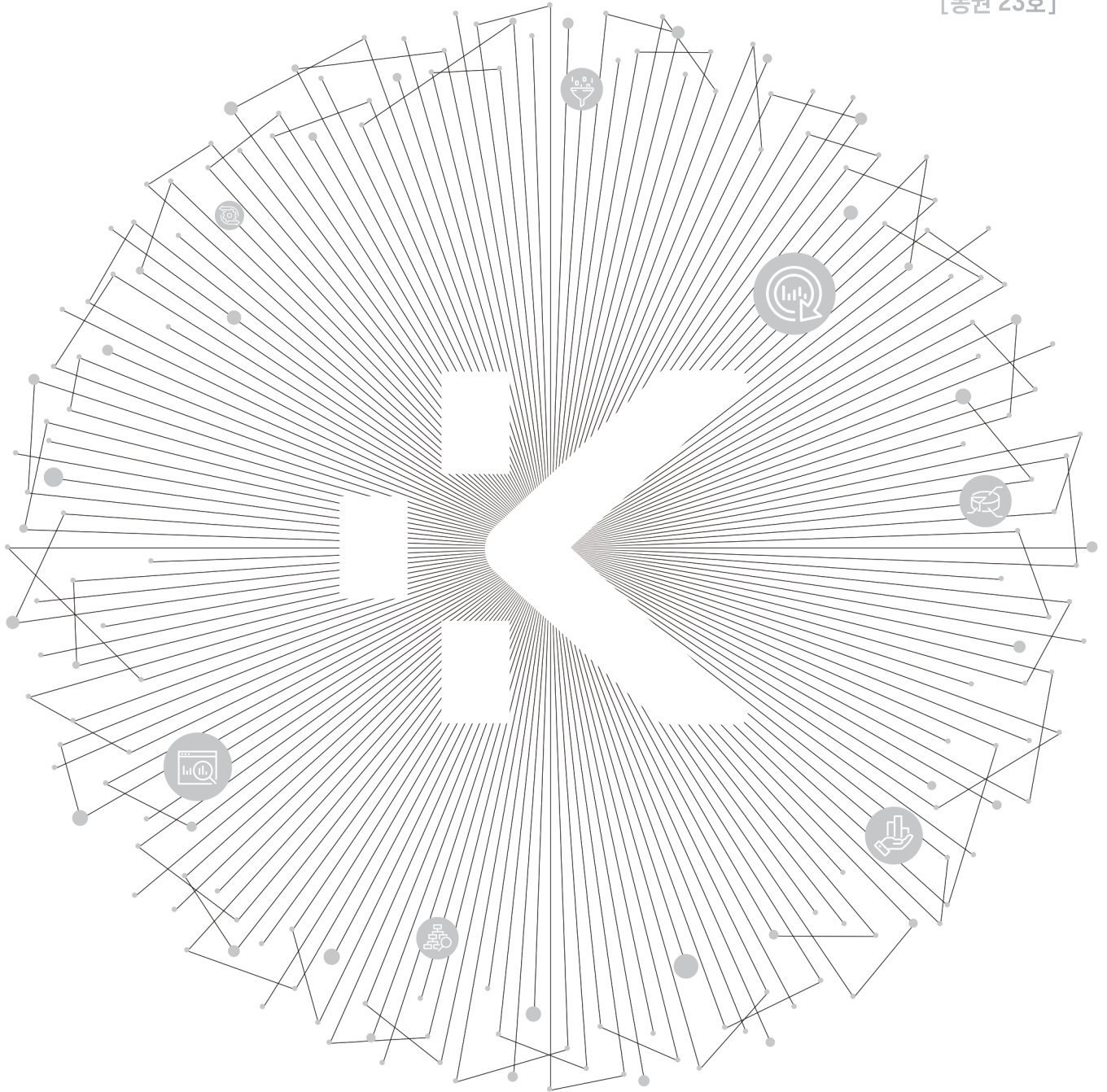


2020

데이터산업 백서

2020 DATA INDUSTRY
WHITE PAPER

[통권 23호]



‘혁신’ 실현과 데이터 경제 ‘선도’

안녕하십니까? 한국데이터산업진흥원 원장 민기영입니다.

2020년 올해는 전 국민이 디지털 전환과 데이터산업 생태계 지각 변동을 체감하는 한 해였습니다. 코로나19 감염병 대유행으로 가계와 기업에서 비대면 경제활동이 선호되며, ICT와 데이터 기반의 비대면 온라인 플랫폼 사용이 급증했습니다. 전통 산업들이 인원 감축 등 체질 개선을 시도하는 동안, 데이터 활용 기업들은 또 한 번의 사업 기회를 포착하고 있습니다. 배달, 전자상거래, 포털 등 대형 온라인 플랫폼 업체와 수익을 공유하는 플랫폼 기반 자영업자들이 증가했고, 과거에 없던 새로운 직업들도 등장했습니다. 전염병 대유행이 산업 간 융합과 혁신을 앞당겨 불러오면서, 디지털 경제가 본격적으로 펼쳐지고 있습니다.

전 산업에서 수집, 가공, 거래, 활용으로 연결되는 데이터는 디지털 경제를 움직이는 핵심 자원입니다. 국내 데이터산업 시장은 지속적인 성장세를 유지하며 2019년 약 17조 원 규모로 성장했고, 전 산업 분야에서 데이터직무를 담당하는 인력만 13만 명을 넘어섰습니다. 디지털 경제의 확대는 금융, 콘텐츠, 의료, 제조 등 다양한 산업에서 데이터를 활용한 신규 비즈니스를 지속적으로 창출할 것이며, 데이터와 기술을 효과적으로 처리할 데이터직무 인력 수요도 지속적으로 늘어날 것으로 전망됩니다.

전 세계 국가 정부와 대규모 데이터를 활용하는 글로벌 기업들은 올 한해에 데이터 활용 및 보호, 수익 창출이 가능한 주요 전략들을 쏟아냈습니다. 데이터산업 생태계 구성원들은 변화의 성장통을 겪으면서 큰 흐름에 동참하고자 노력하고 있습니다. 우리나라 정부도 올해 개정된 데이터3법을 시행해 데이터 활용 권한을 분명히 하고, 데이터 가공·결합·분석을 위한 기본 환경을 마련했습니다. 또한 침체된 국가 경제 위기에 대응하고자 국비 114조 원을 투입한 한국형 디지털 뉴딜 정책을 발표해 데이터댐 건설, 지능형 정부 구현, 스마트 의료와



비대면 교육 인프라 조성 등에 박차를 가하고 있습니다.

2020년, 우리 원은 정부의 보조에 맞춰 데이터 유통 및 활용 환경 조성, 데이터 비즈니스 육성 및 산업 기반 마련, 데이터 혁신인재 양성, 데이터 품질·표준화 등 데이터산업 활성화 지원제도 마련 등을 위해 노력했습니다. 데이터바우처 사업, 데이터활용 사업화 및 해외진출 지원사업, 데이터 안심구역 운영, 본인정보 활용지원 사업, 데이터 전문인력 양성사업이 그 예입니다. 또한 급변하는 데이터 산업에 발맞추고 선제적으로 대응하기 위해 국내외 동향정보 제공, 거래기반 연구와 제도마련 등도 강화했습니다.

이번에 발간하는 「2020년 데이터산업 백서」는 한 해 동안 데이터산업 발전을 위해 추진한 주요 정책과 급변하는 데이터산업 생태계의 변화 및 성장 현황을 수록했습니다. 부족하나마 최초로 2020년 한 해 뉴스 데이터를 수집·분석해 대표 이슈들을 선정하고, 코로나19 영향 등도 살펴봤습니다. 모쪼록 데이터산업 이해관계자분들께 유용한 자료가 되기를 기대합니다.

한국데이터산업진흥원은 걸어온 발자취를 확인·정리하고, 도래할 환경을 예측해 데이터 생태계 혁신과 데이터 경제 선도를 위해 노력하겠습니다. 마지막으로, 본 백서 발간을 위해 애쓰신 모든 분께 깊은 감사의 인사를 드립니다.

한국데이터산업진흥원 원장

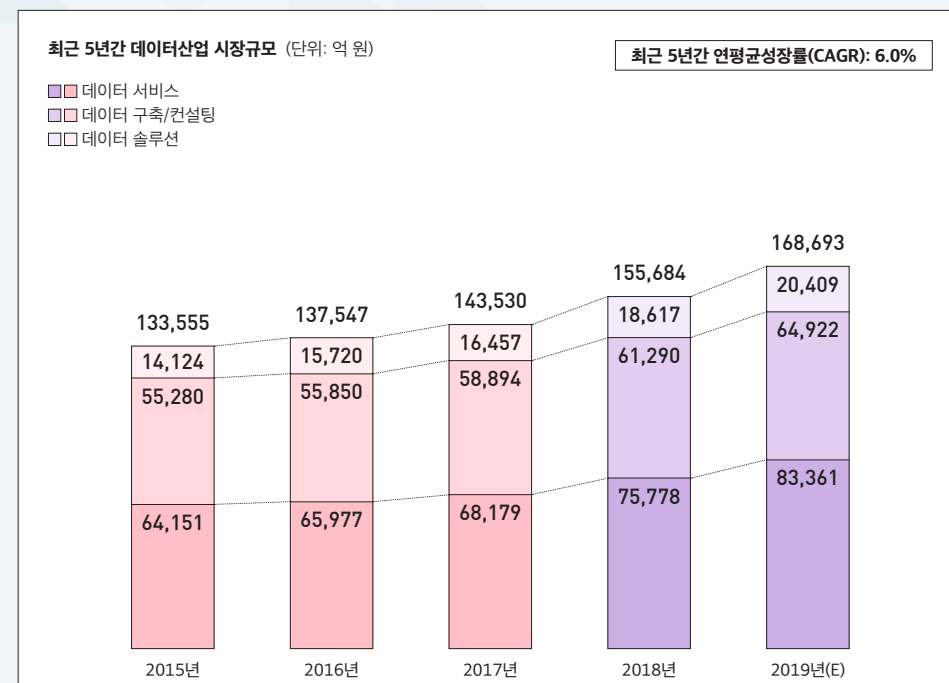
민기영

“2019년 데이터산업 시장규모는 16조 8,693억원, 전년 대비 8.4% 증가 예상”

2019년 국내 데이터산업 전체 시장규모는 2018년 15조 5,684억 원에서 8.4% 성장한 16조 8,693억 원으로 형성돼 2010년 이후 연평균 8.4%의 성장세를 나타내고 있다. 광고 매출 등 데이터와 관련된 간접매출을 제외한 직접매출 규모는 9조 2,094억 원 규모로 형성됐다.

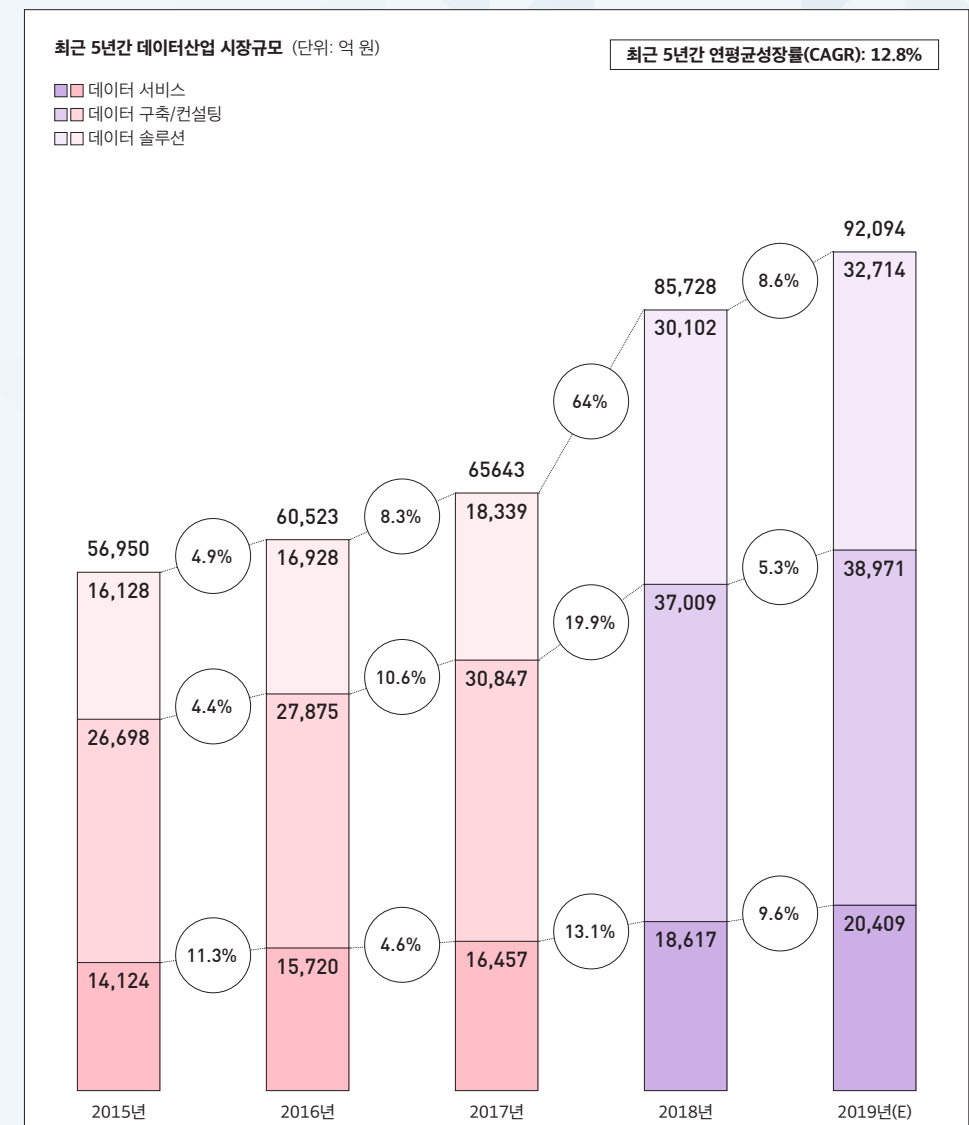
1. 데이터산업 시장규모

2019년 잠정적인 데이터산업 시장규모는 16조 8,693억원, 전년대비 8.4% 증가할 것으로 조사되었다. 분야별 시장규모에서 데이터 거래 및 정보 제공 등을 포함한 데이터 서비스(8조 3,361억원)가 전체 49.4%로 가장 큰 매출 비중을 차지할 것으로 예상되었다.



2. 데이터산업 직접매출 시장규모

2019년 잠정적인 국내 데이터산업 시장규모는 직접매출 기준 9조 2,904억 원으로 조사됐다. 이는 2019년 데이터산업 전체 매출의 55% 수준이며, 전년 대비 7.4% 성장할 것으로 예상된다. 데이터산업의 직접매출 시장규모는 데이터와 직접 관련이 있는 매출 기준이며, 광고 및 시스템 운영 관리 용역 등 데이터 관련 간접매출을 제외한 시장규모를 의미한다.



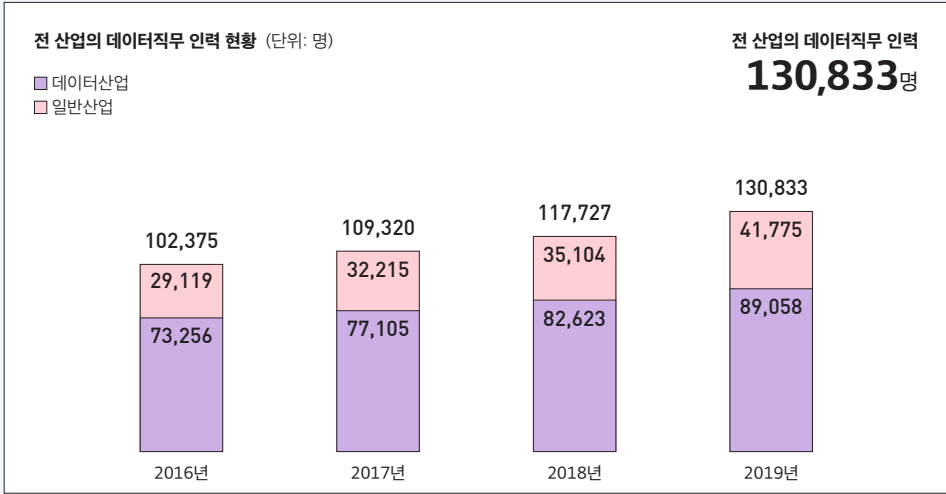
“2019년 데이터직무 인력은 전체산업 13만명 규모, 향후 5년 내 1.3만명 부족 예상”

2019년 데이터산업에서 데이터직무 인력 가운데 가장 높은 비중을 차지하는 직무는 데이터 개발자(35.2%), 데이터 엔지니어(16.4%)로 나타났으며, 향후 5년간 가장 부족할 것으로 나타난 직무는 데이터과학자(부족률 34.2%)와 데이터분석가(부족률 14.8%)인 것으로 나타났다.

1. 데이터직무 인력 현황

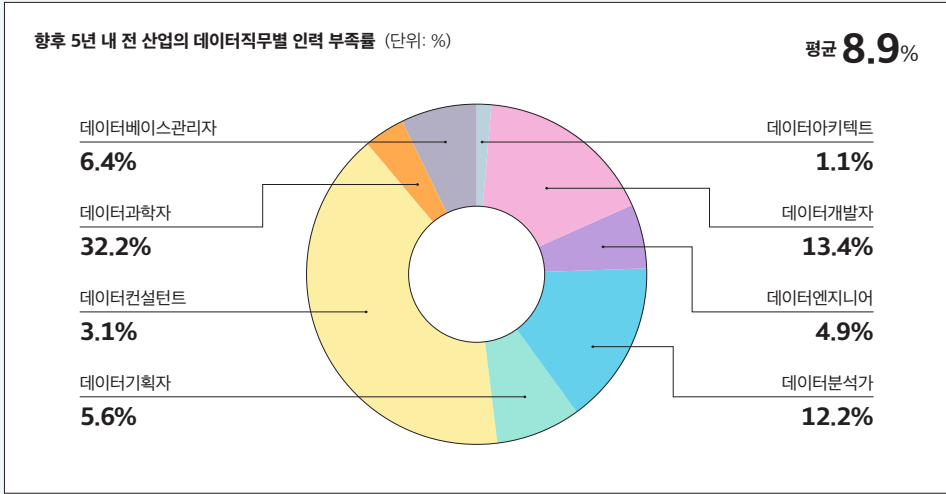
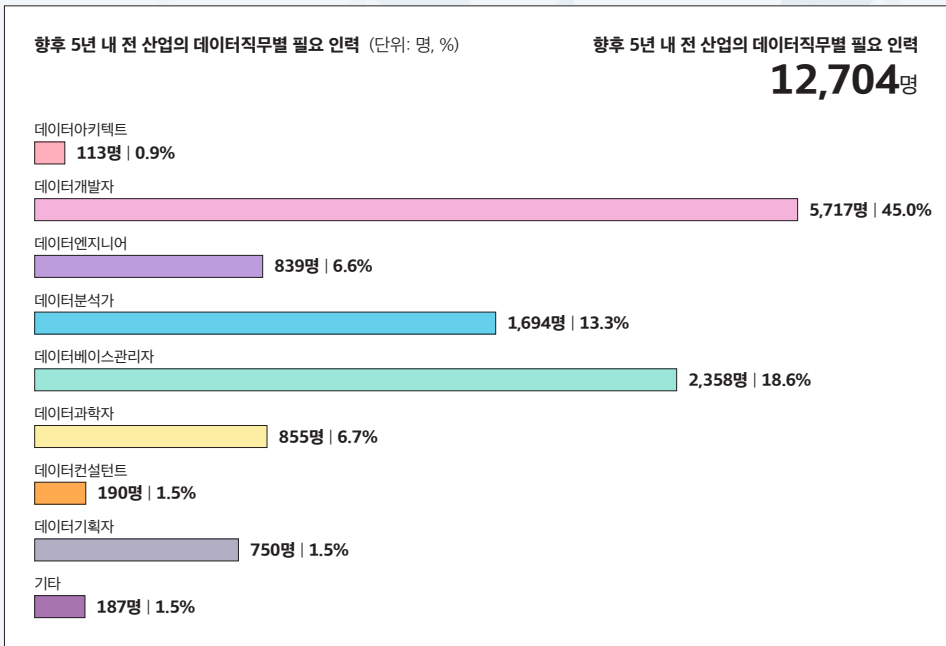
2019년 데이터산업 내 종사하는 모든 인력의 수는 18년 대비 8.4% 증가한 34만 4,672명이며, 그중 데이터직무 인력은 8만 9,058명(데이터산업 전체 인력 대비 25.8%)으로 조사되었다. 금융, 제조, 유통/서비스, 의료 등 일반산업까지 포함한 전 산업 내의 데이터직무 인력은 총 13만 833명으로 나타났다.

데이터직무는 데이터개발자, 데이터엔지니어, 데이터분석가, 데이터베이스관리자, 데이터과학자, 데이터컨설턴트, 데이터기획자, 데이터아키텍트 8개로 구분하였다.



2. 데이터직무 인력 부족률

전 산업에서 향후 5년 내 추가로 필요한 데이터직무 인력은 총 1만 2,704명으로 데이터 직무별 인력 부족률은 평균 8.9%로 조사됐다. 전체적으로 인력이 부족한 가운데 데이터 과학자가 32.2%로 가장 인력난이 심하고 이어서 데이터개발자 13.4%, 데이터분석가 12.2% 순으로 나타났다.



