

데이터 경제 시대, AI 및 빅데이터

2020. 7.

오성탁 지능데이터본부장

NIA 한국정보화진흥원

01

| 디지털 대전환과 데이터경제

02

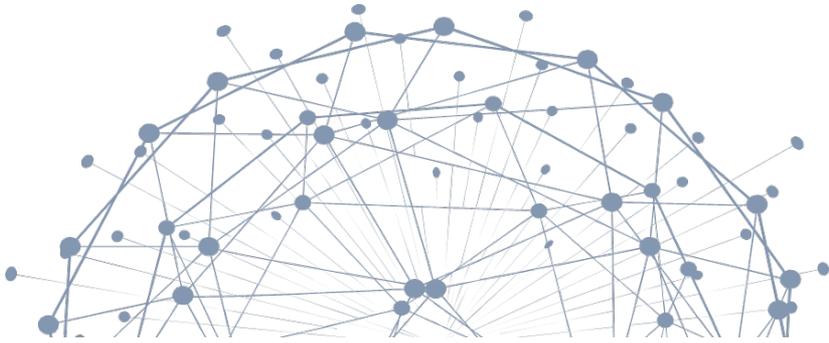
| 국내 추진현황

03

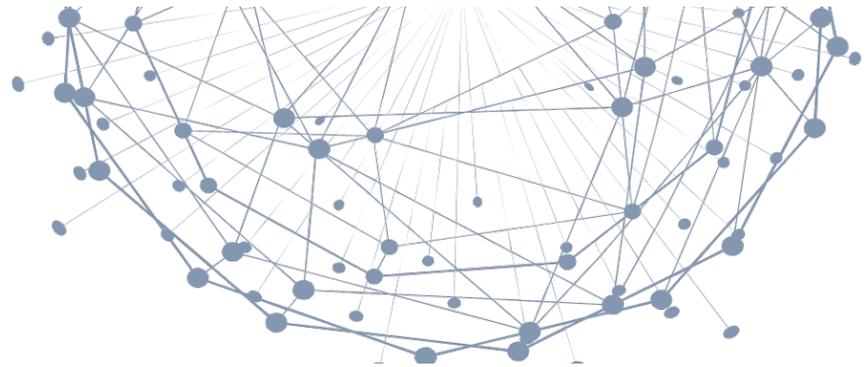
| 추진과제 소개

04

| 새로운 도전



디지털 대전환과 데이터경제

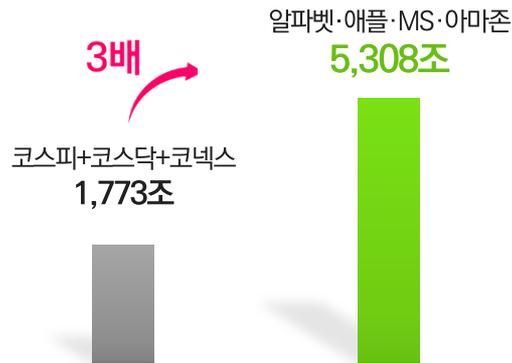


NIA 한국정보화진흥원

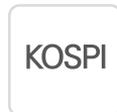
데이터 역량 = 새로운 글로벌 국가 경쟁력

데이터가 경쟁력

- 대규모 데이터를 확보하는 글로벌 ICT기업 시장 주도



1,598조



1,518조

※단위: 원, '20.1.16 기준

융합 및 혁신서비스의 밑거름

- 개인정보 활용 및 데이터 결합을 통한 혁신적인 서비스 및 제품 생산 가능
- 특히 방대한 데이터 축적으로 파급효과가 큰 의료·금융 분야와 타 산업분야 융합을 통해 혁신 성장의 견인차 역할

의료



개인맞춤형 건강 도우미, 시닥터(닥터앤서) 등 활용

금융



비금융데이터 신용평가, 보험금자동청구 등 활용

시도 핵심은 데이터

- 머신러닝 기반 시는 방대한 양질의 학습용 데이터가 성능을 좌우
- 시 경쟁력도 대량의 도메인별 데이터 확보 여부에 따라 결정

세계는 데이터경제로 이행중

데이터 경제

데이터 수집, 저장,
유통, 활용



모든 산업에 적용



경제 가치 창출
(신제품, 서비스,
직업 등)



국가별 주요 정책



Big Data
R&D Strategic Plan
(2016)



Plan for Realizing
Society 5.0 (2017)



Building a
European Data
Economy (2017)



Big Data Industry
Development Plan
(2017)



Data Industry
Facilitation Strategy
(2018)

