원, 내부수익률(IRR)은 5.67%로 나타나 사업 추진의 경제적 타당성을 확보하는 것으로 분석됨

<표 9-48> 국가 연구데이터 공유·활용 체제 구축 사업의 경제성 분석결과

(단위: 억 원, %)

편익(현재가치)	비용(현재가치)	B/C	NPV	IRR
4,325	3,847	1.12	477	5.67%

□ 민감도 분석

- 경제적 타당성 평가과정에 사용되는 편익 및 비용추정에는 불확실성이 내포되어 있어, 이에 대한 대처를 위해 민감도 분석(할인율, 비용, 편익)을 수행함
- 사회적 할인율의 경우 $\pm 1\%p$ 단위로, 편익 및 비용은 $\pm 10\%p$ 단위의 변경요인이 발생하였을 때 경제성 판단 지표들이 어떻게 변화하는지를 검토함
- 할인율을 1%p씩 증감시켜 민감도 분석을 수행한 결과, 할인율이 1.3%p 증가하면, B/C가 0.99로 경제적 타당성을 확보하지 못하는 것으로 나타남
- 편익을 10%씩 증감시켜 민감도 분석을 수행한 결과, 편익이 12% 감소하면, B/C가 0.99으로 경제적 타당성을 확보하지 못하는 것으로 나타남
- 비용을 10%씩 증감시켜 민감도 분석을 수행한 결과, 비용이 13% 증가하면, B/C가 0.99로 경제적 타당성을 확보하지 못하는 것으로 나타남

<표 9-49> 민감도 분석 결과

(단위: 억 원)

구분		B/C	NPV	편익(현재가치)	비용(현재가치)
할인율의 변화	2.5%	1.38	1,694	6,153	4,458
	3.5%	1.24	1,012	5,151	4,139
	4.5%(기준)	1.12	477	4,325	3,847
	5.5%	1.02	60	3,641	3,581
	6.5%	0.92	-264	3,074	3,338
편익의 변화	80%	0.90	-388	3,460	3,847
	90%	1.01	45	3,892	3,847
	100%(기준)	1.12	477	4,325	3,847
	110%	1.24	910	4,757	3,847
	120%	1.35	1,342	5,190	3,847
비용의 변화	80%	1.41	1,247	4,325	3,078
	90%	1.25	862	4,325	3,463
	100%(기준)	1.12	477	4,325	3,847
	110%	1.02	93	4,325	4,232
	120%	0.94	-292	4,325	4,617

9.5. 기대효과 및 파급효과

9.5.1 기대효과

□ 연구데이터 공유·활용의 인식제고 사업

○ 경제적 기대효과

- 연구자가 연구개발 과정에서 필요로 하는 데이터를 손쉽게 수집·공유·활용하여 연구에 투자되는 시간과 비용을 절감하고
- 기존에 수행된 연구개발의 성과물을 활용하여 연구데이터의 불필요한 연구데이터 재생산 및 중복연구를 방지하여 국가연구개발사업의 효율성을 제고함
- 연구자는 국가 연구데이터 플랫폼을 통해 수집된 연구데이터를 공유·활용하여 새로운 산업분야를 개척하고 고부가가치 지식을 창출함
- 수집된 연구데이터를 활용하여 새로운 연구 분야를 개척하고 산업의 발전과 국가경쟁력을 제고함
- 체계적으로 관리되는 연구데이터 및 연구개발 대한 정보를 제공함으로써 연구개발사업의 산업화 과정을 단축하고 효율성을 제고하여 신산업 창출에 기여함
- 연구데이터의 다양성·복잡성으로 인해 시간이 많이 소요되었던 연구개발 및 실용화 기간을 획기적으로 단축시킴

○ 과학기술적 기대효과

- 국가 연구데이터 플랫폼은 연구자 커뮤니티를 형성하여 연구자 간 교류 및 융합연구를 실현할 수 있도록 함
- 연구데이터가 다수의 연구자에게 공유될 경우, 연구자의 자유로운 참여와 전문성으로 연구 결과 및 연구데이터의 품질 향상과 투명성 제고에 기여함
- 국내·외 연구기관에서 수집된 연구데이터 및 메타데이터의 표준화를 통해 연구데이터의 품질을 향상시킴
- 통합하여 수집·관리하는 전문센터의 기능으로 연구데이터의 유실이 방지되며, 공유·활용으로 연구데이터를 지식·자산화 함
- 빅데이터 기술 발전의 현상과 연구데이터의 공유·활용 환경이 상호작용하여 연구개발

생태계를 구축하는 시너지효과를 일으킴

- 연구데이터의 지속적인 수집·관리를 통해 개별 연구기관이 자원별·기관별로 용이하게 연구데이터를 활용할 수 있도록 지원함
- 연구데이터 활용능력 및 과학적 추론 능력을 겸비한 데이터 전문 과학자를 양성함

□ 데이터 및 컴퓨팅 활용 R&D 인재 성장 지원

○ 경제적 기대효과

- 데이터 분석을 위한 기본 교과과정과 전문 강사를 공동 활용하는 교육시스템에 바우처를 활용하여 운영함으로써 비용절감
- 교육현장과 연구, 산업현장 간 공간적·시간적 단절을 극복하고 교육에서 사회로 진입하는 기간을 단축하여 전문 인력을 조기에 확보함

○ 과학기술적 기대효과

- 과학데이터 전문교육 패러다임 변화를 선도하고 4차 산업시대에 부합하는 데이터 전문 과학자를 양성함

9.5.2. 파급효과

- 연구데이터 공유·활용은 연구개발의 주기를 획기적으로 단축시키며, 이를 통해 정부 및 민간 투자를 통한 타 연구개발사업 진출에 기여함
- 다양한 분야와 연령층에서 슈퍼컴퓨팅을 활용할 수 있는 기회를 제공하여 전 국민이 연구데이터 생산에 참여하는 잠재적 연구자로서의 자질을 키움
- 연구데이터의 일괄 수집·관리를 통해 분산적인 정보유통을 중앙 집중화하고, 양질의 데이터를 신속하게 제공·연계하여 범국가적인 연구 연맹 체제를 구축함
- 대학과 출연연구소 센터를 연계하여 실무 적용 및 활용 중심의 빅데이터 교육을 실시하고 현장 전문 인력을 신산업 창출과 동시에, 실무중심의 인재를 사회 전반에 필요한 적재적소에 배치함