

2021. 08

유망시장 Issue Report

자동인식 및 데이터 취득



목 차

I . 개요	01
II . 정책동향	06
III . 기술동향	16
IV . 시장동향	24
V . 산업동향	28

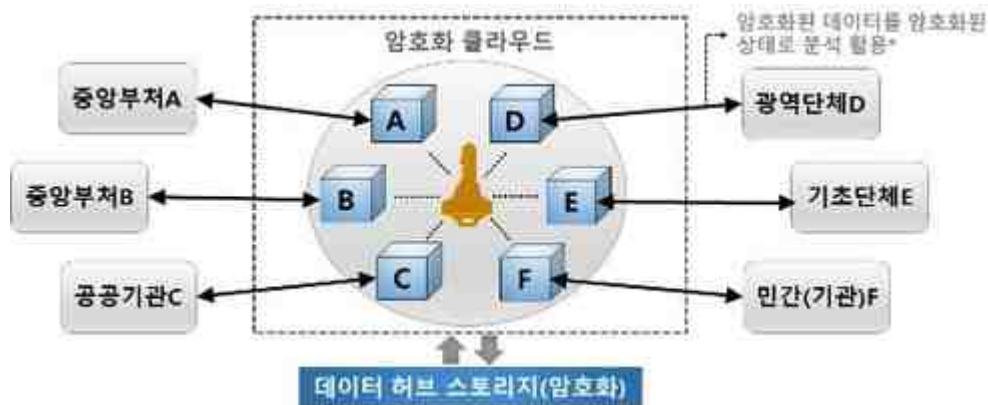
I. 개요

1. 아이템 개요

- 다양한 공공 데이터 수집을 통해 대량의 정형 또는 비정형 빅데이터를 분석함으로써 국민건강 예방 및 정책서비스 설계가 가능하도록 가치있는 공공기관의 정보를 추출할 수 있는 시스템
 - 공공분야의 데이터를 활용하여 국민의 건강, 안전, 복지, 생활 등 다양한 분야에서의 정책설계와 집행에 활용
 - 부처별, 공공기관별 산재되어 있는 공공데이터를 식별하고, 공공분야에 특화되어 있는 데이터를 통합 수집할 수 있는 기술
- 정책의제 설정, 정책설계, 집행의 과정에서 공공데이터의 활용가치가 증가하고 있으며, 공공 빅데이터를 수집하고 가공하여 분석하는 기술에 대한 업무 연관성이 증가
 - 부처별로 산재되어 있거나 중복되어 분야별 데이터의 교차 및 융합 분석이 가능하도록 정책 의사결정을 지원하는 시스템
 - 다양한 분야의 공공데이터를 통합 수집할 수 있어 다양한 정책문제를 확인하고 분석하는 가공 및 분석하는 기술
 - 보유한 공공데이터의 정형 및 비정형 등 데이터의 유형과 특징에 따라 다양한 빅데이터를 수집하고 분석결과에 대한 시각화가 가능한 기술
- 다양한 이해관계자와 정책의 영역이 분야별로 모호해지고 있으며, 협력적이고 융합적인 정책서비스 제공에 정책설계 다양화로 정책 의사결정을 위해서 분야별 공공데이터의 융복합 분석지원이 중요해짐
 - 이해관계자들이 다양해졌고 정책문제도 분야별로 얽혀 복잡해짐에 따라 다양한 분야의 공공데이터의 상관관계를 살펴 정책에 활용해야 함
 - 이종 간 분야의 데이터를 교차 및 상관성을 분석하기 위해서는 이종 간 분야의 공공데이터를 통합 및 일괄 수집해야 하고 분석을 시도할 수 있는 시스템이 필요함
 - 공공 빅데이터를 활용하여 새로운 서비스 및 공공정책에 활용할 기관과 민간 등 연구기관에서 활용 가능

- 데이터를 활용하는 이용자가 개별적으로 공공데이터를 체계적으로 분류하는 과정을 거쳐야 하므로 시간과 비용이 들고 신속한 서비스 개발이 어려움
 - 데이터가 전산화되어 있지 않고 수기로 작성되어 있거나, 정형화되어 있지 않고, 비정형으로 데이터를 생산하고 있는 사례가 많아 이를 정리하고 관리할 수 있어야 함
 - 공공데이터 포털은 일부 제한적인 데이터만을 적재하고 있는 반면에 중복되어 있거나 자료의 생성이 유사한 정도, 융합활용범위 및 추천 등을 통해 공공데이터 활용을 극대화할 수 있는 기능이 요구됨
 - 공공정책 카테고리에 따른 부서별 데이터 활용 정도와 생성현황 식별을 통해 공개할 수 있거나 검색 및 수집을 할 수 있는 데이터의 구별 등 공공데이터 전반에 대한 품질평가와 데이터 보정 등에 대한 정보 제공이 절실함
- 다양한 정책 수요를 가지고 있는 국민의 만족도 높은 증거기반의 정책 및 행정 서비스제공을 위한 기술, 사회, 경제 전반의 파급 및 연쇄효과가 기대됨
- AI 기반 법률 서비스를 제공하는 리걸테크(legal tech) 시장이 폭발적으로 성장할 것으로 추정되고 있음
- 형사소송에서 형량을 예측해주고 법률문서 작성이나 의미 기반 판례 검색 등의 서비스가 AI와 빅데이터 도움으로 가능해짐
 - 이를 위해서는 국내 AI법률 시장 성장을 위해 AI가 학습할 수 있는 재판 판례 등 법률 관련 데이터가 더 많아져야 함
- (기술적 파급효과) 산재하여 있으나 공유되고 있지 않은 부처와 기관별 공공데이터에 대한 검색과 수집을 통한 데이터 관리 기능
- 공공데이터가 필요한 민간, 공공, 연구기관 등 다양한 분야에 활용 가능한 데이터 보정, 중복 및 유사처리 기능
 - 검색을 통한 부처와 기관별 데이터의 융복합 분석이 가능한 활용 범위 및 데이터 형태 추천기능

- (경제적 파급효과) 부처별 기관별로 데이터를 수집하기 위해 투입되는 인력 및 시간을 공공데이터 검색과 수집기능을 통해 효율성을 달성할 수 있음
 - 데이터의 유사성 및 중복성 비교가 가능해짐에 따라 데이터 생성과 관리 측면에서의 조직 및 업무 재설계로 효율적인 조직운영이 가능해짐
 - 공공 빅데이터의 성격이 강하면서 쉽게 공개되기 어려웠던 정보들을 제공함에 따라 일부 독점적(카드데이터, 통신데이터 등)인 공공데이터를 대체하거나 보완할 수 있는 데이터를 제공하여 타 산업 및 데이터 경제를 활성화할 수 있음
- (사회적 파급효과) 다양한 사회문제, 공공의 행위 때문에 생성된(공공기관의 행정문서 및 IoT 데이터 등) 데이터를 활용할 수 있어 양질의 정책 및 공공서비스를 창출할 수 있을 것으로 기대
 - 다양한 공공데이터의 품질과 가치를 높일 수 있어 근거기반의 좋은 정책을 설계하는 데 활용하여, 사회문제를 해결하는 데 많은 도움을 줄 수 있을 것으로 기대



[그림] 공공 빅데이터 허브

※ 출처 : K-통계시스템 구축 추진 계획, 2021, 통계청

2. Value Chain

- (가치사슬) 공공데이터 활용을 필요로 하는 정부부처 및 기관과 민간기업 등에서 공공데이터 검색, 수집, 탐색, 분석, 시각화 결과를 제공하여 공공데이터의 활용과 가치를 극대화하는 시스템
- (후방산업) 부처와 기관별로 산재하여 있는 공공데이터를 탐색하고 수집할 수 있는 기술과 수집된 공공데이터에 대한 분석을 통해 활용 가치를 극대화할 수 있도록 인프라 및 소프트웨어 구축
 - 부처와 기관별로 관리되고 있는 데이터의 유사성과 중복성 그리고 품질 문제를 해결하기 위해 데이터 품질 기술이 연계되어야 함
 - 수집된 공공 빅데이터 활용을 위한 데이터 간 융복합 분석을 추천하거나 분석을 하여 결과를 시각화 해주는 기술 및 방법론의 개발이 필요
- (전방산업) 공공데이터를 활용하는 의료, 복지, 고용, 경제 정책과 민간에서의 시장조사와 마케팅 분석에 활용하는 시스템 및 플랫폼
 - 공공데이터를 통한 정책기획 및 설계, 시장조사와 상품기획, 마케팅 등에 활용될 수 있는 시스템 및 플랫폼 기술