데이터·AI 영역으로, 전통산업 영역으로 IT기업의 변신

IT 기업 → 데이터·AI 영역

NAVER » NBP NAVER BUSINESS PLATFORM | Clova

kt » kt Cloud | GiGA Genie

amazon » aws | alexa

Google » Google Cloud | DeepMind

IBM » IBM Watson

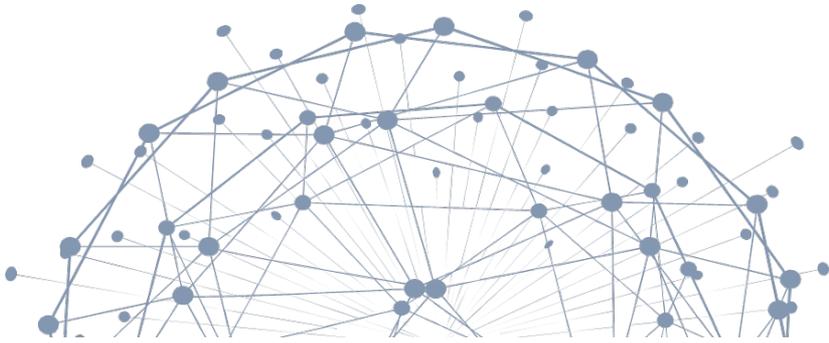
데이터 기업 → 전통 산업

kakao » KakaoTaxi | kakaobank | KakaoPay

Google » WAYMO | SIDE WALK LABS

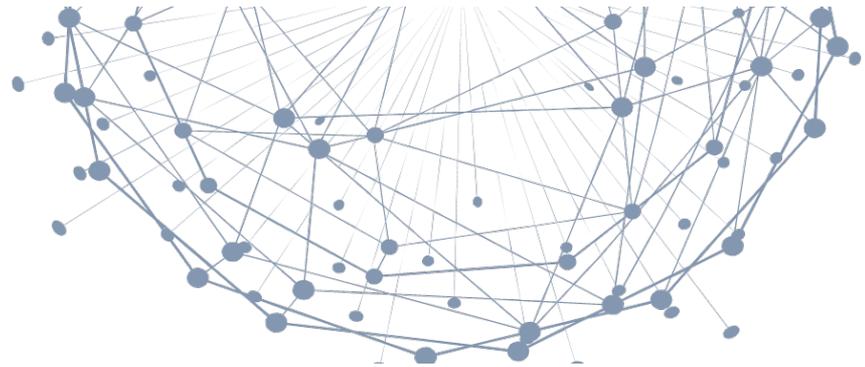
Alibaba.com » ET BRAIN ET 시리 브레인

IBM Watson » 의료 | 금융 | 법률



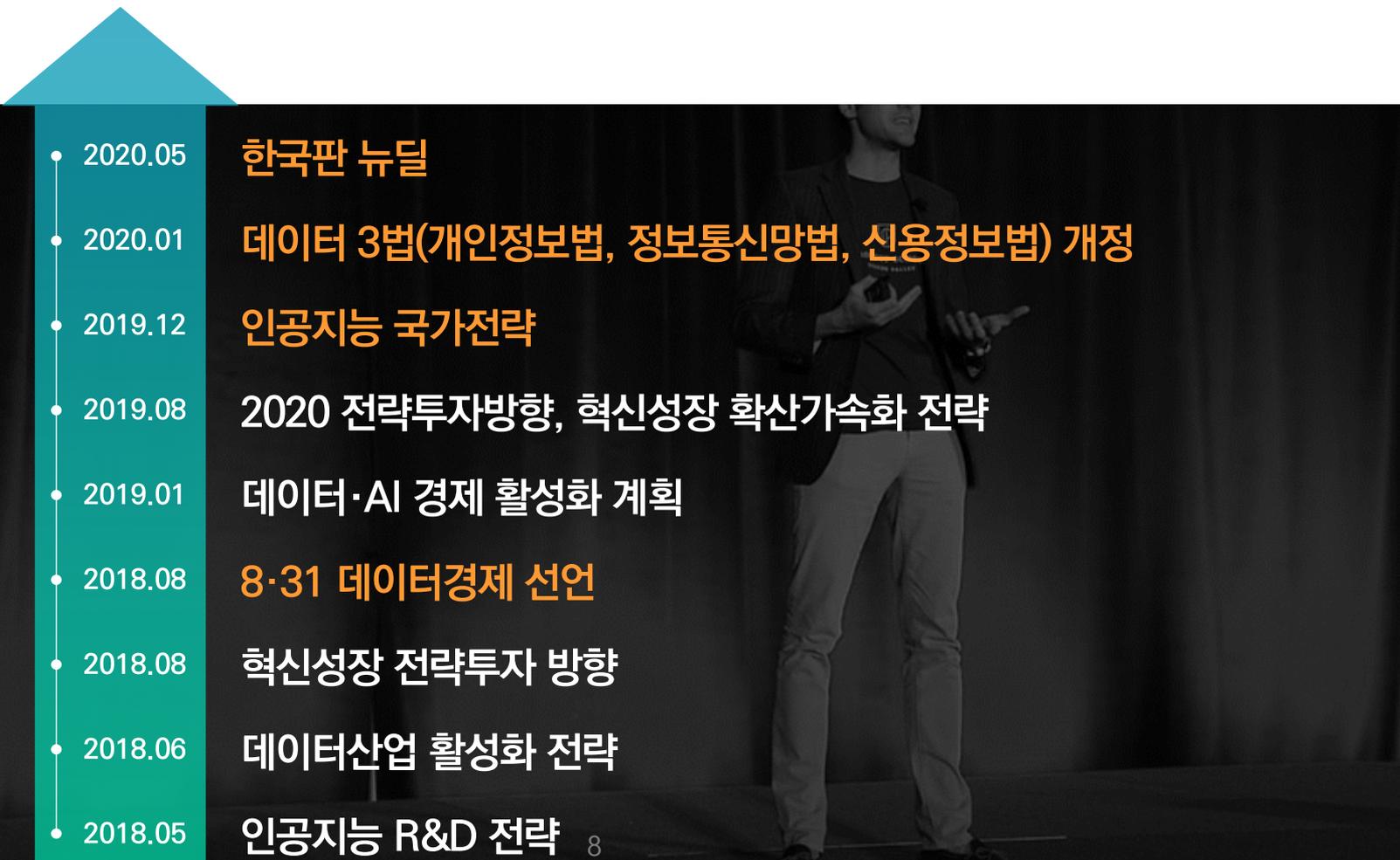
||

국내 추진현황



NIA 한국정보화진흥원

데이터경제가 혁신 성장의 키워드로 확립

- 
- 2020.05 **한국판 뉴딜**
 - 2020.01 **데이터 3법(개인정보법, 정보통신망법, 신용정보법) 개정**
 - 2019.12 **인공지능 국가전략**
 - 2019.08 **2020 전략투자방향, 혁신성장 확산가속화 전략**
 - 2019.01 **데이터·AI 경제 활성화 계획**
 - 2018.08 **8·31 데이터경제 선언**
 - 2018.08 **혁신성장 전략투자 방향**
 - 2018.06 **데이터산업 활성화 전략**
 - 2018.05 **인공지능 R&D 전략**

[DATA]

- ▶ 데이터 경제로의 전환 선언('18.8)
 - ▶ 데이터·AI 활성화 계획('19.1)
 - ▶ 데이터 3법 개정('20.2)
- * 빅데이터 플랫폼(10개), 센터(100개)
구축 완료



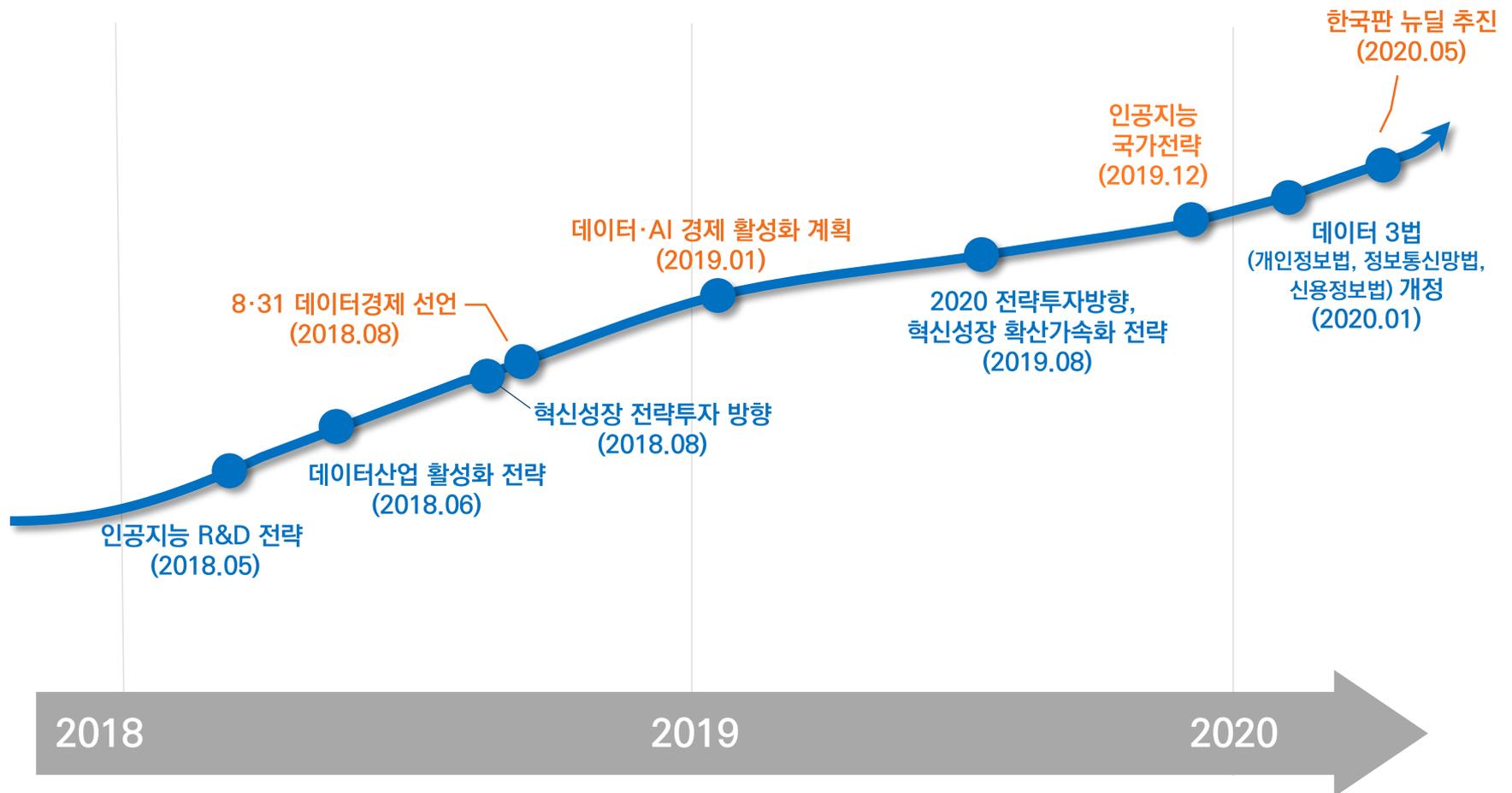
[AI]

- ▶ 인공지능 기본구상('19.10)
 - ▶ 인공지능 국가전략('19.12)
- * AI 스타트업 투자:
'17. 408 → '19. 11월 1,891억원



- ✓ 데이터 이용 활성화를 위한 가명정보 개념 도입
- ✓ 관련 법률의 유사 중복 규정을 정비하고 추진체계를 일원화
- ✓ 데이터 활용에 따른 개인정보처리자의 책임강화
- ✓ 모호한 개인정보 판단 기준의 명확화

데이터·AI 경제로의 이행 가속화



[**D** **N** **A** +]

Data Network AI 역기능, 사회문화

Data-driven Society

대규모 예산 투입으로 데이터 분야 성장 기반 마련

'19년 데이터·AI 분야 약 1,787억 원 투자

구축·개방

빅데이터 플랫폼 및 센터 구축
AI 학습용 데이터 구축

빅데이터 플랫폼
및 센터 구축
743억

AI 학습용
데이터 구축
195억

유통

데이터 구매 및 가공 바우처 사업

데이터 바우처
지원 사업
600억

활용

안전한 개인정보 활용 및 서비스 모델 발굴 지원

개인데이터
활용 서비스
97억

데이터 기업과 산업 전반의 성장 발판 제공

전자신문

2019년 08월 01일 (목)
02면 중상

데이터경제 신바람, 공공·민간 투트랙 견인



김지선 river@
SW 전문기자

정부가 데이터경제 활성화 정책을 올해부터 본격 시행한 가운데 데이터 솔루션, 컨설팅 등 주요 데이터기업 상반기 매출이 전년도 대비 최대 50%까지 증가한 것으로 나타났다.