“코로나19 확산에 따른 데이터산업 글로벌 이슈 분석”

‘코로나19 확산에 따른 데이터산업 글로벌 이슈 분석’은 코로나19 발병 이후인, 2020년 1월부터 2020년 9월까지 발생한 글로벌 ICT 이슈 4만 2932건의 데이터를 바탕으로 분석했다. 전 세계적으로 발생한 코로나19는 봉쇄령, 재택근무, 원격교육 등 데이터산업과 밀접한 새로운 생활 패러다임을 정착시켰으며, 이로 인한 소비패턴 역시 급변하고 있다. 코로나19 발병이 데이터산업에 미친 영향을 분석해, 향후 시장 발전 방향에 대하여 알아본다.

01 코로나19에 따른 데이터산업 기술 발전의 변화

코로나19 발생이 데이터산업 중 가장 큰 영향을 미친 기술 분야는 머신러닝(ML)이다. 코로나19로 인한 일상생활의 극단적인 파괴와 변화는 이전 세대 머신러닝을 통해 발전시킨 학습모델(알고리즘)의 향후 실효성에 대한 불확실성을 경고하고 있다. 가트너(Gartner)를 포함한 글로벌 ICT 매체들은 코로나19 이전 학습된 머신러닝 알고리즘은 전면적인 재검토 작업이 필요하다고 강조하고 있다.

02 개인정보유출과 인권침해 논란에 휩싸인 ‘온라인 수업’

코로나19의 확산 시기에 맞추어 갑작스럽게 도입된 온라인 수업 제도에서는 준비시간 부족으로 다양한 사회적 문제들이 발생하고 있다. 충분한 정보보안체계 없이 시작된 온라인 수업은 개인정보 유출 위험에도 쉽게 노출돼 있으며, 편의성과 효율화를 위해 도입된 얼굴인식기술은 인권과 사생활침해 논란에 휩싸이고 있다.

03 코로나19 기간 ‘정보유출 및 정보보안’ 중요성 대두

코로나19 기간 동안, 사이버공격으로 인한 피해사례 역시 급증하고 있다. 2020년 4월 기준, 코로나19 테마를 악용한 불법 도메인 수는 1만 2258개에 달하며 금융기관과 클라우드를 향한 사이버공격도 각각 238%, 630% 급증했다. 이외 인도네시아와 영국 웨일즈에서 발생한 코로나19 검진 데이터 유출 소식이 사회적으로 이슈화되기도 했다.

04 클라우드 데이터센터 시장 지출 규모 사상 최대치

2020년 1월 이후, 전 세계 데이터 사용량은 평균 30~35% 증가한 것으로 분석되고 있다. 이에따른 데이터센터에 대한 수요 역시 가파르게 증가하였으며, 하이퍼스케일 단위의 데이터센터 설립이 가속화되고 있다. 다만 이러한 성장세는 클라우드 분야에 국한되어 발생하고 있으며, 민간 IT 기업의 투자 지출은 경영악화와 비즈니스의 불확실성 속에 오히려 3% 감소했다.

05 국가별 상이한 ‘코로나19 추적 앱’ 상용화 정책

국가별 코로나19 추적 앱의 개발 동향은 각 국가들의 상이한 개인정보보호정책의 성격을 대변하고 있기도 하다. 국가 통제력이 강한 중국은 QR 코드를 통한 강력한 추적 시스템을 전염병 발병 직후 도입했으며, 싱가포르의 개인 동의하에 활용되는 추적 앱을 개발해 상용화했다. 미국과 유럽에서는 현재까지도 코로나19 추적 앱 사용에 대한 찬반 논란이 지속되고 있다.

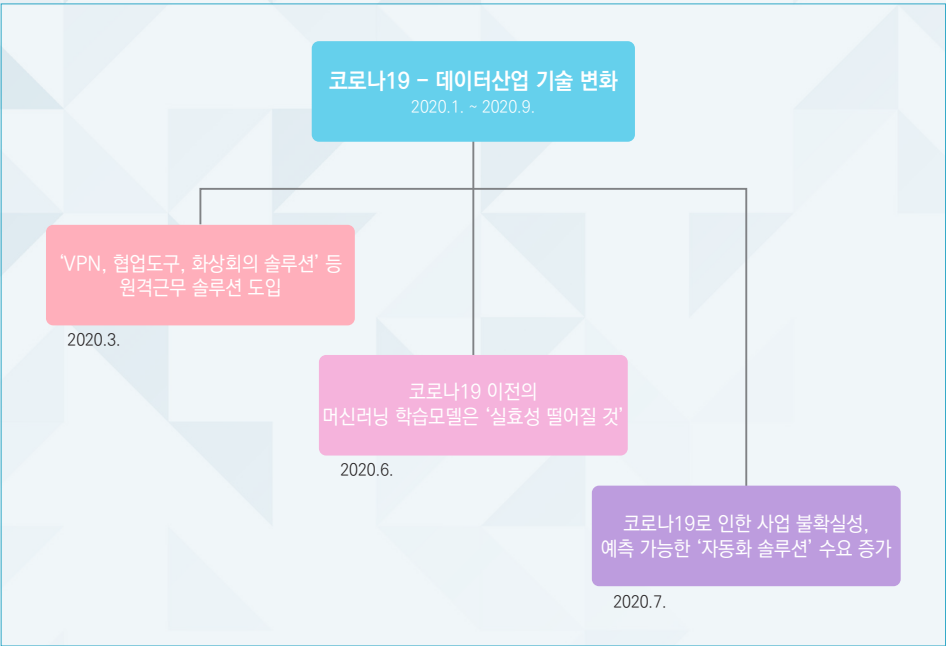
이슈1. 코로나19에 따른 데이터산업 기술 발전의 변화

코로나19 사태로 인한, 데이터산업 분야의 기술 개발 속도가 가속화되고 있다. 각국의 정부와 기업들은 데이터산업에 대한 대규모 투자 계획을 수립하고 있으며, 관련된 다양한 데이터 수집 및 분석 솔루션 제품들이 출시되고 있다.

코로나19가 본격적으로 확산하기 시작한 지난 3월에는 원격근무와 관련된 데이터 솔루션 제품들의 출시가 주를 이루었다. 특히 원격근무 환경 인프라 수준이 떨어지는 유럽을 중심으로 관련 솔루션을 공급하는 기업들의 진출이 본격화되었다.

코로나19가 장기화된 6월 이후로는, 급격한 소비 생활 패턴의 변화로 기존 머신러닝(ML) 학습모델에 대한 실효성 문제가 제기됐다. 소비자의 구매패턴이 코로나19 이전과 이후가 크게 달라짐에 따라 기존 머신러닝 학습모델을 실무에 적용 중인 기업들의 업무방식 재검토가 이뤄지고 있다. 이밖에도 코로나19로 인한 미래 산업과 비즈니스에 대한 불확실성 증가로 향후 갑작스럽게 등장할 수 있는 비즈니스 문제점을 예측할 수 있는 자동화와 AI 솔루션에 대한 수요가 증가할 것으로 전망된다.

[그림 1] 코로나19에 따른 기술 관련 이슈 발생 흐름도



출처 : 글로벌 ICT 뉴스매체 발행 이슈 4만 2932건 분석(2020.1. ~ 2020.9.)

1 ‘VPN, 협업도구, 화상회의 솔루션’ 등 원격근무 솔루션 도입

전염병이 발발하기 시작한 2020년 상반기, 코로나19가 재택근무를 통한 업무혁신 성장의 계기를 만들 것이라는 전망이 나오기 시작했다. 특히 3월 들어 확진자 수가 폭증하기 시작한 유럽과 이탈리아를 중심으로 관련 이슈가 집중되기 시작했다.

가. 유럽 주요국, 전체 기업 중 30% 이상은 원격근무를 위한 기술 인프라 부족

한국과는 다르게 봉쇄령이 내려지기 시작한 유럽에서는 재택근무는 선택이 아닌 필수사항으로 분류됐다. 하지만 IT 기술력이 높은 영국에서도 전체 기업 중 33%는 원격근무를 위한 기술 인프라 보유 수준이 부족한 것으로 나타났다.⁰¹

나. 이그나이트, 원격근무를 위한 핵심 솔루션은 ‘직원 간 협업 솔루션’

미국의 원격근무 솔루션 주요 기업 중 하나인 이그나이트(Egnyte)의 공동창립자 크리스 라히리(Kris Lahiri)는 원격근무를 위한 필수 솔루션으로 ‘협업도구, 화상회의, VPN, 보안 문서 액세스 도구’ 등을 꼽고 있으며, 이 중 협업도구에 대한 중요성을 강조하고 있다.⁰²

2 코로나19 이전의 머신러닝 학습모델은 ‘실효성 떨어질 것’

가트너(Gartner)는 2020년 6월, 코로나 이전에 학습된 ‘머신러닝 모델이 쓸모없게 되었다’라고 발표했다. 코로나19로 인한 일상생활의 극단적인 파괴와 변화로 기존 기계학습모델의 실효성이 크게 떨어질 것이라는 전망이다.

가. 코로나19로 인한 소비생활 패턴 변화, ‘코로나19 이전 학습모델의 유효성은 크게 떨어질 것’

가트너는 특히 이번 발표회에서 머신러닝을 통해 기존에 구축된 알고리즘을 사용하는 기업과 조직들은 기존 방식을 재고하는 것이 현명할 것이라고 경고하고 있다.

머신러닝의 실패사례는 과거 사례에서도 찾아볼 수 있다. 2008년 구글에서 발표한 독감 키워드 검색 빈도 기반의 독감 예측 솔루션은 ‘독감과 연관성이 떨어지는 정부의 공중보건경보 발령과 이로 인한 시민들의 독감 키워드 검색 폭증’으로 오인발표가 이어지며 2013년을 기준으로 서비스가 중단됐다. 미국 다우존스에서 활용했던 투자 솔루션 역시 ‘트위터의 해킹으로 발생한 잘못된 뉴스 기사에 대한 분석’으로 미국 증시 시중가가 1,360억 달러 폭락하는 사태를 야기하기도 했다.

[표 2] 머신러닝의 대표 실패사례

솔루션	시기	내용
구글, 독감예측 솔루션	2008년	알고리즘 독감 키워드 검색 빈도 추적 기반, 유행시기 예측
		실패사례 (2009) 신종인플루엔자(H1N1) 유행 예측 실패, (2013) 실제 독감 발생률 2배에 달하는 예측치 전망
		실패이유 정부의 예측하지 못한 경보 발령 - 2013년 뉴욕시 공중보건경보 발령 - 독감 관련 없는 시민들의 독감 관련 검색 폭증
미국 주식, 투자솔루션	2013년	알고리즘 뉴스 정보 기반, 주식 시장 평가
		실패사례 (2013) 해커의 불법 트윗 오해석으로 다우존스 1,360억 달러 폭락 - 시리아 해커, 트위터 해킹으로 ‘백악관 폭발로 인한 오바마 부상’ 가짜뉴스 발행
		실패이유 예측하지 못한 행동 발생 - 해커의 가짜뉴스 발행

출처 : 인사이드빅데이터(Inside BIGDATA)

01. Censuwide(독일 설문조사기관), 2020. 3.
02. Clare Brown, COVID-19’s impact on the enterprise and remote work, CIO(www.cio.com), 2020.3, www.cio.com/article/3532812/covid-19s-impact-on-the-enterprise-and-remote-work.html

나. 코로나19 이전의 머신러닝 학습모델 ‘손소독제 구매 증가 → 임신 증가’로 잘못 분석
인사이드빅데이터(Inside BIGDATA)는, 이와 같은 머신러닝 알고리즘의 실패는 ‘인간의 행동이
갑자기 변할 때’ 발생한다고 정의하고 있다. 코로나19 발병 이후, 대표될 수 있는 잘못된 분석
사례는 쇼핑 알고리즘에서 찾아볼 수 있다. 코로나19 이전, 머신러닝은 기존에 손소독제
구매가 임신 증가와 연관될 것이라는 모델을 수립할 수 있다. 다만 코로나19 이후에 해당
학습모델을 적용한다면 세계 인구 중 절반이 임신을 했다고 잘못 분석된 결과물을 도출할
것이다.⁰³

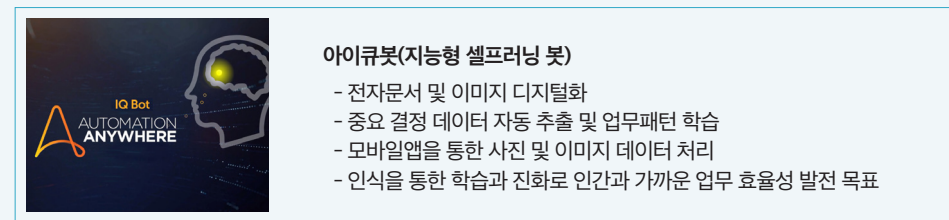
3 코로나19로 인한 사업 불확실성, 예측 가능한 ‘자동화 솔루션’ 수요 증가

가. 확신할 수 없는 미래 수요를 예측하기 위한 ‘자동화 시스템과 로봇 수요 증가’
2020년 7월, IBM은 AI 기반 자동화 기능 향상을 위하여 WDG(WDG Automation)를 인수
결정하고 계약을 체결했다. IBM은 코로나19로 인한 미래의 불확실성을 고려했을 때,
예상하지 못한 문제점을 사전에 분석할 수 있는 AI 기반 자동화 솔루션의 도입이 현시대에
적합할 것이라고 판단했다. 특히 IBM은 해당 기술을 비즈니스 문제와 기회 요인을 감지해
향후 사업 단계를 예측하는 데 활용할 것이라고 발표했다.⁰⁴

나. 로봇공정자동화 분야의 리더 ‘오토메이션 애니웨어’

가트너는 2020년 7월, 로봇공정자동화(Robotic Process Automation, RPA) 분야의 선두기업으로
오토메이션 애니웨어(Automation Anywhere)를 선정했다. 오토메이션 애니웨어는 업계 유일의
웹기반 통합 지능형 문서처리 솔루션인 아이큐봇(IQ Bot)을 설계한 세계 1위의 RPA 기업이다.

[그림 3] 오토메이션 애니웨어의 아이큐봇



출처 : 오토메이션 애니웨어(www.automationanywhere.com)

03. *Where Predictive Machine Learning Falls Short and What We Can Do About It*, InsideBIGDATA(insidebigdata.com), 2020.6.7., insidebigdata.com/2020/06/27/where-predictive-machine-learning-falls-short-and-what-we-can-do-about-it

04. *IBM to Acquire WDG Automation to Advance AI-Infused Automation Capabilities*, Database (www.dbta.com), 2020.7.13., www.dbta.com/Editorial/News-Flashes/IBM-to-Acquire-WDG-Automation-to-Advance-AI-Infused-Automation-Capabilities-141754.aspx

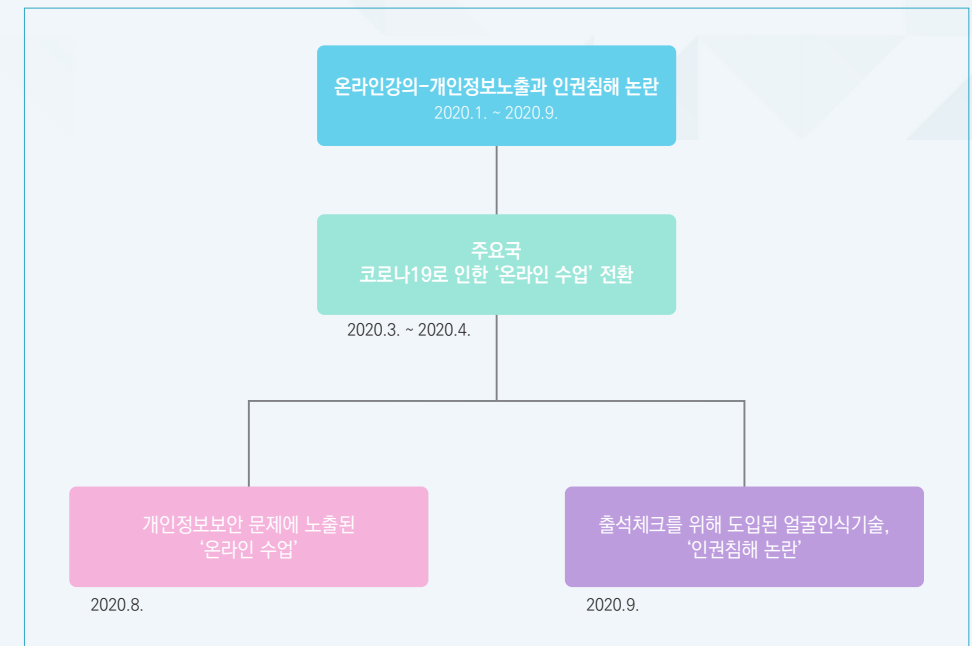
이슈2. 개인정보유출과 인권침해 논란에 휩싸인 ‘온라인 수업’

코로나19 장기화에 따라 전염병 피해국가들의 온라인 수업 제도 도입이 본격화되고 있다. 2020년 3월 이후, 도입되기 시작한 온라인 수업은 아직 발전하지 못한 지원 인프라와 관련 지식 인력 부족으로 어려움을 겪고 있으며, 사회 곳곳에서 우려의 목소리가 제기되고 있다.

이 중 정보보안과 관련된 대비 활동은 주요 선진국들 사이에서도 가장 큰 문제점 중 하나로 꼽히고 있다. 유럽의 IT 선진국인 영국이 이러한 문제를 겪고 있는데, 전체 교사 인구 중 절반 가까이는 ‘교사 본인보다 학생들이 정보보안에 대해 더 잘 알 것이다’라고 설문조사에 응답하기도 했다.

이외 미국의 온라인 수업에 도입하기 시작한 ‘얼굴인식기술’은 사생활 감시, 흑인이나 여성을 잘 식별하지 못하는 문제점 등으로 인권침해 논란에 휩싸이기도 했다. 2020년 9월에 들어서는 50개 주, 1000명의 학부모가 서명한 얼굴인식기술 금지 촉구 서한을 미국 시민단체가 발표하기도 했다.

[그림 4] 코로나19, 온라인 수업 관련 이슈 발생 흐름도



출처 : 글로벌 ICT 뉴스매체 발행 이슈 4만 2932건 분석(2020.1. ~ 2020.9.)

1 주요국 코로나19로 인한 ‘온라인 수업’ 전환

코로나19는 교육에 대한 환경도 급격히 변화시키고 있다. 주요 코로나19 피해국들은 기존 오프라인 방식의 수업 문화를 온라인 방식으로 대체하고 있으며, 전례없는 교육 환경의 변화로 사회 곳곳에서 우려의 목소리와 문제점이 발생하고 있다.

가. 일본 시민 중 84.3%, ‘온라인 수업이라면 코로나19에 안심할 수 있다’

코로나19 확산으로 온라인 수업과 휴교 조치를 가장 서둘러서 도입한 국가는 일본이다. 일본은 지난 2월, 봄학기 개강에 앞서 코로나19 확산을 막기 위한 전국 초중고교 임시 휴교조치를 실행했다. 현지 교육기관인 메가스터디에서 시행한 설문조사 결과에 따르면, 일본 시민 중 84.3%는 ‘온라인 수업 진행으로 코로나19 감염 확산에 안심한다’라고 응답하기도 했다.⁰⁵

나. 호주국립대학교, 학생들 우려 속에서도 온라인 시험 강행

호주도 발빠르게 온라인 수업 문화를 정착해나가고 있다. 코로나19 사태에 봉쇄령으로 대응하고 있는 호주에서는 호주국립대학교가 학점 평가에 반영되는 시험을 온라인 방식으로 진행하겠다고 결정하면서 학생과 학교 사이에 논란이 가중되고 있다.⁰⁶

2 개인정보보안 문제에 노출된 ‘온라인 수업’

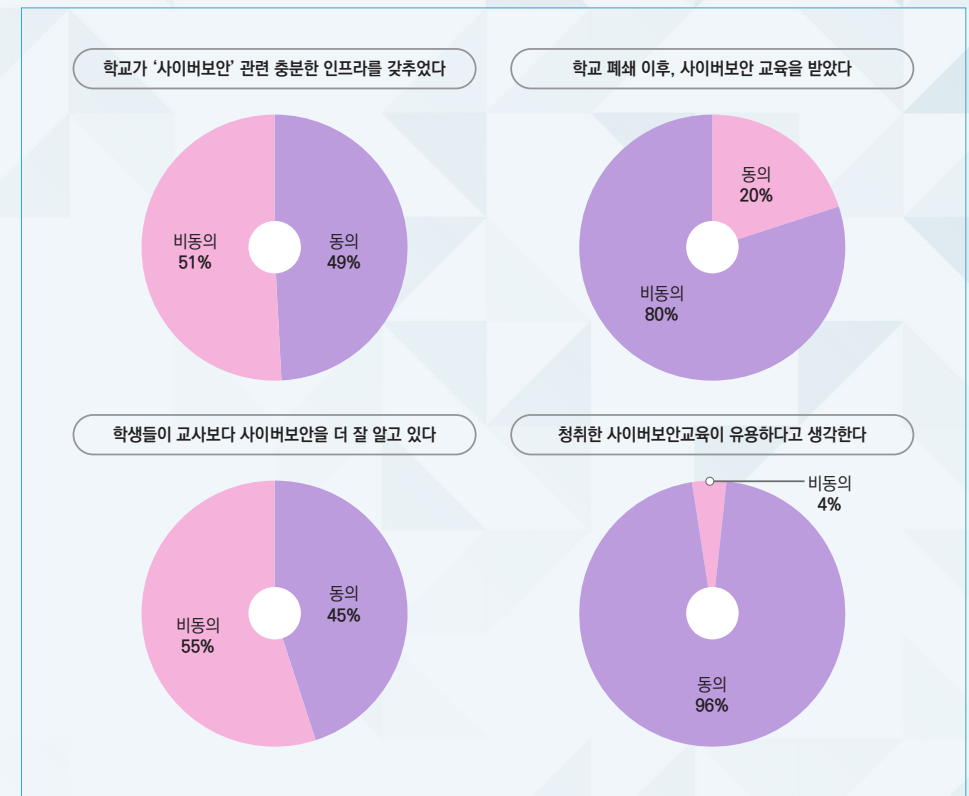
온라인 수업으로 발생할 수 있는 가장 큰 사회문제 중 하나인 정보보안과 관련된 사항은 ‘제도, 인프라, 기술, 전문인력 부재’ 등 모든 부분에 있어 향후 발전과 개선이 필요한 상태다.