[표] 공공 빅데이터 수집 및 분석 시스템 분야 산업구조

후방산업	공공 빅데이터 수집 및 분석 시스템	전방산업
데이터 수집기술, 데이터 품질 및 고도화 기술, 분석 및 시각화 기술, 데이터 관리, 데이터 컨설팅 등	부처와 기관별 공공데이터를 검색, 수집하여 분야별로 데이터 특성에 따른 분석방법론을 추천하고, 분석결과에 대한 시각화를 통해 정책 및 마케팅에 의사 결정해줄 수 있는 시스템	재난, 의료, 교통, 복지, 고용, 경제, 행정 등 근거기반 정책설계 및 분석, 민간분야 시장조사 및 기획과 마케팅 분석 시스템(플랫폼)

※ 출처 : 중소기업 전략기술로드맵 2021-2023(빅데이터), 2020, 중소벤처기업부

- (공공분야) 국민 생활과 안전 및 경제 등 행정과 정책서비스를 극대화하기 위해 필요한 공공데이터의 활용
 - 부처별로 산재하여 있는 중복 및 유사 데이터를 탐색하여 필요에 따라 활용 가능
 - 분야별 공공데이터를 탐색할 수 있어 시간 및 비용을 절감
 - 분야별 공공데이터를 탐색하고 수집 정리할 수 있어 관련 정책에 활용도가 높은 공공데이터를 분류하고 관리할 수 있음
 - 데이터의 유형과 특징에 따라 연계 또는 융합 분석할 수 있는 데이터인지 판단할 수 있도록 지원함
 - 분석방법 등을 지원하여 전통적인 분석방법으로는 알 수 없었던 통찰력과 상관성 높은 정책분석방법을 추천하는 기능
 - 공공데이터뿐만 아니라 다양한 외부 데이터들과도 연계 가능 여부를 체크할 수 있도록 지원
 - 공공데이터의 활용가치가 증가함에 따라 공공데이터를 활용한 행정 업무지원서비스 및 플랫폼 개발 활성화

- (민간분야) 소비자들의 니즈와 유형, 생활양식의 변화와 패턴을 확인하여 상품기획이나 마케팅을 극대화하는 데 필요한 공공데이터 활용
 - 기업 입장에서 고객 및 시장 창출을 위해 공공데이터를 활용하여 대중의 생활양식 및 패턴 등을 확인할 수 있고 신시장 창출과 발굴에 활용
 - 민간 등 기업이 가지고 있는 데이터와 공공데이터와 결합하거나 융합하여 새로운 통찰을 제공할 분석의 기회를 제공
 - 공공의 가치와 연계된 서비스 창출과 분석에 활용

II. 정책동향

1. 국내 정책동향

- 과학기술정보통신부는 데이터 활용을 통한 경제·사회 전반의 혁신 창출을 위해 3월 30일부터 데이터 바우처와 마이데이터 사업 등 ‘20년도 주요데이터 활용사업 공모를 추진

[표] 2020년 주요 데이터 활용 지원 사업 내용

사업명	예산규모	지원규모	지원내용
데이터 바우처	575억 원	1,420개 지원	중소기업, 소상공인, 1인창조기업 등에 데이터 구매 및 가공 지원
마이데이터 실증서비스	97억 원	8개 과제 지원	개인정보 보유 기업/기관과 컨소시움을 통해 다양한 실증서비스 개발 지원
데이터 플래그십	50억 원	10개 과제 지원	빅데이터·AI기반 혁신서비스 개발 지원
중소기업 빅데이터 분석 활용지원	7.5억 원	3개 지자체 지원 (지역별 25개 기업 지원)	중소기업의 빅데이터 활용 방법 및 데이터 수집, 가공, 분석 역량 등 컨설팅·교육

※ 출처 : 대한민국 정책브리핑 브리핑룸, 과학기술정보통신부

- 데이터 3법(개인정보 보호법, 정보통신망법, 신용정보법) 개정안은 2020년 8월부터 시행
 - 추가 정보 없이는 특정 개인을 알아볼 수 없도록 처리한 정보인 ‘가명정보’ 개념의 도입으로 추후 데이터 활용 및 분석에 유리
 - 다수의 부처에 데이터 관련 법들이 산재되어 생긴 중복 규제의 해결을 위한 법 제도 및 감독 기구 일원화하여 규제 완화
 - 데이터 보안 강화 및 개인정보처리 위반 처벌 강화를 통해 ‘활용 범위는 넓히되 처벌은 강하게’ 라는 방식 추진

- 정부는 2020년 7월 ‘한국판 뉴딜 종합계획’에서 10대 과제 중 하나로 ‘데이터 댐’ 발표
 - 분야별 빅데이터 플랫폼 확대, 공공데이터 14.2만 개 신속 개방, AI 학습용 데이터 1,300종 구축 등 빅데이터 확충 추진
 - 분야별 데이터 확충·연계·활용, 5G·AI 융합서비스를 통한 글로벌 선도로 2025년까지 데이터시장 규모는 43조 원으로 성장 예상



[그림] 데이터 댐 개념도

※ 출처 : 한국판 뉴딜 종합계획, 2020, 과학기술정보통신부

- 정보화진흥원(NIA)의 데이터 거래를 위한 10개 빅데이터 플랫폼 사업
 - “데이터를 가장 잘 다루는 나라” 라는 데이터 경제선언(18.8.31) 이후 데이터 고속도로 구축의 일환으로 민간과 공공에서 보유하고 있는 데이터를 오픈하고 유통을 통해 데이터 거래를 활성화시키기 위한 정책을 추진
 - 빅데이터 플랫폼과 센터를 통해 주요 분야별, 다양한 형태의 데이터 수집, 분석, 유통을 지원하고, 주요 기간별 데이터를 체계적으로 생산·관리하도록 추진
 - 빅데이터 플랫폼 및 센터 구축사업에는 금융, 환경, 문화미디어, 교통, 헬스케어, 유통, 통신, 중소기업, 지역경제, 산림 분야에 대한 주요 10개 분야를 선정하여 플랫폼 주관 기관과 데이터를 제공하는 센터 (7개~11개 데이터 기업)를 두고 있음

- 각 센터는 공공 및 민간기관에서 보유하고 있는 데이터를 공개하고 유통하고 판매하는 등 데이터 거래를 통해 빅데이터 생태계를 조성하고, 신규사업 창출, 창업 등 데이터 경제를 활성화할 수 있도록 지원하는 데이터 개방과 혁신을 추진하고 있음
- 과학기술정보통신부는 ‘빅데이터 플랫폼 및 센터 구축 사업’을 위해 2019년 6월 10개 분야 플랫폼 및 100개 전문센터를 선정. 향후 3년 간 1,516억 원을 투입할 계획을 발표하였으며, 2019년에는 총 640억 원 규모로 사업 추진 중

○ 데이터산업진흥원의 데이터 거래를 위한 바우처 사업

- 데이터산업진흥원에서는 데이터 바우처 제도를 통해 데이터 공급 기업과 이에 대한 데이터 수요기업을 매칭하여 공급기업이 보유하고 있는 데이터를 필요에 따라 거래할 수 있도록 정책적으로 지원
- 통신 대기업 등 199곳이 공급기업으로 참여하고 Kdata의 데이터 스토어에 등록하면, 상권분석, 산업별 특허정보 등을 제공하며, 데이터 수집, 분석, 가공 등의 중개서비스를 진행하고, 데이터산업진흥원을 통해 비용을 청구하는 방식
- 중소·벤처기업 등 200개사를 선정하여 데이터 구매 부문과 일반 가공부문, AI가공 부문에 대하여 원하는 데이터 공급자를 검색하여 각 데이터 구매 형태에 따라 1,800만 원에서부터 최대 7,000만 원까지 지원을 받아 바우처를 사용할 수 있음
- 이밖에도 데이터 거래 활성화에 필수인 개인정보보호법 개정 등 비식별조치, 가명 및 익명 정보 개념을 법제화하여 제도적인 활성화 지원이 기대되며, 이를 통해 중국 등 주변국의 빅데이터 시장과 경쟁 우위를 확보하여 한국형 데이터거래소를 성장시키는 구상

○ 데이터 산업 활성화 전략: I-Korea 4.0 데이터 분야 계획 I-Data

- 4차 산업혁명 시대 데이터가 모든 산업의 발전과 새로운 가치 창출의 촉매 역할을 하는 ‘데이터 경제(Data Economy)’로의 패러다임 전환에 입각하여 데이터 이용제도 패러다임 전환, 데이터 가치사슬 전방위 확산, 글로벌 데이터산업 육성기반 조성 등을 추진과제로 발표