공공이 굵직한 사업을 발주해 마중물 역할을 담당하고 민간 기업도 신규 프로젝트를 진행하면서 올해 데이터기업 매출 증대와 산업 성장에 기대된다.

31일 업계에 따르면 주요 데이터 기업 상반기 매출이 지난해 동기 대비 20~50%가량 증가했다.

데이터 기업은 정형데이터, 비정형데이터 등 데이터 관련 솔루션 개발과 컨설팅, 서비스 구축 등을 담당하는 소프트웨어(SW)기업 전반을 의미한다.

데이터 검색, 수집 분석, 구축 등을 담당하는 와이즈넷(20%), 코난테크놀로지(50%), 솔트룩스(20%) 등 주요 기업은 올 상반기에 지난해 대비 높은 성장세를 기록했다. 3사는 지난해 매출 250억원, 110억원, 155억원을 각각 기록했다. 상반기 성장세에 힘입어 연간 매출 역시 급증할 것으로 기대한다.

데이터 컨설팅, 구축 분야 주요 기업인 비투엔과 투이컨설팅도 전년도 대비 상반기에 20%가량 매출이 성장했다. 데이터 분야 스타트업인 크라우드웍스와 빅블류도 상반기 각각 지난해보다 3

데이터 관련 솔루션·컨설팅 기업
상반기 매출 작년보다 20~50% ↑

정부, 관련 예산 1787억 들여 육성
주요 부처 프로젝트 발주 잇달아
금융·제조 민간 사업도 기지개

배~23배 이상 매출이 상승했다.

공공시장이 데이터기업 성장을 위한 마중물 역할을 담당했다.

정부는 지난해 8월 말 부처 합동으로 마련한 데이터 경제 활성화 정책을 발표했다. 이에 맞춰 과학기술정보통신부는 올해 빅데이터 플랫폼, 마이데이터 등 데이터경제 활성화와 관련 사업 예산을 지난해보다 4배 늘어난 1787억원을 투입했다. 과기정통부뿐 아니라 주요 부처에서 데이터 관련 신규 사업을 발주하면서 상반기 산업 전반에 긍정적 영향을 미쳤다.

'빅데이터 플랫폼 및 센터 구축사업-데이터품질관리(와이즈넷, 비투엔)' '데이터 바꾸쳐 지원사업(솔트룩스, 와이즈넷)' '마이데이터활용 컨설팅(투이컨설팅)' '중앙선거관리위원회 국가선거정보공개방체계 구축(비투엔)' '한국전력공사(코난테크놀로지)' 등 주요 공공이 신규 사업을 진행,

기업 상반기 신규 매출 확보에 기여했다.

민간 사업도 늘었다. 삼성화재 데이터활용 품질 최적화 프로젝트(비투엔), SK하이닉스 데이터 거버넌스 컨설팅(투이컨설팅) 등 금융, 제조 등 주요 분야별 신규 사업이 추가됐다.

김영섭 코난테크놀로지 대표는 "정부 데이터 경제 활성화 기조에 맞춰 공공뿐 아니라 민간도 데이터 관련 다양한 사업을 진행한다"면서 "하반기에도 비슷한 기조가 이어져 올해 매출 실적이 향상할 것으로 기대한다"고 말했다.

데이터기업 성장세는 하반기를 넘어 내년까지 지속될 전망이다. 하반기 지역정보개발원, 국민건강보험공단 등 주요 공공이 차세대 프로젝트를 진행하면서 데이터 관련 사업을 신규 발주한다. 올해 금융·통신 등 10개 분야별 빅데이터 플랫폼과 센터 구축을 시작으로 내년부터 데이터 유통이 본격화된다.

조광원 한국데이터산업협회장(비투엔 대표)은 "올해 상반기는 지난해보다 공공 신규 프로젝트가 많이 발주돼 사업기회가 많았고 협회사 전반적으로 성장세를 기록했다"면서 "하반기와 내년에도 공공을 비롯해 민간도 많은 사업이 발주될 것으로 예상돼 데이터기업과 산업 전반이 성장하는 계기를 마련할 것으로 기대한다"고 말했다.

한편 한국데이터산업진흥원에 따르면 데이터 수집, 분석 등 국내 데이터 산업 시장은 지난해 14조원 규모에서 올해 16조원대까지 성장할 전망이다. 박종진 기자 공동취재



〈데이터사업 통합설명회(19.2)〉



〈플랫폼 및 센터 출범식(19.7)〉

민간 기업도 본격적으로 데이터 분야 투자 확대



kt **KT** ICT분야(AI, IoT 등) 3조 9,000억원 투자
- 4차 산업혁명 분야 총 23조원 투자

유통·소비 **유통소비 플랫폼** 데이터거래소 운영
SPC 설립 및 추가 투자('20년, 30억)

한화손해보험 보험+타 분야 데이터 융합
→ 1천 억원 규모 손해보험사 설립

SAMSUNG **삼성카드** 유통업계와 협력 및 데이터 마케팅
→ 908억원 당기순이익 기록('19년 하반기)

MARKET Kurly **마켓컬리** 판매 데이터+ 물류 데이터
→ 매출·물류 예측 시스템 도입(신선식품 폐기율 1%)

농협은행 ICT 인재 100여명 신규 채용
- 절반을 IT·디지털 인재로 뽑는다

카드업계
빅데이터·정보보안 인재에 러브콜

핀테크 바람 타고 은행 취업문도
IT 인재에 '활짝'

금융권 채용 현장에도
디지털 부문 구인 바람

데이터 분야 다양한 일자리 창출 및 법제도 마련

데이터 직무 관련 일자리 창출

- 데이터 플랫폼 및 센터 구축 사업, AI데이터 구축 사업으로 1,161명 신규 고용(NIA 수행사업)

[데이터일자리(예시)]

데이터 설계·분석	데이터 연구개발	데이터 수집·전처리

<대표적인 예 (아마존 Mechanical Turk)>

- 기업 또는 수요자가 온라인을 통해서 업무를 의뢰하고 작업자가 업무를 수행하는 클라우드 소싱 방식의 오픈 플랫폼
- 기계가 하는 것 보다 사람이 하는 것이 더 효율적인 일(예를 들어 사진 속 물체 구별 등)을 파트타임의 형태로 작업 진행
- 생산된 데이터는 머신러닝 학습데이터로 활용
- 현재 전세계 200여 국가에서 등록된 작업자 수가 50만명이 넘고, 이중에 10만명은 가용한 상태이며, 평균 2,000명이 현재 작업 중

데이터 3법 통과

2020년 01월 10일 (금) 종합

KBS

'데이터3법' 국회에서 통과...과기부 "TF 운영 해 데이터 산업 본격 지원"

개인정보보호법

가명정보데이터를 제품이나 서비스 개발에 이용

정보통신망법

온라인상 개인정보보호 규제와 감독권한을 '개인정보보호위원회'로 변경

신용정보법

* 가명정보, 금융분야 데이터를 빅데이터 분석, 이용가능
* 통계작성, 연구 등을 위해 가명 정보 주체 동의 없이 이용과 제공 허용

데이터 이용 활성화를 위한 개인정보보호법 등 '데이터3법'이 국회 본회의를 통과하면서 정부도 TF팀을 만들어 데이터 산업 육성을 본격 지원하겠다고 밝혔습니다.