가. 영국, 교사 중 절반 이상은 ‘학교가 사이버보안 문제를 해결할 수 없을 것’이라고 예상
유럽 IT 선진국 영국에서 이와 같은 현상의 구체적인 예시를 찾아볼 수 있다. 영국은 3월 23일 국가 봉쇄령 이후, 대부분 학교에서 온라인 수업을 채택해 진행중이나 정보보안과 관련된 취약한 인프라와 정부차원의 보안 교육에 대한 시급성이 제기되고 있다.

05. 「新型コロナウイルスでオンライン学習なら安心は84.3% = バンザン調べ =」, ict-enews(ict-enews.net), 2020.3.1.

06. Matt Johnston, ANU commits to online exams, invigilation despite student concerns, itnews(www.itnews.com.au), 2020.4.20, <https://www.itnews.com.au/news/anu-commits-to-online-exams-invigilation-despite-student-concerns-546894>

[그림 5] 영국, 사이버보안 관련 교사 대상 설문조사 결과



출처 : 인터넷매달스(Internet Matters), 교사 1000명 설문조사 결과

인터넷매달스(Internet Matters)에서 실시한 영국 교사 1000명을 대상으로 한 설문조사 결과에 의하면, 전체 응답자 중 51%는 학교가 사이버보안과 관련된 충분한 인프라를 갖추고 있지 않다고 응답했다. 또한 36%는 전년도에는 사이버보안과 관련된 정보가 전무했다고 응답했다. 봉쇄령 이후 사이버보안 관련 교육을 받은 교사는 대상자 중 20%에 불과했다. 더불어 응답자 중 45%는 교사가 아닌 학생들이 사이버보안과 관련된 지식이 더 높을 것이라고 응답하기도 해 관련 문제가 심각함을 알 수 있다. 다만 사이버보안 관련 교육을 받은 교사 중 96%는 해당 교육이 유용하다고 응답해 정보보안교육과 관련된 정부 차원의 투자와 노력이 현재의 문제점 개선에 필수적임을 알 수 있다.

글로벌 IT 보안기업 에셋(ESET)의 마케팅 책임자인 줄리언로버츠(Julian Roberts)는 “사이버 범죄자들은 지속적으로 그들의 정보 해킹 방법론을 발전시키고 있으며, 온라인이 아닌 교실에서도 가상교육에 대한 기술이 적용되고 있음에 따라, 교사와 학생들을 위한 올바른 보안 교육이 매우 중요할 것”이라고 강조하고 있다.

3 출석체크를 위해 도입된 얼굴인식기술, ‘인권침해 논란’

정보보안과 더불어 온라인 수업 진행으로 최근 관심이 집중되고 있는 이슈는 ‘얼굴인식 기술’의 온라인 수업 도입이다. 얼굴인식기술의 온라인 수업 적용은 미국에서 가장 활발하게 이뤄지고 있는데, 최근 50개 주 1000명의 학부모가 동의한 얼굴인식기술 금지 촉구 서한이 비영리단체 파이트포더퓨처(Fight for the Future)에 의하여 발표됐다. 해당 단체는 얼굴인식기술의 도입이 생활감시의 위험과 함께 특정 학생 그룹을 차별하는 데 악용될 수 있음을 우려하고 있다.⁰⁷

얼굴인식기술의 도입을 반대하는 단체들은 현재의 얼굴인식시스템은 여성이나, 피부색이 어두운 사람들을 제대로 인식하지 못한다는 점을 들어 해당 기술이 인종 차별의 수단으로 쓰일 수 있다는 부분을 비판하고 있다.

이슈3. 코로나19 기간 ‘정보유출 및 정보보안’ 중요성 대두

코로나19 사태 발발과 함께 사이버공격은 코로나19를 악용한 방법으로 확장 및 변형되어가고 있다. 2020년 4월 기준, 코로나19와 연관된 사기성 도메인이 1만 2258개, 금융기관 사이버공격은 238% 급증, 금융기관 랜섬웨어 공격 9배 이상 급증, 디지털 사기거래 5% 증가, 클라우드 계정에 대한 사이버공격은 630%까지 급증한 것으로 나타났다.

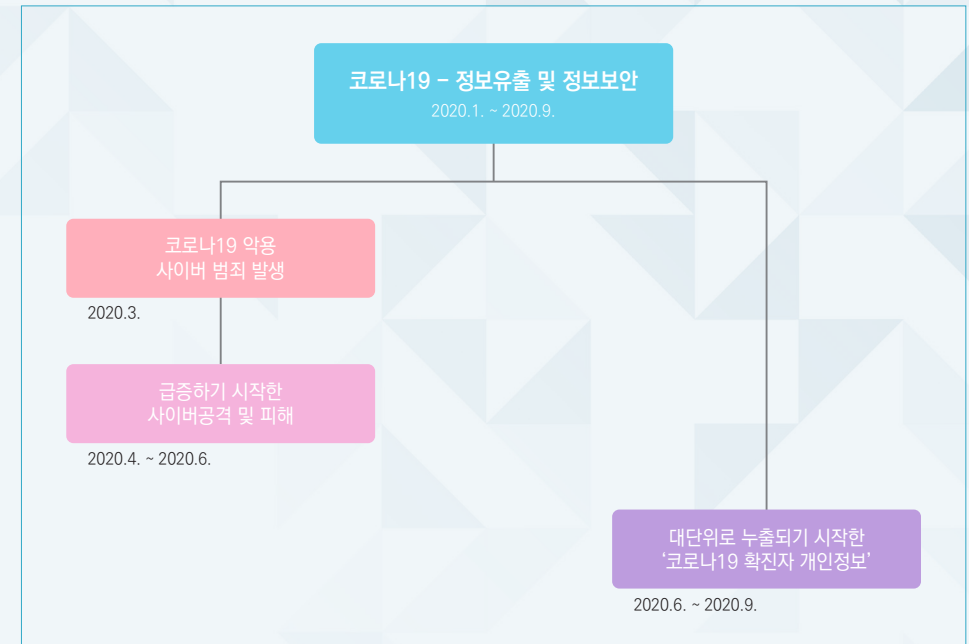
또한 체계적으로 관리되지 못한 코로나19 검진 데이터가 공용 서버와 다크넷⁰⁸에 유출되는 사례가 발생하기도 했다. 2020년 6월 인도네시아의 코로나19 검진자 23만 명에 대한 개인정보를 한화 36만 원에 판매한다는 게시글이 다크넷에 게재되기도 했다. 2020년 9월 영국 웨일즈에서는 안일하게 관리된 검진자 1만 8105명에 대한 개인정보데이터가 보건부 직원의 실수로 공용 서버를 통해 공개되는 해프닝이 발생하기도 했다.

07. Kyle Wiggers, *Over 1,000 parents sign open letter protesting facial recognition in schools*, Venturebeat(venturebeat.com), venturebeat.com/2020/09/16/over-1000-parents-sign-open-letter-protesting-facial-recognition-in-schools

08. 다크넷 : 다크넷 또는 다크 넷은 비표준적인 통신 규약과 포트를 사용하는 특정 소프트웨어, 설정, 또는 허가가 자주 있어야 접속할 수 있는 오버레이 네트워크

09. Alex Scroxton, *Fake antivirus site promises coronavirus protection, delivers trojan*, ComputerWeekly(www.computerweekly.com), www.computerweekly.com/news/252480520/Fake-antivirus-site-promises-coronavirus-protection-delivers-trojan

[그림 6] 코로나19, 정보유출 및 정보보안 관련 이슈 발생 흐름도



출처 : 글로벌 ICT 뉴스매체 발행 이슈 4만 2932건 분석(2020.1. ~ 2020.9.)

1 코로나19 악용 사이버 범죄 발생

생물학적 바이러스인 코로나19를 IT 보안제품이 보호할 수 없는 것은 당연한 사실이다. 다만 코로나19로 인한 불확실한 환경 속에서 이러한 덫에 빠진 사이버범죄 피해자가 속출하고 있다. 또한 코로나(Corona, COVID) 등을 키워드로 한 스팸메일, 악성코드, 악성 도메인이 우후죽순 파생되고 있다.⁰⁹

2 급증하기 시작한 사이버공격 및 피해

가. ‘Coronavirus’ 또는 ‘Covid-19’ 등의 단어가 포함된 악성 도메인 600% 증가
글로벌 시장조사기관 야릭스(Yarix)에 따르면, 코로나19 발병 이후 2020년 4월까지 ‘Coronavirus’ 또는 ‘Covid-19’ 키워드가 포함된 새 도메인 등록이 약 600% 증가한 것으로 나타났다. 특히 코로나19와 관련된 사기성 도메인 및 호스트는 2020년 4월 기준, 1만 2258개가 식별됐다.¹⁰

10. Dominique Filippone, *Covid-19 : 26 000 cybermenaces identifiées par la coalition Cyber Threat*, Lemond informatique(www.lemondeinformatique.fr), <https://www.lemondeinformatique.fr/actualites/lire-covid-19-26-000-cybermenaces-identifiées-par-la-coalition-cyber-threat-79021.html>

[그림 7] 코로나19를 악용해 만든 사이버범죄 웹페이지



출처 : 컴퓨터위클리(Computer Weekly)

나. 코로나19 기간 동안 은행에 대한 사이버공격 238% 증가

금융권에 대한 사이버공격과 그 피해는 더 심각하다. 서버 가상화 업체 브이엠웨어(VM Ware)에 따르면, 코로나19 이후 4월까지 금융기관에 대한 사이버공격이 238% 급증한 것으로 나타났으며, 랜섬웨어로 인한 공격은 약 9배 이상 증가한 것으로 나타났다.

다. 코로나19 확산 이후, 디지털 금융 사기 최대 80%까지 증가할 것으로 전망

증가하는 온라인 금융거래로 디지털 금융 사기에 대한 위험도 급격히 증가할 것으로 전망된다. 글로벌 시장조사기관 트레이드유니온(Trade Union)에 의하면 디지털 사기 거래로 의심되는 거래 비율이 3~4월 5% 가량 증가했으며, 같은 기간 의심되는 사기거래건수는 1억 건 이상일 것으로 보인다. 트레이드유니온은 이와 같은 디지털 사기가 향후 적게는 40%에서 최악의 사태에는 80%까지 증가할 것으로 전망하고 있다.¹¹

11. Sibahle Malinga, *Digital financial fraud predicted to witness rise of up to 80%*, Itweb(www.itweb.co.za), 2020.6.1, www.itweb.co.za/content/VgZeyvJoGA87djX9/zlP3gQ2qGRMnRD1W

라. 클라우드를 대상으로 한 사이버공격 630% 증가

금융권과 더불어 퍼블릭 클라우드도 해커들의 주요 공격 대상이 되고 있다. 맥아피(McAfee)에 따르면 클라우드 계정에 대한 사이버공격은 2020년 4월 기준, 같은해 1월 대비 630% 가량 증가한 것으로 보인다¹² (코로나19 이후, 기업들의 클라우드 서비스 사용량은 50% 가량 증가함).

[표 8] 코로나19 기간 동안 발생한 사이버공격 및 피해 현황

항목	현황	기간	내용
악성 도메인	'Coronavirus' 또는 'Covid-19' 포함 도메인 600% 증가	2020년 2월 ~2020년 4월	피싱, 협박, 악성코드, 거짓 모금 캠페인
	코로나19 테마 악용 도메인 1만 2258개 식별	2020년 3월 ~2020년 5월	
금융기관 사이버공격	금융기관 사이버공격 238% 증가	2020년 2월 ~2020년 4월	자금이체 데이터 유출, 데이터 탈취 몸값 요구
	금융기관 랜섬웨어공격 2월 이후 4월까지 9배 증가	2020년 2월 ~2020년 4월	
디지털 금융사기	디지털금융사기 40 ~ 80% 증가 전망	2020년 6월	아이디 해킹, 기밀문서 유출
클라우드 사이버공격	코로나19 기간 클라우드 사이버공격 630% 증가	2020년 1월 ~2020년 4월	예측하지 못한 행동 발생 해커의 가짜뉴스 발행

출처 : Key4Biz, Itweb, ITpro, Infosecurity-magazine

3 대단위로 누출되기 시작한 ‘코로나19 확진자 개인정보’

가. 인도네시아, 코로나19 검진자 23만 명 정보 다크네티에 유출

2020년 6월, 인도네시아는 ‘인도네시아 코로나19 검진자 23만 명의 개인정보를 420만 루피아(약 36만 원)에 판매한다’는 게시글이 다크네티에 올라오며 비상이 걸렸다. 해당 데이터베이스에는 이름, 주소, 전화번호, 시민권 번호, 진단 날짜, 확진여부 등의 개인정보가 포함된 것으로 나타났다.

나. 웨일즈, 코로나19 확진자 1만 8105명 개인정보 데이터 유출

영국 웨일즈에서도 코로나19 확진자 데이터를 관리하는 보건부 직원의 실수로 1만 8105명의 개인정보가 유출됐으며, 약 20시간 동안 공용 서버를 통해 자료가 일반인에게 공개되기도 했다.

12. John Leyden, *Cloud-based cyber-attacks flaring up during coronavirus pandemic*, Portswigger(portswigger.net), 2020.6.1, portswigger.net/daily-swig/cloud-based-cyber-attacks-flaring-up-during-coronavirus-pandemic

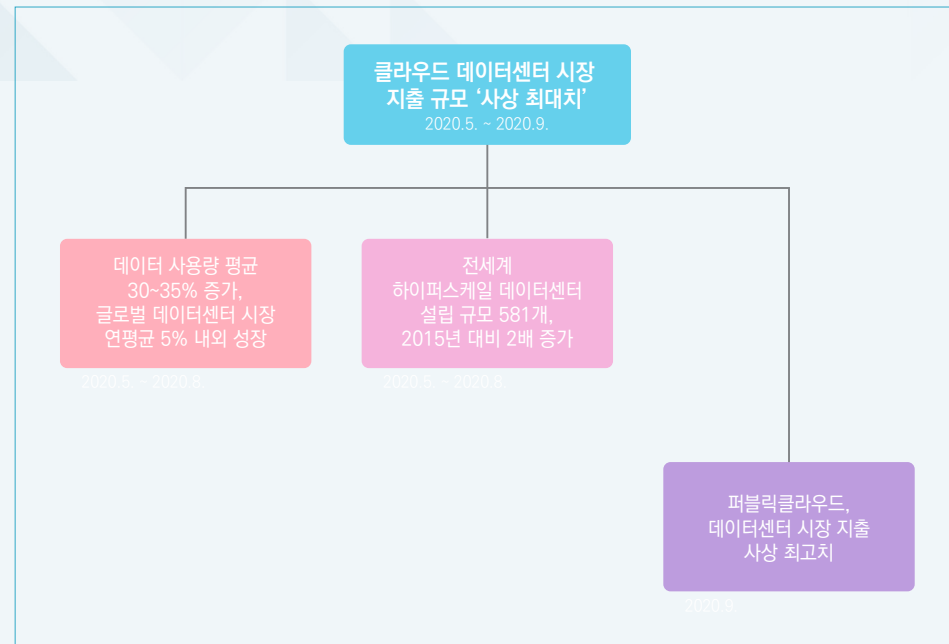
이슈4. 클라우드 데이터센터 시장 지출 규모 '사상 최대치'

코로나19 장기화로 인한 봉쇄령, 재택근무, 원격수업 등은 전 세계적으로 데이터 사용량을 평균 30~35% 증가시켰다. 급작스럽게 증가한 데이터 사용량은 데이터트래픽 문제를 야기했으며, 이는 데이터센터에 대한 수요 증가와 '하이퍼스케일'¹³ 단위의 데이터센터 건설을 촉구하고 있다.

2020년 2분기 기준, 하이퍼스케일 데이터센터 설립수는 글로벌 기준 581개에 달하며, 현재도 176개에 달하는 신규 하이퍼스케일 데이터센터의 투자가 결정되어 설립을 기다리고 있다. 한국은 아직 하이퍼스케일 단위의 데이터센터를 보유하고 있지는 못하지만, 이러한 변화에 맞춰 네이버와 카카오를 중심으로 관련 투자 활동을 계획 중이기도 하다.

데이터센터의 수요는 클라우드 분야에 국한돼 발생하고 있다. 일반 IT 기업의 비즈니스 운영을 위해 설립하는 데이터센터 지출 규모는 오히려 전년 대비 3% 감소한 수치를 기록했다. 코로나19로 인한 기업경영 악화와 투자 불안 심리 증가가 낳은 결과라고 볼 수 있다. 반면 클라우드와 관련된 데이터센터 지출 규모는 전년 대비 25% 증가한 수치를 기록했다.

[그림 9] 코로나19, 클라우드 데이터센터 이슈 발생 흐름도



출처 : 글로벌 ICT 뉴스매체 발행 이슈 4만 2932건 분석(2020.1. ~ 2020.9.)

13. 하이퍼스케일 데이터센터 : 통상 최소 10만 대 서버를 운영하며 2만 2500m² 이상의 규모를 갖춘 데이터센터

1 데이터 사용량 평균 30~35% 증가, 글로벌 데이터센터 시장 연평균 5% 내외 성장 중

코로나19 영향에 따른 원격 작업 증가로 데이터센터에 대한 수요가 급증하고 있다. 코로나19 확산으로 발생한 봉쇄령 및 재택근무(WFH)는 디지털과 온라인을 통한 거래 업무를 증가시켰으며, 줌(Zoom, 화상미팅), SAP(ERP), बैंकिंग과 같은 엔터프라이즈 도구 사용 증가는 데이터센터 시장 수요 증가에 중추적인 역할을 했다.

소비자 측면에서도 넷플릭스(Netflix)와 핫스타(Hotstar)와 같은 OTT 플랫폼¹⁴ 사용 증가가 데이터 트래픽 급증으로 이어졌으며, 이를 위한 강력한 데이터 인프라 체계가 요구되게 됐다.

인도의 세계적인 억만장자이자 히라난다니그룹(Hiranandani Group)의 창립자인 니란잔 히라난다니(Niranjan Hiranandani)는 “코로나19 이후 데이터 사용량은 평균 30~35% 증가했으며, 코로나19 종식 후에도 그 수요는 점차 증가할 것”이라고 강조했다.¹⁵

가. 동유럽, 데이터센터 시장 성장의 50%는 러시아가 주도

2019년, 동유럽 국가에서는 20개 이상의 신규 데이터센터 설립 프로젝트가 진행되었으며 이 중 50% 이상의 투자는 러시아에 의해 진행됐다. 러시아 외 주요 데이터센터 투자 국가는 폴란드와 체코 등으로 분류된다.

나. 영국, 250개 이상의 코로케이션센터¹⁶ 시설을 갖춘 유럽의 데이터센터 선도국

영국은 2019년과 2020년에 30개가 넘는 데이터센터 투자 프로젝트를 진행했다. 이 중 런던의 점유율은 85%에 달한다. 영국은 2020년 1월 31일 유럽연합(EU)에서 탈퇴하며, 기업과 클라우드 서비스 기업 간 데이터 저장에 대한 불확실성이 증가했으며 이를 위한 대대적인 데이터센터 사업 투자를 진행하고 있다.

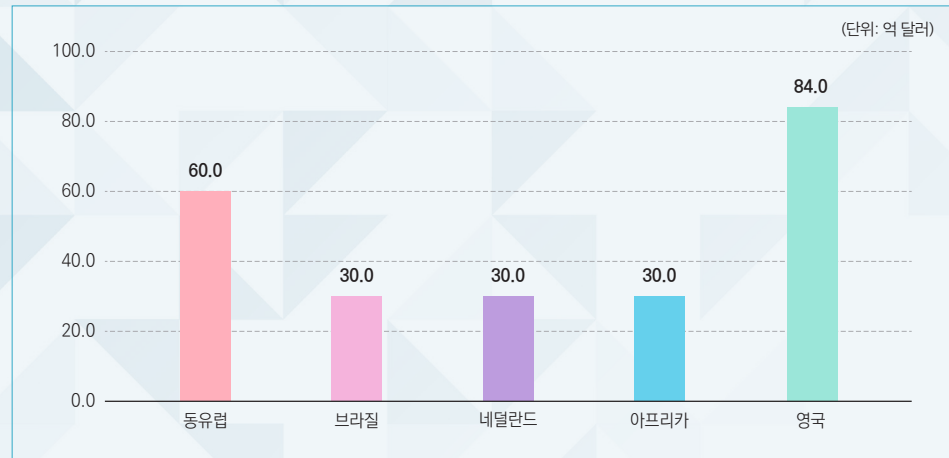
이외에 브라질(코로케이션센터 시장 성장 주도), 네덜란드(독일에 이은 유럽에서 두 번째로 높은 투자 규모), 아프리카(2025년까지 연평균 12% 고성장 전망) 등을 포함한 전 세계에서 데이터센터에 대한 투자 활동이 급증하고 있다.