- 공공데이터의 원칙적 개방(국가안보·개인정보 제외) 등을 통해 공공 부문 핵심 데이터 (통계·연구·참조표준 등) 구축 개방 전면화
 - 데이터의 효율적인 저장 관리를 위한 클라우드 본격 확산 및 데이터 유통 촉진을 위해 민간 공공을 연계한 개방형 데이터 기반을 구축하고 '22년까지 중소 벤처기업 1만 개 이상에 클라우드 도입 지원
 - 빅데이터 산업이 4차 산업혁명을 선도하는 핵심 성장주체로 거듭날 수 있도록 4차 산업혁명 요소기술 융합, 분석 전문인력 양성, 컴퓨팅과워 기반 기업성장 인프라 지원 등을 아우르는 역동적 산업 생태계 조성
 - 빅데이터 전문기업 성장 지원으로 데이터 강소기업 100개를 육성하고 빅데이터 개발에 필수적인 컴퓨팅 자원, 데이터셋 등이 집적된 판교 글로벌 ICT 혁신 클러스터 지속 고도화 및 활용 지원
 - K-ICT 빅데이터센터, K-ICT 클라우드 혁신센터, 글로벌 IoT 시험인증 센터, 정보보호 클러스터, HPC 이노베이션 허브 등 세계 최고 수준의 성능을 갖춘 초고성능 컴퓨터 5호기와도 연계
- 대통령 직속 4차 산업혁명 위원회에서는 데이터 산업 활성화 전략을 발표하여 개인정보 보호와 활용의 균형 및 데이터 경제 활성화 추진
- 정보주체인 개인이 데이터 통제·활용권을 가지게 하고 개인정보의 안전한 활용을 위한 법제화·안심존 구축·신기술 적용 등을 지원하며, 데이터 구축·개방, 저장·유통, 분석·활용의 빅데이터 활용 전 과정의 지원 체계 마련
 - 데이터 주권을 강화하여 개인정보가 해외에서도 안전하게 보호 되도록 국외 이전 중단 명령권, 국외 재이전 시 보호조치 의무 신설 등 제도 정비
 - 데이터 가치 사슬 전주기 혁신을 위하여 데이터 구축 개방(4차 산업혁명 핵심데이터 등, ~'22) → 저장 유통(데이터 거래의 비즈니스화, '18~) → 분석 활용(산업·사회 혁신 활용, '18~) 등 전 과정에 걸쳐 실제 데이터 기반 영역별(의료·교통 등) 국가 빅데이터 지원체계 마련
 - 빅데이터 핵심기술 확보 및 전문인력 확보를 통한 기반 조성 목적
 - 국회 4차 산업혁명특별위원회는 이와 관련하여, 개인정보보호법과 정보통신망법, 신용정보법 등 중복 조항 정비 및 강력한 사후규제를 도입하여 가공에 대한 활용의 안전성을 보장하도록 입법 권고안 발표

2. 해외 정책동향

가. 미국

- 미국 연방 정부는 '19년 1월 정부의 정책 평가에 필요한 증거(evidence)를 '통계'로 규정하는 증거기반정책법(Foundations for Evidence-Based Policymaking Act of 2018) 제정
 - 증거기반정책법은 연방정부의 증거구축활동을 규정하고(제1장) 연방정부가 생산하는 공공데이터의 활용에 초점을 맞춘 OPEN(Open, Public, Electronic, and Necessary) 정부데이터법(제2장)과 비밀정보 보호와 통계적 효율성(제3장)으로 구성하여 데이터와 통계의 유기적인 연계를 강조
 - 10대 원칙은 윤리적 거버넌스, 의식적 설계 및 학습 문화라는 세 범주로 구성되고, 40대 실천 전략은 10대 원칙 실현을 위한 5~10년 간의 중장기 전략
 - 데이터의 가치를 육성하고 공공 부문에의 활용을 촉진하기 위한 실천 전략, 데이터의 관리와 보호를 위한 실천 전략을 효율적이고 적절한 데이터 활용 촉진을 위한 실천 전략으로 구성
- 트럼프 행정부의 주요 정책 중 하나로 추진되던 '클린 네트워크'로 인하여 화웨이와 ZTE 등의 중국 기업을 배제하던 양상에서 불거진 미·중 간의 무역 분쟁이 존재
 - 클린 네트워크는 클린 패스(Clean Path), 클린 캐리어(Clean Carrier), 클린 스토어(Clean Store), 클린 앱(Clean Apps), 클린 클라우드(Clean Cloud), 클린 케이블(Clean Cable)로 구성
 - 다만, 바이든 행정부가 당선되며 클린 네트워크 정책은 앞으로 둔화될 것으로 전망
- 대통령 직속기구인 국가과학기술위원회(National Science and Technology Council: NSTC)에서 연방정부 부처 기관의 ICT 연구개발 프레임워크인 NITRD 프로그램을 운영

- 2018년 빅데이터, 로봇, 인공지능, 고성능 컴퓨팅 시스템 인프라 운용으로 영역을 확대하였으며 2019년 사이버보안을 비롯한 개인정보 등으로 확대 변경하여 시행 중
 - 빅데이터와 인공지능 분야를 강화하기 위해 투자 확대 및 프로그램을 신설하였고, 실제로 예산은 인프라 및 소프트웨어에 비해 증가
- 2009년도부터 시작된 오픈데이터 정책의 대표적 서비스인 Data.gov는 트럼프 정부에 들어서도 계속 진행
- 2019년 6월 말 기준 Data.gov에는 251,906개의 데이터 세트 등재되어 누구나 이 데이터를 활용하여 가치를 제고할 수 있도록 소프트웨어 어플리케이션 함께 제공
- 빅데이터 R&D전략 계획(BDIWG, '16.5)
- 범부처 차원에서 빅데이터 7대 R&D 전략과 18개 세부과제를 제시하여 차세대 빅데이터 기술 역량 확보, 신뢰성 탐구 및 획기적인 발견을 위한 빅데이터 R&D 지원, 빅데이터 혁신이 가능한 사이버 인프라 연구 향상 등이 7대 전략으로 선정

나. 유럽

- 개인정보보호 규정(GDPR, General Data Protection Regulation)
- 2018년부터 시행되어 유럽 의회에서 유럽 시민들의 개인정보 보호를 강화하기 위해 만든 통합 규정
- 공통 유럽 데이터 공간
- 2017년 1월 채택한 유럽 데이터 경제 구축의 후속조치로 서로 다른 부문, 국가 및 분야의 데이터 통합을 제안하고 있으며, 과학적 정보의 접근 및 보존에 관한 2012년 권고안을 개정하고 있음
 - B2B, B2G 환경에서의 민간 부문 데이터 공유에 대한 지침 등을 포함하고 있으며 2018년 5월 일반정보보호규정이 시행됨에 따라 데이터 관리와 개인정보보호를 위한 국제적 프로토콜이 될 가능성이 있어 국내에도 시사하는 바가 큼

- 유럽 47개 주요 교통, 물류, IT 관련기업 및 기관으로 구성된 컨소시엄이 30개월간 1,870억 유로의 예산으로 빅데이터 기반의 Transforming Transport 프로젝트 추진
- 유럽연합은 '15년부터 빅데이터에 대해 향후 5년간 산업 컨소시엄의 20억 유로와 공적 자금 5억을 합친 25억 유로(약 3조 3,870억 원) 투자
 - 아토스(Atos), IBM, 노키아 솔루션 앤 네트워크스(Nokia Solution and Networks), Microsoft(N), 오렌지(Orange Telecom), SAP, Siemens 등의 기업들과 연구기관들 주도로 민관협력 추진 중

다. 영국

- 데이터의 경제적 가치 보고서
 - 영국 재무부에서 발표한 보고서에는 데이터를 중심으로 한 영국 경제 발전의 5가지 과제 발표
 - 1. 데이터 소유권과 통제권에 대한 명확화 2. 개인데이터에 대한 보호 유지 3. 공공부문 데이터의 개방성 4. 데이터 상호운용 및 표준 강화 5. 안전하고 합법적인 데이터 공유 환경 구축
- 빅데이터 역량강화 종합전략 및 정부 변화전략
 - 공공서비스를 위한 정부기관 간의 경계를 허물어 데이터를 효과적으로 사용할 수 있도록 하는 디지털 경제 법안이 제출되었으며 통과되고 나면, 데이터 공유 조항을 통해 전체 정부의 데이터를 효과적으로 사용할 수 있도록 함
 - 정부차원에서 데이터 과학과 분석능력 구축, 데이터 공유 관련 윤리 교육, 국가 데이터 인프라 구축 및 보완성 확보, 데이터 검색 도구 개선, 주요 리포지토리 저장 및 관리 방식 변화 등을 추진함

라. 중국

○ ‘20년 9월 8일 데이터 안보에 관한 글로벌 이니셔티브를 발표

- 미국 정부가 중국산 소셜네트워크(SNS) 틱톡·위챗 사용을 금지한 조치에 대해서 대응하기 위해 발표한 글로벌 데이터 안보 표준안으로 데이터의 국적성을 강조
- 기업은 사업하는 국가의 법률을 존중하고, 기업이 해외에서 취득한 데이터를 그 국가(해외국가)에만 저장하도록 강요해서는 안 됨(4항)
- 다른 국가의 주권, 관할권 및 데이터 관리 권한을 존중하고 타국 데이터에 접근해선 안 됨(5항)

○ 교통관광서비스 빅데이터 활용 시범사업 추진 통지

- 교통, 관광서비스에 빅데이터 활용을 위한 플랫폼 구축 추진

○ 스마트시티 관련 빅데이터 및 클라우드 활용 요강

- 각 지역의 스마트시티 건설 시 빅데이터와 클라우드를 활용한 세부 방안을 발표

○ 빅데이터 산업 발전 계획(‘16~‘20)