개인정보보호법·신용정보법·정보통신망법을 일컫는 데이터3법 개정안은 개인이나 기업이 수집해서 이용할 수 있는 개인 정보 범위를 넓혀 빅데이터 산업을 키우는 내용이 핵심입니다.

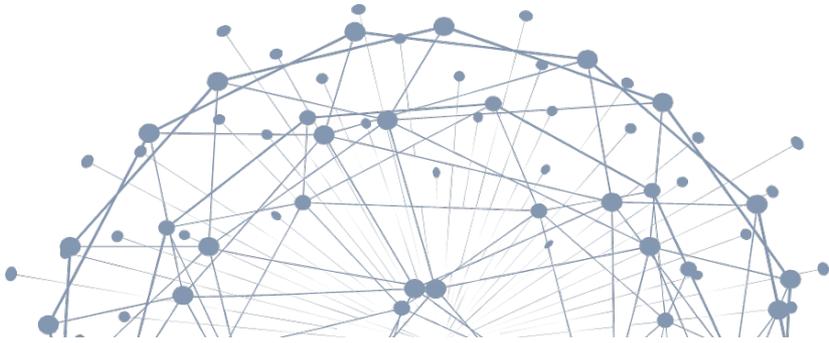
이번 개정안에 의뢰화된 정보에는 개인정보 보호법을 적용하지 않고 가명 정보 개념을 도입하고 개인정보 처리자 책임을 강화하는 내용 등이 포함됐습니다.

이에 대해, 과기정통부는 "다소 모호했던 개인정보의 판단 기준을 명확화했다"며 "데이터의 가명처리를 통해 활용 가능한 데이터의 종류가 다양해지고 새로운 기술 등 활용 분야도 넓어질 것이라고 기대한다"고 밝혔습니다.

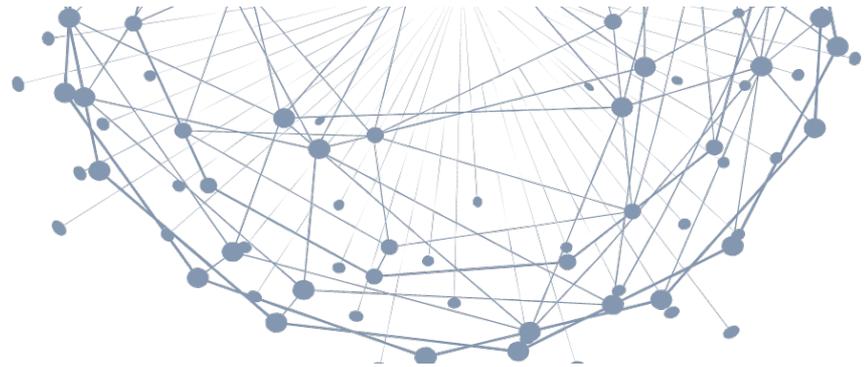
또, 과기부는 금융·의료 분야와 스마트 시티, 자율자동차 등 미래 산업 분야 등 데이터 간 융합과 데이터 산업 육성을 적극적으로 지원할 예정입니다.

이와 함께, 과기정통부는 유럽연합(EU)의 일반개인정보보호법(GDPR) 적정성 평가 승인도 추진하기로 했습니다.

이를 위해, 과기정통부는 각 부처가 참여하는 데이터 경제 활성화 TF를 꾸려 다음 달 안으로 지원 방안을 내놓겠다고 밝혔습니다.



추진과제 소개



NIA 한국정보화진흥원

02 데이터 플래그십 사업

- ▶ 목적: 확산 및 파급 효과가 크며, 다양한 분야(교통·의료·환경·안전 등)에서 활용 가능한 빅데이터 선도 성공사례(Flagship)를 발굴하고 전파
- ▶ 2019년에는 인공지능, 기계학습 등 지능정보기술을 필수적으로 적용하고, 저출산, 버스안전 등 사회문제 해결에 기여할 수 있는 과제를 지원 (경쟁률 7.7:1)

○ 심야버스 노선정책 지원(2013)

KT - 서울시

유동 인구 데이터

버스승객 예측

심야버스노선, 정류장

우리나라 빅데이터 확산의 계기가 된 사업으로 2013년 행정혁신 대통령상 수상

○ 로밍 데이터 활용 감염병 차단 서비스(2016)

KT - 질병관리본부

1차: 도착시 감염병 신고안내

2차: 입국시 자진신고 안내

3차: 위험국가 방문정보 제공

KT

질병관리본부

G20 정상회의 의제 채택(2017), 다보스포럼 제안, 가나 통신사-보건청 등에 수출 (2018)

○ 소형 공동주택 시세 산정 시스템(2017)

빅밸류 - 신한은행

연립다세대 시세산정 빅데이터 시스템

수집체계

빅데이터 분석

시세산정

공인분석

통계학

기계학습

개인 맞춤형 시세평가 서비스

공공데이터의 비즈니스 모델화 성공사례로 20억 규모 투자유치(2018), 금융위 혁신금융서비스 선정(2019)