14. OTT(Over The Top) : 인터넷을 통한 방송 콘텐츠 제공 서비스

15. Kailash Babar, Covid-19 Impact: Demand for data centers surge as remote working catches up, CIO(cio.economictimes.indiatimes.com), cio.economictimes.indiatimes.com/news/data-center/covid-19-impact-demand-for-data-centers-surge-as-remote-working-catches-up/75624084

16. 코로케이션 센터(Colocation Centre) : 소매 고객에게 장비, 공간 및 대역폭을 대여할 수 있는 데이터센터 유형

[그림 10] 주요 국가별 데이터센터 시장 전망치



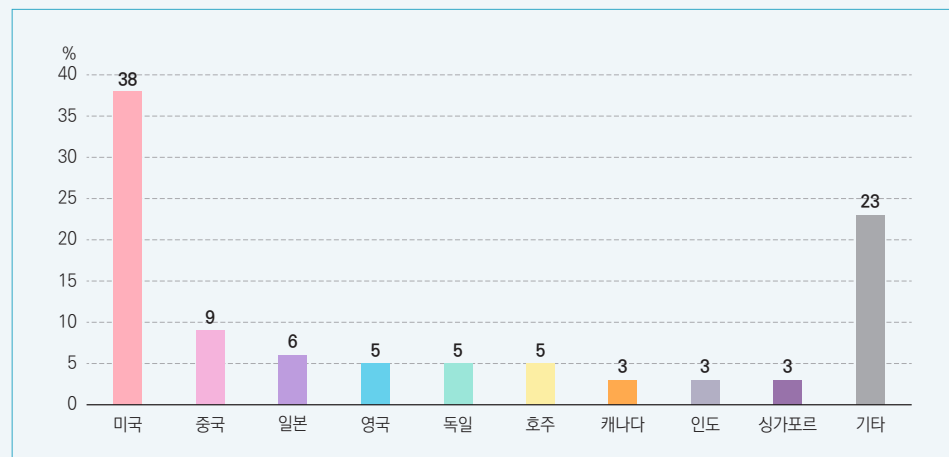
출처 : 아리즈톤(www.arizton.com)

2 전 세계 하이퍼스케일 데이터센터 설립 규모 581개, 2015년 대비 2배 증가

시장조사기관 시너지리서치(Synergy Research) 그룹에 의하면, 2020년 2분기 기준 전 세계에 설립된 하이퍼스케일 데이터센터 수는 2015년 같은 기간 대비 두 배 이상 늘어난 541개에 달할 것으로 보인다. 미국이 전체 시장의 38%를 차지하고 있는 가운데 중국, 일본, 인도, 싱가포르 등 아시아태평양 지역이 가장 높은 성장률을 보이고 있다.

아마존(Amazon)과 구글(Google)은 지난 1년 동안 신규로 설립된 데이터센터 수요의

[그림 11] 2020년 2분기 기준, 하이퍼스케일 데이터센터 국가별 점유율



출처 : 시너지리서치(Synergy Research)

50%를 점유했으며, 마이크로소프트(Microsoft)와 오라클(Oracle)이 그 뒤를 잇고 있다.

그 외 현재 투자가 결정돼 프로젝트가 진행중인 하이퍼스케일 단위의 데이터센터 수도 176개에 달해, 관련 하드웨어 공급 기업 및 데이터센터 운영 사업자를 대상으로 한 사업기회 증가도 당분간 지속될 것으로 전망된다.¹⁷

3 퍼블릭 클라우드, 데이터센터 시장 지출 사상 최고치

이처럼 높은 시장성장률을 기록하고 있는 데이터센터 시장의 성장 속도는 ‘클라우드 데이터센터’ 시장으로 국한되고 있다. 2020년 2분기 기준 데이터센터 시장 관련 총 지출액은 414억 달러에 달하며 전년 대비 7% 증가한 수치다. 데이터센터 시장 성장은 대부분 클라우드 데이터센터에 의해 진행되고 있는데, 클라우드 데이터센터 시장 지출이 전년 대비 25% 증가하며 170억 달러를 기록하는 동안, 델(Dell)·휴렛팩커드(HPE) 등의 일반 IT 기업의 데이터센터 시장 지출은 전년 대비 오히려 3% 감소했다.¹⁸

시너지리서치그룹 수석연구원 존 딘스데일(John Dinsdale)은 “코로나19 영향으로 클라우드 기업들의 데이터센터 지출은 증가했지만, 민간 기업들의 지출은 기업 경영 어려움으로 오히려 감소했다.”고 밝혔다.

이슈5. 국가별 상이한 ‘코로나19 추적 앱’ 상용화 정책

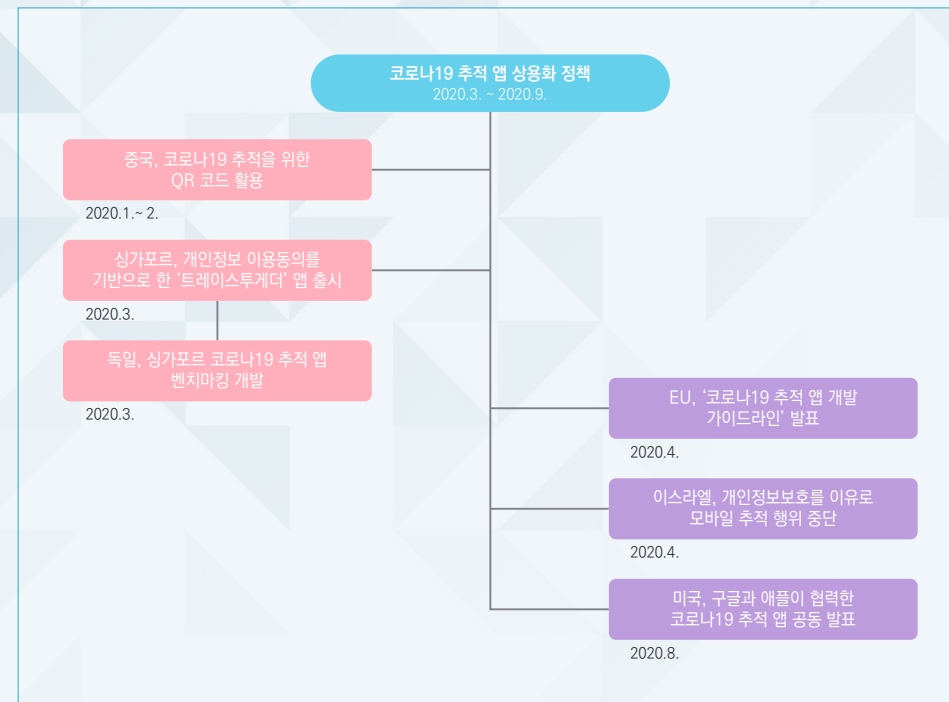
지난 10개월 동안 코로나19 사태 확산 방지를 위한 국가 차원의 코로나 추적 앱 개발이 이뤄지고 있다. 하지만 코로나19 확산 방지라는 공익과 개인정보 보호라는 인권문제가 맞물려 국가마다 서로 다른 코로나19 앱 정책이 운영되고 있다.

먼저 코로나19를 가장 먼저 겪은 중국은 QR 코드 인프라를 통해 코로나19 확진자를 모니터링하고 이동제한을 진행해 확산 통제에 나섰다. 국가 차원에서 코로나19 추적 앱을 배포한 국가는 싱가포르가 최초다. 싱가포르는 트레이스투게더(Trace Together)라는 앱을 발표하여 배포했으며 이후 7월에는 온라인 취약계층을 위한 코로나 추적틀을 배포했다.

17. Synergy : 「全球超大规模数据中心升至541个」, 176个在建, cnBeta(www.cnbeta.com), 2020.07.08, www.cnbeta.com/articles/tech/1000659.htm

18. Why Public Cloud Data Center Spending Is At An All-Time High, crn(www.crn.com), 2020.09.10, <https://www.crn.com/news/data-center/why-public-cloud-data-center-spending-is-at-an-all-time-high>

[그림 12] 코로나19 추적 앱 상용화 정책 이슈 발생 흐름도



출처 : 글로벌 ICT 뉴스매체 발행 이슈 4만 2932건 분석(2020.1. ~ 2020.9.)

한편 유럽에서 독일이 싱가포르 코로나 추적 앱을 벤치마킹해 개발했지만 시민단체들의 거센반발에 부딪쳤으며 덴마크·프랑스·이탈리아 역시 비슷한 상황에 직면했다. 이후 유럽연합집행위원회는 코로나19 추적 앱 가이드라인을 발표해 코로나19 추적이 개인정보보호보다 우선될 수 없다는 점을 밝혔다.

미국은 구글과 애플이 코로나19 앱 개발에 나섰으며 이후 미국 10여 개 주정부에 확대돼 보급이 확산되고 있다.

1 중국, 코로나19 추적을 위한 QR 코드 활용

지난 5월 2일, 중국 국가시장감독관리총국은 코로나19 관련 건강코드에 대한 국가 표준을 발표했다.¹⁹ 국가시장감독총국은 코로나19 건강 QR 코드 표준을 제정함으로써 도시

간 국내 여행을 하는 데에 도움을 줄 것이라고 발표했다. 코로나19 사태를 맞아 중국의 주요 지방정부는 코로나19 감염자 관리와 경제 활성화를 위해 독자적인 QR코드 제도를 시행해왔다. 표준 기준에 따르면 녹색 코드의 경우 도시간 이동이 가능하며, 노란색·빨간색 코드를 보유한 사람은 집에서 자가격리를 해야 한다. 중국 당국은 여행 이력, 발병 지역에서 보낸 시간, 잠재보균자와의 거리를 기반으로 자정에 국민들의 색상 코드를 변경한다. 현재 이 코로나19 건강 QR 코드는 텐센트와 알리바바가 운영하고 있다.

[그림 13] 중국 코로나19 추적 QR 코드 이미지



출처 : SCMP(South China Morning Post)

2 싱가포르, 개인정보 이용동의를 기반으로 한 '트레이스투게더' 앱 출시

싱가포르 정부테크놀로지국(GovTech)은 보건부(Ministry of Health)와 협력해 '트레이스투게더' 코로나19 추적 앱을 출시했다.²⁰ 정부 당국이 일반인에게 공급한 코로나19 추적 앱으로는 트레이스투게더가 세계 최초로 알려져 있다. 트레이스투게더는 구글플레이와 앱스토어를 통해 다운로드 받을 수 있으며, 앱이 설치된 스마트폰이 서로 가까워지면 단말기마다 부여된 ID가 블루투스를 통해 교환된다. 상호간의 스마트폰에 위치한 유저 ID를 기록하는 구조이며, 개인정보는 교환되지 않는다. 코로나19 확진자가 트레이스투게더 앱의 이용자라면 앱 기록을 보건부와 공유해 확진자 가까이에 위치한 스마트폰 소유자에게 연락을 취해 필요한 조치를 전달할 수 있다.

19. Cissy Zhou, *Coronavirus: China gives green light to more domestic travel by unifying its chaotic QR health codes*, SCMP(www.scmp.com), www.scmp.com/news/china/society/article/3082542/coronavirus-china-gives-green-light-more-domestic-travel

20. Dan Koh, *Singapore government launches new app for contact tracing to combat spread of COVID-19*, MobiHealth(MobiHealthnews.com), www.mobihealthnews.com/news/asia-pacific/singapore-government-launches-new-app-contact-tracing-combat-spread-covid-19

트레이스투게더로 방역효과를 거뒀다고 판단한 싱가포르 당국은 7월 트레이스투게더 토큰(Trace Together Token)이라는 휴대용 추적장치를 온라인 취약 계층에게 추가로 배포하기도 했다.

3 독일, 싱가포르 코로나19 추적 앱 벤치마킹 개발

독일 당국은 싱가포르 코로나19 추적 앱 트레이스투게더를 벤치마킹해 ‘코로나완앱(Corona-Warn-App)’을 발표했다.²¹ 독일의 코로나 추적 앱은 트레이스투게더와 마찬가지로 상호간 스마트폰 내 블루투스 교환을 원리로 제작됐으며, 코로나19 확진자와 15분 이상 접촉하는 사람은 자동으로 경고 메시지를 받게 된다.

하지만 독일 코로나19 추적 앱의 배포는 곧바로 국민들의 찬반 논란으로 이어졌다. 과거 나치 정권의 독재와 감시를 경험했던 노년층들은 해당 앱에 대해 반발하고 나섰으며, 전후 세대의 국민들은 대체적으로 코로나19 확산을 막기 위해서는 해당 앱이 필요하다고 주장하고 있다. 독일 보건부 장관은 코로나19 추적 앱 개발에 SAP와 도이치텔레콤(Deutsche Telekom)이 공동 참여했으며, 약 2,000만 유로가 소요됐다고 보고했다. 해당 앱이 실질적인 성과를 거두기 위해서는 전체 국민의 60%가 사용해야 할 것으로 보이지만 실제 다운로드는 44%에 그치고 있다.

4 EU ‘코로나19 추적 앱 가이드라인’ 발표

EU집행위원회는 코로나19 추적 앱 사용에 대한 ‘EU 공통 가이드라인(EU toolbox for the use of mobile applications for contact tracing and warning)’을 제정 및 발표했다. 이번에 발표된 가이드라인은 EU집행위원회가 EU 회원국과 보조를 맞춰 글로벌 팬데믹 대응 긴급조치에 대한 EU 차원의 단계적 해제조치의 일환으로 회원국이 준수해야 할 공통 기준을 담고 있다. 코로나19 추적 앱(디지털 도구)의 사용은 봉쇄령 해제 조치에 필요한 수단이지만, 개인정보 보호라는 유럽의 가치를 포기한 것이어서는 안 되며, 따라서 앱의 사용에 앞서 개인정보 보호조치가 보장돼야 한다고 가이드라인은 전했다. EU집행위원회는 제3자에 의해 처리되는 개인정보의 정확성을 보장하고, 접촉자 추적 앱의 효율성을 위해 블루투스와 같은 기술을 이용하라고 밝혔다. 코로나19 앱 사용을 위한 선결조건은 [표 14]와 같다.²²

21. Day one of using Germany's coronavirus tracing app, DW(dw.com), www.dw.com/en/day-one-of-using-germanys-coronavirus-tracing-app/a-53828730
22. Natasha Lomas, EU lawmakers set out guidance for coronavirus contacts tracing apps, TechCrunch (techcrunch.com), techcrunch.com/2020/04/16/eu-lawmakers-set-out-guidance-for-coronavirus-contacts-tracing-apps/

[표 14] 코로나19 추적 앱 사용에 대한 EU 공통 가이드라인

항목	세부 내용
정책명	EU toolbox for the use of mobile applications for contact tracing and warning
목적	글로벌 팬데믹 대응 조치에 대한 EU 회원국이 준수해야 할 공통 기준 제시
선결조건	· EU의 데이터 보호정책 준수 · 정부 의료당국의 승인과 인지하에 사용 · 사용자의 자발적인 설치원칙, 비사용시 삭제 원칙 · (블루투스 기술 채택 등) 최신의 개인정보보호기술을 적용 · 개인의 익명성 보장 · EU 국가 간 기술의 상호호환성 보장 · 전염병 대응 가이드라인을 준수하는 온라인 톨의 모범사례로 정착 · 안전성 및 실효성 보장

출처: edpb.europa.eu

5 이스라엘, 개인정보보호를 이유로 모바일 추적 행위 중단

한편 이스라엘 당국은 개인정보 문제를 언급하면서 코로나19 확진자 추적을 위해서 통신사에 로밍데이터를 요청하는 행위를 중단했다고 밝혔다. 2020년 3월, 이스라엘 총리는 경찰당국이 통신사로부터 휴대전화 위치데이터를 넘겨받도록 지시했지만 이후 시민단체들이 대법원에 이의를 제기하여 원상 복구하는 계기가 됐다.²³

이스라엘 경찰당국은 통신사로부터 데이터를 추적하는 것이 효율적이고, 지난 한 달 동안 203명에 달하는 격리조치 위반자들을 검거하는 데 결정적인 영향을 주었다고 밝혔다. 하지만 이스라엘 외교 및 국방위원회는 이번 조치의 연장을 불허했으며, 코로나19 추적은 개인정보보호 및 보안 권한에 대해 좀 더 체계적인 지침이 요구된다고 덧붙였다.

6 미국, 구글과 애플이 협력한 코로나19 추적 앱 공동 발표

미국의 스마트폰 운영체계를 양분하고 있는 구글과 애플은 코로나19 환자 추적 앱 개발 톨을 개발했다.²⁴ 해당 앱은 사전 동의(Opt-in) 방식으로 사용자가 설치 시 스마트폰이 블루투스 신호를 통해 가까운 거리의 다른 스마트폰의 기록을 수집해주며, 감염자로 등록된 이용자의 스마트폰 블루투스 신호가 잡히면 경고를 하는 형식으로 운영된다. 구글과 애플은 정부

23. Jaimie Davis, Israel halts mobile-tracking quarantine measures on privacy grounds, Telecom(Telecoms.com), telecoms.com/503911/israel-halts-mobile-tracking-quarantine-measures-on-privacy-grounds/
24. Kif Lewswing, Apple and Google release test version of coronavirus tracing software, CNBC(www.cnbc.com), www.cnbc.com/2020/04/29/apple-and-google-release-test-version-of-coronavirus-tracing-software.html

규정에 따라 신규 앱 개발자가 사용자들의 위치 및 스마트폰 개인정보 수집 기능을 자동으로 활성화 하지 못하게 해, 개인정보 침해를 방지했다고 밝혔다.