- 이미 일상생활에서 광범위하게 활용되는 빅데이터를 제조업 분야로 확대하며, 특히 생산모델 고도화 및 스마트 공장을 추진하는 과정에 빅데이터 응용 필요성 강조
- 빅데이터를 중국의 인구·경제 규모를 활용한 전 산업 발전 기회로 인식하고 10개 이상 글로벌 빅데이터 선도기업, 500개 응용서비스 기업 육성 목표 추진
- ‘20년까지 빅데이터 관련 제품 및 서비스업 매출을 1조 위안까지 확대하겠다는 목표 제시
- 데이터 개방 확대, 플랫폼·오픈소스 기술 지원, 빅데이터 전문 SW 수준 향상, 전문인재의 공급, 데이터거래소 등 생태계 조성

- 정확한 고객 수요 파악과 분석을 통해 R&D의 효율성을 높이는 한편 조립·생산 라인의 경쟁력을 높임으로써 생산예측과 관리를 강화하고 경영 효율화를 제고할 수 있을 것으로 기대

○ 빅데이터의 응용 기술 표준 및 통계 표준 제정

- 통계국은 국가통계국과 텐센트 등 6개 기업 공동으로 연구를 진행하여 정부 통계 데이터를 이용하여 빅데이터 기반 공동 수집·처리·분석·탐색·발표 기술 개발 표준과 통계 표준을 제정

마. 일본

○ 일본 정부는 2018년 6월 ‘데이터 활용의 기반 조성을 위한 관민 데이터 활용 추진 기본 계획’ 수립

- 일본은 5가지 사항을 담은 기본계획을 토대로 세계적인 데이터 활용 국가로 거듭나기 위한 의지 피력

○ 정부 주도로 정부, 지자체, 민간 보유의 모든 데이터를 정리하여 개방할 계획 발표

- ‘20년에 포털 사이트의 형태로 공개할 예정으로, 취합한 빅데이터 자료들을 모두 데이터베이스화 한 후 일반에게 공개하여 다방면에서 인공지능을 활용한 분석 자료로 사용할 예정
- 국가가 보유한 각종 통계데이터, 지도데이터, 기상데이터와 민간에서 보유한 도로 상황 및 교통량과 관련된 데이터, 작물의 재배 상황 정보, 위치정보 등 광범위한 분야의 정보 취합이 목표
- 인공지능 분석이 용이하도록 행정용어에서부터 산업용어까지 빅데이터 내에서의 용어를 통일·정리하여 취합할 예정

1) 1. 관민 데이터 활용의 추진에 관한 기본적인 방침 2. 행정기관의 관민 데이터 활용에 관한 사항 3. 지방공공단체와 기업의 관민 데이터 활용 촉진에 관한 사항 4. 관민 데이터 활용과 관련해 정부가 중점적으로 강구해야 할 시책 5. 관민데이터 활용의 추진에 관한 시책을 종합적이고 동시에 효과적으로 추진하기 위해 필요한 사항

- G20 정상회의에서 데이터베이스 구축에 미국과 유럽의 참가를 요청해 10억 명 이상의 정보가 집적된 초국가적인 데이터베이스 Pool 구축을 최종 목표로 함
- ‘14년부터 추진한 공공기관의 데이터 공개를 통한 정부의 투명성, 경제 활성화, 공공서비스 향상 목적의 ‘빅데이터 활용을 통한 Open Data 정책’의 연장선으로 파악

○ 미래투자전략 2017-Society 5.0의 실현을 위한 개혁

- 빅데이터를 ICT 융합 활성화 촉매로 활용하기 위해 이노베이션·벤처를 탄생시키는 선순환 시스템을 구축하여 빅데이터 산업 집중 투자를 추진
- 건강수명연장, 이동혁명 실현, 공급망 첨단화, 쾌적한 도시 만들기, 핀테크의 5대 신성장 전략 분야에 데이터를 활용한 기반구축
- 새로운 사회의 인프라로서 ‘데이터 기반(현실데이터 플랫폼)’을 구축하고 데이터 활용을 향한 제도 정비, 교육·인재 역량 강화, 혁신벤처 선순환 시스템 구축
- ‘20년까지 국민 데이터 활용추진 전략회의를 중심으로 공공데이터 개방 집중 실시, 데이터 활용 권한을 명확히 하는 기업 간 계약 가이드라인 전개, 데이터 유통을 위한 법제 정비 등 데이터 거래시장의 제도 정비 등을 실시

III. 기술동향

1. 기술범위 및 특징

- 빅데이터는 정형, 비정형을 포함한 다양한 종류의 대규모 데이터로부터 효율적으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석을 지원 하도록 고안된 차세대 기술 및 아키텍처로 이전에는 거대한 데이터 집합 자체만을 의미하였으나, 점차 그 범위가 확대되어, 도구, 플랫폼, 분석 기법 등을 포함한 포괄적 의미로 확대
 - 빅데이터에 대한 정의는 다양하게 설정되고 있으나, 일반적으로 양(Volume), 속도(Velocity), 다양성(Variety), 가치(Value)를 공통된 정의로 규정
- 빅데이터 분석 솔루션은 대용량의 정형 혹은 비정형 데이터로부터 숨겨진 패턴과 알려지지 않은 정보 간의 관계를 찾아내어, 비즈니스 의사결정을 지원할 수 있는 인사이트를 발굴하고 예측하는 소프트웨어 또는 하드웨어를 의미
 - 산업의 투명성 증대, 소비자 니즈 발견, 트렌드 예측, 성과향상, 고객 세분화, 의사결정 지원과 대행, 비즈니스 모델 및 상품/서비스의 혁신 등을 위하여 사용



[그림] 빅데이터 개념도

※ 출처 : 혁신성장동력 시행계획, 2018, 관계부처 합동

- 2018년 데이터 산업 현황 조사 보고서에 따르면, 데이터 산업은 데이터 솔루션, 데이터 구축 및 컨설팅, 데이터 서비스로 구분

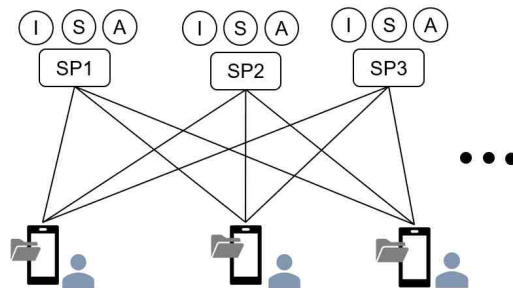
[표] 빅데이터 기술 분류

대분류	중분류	세부 기술
데이터 솔루션	데이터 수집	데이터 검색, 로그데이터 수집, 웹데이터 수집, 데이터통합/연계, 데이터 교환/개방
	DBMS	RDBMS, NoSQL DBMS, 인메모리 DBMS, 기타 DBMS
	데이터 분석	정형/비정형 데이터 분석, 실시간 데이터 분석, 데이터 시각화 분석, 데이터 (전)처리
	데이터 관리	데이터 모델링, 마스터 데이터 관리, 데이터 품질 관리, DB 운영 및 성능 관리
	데이터 보안	DB 보안, 개인 데이터 보안
	데이터 플랫폼	빅데이터 플랫폼 개발 및 공급
데이터 구축 및 컨설팅	데이터 구축	DB 설계/구축, 데이터 이행, 데이터 구축/가공
	데이터 컨설팅	데이터 설계 컨설팅, 데이터 품질 컨설팅, DB 성능 개선 컨설팅, 데이터 거버넌스 컨설팅, 데이터 분석/활용 컨설팅
데이터 서비스	데이터 판매 및 중개	데이터 판매, 데이터 신디케이션
	정보 제공	포털/정보 매개 서비스, 정보 제공 서비스
	데이터 분석 제공	소셜 데이터 분석 정보 제공, 마케팅 데이터 분석 정보 제공, 리스크 데이터 분석 정보 제공, 게타 데이터 분석 정보 제공