● 기획검증 분야

기획부동산 사기예방을 위한 토지거래 위험경보 시스템 개발

한국감정원



딥러닝 영상인식을 활용한 지장물 조사 서비스 개발

업데이터+국토정보공사



상표 침해 예방을 위한 SI 기반 유사상표 이미지 검색 서비스

티쓰리큐+특허정보원



물류센터의 현장관리자를 대체하는 인공지능 시스템 구축

로지포커스



버스 안전운전지수 측정체계 확립과 안전 솔루션 개발

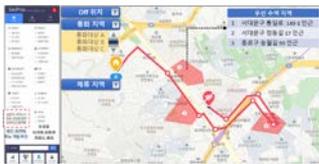
교통안전공단



● 실증확산 분야

실종자 과학수사 정보생성을 통한 범죠평안 플랫폼 구축

KT+경찰청



경기도 농업 가물 예측시스템 구축

경기도



인공지능과 교통 빅데이터를 활용한 혼잡예보 서비스

도로교통공단+UNIST 등



우리아이 시 주치의 서비스

건강보험공단



빅데이터 기반 지능형 하수 관제 서비스 구축

대구환경공단



[참고] 데이터 플래그십 '20년 과제

수행기관	과제명	내용
전자부품연구원/ 질병관리본부	감염병 확산 신속 대응을 위한 데이터기반 역학조사 지원시스템 개발	<ul style="list-style-type: none"> ○ 감염병 확산 신속 대응을 위해 데이터 기반 동선 분석, 감염 위험지역 분석, 감염 네트워크 자동 분석 기능을 제공하는 역학조사 지원시스템 구축
국민건강보험공단	장기요양보험 수급 대상 예측 및 맞춤형 요양 서비스 제공	<ul style="list-style-type: none"> ○ 건강보험 및 장기요양보험의 노인성 질환 및 가족력, 건강검진결과 등과 지자체의 독거, 동반거주 내역 데이터 결합 ○ 장기요양보험 수급 대상 예측 및 맞춤형 요양 서비스 제공
기상청	인공지능 기술을 활용한 기상관측데이터 품질검사 기술 개발	<ul style="list-style-type: none"> ○ 축적된 기상관측자료를 AI를 통해 분석하여 신규 관측 데이터의 오류 자동 판별 및 이상 값 탐지 ○ 정제된 데이터는 방재서비스 등에 활용토록 개방
(주)리얼인사이트	중소상공인을 위한 AI 법률 계약서 리스크 분석 서비스 구축	<ul style="list-style-type: none"> ○ 문서독해 및 텍스트 자동 분석 기술을 활용하여 생활 밀착형 14개 업종(화물운송, 건설, 식음료등)의 국문계약서를 요약, 분석하여 위험조항, 독소조항 및 분쟁 요인을 추출·제시
신용보증기금	『중소기업 데이터 통장』서비스 구축	<ul style="list-style-type: none"> ○ 신용보증, 대출신청 목적으로 제출하는 1회성 데이터를 저장·분석·제공하여 타 금융 및 상거래 업무에 원클릭으로 활용 ○ 수집된 기업데이터를 분석하여 기업진단 서비스 제공
한국감정원	기획부동산 사기예방을 위한 토지이상거래 알람서비스 확산 및 고도화	<ul style="list-style-type: none"> ○ 기획부동산 사기 데이터 기반 사전탐지 기능 개발 ○ 서비스 대상지역의 수도권 전역 확대(기존 5개 → 67개 지역) ○ 사기예방을 위한 실효성 있는 법제도 개선방안 마련
한국법무보호복지공단	출소자의 재범방지를 위한 빅데이터기반 AI 플랫폼 구축	<ul style="list-style-type: none"> ○ 출소자의 재범방지를 위한 사회적응 예측모형 개발 ○ 맞춤형 갱생보호서비스 매칭 시스템 개발 ○ 출소자 통합관리 의사결정을 위한 모니터링 대시보드 개발
한국전자정보통신산업진흥회	포스트 코로나 대응을 위한 전자·ICT 기업 Supply-chain 추천 서비스	<ul style="list-style-type: none"> ○ 기업·특허 데이터 수집과 분석을 통해 중소기업 공급망 진단·추천, 수요처 발굴·추천, 신규제품 개발을 위한 중소·벤처·스타트업에게 맞춤형 기업 추천 서비스 개발
한국해양교통안전공단	연안해역 사고위험도 예측 및 정보제공시스템 개발	<ul style="list-style-type: none"> ○ 해양 교통사고에 영향을 미치는 다양한 영향 요인 분석을 통해 연안 지역 해양 위험도를 도출하고 이를 각 이용자 특성에 맞게 제공하는 “연안해역 사고위험도 예측 및 정보제공시스템” 구축
(주)후후엔컴퍼니	데이터·AI 기반의 실시간 보이스피싱·스미싱 탐지서비스	<ul style="list-style-type: none"> ○ AI 기반 스미싱 탐지 기능 개발, 보이스피싱 탐지 고도화 ○ 통신, 금융, 수사 관련 데이터 결합·활용으로 정보통합 대응체계 구축 ○ 유관기관 신고 및 금융기관 전파를 통한 즉시 대응체계 구축

권역단위

관심 지자체를 선정하여 예산지원, 해당 권역의 중소기업 지원

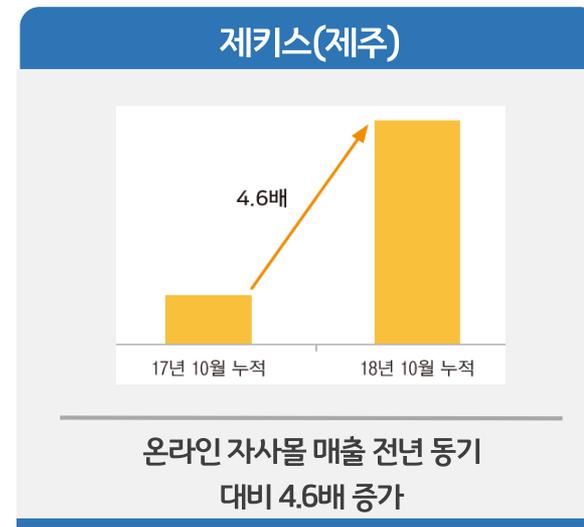
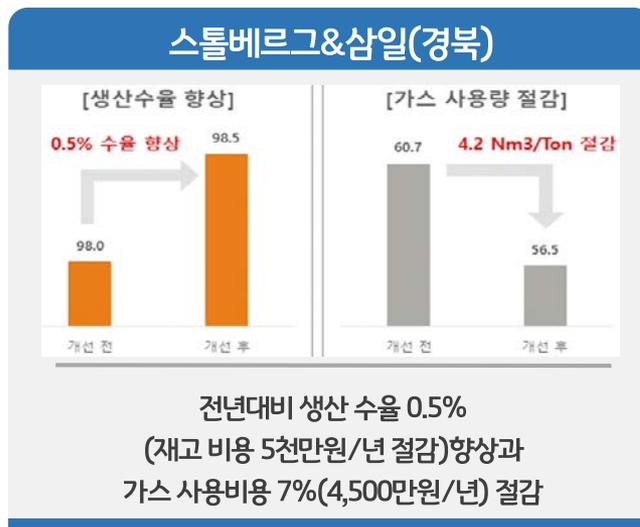
- 예산 매칭이 지역 중소기업에 빅데이터 활용을 통한 성공사례 확산 의지가 강한 기관을 대상으로 추진
- 예산: 정부 + 해당 권역 매칭 (강원 15개, 충북 20개 중소기업)

전국단위

권역과 무관 전국단위 중소기업 지원

- 특정 권역에서만 사업에 지원하여 진행할 경우, 참여하지 않은 권역의 중소기업들은 지원을 받지 못하는 문제 해결
- 예산: 정부 (전국 35개 중소기업)

< 2018년 대표 성과 >



04 판교 빅데이터 센터

대규모 분석인프라와 오프라인 창업/교육 공간 제공
(2016.5월 개소)

〈 빅데이터센터 인프라 〉

K-ICT 빅데이터 센터
교육 인프라
딥러닝 인프라
개발자 인프라
데이터 안전영역
소셜 분석 인프라
분석 인프라
민간 클라우드
사업화 인프라

교육·분석인프라,
교육교재,
실습강사 지원

**비식별
전문기관
운영**

데이터결합,
비식별컨설팅

데이터분석교육,
멘토링

**데이터기반
실습지원**

온·오프라인

**데이터
크리에이터
캠프**



〈판교 스타트업 캠퍼스〉



〈오프랩: 강의장, 입주공간〉



〈회의실, 분석랩〉

빅데이터 플랫폼 및 센터 구축 사업

▶ 빅데이터 플랫폼(10개) 및 센터(100개) 구축·연계

(내용) 데이터 축적-가공-유통 생태계 마련을 위해 ① 빅데이터 센터,

② 빅데이터 플랫폼을 ③ 유기적으로 연계 구축

(일정) 데이터 사업 통합 설명회 개최('19.2.)

10개 빅데이터 플랫폼 및 센터 선정('19.5)

빅데이터 플랫폼 및 센터 출범('19.7)

빅데이터 센터 2차 추가 공모(28개, 112억, '19.7.~8.)

플랫폼 및 센터 감리 사업 발주('19.8, 4개 과제 총 20억 규모)

빅데이터 플랫폼 구축 완료('19.12)