구글과 애플이 코로나19 추적 앱을 공동 발표한 지 6개월 후, 뉴욕과 뉴저지주가 코로나19 추적 앱을 도입하겠다고 밝혔다. 현재 미국 10개 주 정부가 코로나19 추적 앱을 사용하고 있으며 이는 약 7000만 명에 해당하는 수치다.²⁵

25. Kif Lewswing, *States are finally starting to use the Covid-tracking tech Apple and Google built — here’s why*, CNBC(www.cnbc.com), www.cnbc.com/2020/10/03/covid-app-exposure-notification-apple-google.html

참 고 자 료

이슈1. 코로나19에 따른 데이터산업 기술 발전의 변화

- *33% of UK Orgs Lack Tech Infrastructure for Long-Term Remote Working*, Infosecurity, 2020.3.24. <https://www.infosecurity-magazine.com/news/orgs-tech-remote-working/>
- *COVID-19’s impact on the enterprise and remote work*, CIO, 2020.3.16. <https://www.cio.com/article/3532812/covid-19s-impact-on-the-enterprise-and-remote-work.html>
- *‘Machine learning modellen getraind op pre-corona data nuttelos’*, Techzine, 2020.6.24. <https://www.techzine.nl/nieuws/data/443132/machine-learning-modellen-getraind-op-pre-corona-data-nutteloos/>
- *Where Predictive Machine Learning Falls Short and What We Can Do About It*, InsideBIGDATA, 2020.6.27. <https://insidebigdata.com/2020/06/27/where-predictive-machine-learning-falls-short-and-what-we-can-do-about-it/>
- *IBM to Acquire WDG Automation to Advance AI-Infused Automation Capabilities*, Database, 2020.7.13. <https://www.dbta.com/Editorial/News-Flashes/IBM-to-Acquire-WDG-Automation-to-Advance-AI-Infused-Automation-Capabilities-141754.aspx>
- *Automation Anywhere Named a Leader in the 2020 Gartner Magic Quadrant for Robotic Process Automation*, CISION, 2020.7.30. <https://www.prnewswire.com/news-releases/automation-anywhere-named-a-leader-in-the-2020-gartner-magic-quadrant-for-robotic-process-automation-301103293.html>

이슈2. 개인정보유출과 인권침해 논란에 휩싸인 ‘온라인 수업’

- *ANU commits to online exams, invigilation despite student concerns*, ITnews, 2020.4.20. <https://www.itnews.com.au/news/anu-commits-to-online-exams-invigilation-despite-student-concerns-546894>
- 「新型コロナウイルスでオンライン学習なら安心は84.3% = バンザン調べ =」, ICT-enews, 2020.3.1. <https://ict-enews.net/2020/03/01ban-zan/>
- *Under Half of Teachers Think Schools Have “Done Enough” to Tackle Cybersecurity Issues*, Infosecurity, 2020.8.10. <https://www.infosecurity-magazine.com/news/teachers-schools-tackle/>
- *Over 1,000 parents sign open letter protesting facial recognition in schools*, Venturebeat, 2020.9.16. <https://venturebeat.com/2020/09/16/over-1000-parents-sign-open-letter-protesting-facial-recognition-in-schools/>

이슈3. 코로나19 기간 ‘정보유출 및 정보보안’ 중요성 대두

- *Fake antivirus site promises coronavirus protection, delivers trojan*, Computer Weekly, 2020.3.24. <https://www.computerweekly.com/news/252480520/Fake-antivirus-site-promises-coronavirus-protection-delivers-trojan>
- *Domini infetti, crescono del +600% quelli contenenti le parole “Coronavirus” e “Covid-19”*, Key4biz, 2020.4.14. <https://www.key4biz.it/domini-infetti-crescono-del-600-quelli-contenenti-le-parole-coronavirus-e-covid-19/300185/>
- *Attacks on Banks Spike 238% During #COVID19 Crisis*, Infosecurity, 2020.3.15. <https://www.infosecurity-magazine.com/news/attacks-on-banks-spike-238-during/>
- *Cloud cyber attacks up seven-fold during coronavirus pandemic*, ITPro, 2020.3.28. <https://www.itpro.co.uk/cloud/cloud-security/355815/cloud-cyber-attacks-increased-seven-fold-over-coronavirus-pandemic>

- *Digital financial fraud predicted to witness rise of up to 80%*, Itweb, 2020.6.1.
https://www.itweb.co.za/content/VgZeyvJoGA87djX9/zlP3gQ2qGRMnRD1W
- *230K+ Indonesian COVID-19 Patients Personal Information Leaked in the Darknet*, Cybleinc, 2020.6.21.
https://cybleinc.com/2020/06/21/230k-indonesian-covid-19-patients-personal-information-leaked-in-the-darknet/
- *Data on 18,105 coronavirus patients leaks after staffer clicks wrong button*, Cnet, 2020.9.15.
https://www.cnet.com/health/data-on-18105-coronavirus-patients-leaked-after-staffer-clicks-wrong-button/

이슈4. 클라우드 데이터센터 시장 지출 규모 '사상 최대치'

- *Covid-19 Impact: Demand for data centers surge as remote working catches up*, CIO, 2020.5.8.
https://cio.economicstimes.indiatimes.com/news/data-center/covid-19-impact-demand-for-data-centers-surge-as-remote-working-catches-up/75624084
- *Data Center Market Size*, Arizton, 2020.5.~8.
https://www.arizton.com/
- Synergy : 「全球超大规模数据中心升至541个 , 176个在建」, cnBeta, 2020.7.8.
https://www.cnbeta.com/articles/tech/1000659.htm
- *Why Public Cloud Data Center Spending Is At An All-Time High*, CRN, 2020.9.10.
https://www.crn.com/news/data-center/why-public-cloud-data-center-spending-is-at-an-all-time-high

이슈5. 국가별 상이한 '코로나19 추적 앱' 상용화 정책

- *Contact Tracing, le linee guida dei Garanti Privacy Ue per le app anti Covid-19*, Key4biz, 2020.4.15.
https://www.key4biz.it/contact-tracing-le-linee-guida-dei-garanti-privacy-ue-per-le-app-anti-covid-19/300368/
- *Singapore government launches new app for contact tracing to combat spread of COVID-19*, Mobihealth, 2020.3.20.
https://www.mobihealthnews.com/news/asia-pacific/singapore-government-launches-new-app-contact-tracing-combat-spread-covid-19
- *Israel halts mobile-tracking quarantine measures on privacy grounds*, Telecoms, 2020.4.24.
https://telecoms.com/503911/israel-halts-mobile-tracking-quarantine-measures-on-privacy-grounds/
- *Germany aims to launch Singapore-style coronavirus app in weeks*, Reuters, 2020.3.30.
https://www.reuters.com/article/us-health-coronavirus-germany-tech/germany-aims-to-launch-singapore-style-coronavirus-app-in-weeks-idUSKBN21H26Z
- *Apple and Google partner on COVID-19 contact tracing technology*, 2020.4.10.
https://www.apple.com/newsroom/2020/04/apple-and-google-partner-on-covid-19-contact-tracing-technology/

“2020년 데이터산업 국내 정책 및 비즈니스 이슈”

‘2020 데이터 산업 국내 정책 및 비즈니스 이슈’ 조사는 2019년 9월부터 2020년 8월까지 최근 1년 동안 국내에서 발생한 정책 및 비즈니스 측면의 주요 데이터 산업 이슈를 분석하여, 국내 데이터산업의 최근 트렌드를 파악하였다.

기간 내 수집된 이슈 기사 수는 총 383만 5669건에 달하며, 이 중 데이터산업과 관련된 4만 493건의 기사를 바탕으로 가장 많이 빈출된 국내 데이터 산업 이슈를 선정하였다.

아래 내용은 데이터 이코노미, 글로벌 뉴스 트렌드 한국편(한국데이터산업진흥원, 2020.12) 자료를 발췌하여 작성하였다.

디지털 뉴딜(Digital New Deal)과 데이터 3법으로 대표되는 한국의 2020년 데이터 정책

디지털 뉴딜 정책 및 데이터 댐 프로젝트

한국은 7월 14일, 코로나 19 사태로 인한 극심한 경제 침체 극복과 구조적 대전환 대응이라는 이중 과제에 직면한 상황에서 「한국판 뉴딜」 종합계획의 일환으로 ‘디지털 뉴딜’ 정책을 발표하였다. 정부는 ‘디지털 뉴딜’에 2025년까지 국비 44.8조원(총 사업비 58.2조 원)을 투자하여 ①디지털 생태계 강화 및 교육인프라의 디지털 전환, ②비대면 사업 육성, ③SOC 핵심 인프라 디지털화, ④도시·산단·물류 등의 스마트화 및 연관 산업 경쟁력을 제고 등의 주요 사업 목표와 노력을 시도하고 있다. 또한, 정부는 디지털 뉴딜 정책의 일환으로 데이터 산업 인력 확대 및 일자리 창출을 위한 ‘데이터 댐 프로젝트’를 핵심 과제로 삼고 데이터의 양과 질 확장에 따른 새로운 시대에 적응해나가고자 하고 있다.

마이데이터 실증 사업 본격화 · 데이터센터 경쟁 심화 · 코로나19 예방

2020년 6월, 과학기술정보통신부에 의하여 발표된 6개 산업 분야, 8개 사업 프로젝트에 데이터 민간 기업들이 본격적으로 참여하기 시작하며 향후 관련 분야의 데이터 산업 발전 가속화가 기대되고 있다. 이밖에도 네이버, 카카오, 오라클 등 주요 IT 기업들에 의한 국내 데이터센터 확보 경쟁이 보다 심화되고 있는 것으로 나타났으며, 통신사·제조사·병원 간 협력을 통한 ‘언택트 스마트 병원 구축’ 사례가 증가하고 있다.

2020년 데이터 기술은 코로나19 예방에도 큰 역할을 수행하고 있는데, 카이스트와 KT의 주도 하에 코로나19 예방을 위한 데이터 활용 사례들이 나타나고 있음. 포스코·KT·효성 그룹 등에 의한 스마트팩토리개발 사업 확대도 2020년의 주요 이슈로 꼽혔다.

2020년 데이터산업 국내 정책 이슈 랭킹 Top 5

1

데이터3법 시행으로 인한, 한국 데이터산업의 패러다임 변화

- 2020년 8월 5일, 데이터3법 시행
- 금융, 의료 산업 등을 필두로 산업 변화 전망



2

디지털 뉴딜 정책의 핵심, 데이터 댐 프로젝트 추진

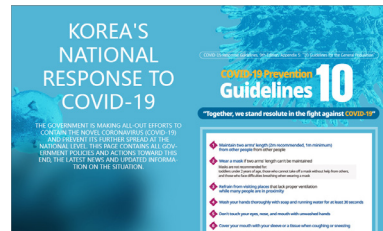
- 2025년까지 총 58.2조 원 투자(국비 44.8조 원)
- 2025년까지 일자리 90.3만 개 창출 목표



3

전세계 모범사례로 인정 받는 한국의 데이터를 활용한 K-방역

- 코로나19 확산 방지를 위한 데이터 활용
- 한국 데이터 활용 방역 사례, 해외에서 벤치마킹



4

한국 공공데이터 정책, 3년 연속 OECD 1위 기록

- 지방정부의 활발한 공공 데이터 개방 확대
- 공공데이터 정책 수준, OECD 국가 중 1위 달성



5

데이터 분석 인력 양성 및 일자리 고용 정책 확대 추진

- 주요 정부기관 '빅데이터 인력 양성 교육' 실시
- 데이터산업 일자리 확대 및 고용 정책 추진



※ 출처 : 「DATA ECONOMY, Global News Trends in Korea」 한국데이터산업진흥원, 2020.12.

2020년 데이터산업 국내 비즈니스 이슈 랭킹 Top 5

1

데이터 민간 기업, 마이데이터 실증 사업 본격 참여

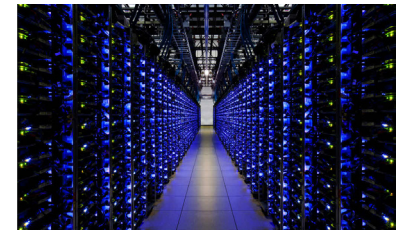
- 과학기술정보통신부 8개 마이데이터 과제 선정
- 의료, 금융, 공공, 생활, 소상공인, 교통 분야 적용



2

국내외 IT 기업, 데이터센터 확보 경쟁 심화

- 네이버 2번째 데이터센터 '각 세종' 10월 착공
- 카카오, 한양대 캠퍼스에 첫번째 데이터센터 설립
- 오라클, 강원도 춘천에 두번째 데이터센터 설립



3

증가하는 데이터 & 헬스케어 융합 사례

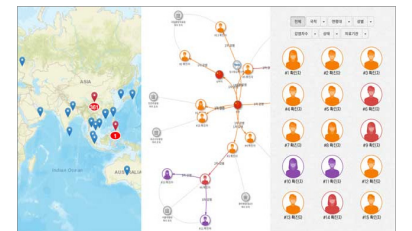
- 주요 기업 및 의료 업계, 디지털 헬스케어 서비스 개발
- 통신사, 제조사, 병원 '연택트 스마트 병원 구축'



4

데이터 기술, 코로나19 예측 및 예방에 활용

- 카이스트, 해외 유입 확진자 수 예측 AI 모델 개발
- KT, AI/빅데이터 기반 감염병 대응 연구 프로젝트



5

스마트 팩토리, 프로젝트 및 협력 사례 급증

- KT, AI 기반 5G 스마트팩토리 비전 출시
- 효성ITX & 포스코ICT, 스마트팩토리 개발 업무협약
- SKT, 5G 기반 스마트공장 구축 지원



2020년 데이터산업 백서

C O N T E N T S



역동적인 데이터산업 생태계 지각 변동과 기회

발간사	002
그림으로 보는 데이터산업 동향	004
코로나19 확산에 따른 데이터산업 글로벌 이슈 분석	008
2020년 데이터산업 국내 정책 및 비즈니스 이슈	031

제1장 데이터 권리와 연결, 그리고 새로운 가치

1. 데이터 권리	044
2. 데이터 연결	049
3. 데이터 가치	052

제2장. 데이터 주도 경제 성장과 발전

1. 데이터 경제 활성화의 본격 시작	054
2. 국내 데이터3법의 시행과 데이터 경제	056
3. 국가 디지털 뉴딜 정책	058
4. 데이터 거래소 운영	059
5. 마이데이터 사업	062
6. 데이터 보안	064
7. 마무리	065

제3장. 새로운 시장 생태계의 지각 변동

1. 데이터 생태계 형성과 성장	066
2. 데이터 생태계의 정의 및 특징	068
3. 데이터 시장 생태계의 역동성과 지각 변동	069

데이터산업 주요 정책 현황

제1장. 국내 데이터산업 정책 현황

1. 한국판 뉴딜 정책과 방향	074
2. 디지털 뉴딜 정책 배경과 주요 내용	075
3. 디지털 뉴딜 대표 과제와 주요 내용	077



데이터산업 시장 현황

제2장. 해외 데이터산업 정책 현황

1. 미국의 정책	089
2. 유럽의 정책	093
3. 일본의 정책	101
4. 중국의 정책	104
5. 동남아시아의 정책	106

제1장. 국내 데이터산업 시장 현황

1. 국내 데이터산업 시장규모	110
2. 국내 데이터산업 부문별 시장규모	113
3. 국내 데이터직무 인력 현황	117
4. 데이터직무 인력 수요	122
5. 마무리	126

제2장. 해외 데이터산업 시장 현황

1. 데이터산업의 시장규모	127
2. 데이터 기업 수	132
3. 데이터 전문인력	133
4. 데이터산업의 경제적 효과	135
5. GDP 대비 데이터 경제적 효과의 비율	137
6. 중장기 데이터 시장 전망	138



산업별 데이터 활용 비즈니스 현황

제1장. 금융 분야 데이터 활용 비즈니스 현황

1. 금융 데이터 시장과 정책 변화	142
2. 금융 데이터 융합을 위한 기술환경 변화	145
3. 데이터 활용과 금융업 비즈니스 사례	148
4. 향후 기대 및 전망	151

제2장. 헬스케어 분야 데이터 활용 비즈니스 현황

1. 디지털 헬스케어 시장규모와 성장	152
2. 디지털 헬스케어 분야의 데이터 활용 확대	153



데이터산업 기술 동향

3. 디지털 헬스케어 데이터와 다른 데이터의 융합 비즈니스	157
4. 향후 전망과 기대	158

제3장. 제조 분야 데이터 활용 비즈니스 현황

1. AI와 접목한 제조업 시장 동향	160
2. 주요 기업 동향	161
3. 제조업에서 AI 도입의 도전	162
4. 제조업에서 AI 적용 분야	166
5. 제조업에 AI 주요 도입 사례	167
6. 향후 제조업의 AI 적용 전망	169

제4장. 통신 분야 데이터 활용 비즈니스 현황

1. 국내외 통신시장 현황	171
2. 통신분야 데이터 비즈니스 동향	174
3. 향후 전망	179

제5장. 광고 분야 데이터 활용 비즈니스 현황

1. 개인데이터를 기반으로 한 광고 사업의 성장	181
2. 데이터를 기반으로 한 광고 시장의 진화	185
3. 데이터 기반 광고 시장의 미래	189

제6장. 교육 분야 데이터 활용 비즈니스 현황

1. 세계 에듀테크 산업 및 기업 동향	192
2. 에듀테크 정책 추진 동향	196
3. 에듀테크 발전을 위한 제언	199

제7장. 유통·물류 분야 데이터 활용 비즈니스 현황

1. 유통 물류 서비스의 혁신	200
2. 생활 물류서비스의 성장, 새벽배송과 배달 대행 서비스	203
3. 데이터 분석 기반의 유통 및 물류 서비스 혁신과 이슈	206
4. 데이터 기반의 유통 및 물류 산업의 혁신 가속화	208

제1장. 데이터 생산 및 수집 기술

1. 데이터 생산 및 수집 기술 개요	212
2. 데이터 생산	213
3. 데이터 수집	215

제2장. 데이터 저장 및 관리 기술

1. 데이터 기술 변화의 동인	222
2. 클라우드 DBMS 방향성과 한계	223
3. 클라우드 DBMS 기술의 필요 요소	226
4. DBMS 진화의 선택업, 클라우드, AI	230

제3장. 인공지능 기술

1. 인공지능 개요	231
2. 딥러닝	232
2. 자연어처리	235
3. 엣지 컴퓨팅	237

제4장. 데이터 보안과 비식별 기술

1. 데이터 보안 기술의 새로운 흐름	239
2. 개인정보 보호 및 활용 시대의 절차와 기술	243

제5장. 데이터 분석 및 시각화 기술

1. 데이터 바로 보기	250
2. 데이터와 대화하기	251
3. 데이터 시각화 시장 동향	251
4. 데이터 시각화 기술 동향	252
5. 데이터 시각화의 미래	254

2020년 데이터산업 백서 집필진	255
--------------------	-----



[그림 1-1-1]	비즈니스 모델 기반 마이데이터 비즈니스 기획 절차	053
[그림 1-2-1]	빅데이터 플랫폼 통합 데이터지도	055
[그림 1-2-2]	데이터3법 개정으로 바뀌는 산업들	057
[그림 1-2-3]	데이터바우처 지원 사업 구조	060
[그림 1-2-4]	데이터 거래소의 거래 절차	062
[그림 1-2-5]	마이데이터 활용 구조	062
[그림 1-3-1]	데이터 생태계 라이프 사이클	069
[그림 1-3-2]	데이터 기술 관점에서 생태계 연계 및 통합성 강화	070
[그림 1-3-3]	데이터 생태계의 데이터 순환구조와 역동성	071
[그림 2-1-1]	한국판 뉴딜(2020-2025)	075
[그림 2-1-2]	디지털 뉴딜의 4대 분야와 목표	075
[그림 2-1-3]	디지털 서비스 전문계약 제도 운영방식	080
[그림 2-1-4]	차세대 지능형 교통시스템(C-ITS) 체계도	082
[그림 2-1-5]	디지털 트윈의 구성	082
[그림 2-2-1]	미국 연방 데이터 전략의 2020년 실천계획 개요	092
[그림 2-2-2]	독일 GAIA-X 프레임워크	100
[그림 3-1-1]	최근 5개년(2015~2019(E)) 데이터산업 시장규모	111
[그림 3-1-2]	최근 5개년(2015~2019(E)) 데이터산업 부문별 시장규모 비중	112
[그림 3-1-3]	2019년(E)~2025년(P) 데이터산업 시장 전망	112
[그림 3-1-4]	최근 5개년(2015~2019(E)) 데이터 솔루션 시장규모	113
[그림 3-1-5]	2019년(E) 데이터 솔루션 중분류별 시장규모 비중	114
[그림 3-1-6]	최근 5개년(2015~2019(E)) 데이터 구축·컨설팅 시장규모	115
[그림 3-1-7]	2019년(E) 데이터 구축·컨설팅 중분류별 시장규모 비중	115
[그림 3-1-8]	최근 5개년(2015년~2019년(E)) 데이터 서비스 시장규모	116
[그림 3-1-9]	2019년(E) 데이터 서비스 중분류별 시장규모 비중	117
[그림 3-1-10]	최근 5개년(2015~2019년) 데이터산업 전체 인력 현황	118
[그림 3-1-11]	2019년 데이터산업 인력 및 데이터직무별 인력 비중	119
[그림 3-1-12]	2019년 전 산업의 데이터직무 인력 현황	120
[그림 3-1-13]	향후 5년 내 데이터산업 부문별 데이터직무 필요 인력	122
[그림 3-1-14]	향후 5년 내 데이터산업의 기술등급별 필요 인력 비중	123
[그림 3-1-15]	향후 5년 내 데이터산업 부문별 데이터직무 인력 부족률	123
[그림 3-1-16]	향후 5년 내 데이터산업·일반산업·전산업 데이터직무 필요 인력	124
[그림 3-1-17]	향후 5년 내 전 산업 내 데이터직무 인력 부족률	124
[그림 3-1-18]	향후 5년 내 전 산업의 필요 인력 기술등급별 비중	125

[그림 3-2-1]	2017~2022년 데이터 기반 솔루션 전체 시장규모	128
[그림 3-2-2]	전 세계 대륙별 데이터 기반 정보서비스 시장 비교	129
[그림 3-2-3]	코로나19에 따른 2020년 데이터 기반 정보서비스 시장규모 변화	130
[그림 3-2-4]	2016~2020년 디지털 데이터 시장규모	131
[그림 3-2-5]	2016~2020년 디지털 데이터 기업 수	132
[그림 3-2-6]	2016~2020년 데이터 전문인력 수	133
[그림 3-2-7]	2016~2019년 경제적 효과 중 직접 효과	136
[그림 3-2-8]	2014~2018년 경제적 효과: 간접 후방 효과	136
[그림 3-2-9]	2017~2019년 GDP 대비 경제적 효과의 비율 (직접 효과 + 간접 후방 효과)	138
[그림 3-2-10]	2019년~2025년 EU의 중장기 데이터 시장 전망	139
[그림 4-1-1]	디지털 금융시장의 성장속도	143
[그림 4-1-2]	금융데이터거래소 소개	144
[그림 4-1-3]	클라우드 도입의 편익과 리스크	146
[그림 4-1-4]	금융분야 AI 활성화 워킹그룹 구성안	148
[그림 4-3-1]	제조기업 시장규모	161
[그림 4-3-2]	제조업의 AI 도입·구축 과정에서 어려움	164
[그림 4-4-1]	모바일 트래픽 추이	172
[그림 4-4-2]	국내 이동통신 가입자 및 ARPU 전망	172
[그림 4-4-3]	2019년 디지털 광고 지출이 최소 50%에 달하는 국가	175
[그림 4-4-4]	통신사 데이터를 활용한 코로나19 확산의 징후 방송사례	177
[그림 4-5-1]	2019~2024년 전 세계 디지털 광고비	182
[그림 4-5-2]	2019년 디지털 광고 지출이 최소 50%에 달하는 국가	182
[그림 4-5-3]	2014~2018년 미국 프로그래매틱 디지털 광고비	183
[그림 4-5-4]	2018년 유럽 전체 디지털 광고 대비 프로그래매틱 광고비 비중	183
[그림 4-5-5]	구글, 페이스북, 네이버의 기업별 전체 매출 대비 광고 매출의 비중	184
[그림 4-5-6]	2019, 2020 구글과 페이스북 vs. 아마존의 디지털 광고 점유율	184
[그림 4-5-7]	가격 모델에 따른 2005~2018년 인터넷 광고 수익	186
[그림 4-6-1]	세계 에듀테크 시장 전망과 교육 시장규모	193
[그림 4-6-2]	국내 에듀테크 시장규모 및 성장률	193
[그림 4-6-3]	미래교육 전환을 위한 데이터 활용 정책 방향	198
[그림 4-7-1]	온라인 쇼핑물 거래 규모	201
[그림 4-7-2]	택배 물동량 및 이용 현황	202