※ 출처 : 2018 데이터 산업 현황 조사, 2019, 한국데이터산업진흥원

2. 국내/외 기술 Trend

- 정부 차원에서의 법적인 대응과 더불어 기술적으로 인터넷 데이터 집중화 문제를 해결하고자 하는 연구들이 미국, 유럽 등을 중심으로 추진
 - 기술적인 관점에서 인터넷 서비스 기업에 의한 데이터 독점 현상은 기본적으로 인터넷 서비스의 가장 기본이 되는 웹 기반 서비스의 Client-Server 구조에 기인
 - 이 구조 하에서 수많은 사용자(Clients)들은 서비스의 무료 사용의 대가로 자신의 개인 정보를 인터넷 서비스 기업(Server)에게 제공하며, 인터넷 서비스 기업은 사용자의 식별자, 데이터 등 모든 정보를 독점적으로 관리
- 각 사용자는 각각의 서비스 제공자들(Service Providers: SPs)로부터 중앙집중 형태로 타 기업에 배타적인 서비스를 제공받음
 - 각 서비스 제공자는 서비스의 기본 요소인 사용자들에 대한 식별(I: Identification), 데이터 저장(S: Storage), 서비스 응용(A: Application) 기능을 모두 개별적으로 수행
 - 사용자의 모든 정보는 사용자의 의사와는 무관하게 서비스 제공자의 서버에 독점적으로 저장 및 관리되며, 따라서 서비스 제공자들 간의 사용자 데이터 공유도 어려운 구조
 - 이러한 특정 사업자의 사용자 데이터 독점 현상을 기술적으로 해결하기 위해서는 현재 인터넷 서비스 제공자의 서버에 독점적으로 집중되어 있는 식별, 저장, 응용 관련 데이터 관리를 어떻게 사용자의 주권과 프라이버시 보장이 가능하도록 구조적으로 바꿀 것인가가 핵심 이슈



[그림] 클라이언트-서버 기반의 인터넷 서비스 구조

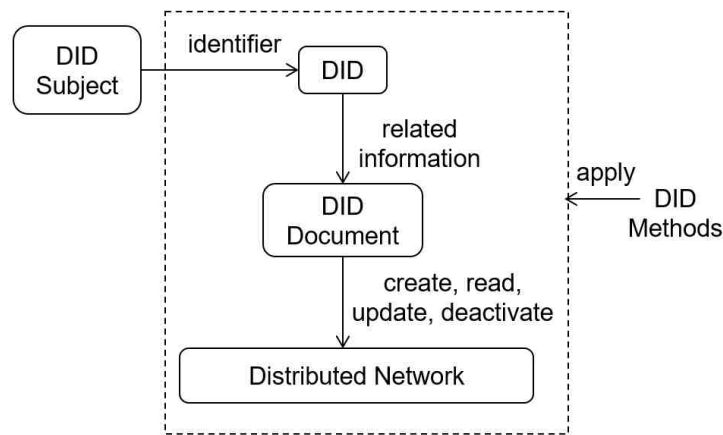
※ 출처 : 데이터기반 사회에서 데이터 주권 이슈와 대응기술 동향, 2020, 한국전자통신연구원

- 데이터 독점에 대한 세계적인 문제 인식 확산에 따라 최근 이에 대한 기술적 대응 방안 논의와 탈중앙형(Decentralized) 구조를 기반으로 하는 새로운 기술 제안들이 전 세계적으로 활발히 이루어지고 있으며 관련 스타트업 기업들도 다수 생겨나고 있음
 - 이 중 가장 주목을 받고 있는 대표적인 기술로는 식별 체계 관점에서 사용자 주권 보장을 위한 DIDs(Decentralized Identifiers)와 서비스 제공자로부터 데이터 저장을 독립시키기 위한 Solid(Social Linked Data)가 있음
 - 특히, 두 기술은 모두 웹 표준을 담당하는 W3C(World Wide Web Consortium)에서 진행 중인 표준화 작업과 밀접하게 관련되어 개발되고 있다는 점에서 다른 기술들에 비해 더 유력한 기술로 간주됨

가. DIDs

- 식별(Identification)은 물리적 세계에서 주민등록 번호나 여권과 같은 역할을 사이버 공간에서 제공하며, 사용자가 인터넷 상에서의 자신의 데이터를 제어할 수 있는 가장 기반
 - 현재 인터넷 서비스에서 사용자의 식별은 각 인터넷 서비스 제공자에 의해서 배타적으로 주어지며, 이는 사용자의 모든 서비스와 데이터가 특정 서비스 제공자에 속박되는 원인
 - DIDs는 이와 같이 다수의 사용자가 특정 인터넷 서비스 제공자에 의해 속박되는 현재의 구조를 탈피하여 서비스 제공자에 독립적인 사용자 ID를 기반으로 다양한 서비스를 제공받을 수 있도록 해주는 탈중앙형 식별 체계
- DIDs는 검증 가능하며 분산형 디지털 식별을 제공하는 새로운 형태의 식별 체계로 정의되며 어떠한 중앙집중형 등록소(Registry), 식별제공자(Identity Providers) 또는 인증기관(Certificate Authority)과도 독립적으로 제어, 구현될 수 있도록 하는 것을 목표로 함

- 따라서 기존의 중앙집중형 방식에 비해 특정 기관에 대한 의존이나 단일 고장점(Single Point of Failure)에 대한 문제없이 인터넷 서비스를 위한 신뢰기반을 구축할 수 있다는 장점을 가짐
- DID는 식별하는 대상인 DID Subject, 식별자인 DID(Decentralized Identifier), DID에 관련된 정보를 포함하는 DID Document, DID Document가 저장 관리되는 분산 네트워크(Distributed Network) 및 DID와 DID Document가 처리되는 방식을 규정하는 DID Method로 구성



[그림] DIDs의 구성

※ 출처 : 데이터기반 사회에서 데이터 주권 이슈와 대응기술 동향, 2020, 한국전자통신연구원

- 먼저, DID Subject는 DID에 의해서 식별되고 DID Document에 의해서 기술되는 엔티티(Entity)로 사람, 기관, 사물 등이 될 수 있음
- DID는 DID Subject와의 신뢰성 있는 상호작용이 가능하도록 DID Subject와 DID Document를 관계 지어주는 URL 형태의 ID(Identifier)로 DID의 형식은 세 부분²⁾으로 이루어진 문자열로 구성
- 문자열의 첫 부분은 ID가 DIDs 기반의 ID임을 나타내며, 두 번째 부분은 DIDs에 사용된 DID Method의 이름이며, 마지막 부분은 DID Method에서 규정하는 형식에 따른 DID Subject의 ID를 표시
- 예로서 DID Method로 IPFS가 사용되고 IPFS ID가 “123456abcdef”로 주어지는 경우 DID는 “did:IPFS:123456abcdef”와 같이 표현됨

2) ① URL scheme ID(did), ② ID for DID Method, ③ DID Method-specific ID

- DID는 적절한 해석 절차에 따라 DID Document로 매핑
- DID Document는 DID를 어떻게 사용할 것인가를 기술하는 간단한 문서로 해당 DID와 관련된 다음과 같은 정보를 포함
 - 암호 기반으로 해당 엔티티를 인증하기 위한 공개키와 같은 정보와 증명 방법
 - 엔티티와 상호작용하기 위해 사용되는 서비스들
- 분산 네트워크는 실제로 DID Document가 저장 관리되는 장소로서 블록체인과 같은 분산원장기술(Distributed Ledger Technology: DLT)이나 DHT(Distributed Hash Table)와 같은 분산 네트워크 기술 등으로 구현
 - DID Method는 특정한 DIDs 기술이 분산 네트워크 상에서 어떻게 실제 구현되는지를 정의
 - DID Method는 어떻게 DID가 해석되고 비활성화되며, 어떻게 DID Document가 작성되고 갱신되는지에 대한 구체적인 방법을 포함하며 현재까지 40여 개의 DID Method가 정의
- 웹은 현재 대부분 인터넷 서비스의 기반이 되는 플랫폼 기술
 - 팀버너스리가 최초로 개발한 웹은 누구나 자신의 홈페이지를 가지며, 누구나 자신의 생각을 게시할 수 있으며, 자신의 데이터를 자신이 소유하며, 각 노드들이 링크로 연결되는 분산 네트워크로 설계
 - 그러나 전술한 바와 같이 현재 웹 기반 서비스는 소수의 기업들에 의해서 독점되고 있는 상황
 - 팀버너스리는 자신이 설계한 웹이 원래의 목표와는 달리 현재 불평등과 분열의 엔진으로 작용하고 있다고 인정하고 자신이 최초 목표로 했던 탈중앙형 웹 구축을 위해 Solid라는 새로운 응용 플랫폼을 제안
 - Solid는 오픈소스 프로젝트의 일환으로 웹의 최초 목표인 모든 사람이 자신의 데이터에 대한 완전한 제어권을 확보할 수 있도록 보장
 - 즉, 어떤 사용자나 그룹이 그 데이터에 접근 가능하며, 어떤 응용을 사용할 것인지에 대한 선택권을 구조적으로 사용자에게 제공