〈 데이터사업 통합설명회('19.2) 〉

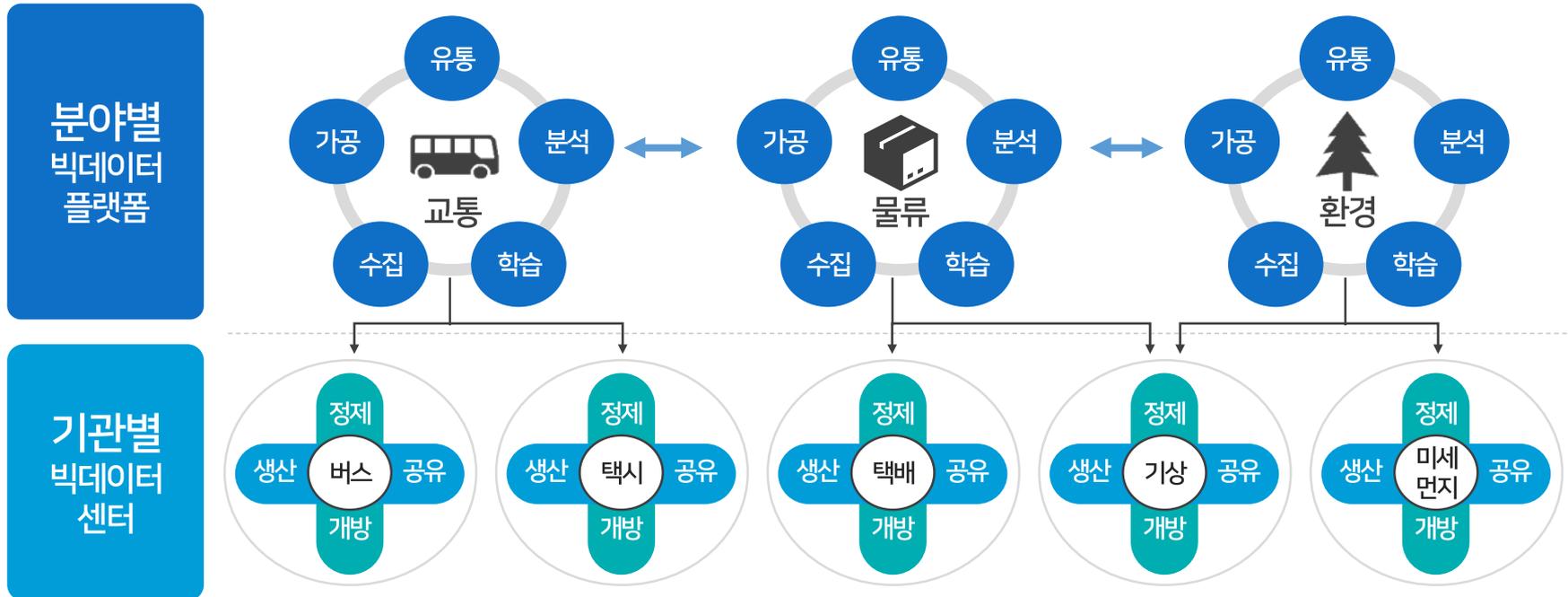


〈 플랫폼 및 센터 출범식('19.7) 〉

(플랫폼) 주요 분야별로 각종 데이터의 수집·분석·유통을 지원

※ 금융, 환경, 문화·미디어, 교통, 헬스케어, 유통, 통신, 중소기업, 산림, 지역경제

(센터) 중소기업·대학 등 주요 기관별로 데이터를 체계적으로 생산·관리



[참고] 빅데이터 플랫폼 및 센터 선정 현황

분야	플랫폼 주관	센터 구성	플랫폼 주요내용
통신	케이티	14개 (비씨카드, 인터넷진흥원, 인천테크노파크, 소상공인연합회, 경기대, 어메이징푸드솔루션, 두잉랩, 코난테크놀로지, 오픈메이트, 넥스트이지, 네스, 한국스마트그리드사업단, 제로투원파트너스, 더큰니움엠티엔)	- 유통인구, 상권, 카드소비, 관광, 교통카드정보, SNS 등의 데이터를 융합하여 상권분석 서비스 및 생활인구 분석 서비스 등 제공
헬스케어	국립암센터	10개 (삼성서울병원, 연대 세브란스병원, 건양대병원, 전북대병원, 대구가톨릭대병원, 서울대학병원, 분당서울대학병원, 화순전남대병원, 아주대학교병원, 길병원, 가톨릭대병원)	- 10대 암종별 임상데이터를 융합하여 암 진단·치료 의사 결정 및 항암 치료제 연구개발 등 활용
문화 미디어	문화정보원	11개 (국립중앙도서관, 국민체육진흥공단, 문화예술위원회, 아놀자, 청소년활동진흥원, 부산정보산업진흥원, 레드테이블, 레드타이, 원투씨엠, 데이터마케팅코리아, 티엔엠에스)	- 문화, 숙박, 레저, 음식, 상권, 도서출판 등의 데이터를 융합하여 문화여가 종합정보 서비스 및 한류 비즈니스 매칭서비스 등 제공
교통	교통연구원	10개 (울산시, 포항테크노파크, 아이나비, SKT, KT, KST플레이스, KCB, 성남시청, 진주시청, 대전시청)	- 실시간 교통량, 대중교통, 열차, 고속도로, 네비게이션, 블랙박스, 유동인구, 주차 등의 데이터를 융합하여 도로 및 대중교통 개선 서비스 및 스마트시티 지원 서비스 등 제공
환경	수자원공사	10개 (기상산업기술원, 국립생태원, 환경정책평가연구원, 지질자원연구원, GDS컨설팅그룹, 그린에코스, 아이렉스넷, 한국과학기술원, 순천향대학교, 노바코스, 데이터웨이)	- 물, 기상기후, 미세먼지, 지질재해, 생태자원, 화학물질, 환경SNS 등의 데이터를 융합하여 맞춤형 수질정보 서비스 및 대기질 야외활동 추천 서비스 등 제공
금융	비씨카드	10개 (노타, 닐슨코리아, 다음소프트, SBCN, 망고플레이트, 해빗팩토리, KT, 한국감정평가사협회, 기용정보통신, 한국금융솔루션, 한국신용정보원)	- 금융대출, 보험, 증권, 비금융(통신, 소셜, 유통, 미디어, 상권) 데이터를 융합하여 소상공인 창업 지도 서비스 및 국민 금융생활 플래너 서비스 등 제공
유통 소비	엠비엔	10개 (나이스디앤알, 다음소프트, 데이블, 로플랫, 빌트온, 식신, 온누리에이지엔씨, 지인플러스, KCB, 우편사업진흥원)	- 유통상품, 카드결제, 택배송장, 통신, 부동산, 상권, 물류, 맛집, 중고차시세, SNS 등의 데이터를 융합하여 라이프 스타일별 선호 외식업종 서비스 및 지역별 온라인 상품 구매정보 서비스 등 제공
중소기업	더존비즈온	10개 (빅밸류, 한화손해보험, 한국생산성본부, 한국무역정보통신, 와이즈넷, NICE평가정보, 산업기술진흥협회, 인크루트, 녹색기술센터, 선도소프트)	- 중소기업 회계정보, 부동산, 보험계약, 기업 고용복지후생, SNS 등이 데이터를 융합하여 기업 경영정보 분석 서비스 및 일자리 수요예측 서비스 등 제공
산림	임업진흥원	10개 (비글, 아로정보기술, 인포보스, 산림복지진흥원, 시선아이티, 우림엔알, 삼아항업, 우림인포텍, 마케팅크, 한국한의학연구원, 케이웨더)	- 임업, 등산로숲길자전거, 교통, 산악기상, 산림재해, 항공영상 등의 데이터를 융합하여 트래킹 서비스 및 산림재해 예측 서비스 등 제공
지역경제	경기도청	8개 (경기콘텐츠진흥원, 경기일자리재단, 더아이엠씨, 경기신용보증재단, 한국기업데이터, 국토연구원, 한양대학교, 한국생산기술연구원)	- 지역화폐 결제정보, 기업정보, 일자리, 신용평가, 카드사정보, 경기도 인구주거환경 등의 데이터를 융합하여 지역 소비패턴 분석 서비스 및 맞춤형 일자리 매칭 서비스 등 제공