[그림 4-7-3]	코로나19 이후 온라인 쇼핑 및 택배물동량 변화	203
[그림 4-7-4]	택배 물류서비스와 말단 배송 서비스 다양화	204
[그림 4-7-5]	새벽배송 시장의 변화	205
[그림 4-7-6]	새벽배송 시장의 변화	205
[그림 5-1-1]	IoT 플랫폼 구성	214
[그림 5-1-2]	IoT 기반 서비스 데이터 저장 및 처리를 위한 SW 기술요소	215
[그림 5-1-3]	엔터프라이즈 기업의 EAI 구성	216
[그림 5-1-4]	데이터 3법 개정에 따른 오픈 API 플랫폼 필요성	218
[그림 5-1-5]	오픈 API 플랫폼 구성도	219
[그림 5-1-6]	글로벌 APIM 시장 현황	220
[그림 5-2-1]	국내 클라우드 시장 현황	223
[그림 5-2-2]	DB 서버 확장의 한계	224
[그림 5-2-3]	데이터 통합의 한계	224
[그림 5-2-4]	현재의 MSA 구조	225
[그림 5-2-5]	5티어 아키텍처와 2티어 DB 구조	226
[그림 5-2-6]	DB 노드 간 경합 구조	227
[그림 5-2-7]	가상 DB 구조	227
[그림 5-2-8]	다양한 데이터 형태 통합	228
[그림 5-2-9]	지능화한 DB 관제	229
[그림 5-2-10]	5G DBMS로의 발전	230
[그림 5-4-1]	가명 처리 단계별 세부 절차도	246
[그림 5-4-2]	가명정보 결합, 반출 절차	248
[그림 5-4-3]	익명화 처리 2단계	249
[그림 5-5-1]	쿼리 기반 시각화와 연관 분석 시각화 비교	252

[표 2-1-2]	과학기술정보통신부의 데이터 댐 관련 7대 핵심 사업	078
[표 2-1-3]	스마트 의료 인프라의 현재와 미래	080
[표 2-1-4]	디지털 뉴딜 4개 분야 12대 과제	083
[표 2-2-1]	유럽 데이터 전략의 4가지 축 개요	094
[표 2-2-2]	영국의 국가 데이터 전략의 주요 내용	098
[표 3-1-1]	최근 5개년(2015~2019(E)) 데이터산업 시장규모 총괄	111
[표 3-1-2]	최근 5개년(2015~2019(E)) 데이터 솔루션 중분류별 시장규모	114
[표 3-1-3]	최근 5개년(2015~2019(E)) 데이터 구축·컨설팅 중분류별 시장규모	115
[표 3-1-4]	최근 5개년(2015~2019(E)) 데이터 서비스 중분류별 시장규모	117
[표 3-1-5]	최근 5개년(2015~2019년) 데이터산업 전체 인력 현황	118
[표 3-1-6]	최근 5개년(2015~2019년) 데이터산업 부문별 데이터직무 인력 현황	119
[표 3-1-7]	최근 4개년(2016~2019년) 전 산업의 데이터직무 인력 현황	120
[표 3-1-8]	2019년 전 산업의 데이터직무별 인력 현황	121
[표 3-1-9]	최근 5개년(2015~2019년) 국내 주요 산업별 시장규모 추이	126
[표 3-2-1]	2016~2020년 유럽연합의 데이터 전문인력 수	134
[표 4-1-1]	AI 기반 금융서비스 분류	147
[표 4-1-2]	데이터 활용과 금융 서비스 예시	150
[표 4-2-1]	비즈니스에서 많이 활용되는 데이터 원천	153
[표 4-2-2]	임상진료 데이터 활용 사례	154
[표 4-2-3]	유전체 포함 오믹스 데이터 활용 사례	155
[표 4-2-4]	개인 건강 데이터 활용 사례	156
[표 4-2-5]	청구 데이터 활용 사례	156
[표 4-2-6]	디지털 헬스케어 데이터와 타산업 융합 비즈니스 사례	157
[표 4-3-1]	IT 서비스 대기업의 스타트업 인수 현황	161
[표 4-3-2]	B2C AI와 B2B AI 비교	163
[표 4-3-3]	제조 데이터 분석의 어려움	163
[표 4-3-4]	제조업에서 AI 기술 주요 적용 분야	166
[표 4-3-5]	제조업 AI 적용사례 요약	168
[표 4-6-1]	국외 에듀테크 서비스 동향	194
[표 4-6-2]	국내 에듀테크 서비스 동향	195
[표 5-1-1]	사물인터넷 데이터 생산 방식	214
[표 5-4-1]	신용정보법에 따른 개인정보의 개념과 활용 가능 범위	244
[표 5-4-2]	식별 정보와 식별 가능 정보(예시)	245



[표 1-1-1]	데이터 권리의 주체 비교	046
[표 1-1-2]	개인데이터 규제의 적용 대상 비교	047
[표 1-1-3]	개인데이터 통제권 비교	047
[표 1-1-4]	개인데이터 법규 위반에 따른 제재 비교	048
[표 1-2-1]	디지털 뉴딜 4대 분야 12개 추진과제	058
[표 1-2-2]	2020년 데이터바우처 사업 지원 규모	060
[표 1-2-3]	2020년 본인정보 활용지원(MyData) 실증서비스 선정과제 목록	063
[표 1-3-1]	데이터 생태계의 구성 요소와 촉진 역할	067
[표 2-1-1]	디지털 뉴딜 4대 분야 12개 추진과제와 목표	076

제 1 부

역동적인
데이터산업 생태계
지각 변동과 기회

제1장.
데이터 권리와 연결, 그리고 새로운 가치

제2장.
데이터 주도 경제 성장과 발전

제3장.
새로운 시장 생태계의 지각 변동

제1장. 데이터 권리와 연결, 그리고 새로운 가치

필자: 김인현
(투이컨설팅 대표)

데이터 경제의 진전에 따라 데이터 권리의 법제화가 전 세계적으로 진행되고 있다. 데이터주체인 개인의 통제권을 좀 더 보장하는 추세다. 세부적으로 나라마다 차이는 있지만 데이터 이동권이 핵심이 되고 있다. 데이터 이동권에 따라 새로운 데이터 비즈니스가 탄생할 수 있는 길이 열렸고, 데이터 연결이 확대되고 있다. 데이터 거래와 데이터 결합에 의해 고객을 완전하게 이해할 수 있다. 오픈 API를 이용한 데이터 접속이 활발해짐에 따라 서비스가 고객 니즈 중심으로 재편되고 있다. 데이터의 가치는 상대적이다. 데이터에서 가치를 찾아내 활용하는 것은 전통적 방식으로서는 불가능하다. 비즈니스 모델링, 서비스 디자인, 애자일 피드백 등의 방법론을 적용해야 한다.

1. 데이터 권리

가. 데이터 권리가 떠오른 배경

오성 이항복의 감나무 일화는 유명하다. 이항복 집의 감나무 가지가 옆집 권율 대감 집으로 넘어갔다. 권율 대감 집의 하인들이 그 나무의 감을 따먹자, 이항복은 권율 대감집을 찾아가 창호지에 주먹을 밀어 넣고 물었다. “이 방 안에 들어온 손은 누구 것입니까?” 권율 대감은 대답했다. “그야 자네 손이지.” 이후로 권율 대감의 하인들은 함부로 이항복의 집에서 넘어온 감을 따먹지 못했다.

다시 생각해보자. 감나무 가지가 옆집으로 넘어가면 잎이 떨어질 것이고, 하인들은 이를 청소해야 할 것이다. 그럼에도 그 감나무의 감을 완전히 이항복 집의 것으로 인정하는 것이 옳을까? 개인데이터의 권리도 같은 관점으로 생각할 수 있다. 개인데이터는 개인이 생성하지만, 실제로 데이터를 처리·가공하는 주체는 데이터를 다루는 기업 또는 기관이기

때문이다.

개인데이터의 권리를 어디까지 인정해야 할 것인가를 놓고 다양한 논의가 있었다. 2014년, 우리나라에서 카드3사의 개인데이터 유출사건이 발생했다. 이후 개인데이터 보호를 강화하는 방향으로 제도와 정책이 펼쳐졌다. 대형 할인점인 홈플러스는 2011년부터 2014년까지 개인정보 712만 건을 보험회사들에게 판매해 231억 원의 추정 손익을 올렸다. 하지만 손해배상으로 판결된 금액은 8,465만 원에 불과했다. 이 사건은 소비자 단체들이 기업의 개인정보 활용에 경각심을 갖게 된 계기가 됐다.

이러한 배경으로, 우리나라는 세계에서 개인데이터 보호가 가장 강력한 나라가 됐다. 디지털 기술의 급속한 발전으로 데이터가 양산되기 시작했고, 기업과 기관들은 대량의 개인데이터를 확보하게 됐다. 기업과 기관들은 개인데이터를 활용하면서 비용 절감뿐 아니라, 그 개인(고객)에게도 더 편리한 맞춤형 서비스를 개발·제공할 수 있게 됐다.

나. 데이터 권리의 제도화

모든 국가는 데이터경제를 뒷받침할 수 있도록 데이터 권리를 제도화하고자 노력해왔다. 그 결과 2018년 5월 25일, EU(유럽 연합)는 세계 최초로 개인데이터 보호를 강화한 GDPR(General Data Protection Regulation)을 시행했다. 미국 캘리포니아주는 개인정보보호법인 CCPA(California Consumer Privacy Act)를 2020년 1월부터 시행했다. CCPA는 미국 역사상 가장 포괄적인 개인정보보호법으로 평가받고 있다. GDPR과 CCPA는 정보 주체인 개인의 자기 정보 통제 권한의 범위와 개인데이터 사용 조직의 의무를 명확하게 함으로써 데이터 활용을 확대할 수 있는 기반을 제공했다는 평가를 받는다.

우리나라는 2020년 1월 9일에 데이터3법이 국회본회의를 통과했다. 데이터3법은 정보통신망법(정보통신망 이용촉진 및 정보보호 등에 관한 법률, 이하 정보통신망법), 신용정보법(신용정보의 이용 및 보호에 관한 법률, 이하 신용정보법), 개인정보보호법 등을 말한다. 이 중에서 데이터주체인 개인의 권리를 규정한 법률은 신용정보법과 개인정보보호법이다. 신용정보법은 정보주체의 데이터 이동권을 규정함으로써, 최초의 마이데이터 관련 법률이 됐다. 하지만 신용정보법의 대상인 신용정보는 금융 및 상거래 등에 관한 정보로 한정돼, 의료·공공·교육 등 타 분야에는 적용되지 않는다.

GDPR, CCPA, 신용정보법 등에서 정하고 있는 정보주체의 권리는 ‘소유권’이 아닌 ‘통제권’이라고 보는 것이 맞다. 우리나라 민법은 ‘소유자는 법률의 범위 내에서 그 소유물을 사용, 수익, 처분할 권리가 있다.’고 규정하고 있다. 소유권의 권능은 사용, 수익, 처분 등을 포함한다. 개인데이터는 개인이 동의했더라도, 동의를 받은 조직이 사용·수익·처분 등을 마음대로 할 수는 없다. 또한 데이터주체인 개인도 완전한 소유권이 아닌 자기 데이터를

통제할 수 있는 권리를 갖는다.

데이터 권리를 정의하기 위해서는 데이터주체, 데이터 권리범위, 통제 대상 조직, 위반 시 제재 등을 명확하게 해야 한다. 신용정보법, GDPR, CCPA 등은 큰 맥락에서는 정보주체인 개인의 권리를 더 강력하게 보장한다는 점에서는 같지만, 세부적으로는 다른 점도 있다. 개별 기업은 영위하고 있는 사업이 해당 법규에 영향을 받는지 판단하고, 해당될 경우 상응하는 조치를 해야 한다. 글로벌 비즈니스를 하고 있는 기업이라면 GDPR과 CCPA의 데이터 권리 규제에도 대응해야 할 것이다.

다. 각국의 데이터 권리 제도 비교

데이터 권리 주체의 대상에 대한 정의는 각 국의 법제도에서 일부 차이가 있다. 우리나라 개인정보보호법에서는 정보주체를 ‘처리되는 정보에 의해 알아볼 수 있는 사람으로서 그 정보의 주체가 되는 사람’으로 정의하고 있다. 신용정보법에서 본인신용정보의 대상인 정보주체는 ‘처리된 신용정보로 알아볼 수 있는 자로서 그 신용정보의 주체가 되는 자’로 정의하고 있다. 신용정보의 주체는 개인과 법인이 해당될 수 있지만, 본인신용정보관리업에서 대상은 개인으로 한정하고 있다. CCPA는 가계도 포함함으로써, IoT 등 센서 데이터도 마이데이터 대상에 속하는 효과를 거뒀다.

[표 1-1-1] 데이터 권리의 주체 비교

구분	권리 주체
신용정보법	개인인 신용정보주체(살아 있는 개인)
GDPR	EU 거주자(살아 있는 개인)
CCPA	미국 캘리포니아에 거주하는 캘리포니아 주민 (자연인뿐 아니라 가계도 포함)

비즈니스를 영위하는 모든 기업과 대국민 서비스를 하는 모든 조직은 필연적으로 개인데이터를 취득하고 관리하며 활용하게 된다. 개인데이터를 보유하지 않은 조직은 거의 없을 것이다. 따라서 법에서 어디까지를 규제 대상으로 할 것인가가 중요한 문제가 된다. 우리나라 신용정보법에서는 개인 신용정보를 제공하거나 이용하는 조직으로 대상을 정하고 있다. EU의 GDPR은 EU 지역 거주자를 대상으로 비즈니스를 제공하는 모든 조직을 그 대상으로 한다. CCPA는 공공 조직은 제외하고 있으며, 영리 조직 중에서 특정 기준을 충족하는 조직으로 대상을 한정하고 있다.

[표 1-1-2] 개인데이터 규제의 적용 대상 비교

구분	규제 적용 대상
신용정보법	신용정보를 제공한 이용자 및 신용정보를 보유한 공공기관 (대통령령으로 정의)
GDPR	EU에서 사업장을 운영하며 개인정보 처리와 EU 거주자에게 재화나 서비스를 제공하고, EU 거주자의 EU 내 행동을 모니터링하는 기관과 조직 (공공기관 및 비영리 부문 등을 포함해 모든 조직에 적용)
CCPA	연간 총 매출액 2,500만 달러를 초과하고 연간 5만 명(개) 이상의 상업적 목적의 개인정보를 처리하며 연간 매출액의 50% 이상이 캘리포니아 소비자 정보 판매로 발생하는 조직 (일정 조건을 갖춘 영리 사업자에게만 적용)

개인의 자기 데이터 통제 권리를 체계적으로 정의한 사례는 EU의 GDPR이 최초다. GDPR은 데이터 권리를 8가지로 정의했다. 디지털 시대의 개인이 가져야 할 권리를 정의했다는 점에서 의미가 크다. 특히 정보이동권을 규정함으로써, 개인데이터 서비스를 전문으로 하는 제3의 기업 탄생을 준비하고, 이를 통해 개인이 자신의 데이터를 더 잘 활용할 수 있도록 했다는 점이 특징이다.

우리나라 신용정보법은 GDPR의 데이터 권리를 모두 반영하고 있다. 미국의 CCPA에서 규정한 데이터 권리는 판매중단지시권⁰¹과 서비스평등권⁰²을 그 특징으로 한다. GDPR은 포괄적으로 개인 동의를 사전에 요구하고 있는 반면, CCPA는 데이터 판매에 대해 사후적으로 중단을 요청할 수 있는 권리로 정했다.

[표 1-1-3] 개인데이터 통제권 비교

데이터 권리	신용정보법	GDPR	CCPA
정보를 제공받을 권리	○	○	○
정보에 접근할 권리	○	○	○
정보 정정권	○	○	
정보 삭제권(잊혀질 권리)	○	○	○

(계속)

01. 판매중단지시권 : 개인정보의 판매 중단을 지시할 권리로 옵트아웃(Opt-out) 권리를 의미
02. 서비스평등권 : 데이터주체인 개인이 자신의 데이터에 대해 어떠한 요구나 권리행사를 할 때, 서비스를 차별화해 제공하면 안 된다는 규정

데이터 권리	신용정보법	GDPR	CCPA
처리 제한권	○	○	
개인정보 이동권	○	○	○
반대할 권리(OPT IN)	○	○	
판매중단지시권리(OPT OUT)			○
자동화처리 선택권	○	○	
서비스 평등권			○

개인데이터 권리 보장을 위해서는 권리를 지키지 않을 경우, 이에 대한 분명한 제재가 필요하다. 우리나라는 형사처벌, 과징금, 손해배상, 과태료 등의 제재를 사안별로 정의하고 있다. GDPR은 강력한 벌금 규정을 두고 있다. 반면 CCPA는 형사벌 제도는 도입하지 않은 대신에 소비자에 대한 손해배상 제도를 규정하고 있다. 개인데이터 권리 침해에 따른 손해배상이 개인에게 직접 돌아가기 때문에 개인이 자신의 피해 내용에 대해 더 용이하게 보상받을 수 있는 제도를 갖추는 것이 중요하다.

[표 1-1-4] 개인데이터 법규 위반에 따른 제재 비교

구분	제재 규정
신용정보법	과징금은 전체 매출액의 100분의 3이하와 50억 원 이하 손해배상 책임과 사안별 과태료(3,000만 원 ~ 1억 원) 양벌 규정 적용
GDPR	전 세계 연간 총 매출의 최대 4% 또는 2,000만 유로 중 큰 금액으로 벌금 부과
CCPA	형사벌 제도는 도입하지 않음 소비자 집단 소송, 법정 손해배상, 실제 손해배상, 민사 벌금제도 (손해배상액은 소비자당 100 ~ 750 미국 달러와 실제 손해배상액 중 높은 금액으로 하되 상한은 없음)

데이터 권리는 개인으로 이동하고 있다. 정확하게는 개인과 조직의 데이터 권리의 균형으로 수렴하고 있다. 이전까지는 개인이 자신의 데이터를 통제하고, 자신의 이익을 위해서 활용하기보다는 기업이 기업의 이익을 위해 데이터를 사용해온 것이 사실이다. 개인이 자신의 데이터 권리를 실질적으로 행사하는 시대는 2020년부터 열리고 있다. 개인데이터와 관련된 비즈니스들이 탄생하고, 개인은 자신의 데이터로 더 큰 혜택을 받는 시대가 다가오고 있다.

2. 데이터 연결

가. 데이터3법과 데이터 연결

데이터3법이 발효된 이후 두드러진 현상은 데이터 거래가 크게 늘어났다는 점이다. 금융보안원은 2020년 5월에 금융 데이터거래소를 출범시켰다. 출범 이후 4개월 동안 월 평균 거래액은 1억 100만 원 수준이었다. 데이터3법이 발효된 8월 12일부터 9월 7일까지 발생한 거래액은 3억 5,700만 원으로 이전과 비교하면 세 배 이상 늘어난 규모다. 법이 발효된 이후 시행령과 가이드라인 등이 이어서 발표됐다. 아직 데이터 거래 개념이 생소한 상황임에도 이러한 거래 증가 추세가 시사하는 바는 크다.

데이터는 상거래의 결과로 발생한다. 데이터가 생성·보관되는 기업은 자신이 담당할 거래에 대한 데이터만 확보할 수 있다. 카드회사는 카드의 결제 내역을 알 수는 있지만, 무슨 물품을 얼마나 샀는지는 알 수 없다. 반대로 카드 가맹점은 구매 내역은 알 수 있지만, 고객이 얼마나 결제를 충실히 하는지는 알 수 없다. 구매와 결제 데이터를 통합해 분석할 수 있다면, 카드회사는 고객이 선호하는 물품을 알고 구매력을 판단할 수 있게 된다. 분석 결과를 토대로 마케팅 성과를 크게 높일 수 있다.

데이터 융합은 원래의 목적을 지키고 개인데이터 보호를 준수할 수만 있다면, 데이터 주체인 개인에게도, 상품과 서비스를 제공하는 기업에게도 좋은 일이다. 개인은 더 편리한 서비스를 받을 수 있고, 기업은 경영 활동의 효율과 효과를 높일 수 있다. 데이터 융합이 이뤄지기 위해서는 데이터를 주고받을 수 있는 연결이 필요하다. 데이터를 주고받는 방법은 두 가지가 있다. 첫째, 데이터를 한 번에 주고받는 방법이다. 이를 데이터 거래라고 부른다. 데이터 보유 기업은 데이터 구매 기업으로부터 판매 대금을 받을 수 있다. 가명화 기술을 사용한다. 둘째, 데이터를 거래 수준에서 실시간으로 주고받는 방법이다. 이는 데이터 접속이라고 부른다. 오픈 API(Open Application Programming Interface)를 사용한다.