- Solid는 기본적으로 팀버너스리가 제안한 링크드 데이터(Linked Data) 원칙에 기반하며, 사용자가 소셜 미디어를 위한 데이터 소유권을 갖도록 권한을 부여
 - 이를 위해 사용자 데이터를 애플리케이션으로부터 분리하고, 데이터의 완전한 제어를 소유자에게 분산된 방식으로 제공하도록 설계
 - 사용자 데이터를 애플리케이션으로부터 분리하기 위해 Solid는 Pod(Personal Online Datastore)의 개념을 도입
 - 사용자의 Pod는 사용자의 모든 정보를 독립적으로 저장하며, 다양한 형태로 구현될 수 있음
 - Pod에 있는 정보에 접근하기 위해서는 모든 사용자는 URI 형태의 Web ID 기반으로 인증을 받아야 함
- 사용자의 데이터 제어를 위해 Solid는 데이터 소유자가 지정한 인증 정책을 기반으로 Pod에서 데이터에 대한 액세스 제어를 할 수 있도록 지원
 - 즉, Pod 내 정보에 대한 요청에 대한 응답은 사용자가 작성한 ACL(Access Control List) 파일에 기반하여 분산 액세스 제어 메커니즘인 WAC(Web Access Control)에 의해 제어
- Solid 플랫폼에서 사용자는 자신의 데이터와 프로파일을 자신의 Pod에 저장함
 - Pod는 Pod용 서버 등 다양한 형태로 구현될 수 있으며 하나의 사용자가 다수의 Pod를 보유할 수도 있음
 - Solid 기반의 응용을 사용하기 위해 사용자는 먼저 응용 제공자로부터 응용을 자신의 클라이언트로 다운로드 받고 설치된 응용은 사용자의 Pod로부터 식별에 관련된 프로파일을 수신
 - 이 프로파일 정보를 이용하여 응용은 사용자의 Pod나 다른 사용자의 Pod로부터 필요한 데이터를 가져올 수 있고, 데이터를 가져올 때 필요에 따라 데이터에 대한 접근 제어를 위한 인증이 수행

- Solid에 적용되는 프로토콜은 기본적으로 W3C 표준 기술이 사용되고, Solid에 적용되고 있는 대표적인 W3C 표준 기술은 다음과 같음
 - RDF(Resource Description Framework) 1.1: Solid data model
 - WebID 1.0(Web Identity and Discovery): Universal username/IDs for Solid applications
 - FOAF Vocabulary 0.99: FOAF language used in WebID file and Access Control List
 - WebID-TLS: Universal identification mechanism
 - LDP(Linked Data Platform) 1.0: A set of HTTP operation on Web resource

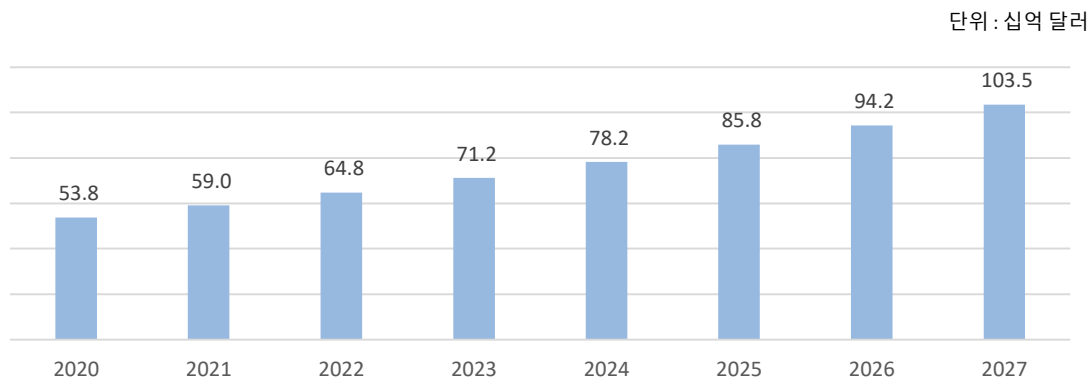
- 현재 인터넷 서비스는 대부분 기존 중앙집중형 웹 기반으로 구축되어 있고 따라서 이를 Solid 기반으로 다시 구축하는 것은 기술적인 부분뿐만 아니라 Solid 기반의 새로운 생태계 구축을 필요로 함
 - 팀버너스리는 이를 위해 Inrupt라는 스타트업을 만들고 Solid 활성화를 위한 인프라 구축을 지원
 - Inrupt는 Solid 기반 응용 개발에 대한 지원과 상업적인 활용 및 생태계 구축을 주 목표로 하며 Inrupt에 의해서 현재까지 개발되어 공개된 주요 Solid 기반의 응용은 Solid Weather, SolidVC, Solid Calendar 등

IV. 시장동향

1. 글로벌 시장

○ 글로벌 자동 식별 및 데이터 획득 시장은 2020년 538억 달러에서 연평균 9.8% 성장하여 2027년에 1,034.7억 달러로 성장 전망

- Automatic Identification and Data Capture(AIDC) 장치는 자동으로 개체 또는 이미지를 식별하고, 데이터를 캡처하고, 인간의 간섭 없이 컴퓨터 시스템에 직접 저장
- AIDC 기술은 소매, 제조, 운송 및 물류, 정부, 의료 및 접대를 포함한 여러 산업 분야의 재고 및 자산 관리, 보안 및 워크플로우 관리에 사용 중
- AIDC 시장의 성장은 주로 바코드, RFID 및 생체 인식과 같은 첨단 기술의 수용이 증가함에 따라 좌우되며 AIDC 장치에 대한 소비자 수요 증가, 스마트폰 및 온라인 쇼핑의 사용 증가, 데이터 보안의 증가 감각 등이 주요한 시장 성장 요인
- 그러나 연결 인프라의 부재와 첨단 기술 채택의 의지가 이 시장의 성장을 저해하고 있음



[그림] Global Automatic Identification and Data Capture Market

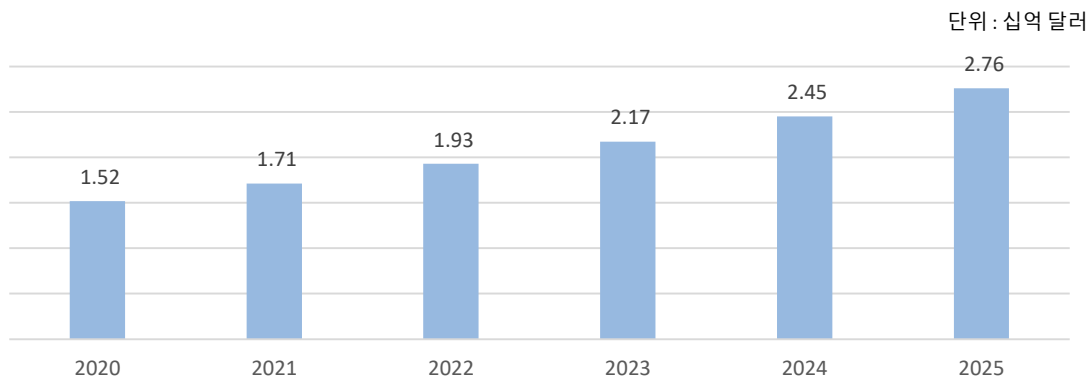
※ 출처 : Global Automatic Identification and Data Capture Market, 2021, meticulousresearch

○ 글로벌 공공/개인 데이터 관리 플랫폼(DMP) 시장 규모는 2025년까지 27억 6천만 달러에 이를 것으로 예상되며, 2020-2025년 동안 12.69%의 CAGR로 성장

- CRM 데이터, POS 데이터, 모바일 웹 및 앱, 소셜 네트워크, 웹 분석 도구 등과 같은 소스에서 데이터를 수집해야 하는 필요성이 증가함에 따라 데이터 관리 플랫폼 시장 성장
- 기업에서 데이터의 사용이 증가함에 따라 효율적인 데이터 보호 요구사항이 증가하고 있으며, 이로 인해 인공지능 기술에 대한 필요성이 증가

○ 대부분의 조직은 웹사이트, 모바일 장치 및 사물 인터넷(IoT)이 구동하는 장치와 같은 다양한 포인트의 데이터를 수집

- 다양한 획득 포인트에서의 데이터 수집은 AI의 통합으로 이어지는 효율적인 데이터 보호에 대한 수요로 연결되며, 인공지능의 사용이 증가함에 따라 조직의 IT 팀은 데이터 보호 요구가 증가
- 이러한 발전은 데이터를 효율적으로 분석하기 위한 증강 분석(augmented analytics)의 도입이 필요
- 효율적인 비즈니스 프로세스에 대한 중요한 결정을 내리기 위해서 증강 분석은 비즈니스 인텔리전스와 데이터 공유를 촉진하여 데이터 관리 플랫폼 시장을 더욱 촉진