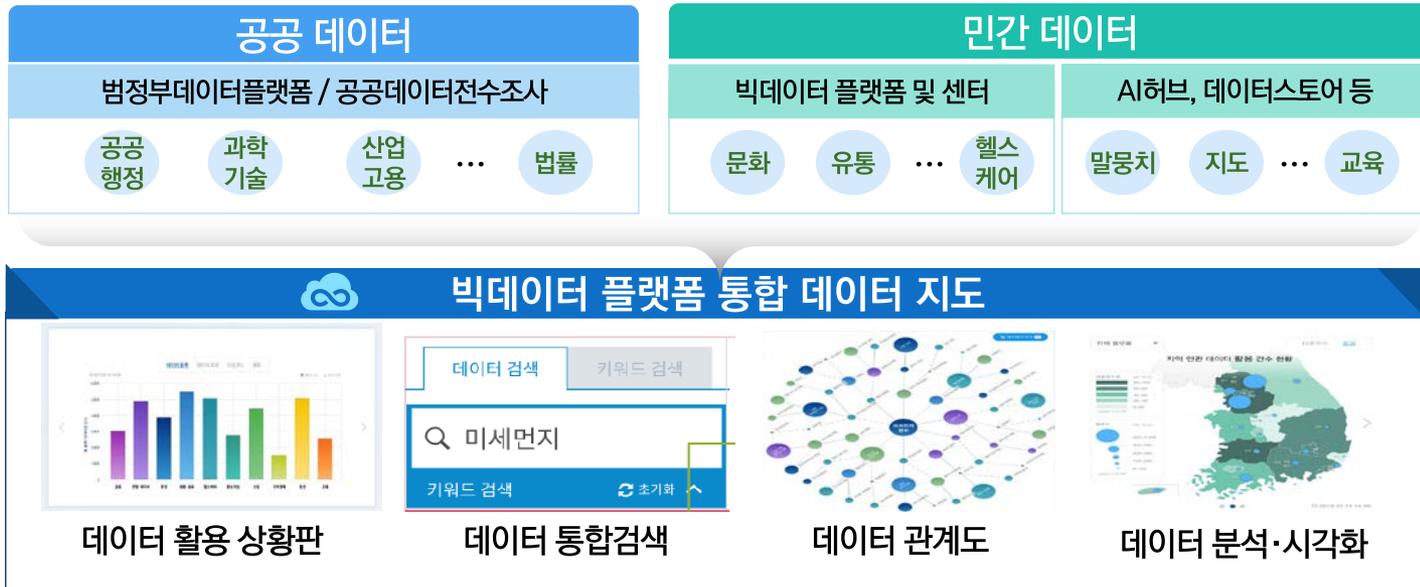
※ 총 기관(기업)수 : 198개

[참고] 플랫폼 데이터 개방 사례

분야	센터명	데이터 개방 사례
금융	기웅정보통신	관리비 납부여부, 거주일수 등 아파트 관련 데이터 공개 및 혁신서비스(비금융 신용등급 체계) 출시 예정
문화미디어	티엔엠에스	고가의 시청률 데이터, 광고집행 데이터 등을 스타트업 대상 무료 개방(15종)
산림	비글	SKT에 트래킹 통신 감도데이터 제공하여 산악지역 5G커버리지 확대
	인포보스	유전체 정보 개방을 통해 신약개발 소요기간 단축
통신	어메이징 푸드솔루션	삼성병원이 보유중인 데이터와 연계·융합하여 수술환자들을 위한 식이 데이터 생성·제공 (ex. 학술연구활용, 환자 영양식 제공 등)

국민과 기업이 데이터를 한 곳에서 쉽게 검색할 수 있도록 하는 공공·민간 통합 데이터 지도 구축

- ☑ 빅데이터 플랫폼 통합 데이터 지도 구축 ('20.3월)
- ☑ 민간 분야* 데이터 지도 구축('20.12월)
* 10개 플랫폼 및 100개 센터와 시허브, 데이터 스토어 등의 데이터
- ☑ 공공·민간 통합 데이터 지도 구축('21)



사업기본방향



- 이미지, 영상, 음성, 텍스트 등 AI개발에 공통적으로 활용 가능한 범용성 높은 학습용 데이터 우선 구축, 확산
- 중소벤처, 스타트업 등이 AI응용서비스 개발에 활용 가능한 기반 데이터셋 구축
- 학습용 데이터셋 구축에 활용할 수 있는 데이터 저작 도구의 공개 및 배포를 통한 데이터 확산 기반 마련
- 구축된 데이터셋을 활용한 AI 응용서비스 시범모델 개발과 AI 허브를 통한 데모 제공으로 데이터 활용 모델 제시
- 일자리 창출 및 AI 데이터 전문가 육성을 위해 지원 예산 클라우드소싱 기반 데이터 구축·개발 인력 투입

06 AI데이터 구축 사업

AI 제품·서비스 및 기술 개발에 활용가치가 높은 대규모 AI 학습용 데이터 구축 및 개방, 응용 개발



실서비스 제공
(API 개발)



데이터 가공 (전처리)

데이터 라벨링, 구분·선별, 포맷 변경, 결합, 변형 등

데이터 획득

IoT, mobile 등을 통해 획득한 이미지, 영상, 텍스트 등

클라우드 · 고성능컴퓨터
GPU

인공지능 서비스는 데이터에서 시작되어 모델을 생성하고 서비스 제공
개발 단계 : ① 데이터 획득, ② 데이터 가공, ③ AI 모델 생성, ④ AI 서비스

AI 핵심 기술 트렌드 분석으로 ‘범용 데이터 집중 구축’, ‘원천 데이터 확보 기반한 대용량 데이터 구축’ 등 전반적인 과제 방향성에 대한 근거 확보

AI 기술 트렌드 반영한 AI 데이터 구축 방향성

핵심 기술 트렌드

데이터 트렌드 특성

데이터 구축 방향성

AI 알고리즘 트렌드

Transfer Learning

Comprehensive Learning

Reinforcement Learning

GAN

대용량 범용 데이터

융합 데이터

중간 수준, 최소한의 라벨링

알고리즘 기반 데이터 생성

지능별 AI 기술 트렌드

언어 지능

시각 지능

음성 지능

융합 지능

분석 → 생성 중심

Still Image → 동영상

사람 행동 연구 활발

3D 영상 니즈 증가

음향 니즈 증가

1인 발화 → 대화 데이터

복합 타입의 데이터

1 범용 데이터 구축의 확대
(예: 한국형 OCR, 안면 이미지 추가 확대 등)

2 원천 데이터 확보로 대용량 데이터 구축 (예: 대용량 데이터, 한국인 음성 데이터 등)

3 기본적인 가공만 진행된 raw 데이터 위주 제공

4 동영상 중심의 Visual 데이터 구축, 3D 데이터 본격 구축
(예: 사람 행동 영상 확장, 3D 사물 이미지)

5 음향, 음성 데이터 본격 구축
(예: 도시 소리 데이터, 대화 음성 데이터 확장)

6 시각+Text 결합된 형태의 융합 데이터 구축 (예: VQA)

.....

AI 기술 및 제품 서비스 개발에 필수적인 AI데이터, AI알고리즘, 고성능 컴퓨팅 자원을 국내 중소벤처, 스타트업이 쉽게 활용할 수 있도록 제공하는 **원스톱 플랫폼**

AI데이터 구축·개방

- 시급성·유망성이 높은 분야 AI 학습용 데이터 구축·개방
- 글자체, 안면 이미지 등 10개 분야 선정

AI알고리즘

- 정부 AI R&D 결과물(AI모델) → AI 서비스용 open API로 제공
- AI HUB 내 AI API 사용 신청을 통해 상시 제공 운영 중

고성능 컴퓨팅 자원

- GPU 기반 고성능 컴퓨팅 자원 → 중소·벤처, 스타트업에 제공
- 자원 신청은 매년 초 접수를 받아 선정결과 별도 통보



법률, 특허 등 11종의 AI 학습용 데이터를 구축하여 '18.1월 4종, '19.1월 3종, '19.5월 4종 등을 AI HUB를 통해 단계적으로 개방