나. 데이터 거래

고객을 완전하게 파악하려면 ‘누가, 언제, 어디서, 무엇을, 어떻게, 왜?’에 대한 데이터를 가져야 한다. ‘누가, 언제’라는 정보는 대부분의 기업이 확보할 수 있다. 다만 ‘누가’에 대한 데이터의 상세화 수준은 업종에 따라 크게 다르다. 여신과 보험 등의 금융상품 판매를 위해서는 개인에 대한 상세정보를 알아야 한다. 상거래 회사들은 개인을 식별할 수 있는 정도의 데이터만 갖고 있다. ‘어디서’에 대한 데이터는 고객의 위치정보를 알 수 있는 통신사가 유리하다. ‘무엇’에 대한 데이터는 금융거래의 경우는 금융회사가 유리하고, 상거래의 경우는 인터넷 쇼핑 또는 가맹점이 유리하다. ‘왜’는 고객과 상담을 하거나, 고객 접점을 담당하고 있는 회사들이 알기

쉽다. 최근 비대면 거래 비중이 증가함에 따라 인터넷 상거래 업체들이 '왜'를 파악하는 데 더 유리한 위치에 있다.

데이터 거래는 파편화된 고객 데이터를 융합해 완전한 답을 찾는 방향으로 진행되고 있다. 신용카드사의 고객 정보와 결제 데이터, 편의점 등 상거래 업체의 물품 구매 내역, 통신사의 위치정보 등을 결합하면 고객 행동을 정확하게 파악할 수 있는 데이터를 갖게 된다. 금융 그룹사가 협력해 은행, 증권, 보험, 카드 등의 데이터를 결합하면 개별 고객의 금융 자산과 흐름을 완전하게 파악할 수 있다. 은행의 고객 소득 및 자산 정보, 물류회사의 온라인 채널 택배정보, 통신사의 IPTV 시청정보 등을 융합하면 개별 고객의 성향과 소비행태, 상권 분석 등이 가능하다.

데이터 융합을 위해서는 데이터 거래와 데이터 결합이 이뤄져야 한다. 데이터 거래는 한국데이터거래소와 금융데이터거래소 등이 담당하고 있다. 데이터 융합은 데이터전문기관을 통해서만 할 수 있다. 데이터전문기관은 기업들이 신청한 데이터를 안전하게 결합한 후 정보 주체를 알아볼 수 없도록 익명 또는 가명처리해 전달하는 역할을 한다. 데이터전문기관으로 신용전문기관과 금융보안원이 지정됐다. 민간 데이터전문기관은 추후 심사를 거쳐 선정할 계획이다.

데이터를 활용해 혁신적인 서비스를 개발하려고 하지만, 데이터를 확보하기 위한 자금이 부족한 기업들을 위해서 데이터바우처 제도가 운용되고 있다. 데이터바우처 지원 사업은 중소기업, 소상공인들이 데이터를 확보할 때 데이터 구입비용을 지원한다. 데이터 구매바우처와 데이터 가공바우처 두 종류가 있다. 데이터 구매바우처는 데이터 구매 비용을, 데이터 가공바우처는 데이터 처리 비용을 지원한다. 데이터바우처 사업은 과학기술정보통신부의 사업으로 한국데이터산업진흥원이 주관하고 있다.

다. 오픈 API를 통한 데이터 접속

관세청은 2015년부터 통관·무역·물류 분야의 공공데이터를 오픈 API로 연결해주는 서비스를 시작했다. 2020년 6월 한 달 동안 API 이용 건수는 사상 최대인 15억 건을 돌파한 것으로 발표됐다. 관세청이 제공하는 오픈 API는 모두 30종이며, 올해 말까지 40종으로 늘릴 계획이다. 서울대학교 산학협력단의 연구 결과에 따르면, 오픈 API를 이용하는 기업들의 월간 물류비용 절감효과는 4,725억 원으로 추정된다. 행정안전부는 주민등록등초본, 병적증명서 등의 전자증명서를 오픈 API방식으로 제공해 개인이 자신의 스마트폰에서 전자증명서를 제출할 수 있도록 하는 프로젝트를 2020년말까지 완료할 계획이다.

오픈 API는 프로그램 간의 데이터 전달로 이뤄지며, 사용자는 오픈 API를 통해서 다른 프로그램을 사용할 수 있게 된다. 오픈 API는 데이터에 발을 달아주는 것과 같다. 오픈 API로

개방하지 않으면 그 데이터는 보관하고 있는 회사의 애플리케이션만 이용할 수 있게 된다. 오픈 API는 데이터 보유조직들이 실시간으로 거래를 연계해서 처리할 수 있도록 한다. KB손해보험은 아파트 주소와 동 호수만 입력하면 공공기관이 제공하는 오픈 API를 이용해 건축물 대장 등을 확인해 보험료를 바로 산출해주는 서비스를 한다. 우리은행은 네이버와 오픈 API로 협업해 고객에게 영업점 번호표 발급 서비스를 한다. 카드사의 미래 비즈니스는 결제 수수료 기반에서 오픈 API 사용료로 바뀔 것이라는 전망도 있다.

신용정보법에 의해서 본인신용정보관리회사가 등장하게 되면 개인의 신용정보를 보유하고 있는 모든 조직은 고객이 원할 경우, 본인신용정보관리회사들에게 고객의 신용정보를 전송해야 한다. 해당 기업의 의사와 상관없이 기업과 기업의 연결은 불가피하다. 연결은 오픈 API로 데이터를 전송하는 형태가 될 것이다. 신용정보 보관 조직은 본인신용정보관리회사에 데이터를 오픈 API 형태로 제공해야 한다. 대부분의 금융회사들이 이에 해당한다. 이를 위해 금융회사들은 개인신용정보 전송을 위한 오픈 API를 개발한 뒤 API 적합성 테스트를 받아야 한다. 본인신용정보관리회사는 API 호출을 위한 자격증명을 발급받아서 자체 서버에 적용해야 한다. 금융보안원은 이를 위한 마이데이터 테스트베드 구축을 추진하고 있다. 2020년 10월에 착수해 2021년 2월에 시스템 구축을 완료할 예정이다.

비즈니스 성장동력 차원에서도 오픈 API는 중요한 역할을 한다. 데이터가 디지털 경제의 자원이라면, 데이터로 수익을 창출하는 수단이 오픈 API이기 때문이다. 오픈 API는 보유하고 있는 데이터의 가치를 실현한다. 또한 서비스 혁신에도 적용된다. 디지털 트랜스포메이션은 현재의 비즈니스 모습에서 디지털 비즈니스 모습으로 바꾸는 것을 의미한다. 변화의 큰 흐름은 언번들링(unbundling)과 리번들링(rebundling)이다. 과거의 서비스는 기업의 경계 안에 머물러 있었다. 언번들링 현상에 따라 기업의 사업 경계가 무너지기 시작했다. 서비스가 기업 밖으로 나오게 된 것이다. 리번들링은 해체된 서비스들이 고객의 니즈를 중심으로 다시 묶이는 것을 뜻한다. 예를 들면 지불결제와 상거래가 묶여서 쇼핑 단계를 간단하게 하는 것이다.

오픈 API를 도입·적용하는 것은 데이터 경제의 입장권을 사는 것과 같다. 생태계의 고객, 파트너, 공급자 등과 데이터 연결은 오픈 API로 이뤄질 것이기 때문이다. 오픈 API가 적용돼 있지 않으면 데이터 경제 생태계에서 외톨이가 될 수밖에 없다. 앞으로 잘 만들어진 오픈 API가 고객을 유입시키고, 상품 판매를 촉진하는 역할을 할 것이다. 오픈 API는 불특정 다수와 데이터 접속 방식을 약속하는 것이기 때문에 글로벌 표준, 국가 표준, 산업 표준을 준수하는 것이 중요하다.

3. 데이터 가치

가. 데이터 가치의 의미

“내가 그의 이름을 불러주기 전에는
그는 다만
하나의 몸짓에 지나지 않았다
내가 그의 이름을 불러주었을 때
그는 나에게로 와서
꽃이 되었다

- 김춘수, ‘꽃’

데이터는 그 자체로는 가치를 갖지 않는다. 데이터를 이해하고, 잠재된 효용을 찾아내면 가치를 갖게 된다. 이런 점에서 김춘수 시인의 ‘꽃’ 시구가 주는 의미와 일맥상통한다. 데이터의 가치는 상대적이다. 재고 추이 데이터는 상인의 신용정보 판단에는 유용하지만, 좋아할 만한 영화 추천에는 쓸모가 없다. 대규모의 콜센터 데이터는 인공지능을 다룰 줄 아는 환경을 갖춘 데이터 사이언티스트에게는 좋은 선물이지만, 서버 운영자에게는 쓸모없을 수 있다. 관건은 누가 어떻게 데이터를 바라보는가이다.

지금 활동하는 대부분의 기업은 산업경제 또는 정보화시대에 창립돼 성장해왔다. 2020년은 데이터 경제가 본격화되는 시점이다. 이전까지 기업의 성장 패러다임이나 전략의 변화가 필요하다. 하지만 기존의 경영자, 임원, 전문가들은 데이터의 가치를 찾아내는 데 익숙하지 않다. 자동차 제조사의 주산물은 자동차였다. 하지만 데이터 경제가 본격화되면 자동차 제조회사의 주산물은 모빌리티 데이터가 되고, 자동차는 다만 모빌리티 데이터를 얻기 위한 수단이 될지도 모른다. 데이터를 중심으로 비즈니스를 바라볼 수 있어야 데이터의 가치를 찾아낼 수 있다.

나. 데이터 가치를 찾아서 실현하는 방법

전통적 기업은 프로세스 혁신에 치중했다. 데이터가 디지털 경제의 원유라고는 하지만 어떻게 다루어야 하는지는 알지 못한다. 전통적 기업의 혁신은 엔지니어링 접근법을 택해 왔다. 가장 먼저 하는 일은 요구사항이 무엇인가이다. 그리고 요구사항을 좀 더 잘 다루기 위한 목표 모델을 찾는 방식이다. 하지만 데이터 기반의 혁신은 주어진 요구사항에서 출발하지 않는다. 데이터가 있으므로 새로운 사업모델과 혁신을 시작하는 것이다. 기존의

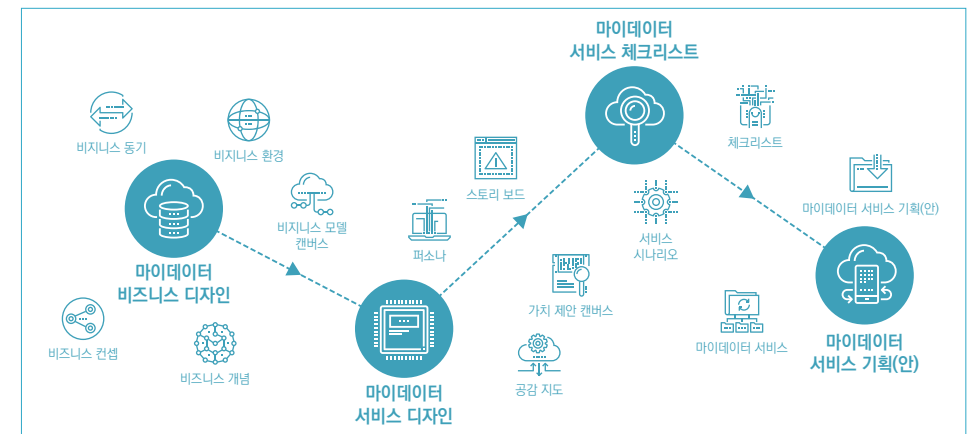
혁신 방법론으로는 해결책을 얻을 수 없다. 데이터를 ‘꽃’으로 부를 수 있는 접근방법이 필요하다.

1) 비즈니스 모델링

이미 보유하고 있는 또는 확보할 수 있는 데이터 자산 현상을 조사한다. 데이터 자산과 관련된 실체들의 애로 사항(pain point)과 잠재 기회(gain point)를 중심으로 비소비를 분석한다. 비소비가 해결됐을 때의 가치 곡선지도를 작성해 비즈니스 모델의 가치를 확인한다.

목표 가치 곡선을 달성할 수 있는 비즈니스 모델을 기획한다. 비즈니스 모델은 비즈니스 모델 캔버스와 인과관계도(Casual Loop Diagram, CLD)로 그린다. 비즈니스 모델의 타당성을 체크리스트를 토대로 검토하고 보완한다.

[그림 1-1-1] 비즈니스 모델 기반 마이데이터 비즈니스 기획 절차



2) 서비스 디자인

비즈니스 모델의 목표 고객 몇 개의 퍼소나(persona)로 구분한다. 퍼소나별로 공감지도, 가치 제안 캔버스, 스토리보드 등을 작성하는 과정을 통해 서비스를 구체화한다. 서비스 시나리오를 작성하고, 프로토타입을 통해서 확인한다. 최종 체크리스트로 검토하고 보완을 거쳐서 서비스 기획안을 완성한다.

3) 애자일 피드백

요구사항에서 출발한 것이 아니므로 서비스를 출시한 이후 고객과 시장의 반응을 빠르게 읽고 피드백해야 한다. 이를 위해 모니터링 데이터 수집, 성과 지표 분석, 백로그 관리 등을 수행한다. 데이터옴스(DataOps)를 도입하는 것이 바람직하다.

제2장. 데이터 주도 경제 성장과 발전

필자: 박영일
(강원연구원 원장)

2020년은 데이터 경제 활성화를 위한 제반 정책들이 본격 추진된 해라고 할 수 있다. 빅데이터 플랫폼을 본격 가동하면서 이를 통한 혁신 신규 서비스가 본격적으로 창출되기 시작했다. 2020년 8월, 데이터3법 개정의 시행으로 마이데이터 산업 발전과 데이터 거래의 활성화가 가속화하고 있다. 전 산업의 디지털 대전환을 가속화하기 위해 디지털 뉴딜 정책이 확대 추진되고 있다. 데이터바우처, 중소기업의 빅데이터 활용·분석 활성화 지원, 데이터 플래그십 사업 등 다양한 데이터 유통 촉진 정책이 전개되고 있다. 한편 디지털 경제에서 다양한 첨단 신기술들이 적용됨에 따른 사이버 보안의 우려를 예방하기 위한 데이터 보안에 대한 관심과 노력이 강조되고 있다.

1. 데이터 경제 활성화의 본격 시작

가. 데이터 권리가 떠오른 배경

세계는 데이터를 미래 경제를 견인할 핵심 동력으로 보고, 데이터 시장을 주도하기 위해 사활을 걸고 있다. 이에 따라 글로벌 ICT 기업의 부상이 더욱 뚜렷해졌다. 10년 전, 미국 증시 상위 10개 기업 중 IT 및 디지털 관련 기업이 2개였던 것에 반해 2020년 8월에는 5개로 늘었다. 미국, 중국의 각 상위 5개 기업의 시가총액 합은 약 8,092조 원과 2,211조 원이다. 이는 한국 정부 올해 본 예산(512조 원)의 각각 16배, 4배에 달한다.⁰¹

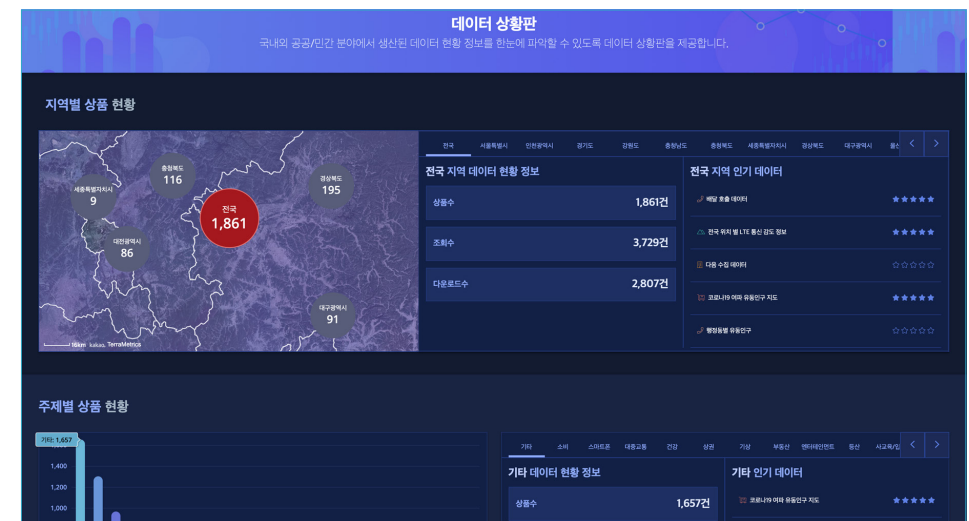
국가별 대응도 예사롭지 않다. 미국과 중국, EU 등의 주요 국가는 데이터 경제를 향한 패권 경쟁에 돌입했다. 국가별 정부 차원에서 데이터를 활용해 시장경제 확대를 꾀하는

01. 「주요국 ICT 기업의 시가총액 변화와 디지털기업의 부상」 보도자료, 전국경제인연합회, 2020.8.10.

동시에 데이터 안보를 강조하는 정책을 다수 내놴다.⁰² 또한 목적 외 데이터 이용과 무분별한 수집, 외부 데이터 유출을 정의하고, 자국 경제 부흥을 꾀하고 국민을 보호할 수 있는 대응 정책들을 빠르게 내놓고 있다.

한국 역시 2020년에 들어서 그동안 축적해 온 데이터 인프라를 기반으로, 데이터 개방 및 활용을 확대해 경제 활성화를 본격화했다. 앞선 2019년에는 구축한 10개 분야의 빅데이터 플랫폼의 1400여 종 데이터를 전면 개방·유통하고, 이를 기반으로 한 다양한 혁신 서비스 발굴을 시작했다. 또 신뢰할 만한 양질의 데이터를 유통하기 위해 빅데이터 플랫폼 간의 연계 공통기준을 마련해 플랫폼 간 연계 촉진에 나섰다. 2020년 3월 31일에는 빅데이터 플랫폼 통합 데이터지도 서비스도 개시했다.

[그림 1-2-1] 빅데이터 플랫폼 통합 데이터지도



출처: <https://www.bigdata-map.kr/>, 2020.08.26. 접속

2020년 1월 9일, 국회 본회의를 통과한 ‘개인정보보호법’, ‘정보통신망 이용촉진 및 정보보호 등에 관한 법률’, ‘신용정보의 이용 및 보호에 관한 법률’ 등 이른바 데이터3법 개정 법률안이 2020년 2월 4일에 공포됐다. 그로부터 6개월이 지난 2020년 8월 5일부로 시행됐다.⁰³

02. 유럽, A European Strategy for Data 및 White paper on Artificial Intelligence

중국, 신인프라 육성 정책 및 중화인민공화국 데이터보안법 초안

미국, 캘리포니아주 소비자 프라이버시법(CCPA) 및 안면인식사용기술 관련 법안

03. 단, ‘신용정보의 이용 및 보호에 관한 법률’ 개정안의 일부 규정들은 공포 후 1년 또는 1년 6개월 내 대통령령이 정하는 날에 시행될 예정

데이터3법에 근거해 활용 가능한 데이터의 종류가 다양해지고 데이터의 가치가 제고됨에 따라 다양한 분야에서 새로운 혁신 서비스 창출을 활성화할 것으로 기대된다. 아울러 고품질의 데이터 확보를 위한 다양한 데이터 보유 민간기업·기관들의 참여 확대 유도, 기업들의 데이터 활용 지원, 특히 중소·벤처기업들의 데이터 구매 및 가공 서비스 활용 지원과 이를 통한 새로운 서비스 창출 지원, 데이터 관련 스타트업 발굴·육성 등 다양한 지원 정책을 통해 민간 부문의 데이터 비즈니스가 확충되고 있다.

이 관점에서 2020년은 가히 '데이터 경제 활성화의 원년'으로 기록될 만하다.

2. 국내 데이터3법의 시행과 데이터 경제

가. 데이터3법의 주요 내용

데이터3법 개정법은 기존의 개인정보 보호 관련 법률의 내용과 체계에서 크게 탈피한 모습이다. 개인정보 관련 체계를 개인정보, 가명정보, 익명정보로 명확히 하고, 가명정보의 처리에 관한 상세한 규정을 신설했으며, 개인정보보호위원회를 일원화한 감독기구로 설치했다. 또한 정보통신 서비스 제공자들의 개인정보 보호 관련 조항들을 '개인정보보호법'으로 이관했다. 아울러 기존에 금융회사들에 주로 적용됐던 신용정보법에 의한 개인정보 보호문제가 일반 상거래에서도 적용된다는 점을 감안해, 일반 상거래에서의 신용정보법 관련 조사 등의 권한을 금융위원회 및 금융감독원이 아닌 개인정보보호위원회로 이관했다.

나. 데이터3법이 데이터 경제에 미치는 영향과 과제

데이터3법의 의의는 활용 가능한 데이터의 범위를 확대하면서도 명확히 함으로써 데이터 혁신 서비스를 창출할 수 있는 법적 기반을 마련했다는 데에 있다. 개정법은 가명정보를 활용할 수 있는 근거를 확대 규정했다. 즉 개인정보처리자가 통계작성, 과학적 연구(기술의 개발과 실증, 기초연구, 응용연구 및 민간 투자 연구 등 과학적 방법을 적용하는 연구), 공익적 기록 보존 등을 위해 정보주체의 동의 없이 가명정보를 처리할 수 있도록 했다.