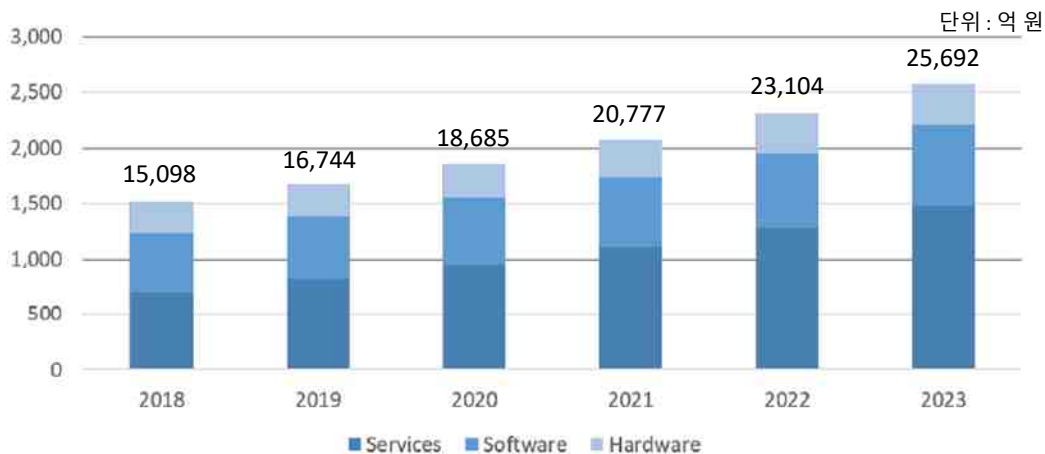


[그림] Global Data Management Platform(DMP) Market

※ 출처 : Global Data Management Platform(DMP) Market, 2021, industryarc

2. 국내 시장

- 국내 빅데이터 및 분석 시장은 연평균 11.2% 성장해 2023년 2조 5,692억 원 규모 전망
 - 예측 기간 동안 전체 시장에서 IT 및 비즈니스 서비스가 연평균 16.3% 성장하며 가장 큰 비중을 차지할 것으로 예상
- 기업 및 공공기관 등 모든 산업에서 비즈니스 혁신 및 인사이트 도출을 위한 데이터 활용의 중요성 대한 인식이 높아지고 있음
 - 이에 기업은 디지털 트랜스포메이션(DX)을 위한 노력을 지속적으로 기울이고 다양한 형태의 클라우드 환경 전환을 활발히 고려 중
 - 이를 성장 동인으로 빅데이터 및 분석 시장은 그 규모와 가능성에 있어 꾸준한 발전을 보이고 있으며, 빅데이터 기반의 고급 분석 및 인공지능 시스템 구축을 위한 데이터의 필요성 증가 요인에 의해 관련 시장은 향후 지속적 성장이 예측
 - 또한 기업은 빅데이터 및 분석 솔루션을 통해 다양한 형태의 데이터 인사이트를 실시간으로 발굴하고 복잡한 분석 및 비즈니스 의사결정을 자동으로 해결함으로써 지속적인 수익 창출과 경쟁 우위를 확보할 수 있다고 인식 중



[그림] 국내 빅데이터 및 분석 시장

※ 출처 : 국내 빅데이터 및 분석 2019-2023 시장 전망, 2020, 한국IDC

- 국내 법률 서비스 산업의 매출 규모는 2018년 기준 약 5조 9,334억 원, 2006년 이후 지속적인 증가세를 보이는 중³⁾
 - 2006년 약 2.17조 원에서 연 5~7% 성장하여 2018년에는 2배 이상 증가한 수치
 - 같은 기간 법률시장의 송무 사건 증가가 연평균 2%에 불과해 전체 법률시장의 매출액 및 과세표준액 증가는 비송무, 즉 법률자문 및 비소송 분야가 주도한 것으로 평가됨
- 한국고용정보원 자체조사 결과 국내 법률 관련인들은 국내 법률시장이 향후 5년 내 더 커질 것이며 10년 내 시장에서는 새로운 법률서비스와 스타 변호사, 인공지능 판단 등이 나타날 것으로 예상하였음



[그림] 국내 법률 서비스 시장 전망 설문조사 결과

※ 출처 : 법률직의 미래 직업세계, 2021, 한국고용정보원

3) 국세청 과세표준 신고액 기준

V. 산업동향

1. 글로벌 산업동향

- 빅데이터 플랫폼 기업들은 대용량 데이터의 실시간 처리, 다양한 파일 시스템에 대한 접근성 강화, 클라우드와 오픈 소스 기반 빅데이터 솔루션 연계 등을 통해 기업 생태계 형성 및 산업 성장에 자원과 역량을 집중
 - 주요 기업들은 분산형 병렬처리 및 실시간 데이터 처리를 위해 하둡 등의 솔루션을 활용하고 있음
 - 아마존은 자사의 AWS(Amazon Web Service) 클라우드 서비스와 오픈 소스 빅데이터 솔루션의 연계를 통해 차별화된 플랫폼 환경 구축 추진
 - 구글과 애플은 모바일 기반 고객 데이터와 자사의 분석 플랫폼을 기반으로 빅데이터 분석 환경 마련
- 빅데이터 플랫폼 기업의 핵심 기능을 크게 빅데이터 관리와 분석으로 구분 가능
 - (빅데이터 관리 분야 플랫폼 기업) 하둡을 기반으로 데이터 수집, 분산 컴퓨팅 및 병렬 처리에 집중
 - (빅데이터 분석 분야 플랫폼 기업) 데이터 마이닝, 텍스트 마이닝, 인공지능 알고리즘 등의 분석 툴을 활용하여 가치 있는 솔루션 제공
- Google, Amazon, MS 등 글로벌 기업들은 빅데이터 관리 및 분석 관련 자체 플랫폼을 바탕으로 자사의 기업 고객에게 토탈 솔루션을 제공

[표] 빅데이터 플랫폼 제품과 기업 현황

구분	빅데이터 관리	빅데이터 분석
주요 제품·기술	하둡, HDFS, 맵리듀스, Spark, 데이터 수집, 분산 컴퓨팅, 분산 파일 시스템, NoSQL, 병렬DBMS	Mahout, R, Spotfire, Tableau 통계분석, 데이터·텍스트 마이닝, 자연어 처리, 기계학습
주요 기업	Google, IBM, MS, Amazon, Oracle, SAP, Intel, Cisco, Splunk, Teradata 등	Google, IBM, MS, Amazon, Oracle, SAP, TIBCO Spotfire, Tableau Software 등

※ 출처 : 중소기업기술정보진흥원

- 2019년 4월 기준 전 세계 빅데이터 애널리틱스 분야에 약 3,800개의 스타트업이 존재하고, 우수한 기술력을 지닌 스마트업 대상 인수합병이 활발하게 전개되고 있음
 - 2019년 4월 1일 기준 3,821개의 빅데이터 애널리틱스 스타트업과 19,613명의 엔젤투자자가 있으며 한 기업당 평균 투자금액은 460만 달러로 나타남
 - 2017년 기준 인수합병 주요 기업으로는 구글에 인수된 빅데이터 예측 분석 플랫폼 기업 Kaggle, 2016년에는 오라클에 인수된 클라우드 기반 에너지 관리 플랫폼 기업 OPwer 등이 대표적임
- 최근 각 나라 정부는 공공 서비스에 빅데이터를 도입하기 위한 방법을 모색 중
 - Statista의 조사에 따르면, 빅데이터 도입 계획이 가장 많은 분야는 교육 분야로 나타났으며 이미 빅데이터를 도입한 비율은 26.5%, 향후 도입할 계획이 있는 기관의 비율은 73.5%로 나타났음
 - 주정부 및 연방정부의 85%가 이미 빅데이터를 도입했거나 향후 도입할 계획이 있다고 응답하였고 앞으로 공공 서비스 부문에서의 빅데이터 도입이 더욱 활발해질 전망이다
- 미국 환경보호청(EPA, Environmental Protection Agency)은 환경 규제 대상의 시설 규정 준수 사항 평가에 빅데이터를 활용하고 있음
 - 환경보호국은 80만 개의 규제 대상 시설로 구성된 중앙 데이터베이스 ‘ECHO’를 구축하고, 대상 시설의 규정 준수 기록 등에 대한 데이터를 축적하여 활용하고 있음
- 헬스맵(HealthMap)은 빅데이터를 활용하여 전 세계 질병 현황을 제공하는 서비스임
 - 전 세계 15개 언어의 공식문서와 뉴스, SNS, 국제기관 웹사이트를 분석하여 질병 전파 현황을 한 눈에 볼 수 있는 서비스를 제공함

- (에이젠테크놀로지) 옥스퍼드대 출신 연구진이 2014년 설립한 금융 계약·분석 AI 솔루션 기업으로 특히 텍스트에서 의미 있는 정보를 분석하는 컴퓨터과학의 한 분야인 자연어처리(NLP) 기술에 특화
 - 사용자가 질문을 남기면 학습한 문서를 토대로 분석하고, 문서를 작성하며 골드만삭스·테마섹 등으로부터 2019년 3,700만 달러 규모 투자를 유치
 - 에이젠테크 측에서는 자사의 AI가 금융 보험 분야 주요 계약 형태를 모두 학습하여 문서를 이해하는 정확도가 98%에 달한다고 밝힘