법률

교통사고, 퇴직금 등 생활관련 법령, 판례, 사례, 법률용어, 생활용어 등의 관계정보를 구조화한 지식 데이터



특허

국내 전기전자, 기계화학분야에 출원·등록된 특허의 특허정보, 심사정보, 특허기술용어 등의 관계정보를 구조화한 지식 데이터



한국형 이미지

다양한 각도와 환경에서 촬영한 한국인 400명의 얼굴 이미지와 100여종의 한국음식 이미지 데이터



일반 상식

한국어 위키백과에서 추출한 객체정보와 관련정보의 관계를 구조화한 지식 데이터



관광

국내 주요 관광지의 식당, 시설, 표지판 등을 촬영한 이미지와 다국어 POI 정보를 결합한 데이터



농업

국내 농업관련 영농정보, 상담정보, 지원사업정보 등의 관계를 구조화한 지식 데이터와 농작물 2종의 병충해 이미지 데이터



헬스케어

녹내장, 당뇨망막증 등 주요 안과질환의 안저 촬영 이미지와 전문의 소견을 결합한 데이터



한국어 음성

2~3명이 서로 자연스럽게 연속하여 발화하는 한국어 음성과 발화 내용을 전사한 데이터



한국어 대화

음식점, 소매점, 학원 등 소상공인 업종에서 자주 발생하는 상황과 질문·답변·의도 등을 구조화한 대화 데이터



멀티 모달

영상 내에 사람의 감정, 행동, 상황, 대화 내용 등의 멀티모달정보를 결합시킨 동영상 데이터



기계 독해

다양한 질문과 정답, 정답을 포함하고 있는 지문의 쌍으로 구성된 지능형 QA용 텍스트 데이터



한국어-영어 번역 말뭉치

한국어 인공지능 번역 기술개발 및 성능강화를 위한 한국어-영어 병렬 말뭉치 데이터 150만 문장



사람 동작 영상

사람의 동작·자세·행동 인식기술 개발을 위해 다양한 조건에서 사람 동작 영상 데이터 20만 클립



한국형 사물 이미지

한국형 객체, 장소, 상황 인지기술 개발 및 성능 강화를 위한 사물/거리 등 이미지 데이터 360만장



한국인 안면 이미지

여러 각도, 조도 등의 환경 하에서 안면인식·식별 성능강화를 위한 한국인 얼굴 데이터 1,800만장



한국어 글자체 이미지

한글 광학글자인식(OCR) 성능개선을 위한 한글 글자체(손글씨 및 인쇄체) 이미지 500만장



위해물품 엑스레이 이미지

위험물·도구 자동판별 기술개발 및 성능개선 위한 범죄 도구, 반입금지물품 등의 X-ray 이미지 44만장



인도 보행 영상

보행지원기술 개발을 위한 인도 보행 및 인도 위 객체(사람, 자전거, 펜스 등) 라벨링 데이터 200시간



질병진단 이미지

국내에서 유병률, 중증도 등이 높은 주요 질환 관련 진단 이미지(초음파 영상 등) 및 결과 데이터 3만장



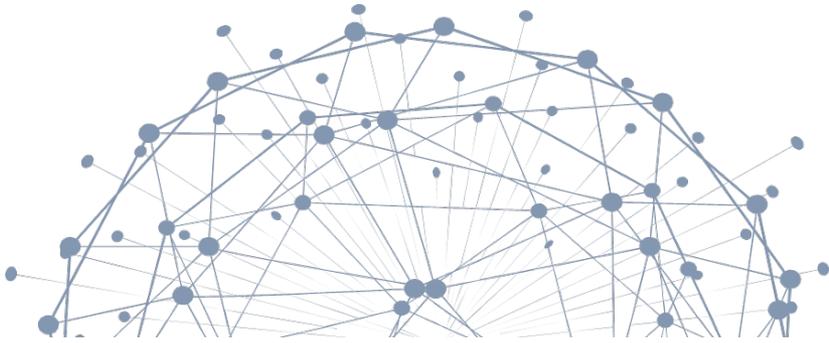
멀티모달(융합) 영상

영상내 상황과 연관된 인물, 객체, 감정, 상황정보 등 여러 형식 정보가 포함된 멀티모달 영상 데이터 100시간



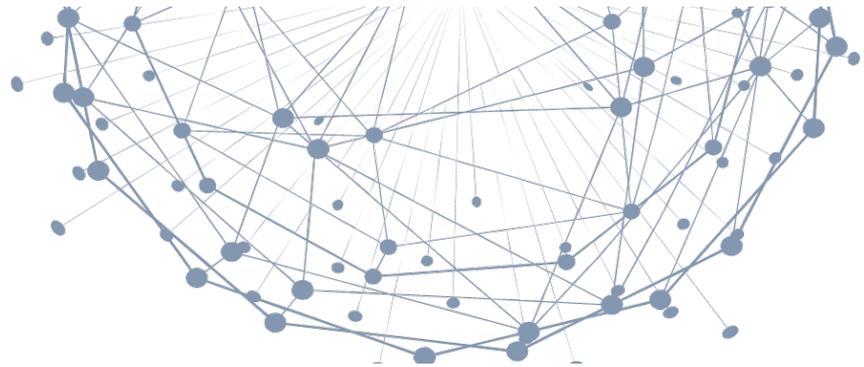
이상행동 CCTV 영상

주야간 적용가능한 이상행동 지능형 탐지기술개발 및 성능 강화를 위한 CCTV 영상 데이터 700시간

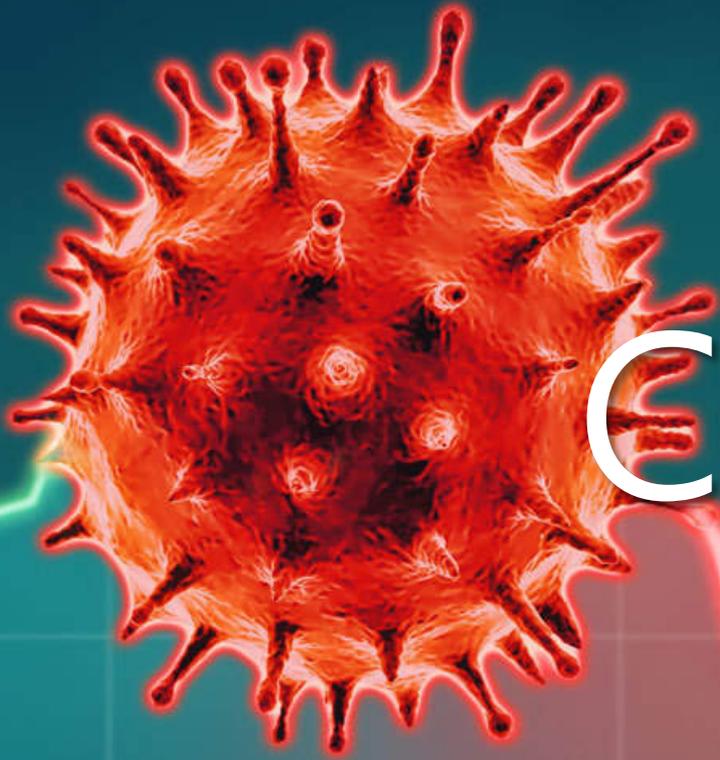


IV

새로운 도전



NIA 한국정보화진흥원



COVID 19

CORONAVIRUS

국가 최대 현안



경제회복

코로나19 경제위기 극복



일자리창출

4차산업혁명 미래 대응



미래사회대응

비대면 디지털 라이프

코로나19 위기 극복, 경제활성화·일자리 창출 한국판 뉴딜



디지털 인프라 구축(5G·데이터),
비대면 산업육성(의료·교육·유통),
스마트SOC(도시·산단·교통망·노후SOC 등)



온실가스 감축 등 기후변화에 대응
에너지·자원순환 경제 활용 녹색기술 개발
기후변화 대응 산업으로 신성장 동력 확보

한국판 뉴딜, 3년31조...일자리 55만개 창출(비상경제대책회의, 6.1)

비상경제 중양대책본부, 「한국판 뉴딜 추진방향」 발표('20.5.7.)

1. 디지털 인프라 구축

데이터 수집·활용
기반구축

- 데이터 쏠주기 인프라 강화
- 국민체감 핵심 6대 분야 데이터 수집·활용 확대

5G 등 네트워크
고도화

- 5G 인프라 조기 구축
- 5G+ 융복합 사업 촉진

AI 인프라 확충 및
융합확산

- AI 데이터·인프라 확충
- 쏠산업으로 AI 융합 확산

2. 비대면 산업 육성

- 비대면 서비스 확산 기반 조성
- 클라우드 및 사이버안전망 강화

3. SOC 디지털화

- 노후 국가기반시설 디지털화
- 디지털 물류서비스 체계 구축

대규모 데이터 뉴딜로 많은 일자리를 만들자

AI 학습용 데이터 전면 구축

범용성(이미지, 영상, 음성, 텍스트 등), 공공성, 한국 특수성을 반영한
학습용 데이터셋 구축, 확산
※ 누구나, 바로, 집에서 할 수 있는 클라우드소싱 방식으로 운영

5개 플랫폼 50개 센터 추가 구축

새롭게 5개분야 플랫폼 추가 구축 및 기존 플랫폼과 연계

공공데이터 개방, 품질관리 인턴십

중앙, 지방, 공공기관(730여개) 에 공공데이터 전면개방 확대 및
품질개선을 지원하는 인턴십

공공데이터 바우처 제공

중소기업, 스타트업 등과 행정, 공공기관의
공공데이터 구축, 가공 수요를 매칭하여 바우처 제공

감사합니다.

NIA 한국정보화진흥원