- (루미넨스) 2015년 영국 케임브리지에 설립된 법률 전문 AI 기업으로 영국 대형 로펌 상위 5개 중에서 4곳을 고객사로 두고 있음
 - 루미넨스가 제공하는 AI 소프트웨어는 세계 각국 법률과 정책, 판례 등을 제공하고 분석
 - AI는 문서를 분석해 위험 요소를 짚어주고, 각 판결문과 정책 요소를 시각화해서 보여주며, 법률에 대한 리뷰를 제공
 - 사용자들은 루미넨스의 서비스를 통해 유럽 개인정보보호법(GDPR) 준비 시간을 80% 정도 줄였다고 밝힘
 - 자체 인력으로 법적 매뉴얼을 준비할 때 사전 작업에만 9주가 소요된다고 예측되었지만 루미넨스를 사용해서 단 2명으로 7일 안에 작업을 완료 가능

2. 국내 산업동향

- 빅데이터의 개념적 모호성으로 인해 공공빅데이터는 실체를 가진 개념이라기보다 오픈데이터, 정부데이터, 빅데이터의 성격을 가진 ‘공공부문에서 공개한 빅데이터’에 가까움
 - 오픈데이터는 목적에 관계없이 누구나 어디에서 자유롭게 접근이 가능하고 공유 및 활용할 수 있는 데이터를 의미
 - 저작권과 같은 제한적인 규칙을 적용하지 않기 때문에 법적 개방성의 특징을 가지며, 특정 소프트웨어에서만 확인이 가능한 형태가 아니기 때문에 기술적 개방성을 가짐
 - 공공데이터는 정부데이터 가운데 오픈데이터 성격을 갖는 데이터로 “데이터베이스, 전자화된 파일 등 공공기관이 법령 등에서 정하는 목적을 위하여 생성 또는 취득하여 관리하고 있는 전자적 방식으로 처리한 자료 또는 정보”
 - 2013년 ‘정부3.0’을 추진하면서 제정된 공공데이터법에 따라 국가 및 지방자치단체는 ‘공공데이터의 제공 및 이용 활성화에 관한 기본계획’을 수립하고 관련 정책을 추진
 - 정부의 데이터 개방이 원활하게 이루어질 경우 새로운 제품과 서비스에서 데이터의 재활용이 이루어질 수 있음
 - 사회적 가치 실현에 있어 혁신적인 솔루션을 탐색할 수 있고, 공공행정 부문에 있어 데이터를 공유하여 운영의 효율성을 증대시킬 수 있음
 - 더불어 사회적 데이터 활용과 관련한 정치 및 사회생활에 시민참여를 촉진시킬 수 있고 나아가 정부의 투명성을 제고
- 데이터는 생성 주체와 공개 여부에 따라 구분하여 유형화할 수 있음
 - 이를 종합해본다면 공공빅데이터는 영리 빅데이터 수준의 가치를 가지고 있을 것으로 보임
 - 다만 개인정보 등과 같은 제약조건으로 인해 개방되지 못하거나, 개방되더라도 소수에 그쳐 아직은 활용도가 떨어질 수 밖에 없음
 - 그럼에도 미개방 공공 빅데이터를 개방하고 영리 빅데이터의 거래 활성화를 지속적으로 유도한다면 접근성 및 개방성이 개선될 수 있을 것



[그림] 빅데이터, 오픈데이터 및 정부데이터에 따른 유형

※ 출처 : 서울시 공공빅데이터 활성화 방안 연구, 2019, 최봉

- 공공빅데이터는 다양한 이해당사자가 참여하여 활용하기 때문에 민간 대비 관련 산업 생태계에 미치는 파급효과가 큼
 - 민간 영리 빅데이터는 통신사, 카드사, 신용정보사, 포털업체 등 소수의 민간기업이 정보를 독점하고 있어, 활용성이 높음에도 접근과 개방성이 취약한 상황
 - 앞으로도 기업들이 데이터를 그들만의 자산으로 인식하여 개방에 있어 소극적인 자세를 취할 것
 - 이러한 환경 속에서 공공빅데이터는 빅데이터 획득에 대한 차별을 제거해줄 수 있고 관련 산업 발전의 거대 인프라 역할을 수행
- 국내 빅데이터 생태계는 빅데이터 전문기술을 보유한 스타트업이 부족하며, 글로벌 수준의 빅데이터 기술혁신 기업도 전무한 상황
 - 미국의 CRN사(2017)에 의하면, 글로벌 100대 빅데이터 기술혁신 기업 중 국내 기업은 없는 것으로 나타남

- 우리나라는 빅데이터 스타트업 지원을 목적으로 인프라 지원 강화 및 맞춤형 성장 지원 등을 추진
 - 빅데이터 인프라 지원을 위한 판교 글로벌 ICT 클러스터를 추진 예정
 - 정부는 K-Global DB-Stars 사업을 통해 빅데이터 활용 스타트업을 선발, 컨설팅·투자 유치 등의 지원 확대
- 공공 빅데이터를 활용한 대표적인 법률 서비스 플랫폼 ‘로톡’은 2020년 말 기준 변호사 3,844명, 누적 상담수 40만 6,930건을 기록하는 등 높은 성장세를 보이는 중
 - 스마트폰 어플리케이션을 통해 상황에 맞는 상담사례 및 변호사를 찾아 법률상담이 가능하며 변호사의 전문성, 수임료, 상담후기, 해결 사례 등이 투명하게 공개
- 로톡은 대한변호사협회와의 갈등으로 인한 법정 공방을 펴고 있으며, 법무부와 국회, 현재에까지 범위가 확장될 정도로 갈등이 깊어지는 중
 - 2021년 대한변호사협회에서는 로톡과 같은 서비스가 변호사법을 위반하였다며 로톡 가입 변호사들을 징계처리 예고
 - 로톡은 대한변호사협회의 관련 규정에 대해 헌법소원 및 공정위 제소를 하였으며, 법무부와 국회에서는 소비자 선택권을 침해한다는 이유로 로톡의 입장을 대변하고 있음



[그림] 로톡 서비스 화면

※ 출처 : 로톡 홈페이지

○ (인텔리콘) 법률인공지능 기업으로 지능형 법률판례 검색기 ‘유렉스’를 개발하여 한국·일본 특허 출원, 국제 PCT(Patent Cooperation Treaty) 출원 등 세계 시장 진출

- 인텔리콘연구소는 ‘시각화 내비게이션 방식의 법률정보 시스템 및 법률추론 방법에 관한 특허’로 산업부 장관상을 수상
- 아이리스(i-LIS, Intelligent Legal Information System) 엔진은 기계독해 및 법률추론, 자연어 처리(NLP, Natural Language Processing), 머신러닝 기술 등이 융합된 ‘추론형 검색 기반 법률·판례 네트워크’ 시스템
- 유렉스는 이 시스템을 기반으로 일상어로 표현된 질문의 의미를 법률적으로 이해해 정보를 제공하며 법률전문가의 사고 패턴을 반영한 법률 네트워크 지식 구조를 보임



[그림] 유렉스 이미지

※ 출처 : IT조선

- (메디팔) 만성질환자를 위한 온라인 개인간호사 구독 서비스를 운영하는 스타트업으로 ‘제8회 범정부 공공데이터 활용 창업경진대회’에서 국무총리상(최우수상)을 수상하였음
 - 메디팔은 개인간호사 코치를 통해 올바른 처방 이행과 건강관리 습관 형성을 돕는 서비스로 만성질환자가 온라인 어플리케이션을 통해 진단명 등 간단한 정보를 입력하면 간호사와 1:1로 매칭
 - 이후 개인간호사는 정기 전화 상담, 실시간 채팅 상담 등을 통해 환자의 생활 패턴을 파악하고 맞춤형 컨설팅을 통해 일상생활 속 행동 변화를 만들어감
 - 메디팔은 본격적인 모바일 어플리케이션 개발 이전, 서비스를 압축한 형태의 필드테스트를 진행하여 약 60명을 대상으로 5점 만점에 4.7점의 높은 평가를 받았음
 - 메디팔은 향후 보건의료 빅데이터에 기반한 인공지능 알고리즘을 통해 환자의 질병 상황과 개인 성향을 고려한 개인맞춤형 서비스를 제공하는 동시에, 자동화 기술을 개발하여 확장성 확보에 주력할 계획



[그림] 메디팔 서비스

※ 출처 : 메디팔 홈페이지

[참고문헌]

- 빅데이터 기반 공간계획 기술 도입의 경제적 파급효과, 2021, 한국건설기술연구원
- 2021 공공 고객을 위한 IT 산업 전망, 2021, IBM
- 중소기업 전략기술로드맵 2021-2023(빅데이터), 2020, 중소벤처기업부
- 데이터기반 사회에서 데이터 주권 이슈와 대응기술 동향, 2020, 정희영
- AI 국가 경쟁력 확보를 위한 AI 데이터 생태계 조성방안 연구, 2020, KISDI
- 데이터 경제 축진의 매개체, 공공 주도의 오픈데이터 플랫폼, 2019, 딜로이트
- 서울시 공공빅데이터 활성화 방안 연구, 2019, 최봉
- 빅데이터 플랫폼의 산업생태계 현황과 주요 이슈, 2019, ETRI
- 공공부문 디지털 트랜스포메이션 현황과 전망, 2018, KISA