또한 익명정보는 개인정보보호법의 적용 대상에서 명확히 제외했고, 당초 수집 목적과 합리적으로 관련된 범위 내에서 이미 수집·보유하고 있는 개인정보를 정보주체의 동의 없이 수집·이용·제공할 수 있도록 함으로써 데이터 경제 활성화에 큰 축을 세우게 됐다. 다만 정보의 활용 조건 및 적용 범위에 산업 목적이 포함되는지 등이 명확하지 않아 업계의 쟁점이 되고 있다. 그 결과에 따라 정보의 실제 활용 범위가 달라질 수 있다는 지적이 있어, 이에 대한

면밀한 검토와 대응이 필요하다.

데이터3법 개정법에 따라 데이터를 다루고 있는 산업에서는 많은 변화가 예상된다. 주요 변화를 네 가지로 요약하면 다음과 같다.

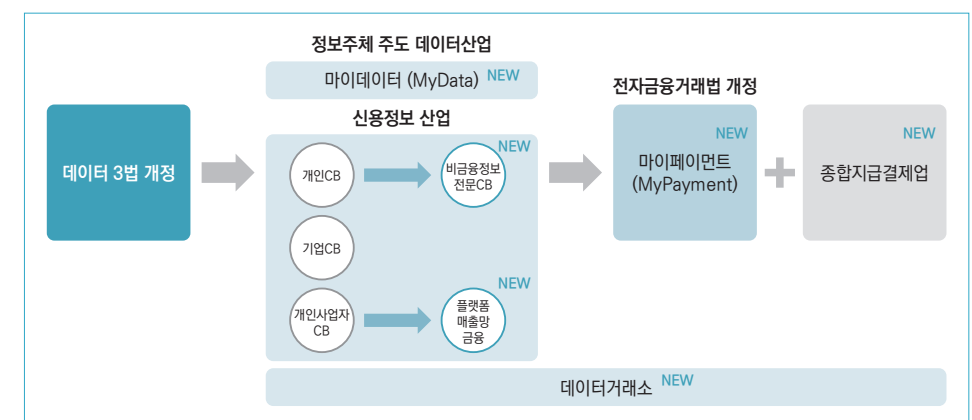
첫째, 마이데이터(MyData) 산업이 본격화할 것이다. 마이데이터 사업자는 은행, 보험회사, 카드회사 등의 개인신용정보를 통합해 개인의 정보관리를 돕고, 맞춤형 상품 추천과 금융상품 자문 같은 새로운 혁신 서비스를 제공할 수 있게 된다.

둘째, 데이터 거래의 활성화로 다양한 산업 생태계가 전개될 것이다. 데이터 거래소에서 데이터를 구매해 기존 사업들을 고도화하거나 새로운 데이터 관련 산업에 진출하는 등 새로운 비즈니스 기회가 늘어날 것이다.

셋째, 신용평가 시장이 확대돼 개인사업자 신용조사업(Credit Bureau, CB), 비금융 데이터를 활용하는 비금융정보 전문 신용정보조사업 등이 탄생하게 되고, 결과적으로 금융자금 조달 체계에 변화를 가져올 것이다.

넷째, 핀테크 산업 영역의 확대와 기존의 금융 서비스 전달 체계에 변화를 가져올 것이다. 기업고객 데이터의 가치가 핀테크로 연결되고, 오픈뱅킹이 본격화하면서 새로운 사업자들의 등장이 예고되고 있다. 데이터 개방과 데이터 이동권의 보장이 완벽해지면, 많은 기존 금융 서비스들이 핀테크 시장으로 이동할 것으로 예상된다.

[그림 1-2-2] 데이터3법 개정으로 바뀌는 산업들



출처: 김소혜, 「데이터 비즈니스의 시작, 승부는 지금부터」, 한화투자증권, 2020.5.25.

3. 국가 디지털 뉴딜 정책

가. 디지털 뉴딜 정책의 진행 방향

정부는 2020년 5월 7일, 제2차 비상경제 중앙대책본부회의를 하여 한국판 뉴딜 정책의 추진을 가시화했다. 2020년 7월 14일에는 디지털 뉴딜을 주요 사업으로 선정한 ‘한국판 뉴딜 종합계획’을 발표했다.

디지털 뉴딜 정책은 코로나19 전염병에 따른 산업 생태계의 급속한 변화 양상 속에서 전 산업 분야의 디지털 전환(Digital Transformation) 촉진, 산업 혁신 가속화, 이를 통해 국가경쟁력의 유지·향상 도모를 목표로 한다. 데이터 경제에 발맞춰 데이터와 네트워크, 인공지능 기술을 기반으로 디지털 전환을 본격화하고, 새로운 비즈니스 모델을 창출하거나 기존의 비즈니스 모델을 재정립하겠다는 것이다.

【표 1-2-1】 디지털 뉴딜 4대 분야 12개 추진과제

D.N.A. 생태계 강화	교육 인프라 디지털 전환	비대면 산업 육성	사회간접자본(SOC) 디지털화
① 데이터 구축·개발·활용 ② 전 산업 5G·AI 융합 확산 ③ 5G·AI 기반 지능형(AI) 정부 ④ K-사이버 방역 체계	⑤ 초중고 디지털 기반 교육 인프라 조성 ⑥ 전국 대학, 직업훈련기관 온라인 교육 강화	⑦ 스마트 의료·돌봄 인프라 ⑧ 중소기업 원격근무 확산 ⑨ 소상공인 온라인 비즈니스 지원	⑩ 4대 분야 핵심인프라 디지털 관리체계 구축 ⑪ 도시·산단 공간 디지털 혁신 ⑫ 스마트 물류체계 구축

출처: 「디지털 뉴딜, 코로나 이후 디지털 대전환을 선도합니다(보도자료)」 과학기술정보통신부, 2020.7.5.

정부는 디지털 뉴딜 사업에 2025년까지 총 58조 2,000억 원을 투자해, 총 90만 3000개의 일자리를 창출하는 등 디지털 대전환을 선도할 것을 천명했다.

나. 디지털 뉴딜의 키워드

디지털 뉴딜의 핵심은 D.N.A.로 대표되는 데이터(Data), 네트워크(Network), 인공지능(AI) 생태계를 강화하는 것이다. 특히 데이터 인프라의 구축과 개발, 활용을 통해 새로운 혁신산업을 창출하는 것을 강조한다. 그 중 대표적인 ‘데이터 댐’ 조성은, 공공과 민간의 데이터들을 데이터 댐에 모아 이를 결합·가공해 새로운 AI 서비스를 개발하고, 신산업을 창출하는 선순환 결과를 기대하는 사업이다.

디지털 뉴딜의 핵심 성장동력이자 주요 자원은 데이터다. 궁극적으로 교육 인프라의 디지털 전환, 비대면 라이프스타일 확산에 따른 비대면 산업 육성, 사회간접자본(SOC)의 디지털화와 디지털 전환으로 경제·사회의 대전환을 추구한다.

다. 디지털 뉴딜의 예상 성장 산업

디지털 뉴딜을 통해 성장이 예상되는 산업은 인프라 서비스(Infrastructure as a Service, IaaS)·플랫폼 서비스(Platform as a Service, PaaS)·소프트웨어 서비스(Software as a Service, SaaS) 등의 클라우드 서비스 산업, 원격진료 등 디지털 헬스케어 산업, 스마트 스토어 등의 지급결제 산업, 디지털 콘텐츠 산업, 정보보호 산업 등이다.

4. 데이터 거래소 운영

가. 데이터 유통 지원 시책의 추진

데이터 거래 활성화를 위한 전제조건은 빅데이터 유통시장의 원활한 조성과 이용 기업들에 대한 지원 확대다. 최우선 과제는 빅데이터 유통 활성화를 위한 시장 생태계의 조성이다. 데이터 거래를 활성화하기 위해서는 데이터 공급자가 양질의 데이터 집중 제공은 물론, 데이터 수요자가 원하는 정보를 손쉽게 찾아 안전하고 편리하게 데이터를 확보할 수 있도록 해야 한다. 또한 기업들이 이종(異種) 데이터와 보유한 데이터를 용이하게 결합·분석해 신규 서비스 개발에 활용할 수 있도록 지원 시책을 적극 펼쳐야 한다.

1) 데이터 바우처 지원

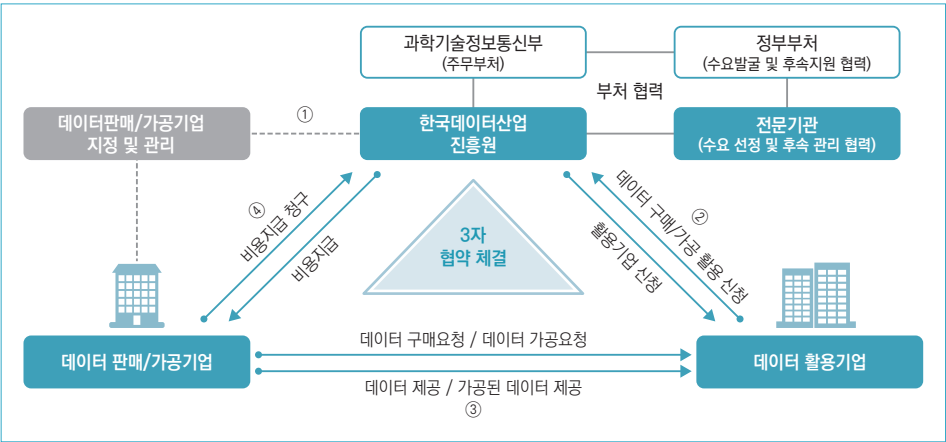
데이터 유통 환경 조성을 위한 중요한 시책 가운데 하나는 ‘데이터 바우처 지원’이다. 우리나라의 데이터 시장은 아직 데이터 가격에 대한 공감대가 부족⁰⁴해 구매자가 거래에 적극적으로 나서기 어려운 실정이다. 따라서 초기 데이터 거래의 활성화를 위해서 데이터 바우처를 지원하고 있다.

2019년부터 추진돼 온 데이터바우처 지원 사업은 2020년에는 총 575억 원의 예산이 투입됐다. 중소기업, 소상공인, 1인창조기업등에 총 1420개의 바우처를 지원하고 있다. 특히 2020년에는 더 다양한 분야에서 데이터 기반 혁신 창출을 유도하고자 9개 정부부처⁰⁵ 및 소속 전문기관과 함께 바우처의 수요를 발굴하고, 각 부처 사업과 연계해 우수 사례를 창출·공유한다. 또한 데이터3법 개정⁰⁵에 따라 가명·익명화(비식별) 처리를 가공업무에 포함하고, 분야별 가명정보 가공 수요를 집중 발굴·지원한다.

04. 「ICT 기반 신산업 발전을 위한 데이터 거래 활성화 방안」, 정보통신정책연구원, 2018.10.

05. 행정안전부, 금융위원회, 국토교통부, 문화체육관광부, 산업통상자원부, 중소벤처기업부, 환경부(기상청), 농림축산식품부, 해양수산부

[그림 1-2-3] 데이터바우처 지원 사업 구조



출처: 「데이터 경제에 활력을 불어넣을 데이터 활용 지원 본격 추진(보도자료)」, 과학기술정보통신부, 2020.3.30.

한편 2020년에는 코로나19 대응을 위한 우선지원제도를 마련해 감염병 치료나 예방, 확산 방지 등을 위한 연구나 서비스 개발에 50억 원을 별도로 배정하고 대학 및 연구기관들도 지원했다. 데이터 바우처 사업의 지원 분야는 유통·수출, 헬스·복지, 문화·교육, 교통·공간, 환경·기후, 금융, 중기·제조, 농림·수산, 공공·지역, 통신 등이다. 데이터 구매에 최대 1,800만 원씩 700건, 데이터 가공에 일반 분야는 최대 4,500만 원, AI 분야는 최대 7,000만 원씩 720건을 지원했다.

[표 1-2-2] 2020년 데이터바우처 사업 지원 규모

사업명	예산규모	지원 규모	지원 내용
데이터바우처	575억 원	1420개 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·AI 기반 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
합계	729.5억 원		

출처: 「데이터 경제에 활력을 불어넣을 데이터 활용 지원 본격 추진(보도자료)」, 과학기술정보통신부, 2020.3.30.

2) 중소기업 빅데이터 분석·활용 지원

이외에도 ‘중소기업 빅데이터 분석·활용 지원’ 사업이 있다. 이는 중소기업들이 빅데이터를 활용해 기업 경쟁력을 강화하고 새로운 비즈니스 가치를 창출할 수 있도록 지원하는 사업이다. 지방자치단체와 손잡고 지역별 중소기업과 빅데이터 전문기업 간 연계·매칭 등 맞춤형 컨설팅을 제공한다. 2020년에는 3개 지자체와 약 75개의 중소기업을 지원하는 데 7억 5,000만 원의 예산이 투입됐다.

3) 데이터 플래그십

데이터 플래그십은 빅데이터, AI 기술을 활용해 사회 현안 해결 및 비즈니스 혁신을 위한 제품 및 서비스 개발을 지원하는 사업으로, 지난 2013년 시작됐다. 그동안 서울시 심야버스 노선 수립, 전국 건물 대상 전기화재 위험도 평가, 실종자 과학수사 지원 등 다양한 분야에서 빅데이터 활용 우수 사례를 많이 만들어 냈다.

2020년에는 50억 원의 예산을 투입해 공공기관, 민간기업 구분 없이 총 10개 과제를 지원하고 있다. 특히 코로나19로 인한 사회적 피해가 큰 상황을 감안해, 질병관리본부의 감염병 역학조사를 위한 데이터 분석 기능 개발, 역학조사 지원 시스템 고도화를 긴급과제 형식으로 지원했다.

나. 데이터 거래소의 발족

데이터 거래소는 데이터 공급자와 수요자를 상호 매칭해 비식별정보, 기업정보 등의 데이터를 거래할 수 있는 중개 시스템을 말한다. 여기서 데이터 검색, 계약, 결제, 분석 등 데이터 유통 전 과정을 원스톱으로 지원한다. 거래소 시스템만으로 데이터의 모든 거래 절차를 진행할 수 있도록 설계된다. 또 수요자 중심의 거래 시스템도 가능토록 해, 수요자가 원하는 데이터를 다수 또는 특정의 공급자에게 제공을 요청할 수 있도록 할 계획이다.

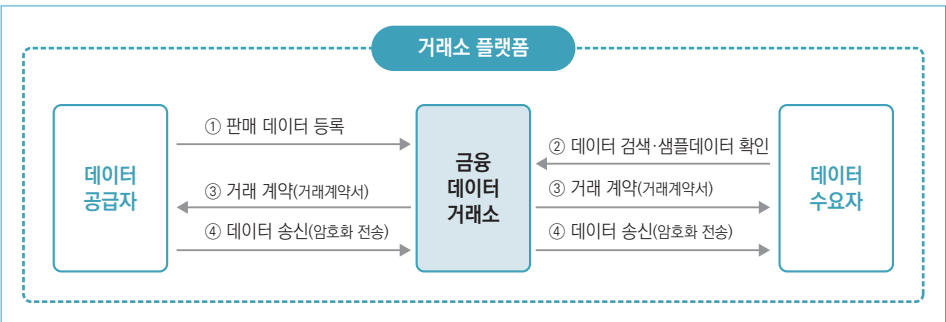
데이터 거래소는 다음과 같은 중요한 해결과제를 안고 있다.

현재 공공데이터 거래 플랫폼은 금융데이터 거래소와 빅데이터 플랫폼, 데이터스토어 등 10개 이상이며 지자체 등에서 별도 운영하기도 한다. 민간 데이터 거래 플랫폼 역시 SKT, LG CNS 등 주요기업이 운영하는 플랫폼만 10개 이상이다. 벤처기업들의 플랫폼까지 합치면 더욱 많이 산재돼 있어 거래소 통합과 운영방식에 대한 검토가 필요하다.⁰⁶

대표적으로 금융분야 데이터 거래소는 금융보안원이 운영을 맡아, 금융회사, 핀테크·빅테크 기업들이 데이터 유통·결합·사업화와 금융과 통신 융합데이터를 제공한다.

06. 「“흩어진 데이터 한곳에”… 민간 통합 거래소 추진한다」 보도자료, 조선비즈, 2020.8.7.

[그림 1-2-4] 데이터 거래소의 거래 절차

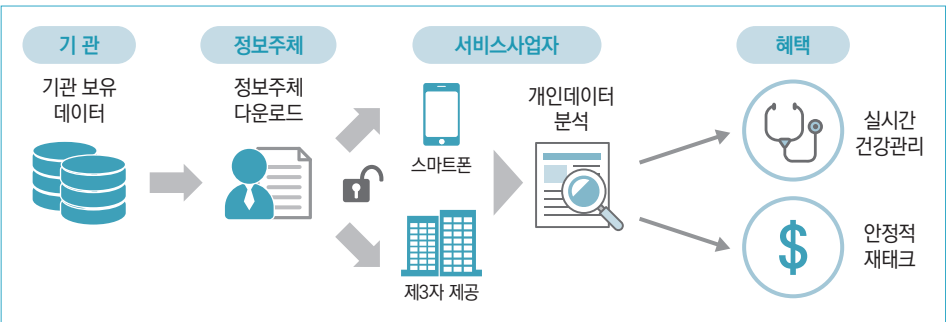


출처: 「데이터 경제 활성화를 위한 금융분야 ‘데이터 거래소’ 출범(보도자료)」, 금융위원회, 2020.5.11.

5. 마이데이터 사업

마이데이터(MyData)는 서비스 기업에게 위임했던 개인정보 처리 권한을 개인에게 재위임하는 모델이다. 유럽을 중심으로 개인 주도의 데이터 유통 생태계 형성을 목표로 제도화한 개념이다. 현재 개인 데이터는 은행, 카드회사, 병원 등에서 각기 분산해 관리하고 있지만, 개인의 정보를 누가 얼마나 갖고 있고 또 어떻게 활용하고 있는지 모르는 구조다. 그러나 마이데이터가 제도화되면 자기 정보를 분석해 유리한 의사결정을 하거나 서비스 기업을 주도적으로 선택해 최적의 서비스를 제공받을 수 있게 된다. 특히 우리나라와 같이 개인정보 동의와 관련된 규제가 높고 규제 완화가 쉽지 않은 경우에는 마이데이터 제도의 적극 추진이 데이터 경제 활성화의 관건이 될 수 있다.

[그림 1-2-5] 마이데이터 활용 구조



출처: 「데이터 산업 활성화 전략」, 4차산업혁명위원회, 2018.6.

지난 2018년 6월 ‘데이터 산업 활성화 전략’이 추진되면서 데이터 이용제도의 패러다임 전환 전략이 가시화됐다. 2019년에는 개인정보 관련 법 개정 없이도 시행 가능한 의료, 금융,

에너지 등 5개 분야에서 마이데이터 제도의 우선적 도입이 추진됐고, 8개의 마이데이터 활용 서비스가 발굴·추진됐다. 그 기세를 이어 2020년에는 ‘마이데이터 플랫폼⁰⁷ 시대’로의 본격 진입을 정부가 선언하기에 이르렀다. 2020년에는 지원 분야를 개인데이터의 양적·질적 활용 가능성이 풍부한 6개 분야⁰⁸ 9개 실증 서비스 과제로 확대했다. 특히 정보주체 중심의 데이터 유통·활용체계를 조성하는 데 중점을 뒀다.

[표 1-2-3] 2020년 본인정보 활용지원(MyData) 실증서비스 선정과제 목록

분야	과제명	수행기관		서비스
		주관	참여	
의료	분산원장증명(DID) 기반 의료 마이데이터 유통 플랫폼	엔디에스	유투바이오, 장원의료재단, 약학정보원, 교보생명, 진원온원	전자 처방전을 개인 데이터 저장소에 저장해 스스로 정보를 관리, 중복 검사를 억제, 맞춤형 자가 건강 관리 서비스 제공
의료	마이데이터 플랫폼 기반의 개인 의료·건강 데이터를 활용한 맞춤형 홈케어 서비스	평화이즈	카톨릭대 산학협력단, 경희의료원, 커넥타젠, 메디플러스솔루션, 미셀푸드	앱을 통한 개인의 건강검진 및 의료 처방 데이터 관리, 유전자 분석과 정신 건강·생활습관 관리 등 개인 맞춤형 홈케어 서비스 제공
금융	모빌리티 데이터를 활용한 신용평가 및 금융상품 개발	핀테크	SK텔레콤, SK에너지, SK네트웍스, 전북은행, 광주은행, JB우리캐피탈	개인의 모빌리티 데이터 분석 데이터와 공공·금융정보를 결합해 신용평가, 금융상품 서비스 제공
금융	금융 서비스를 연결하는 온디바이스 기반 마이데이터 결합 플랫폼	NH농협은행	에스엔피랩, 농협생명보험, 농협손해보험, NH투자증권, NH농협캐피탈, NH저축은행	개인 데이터를 행동 순서에 따라 수집·결합해 개인정보 분석을 통한 맞춤형 금융상품 추천
교통	포스트 코로나19 대비 공공 교통수단 클린 이용 서비스	서울특별시	BC카드, 네이앤크퍼니, KST모빌리티, 유아이네트웍스, 한국교통연구원, 코리아크레딧뷰로	공공 교통수단에서 수집된 데이터를 기반으로 재난 알림 서비스 확대 및 통합 경로 데이터를 통한 맞춤형 서비스 제공
생활	직장인 맞춤 웰니스 서비스: M-Box	한컴위드	고려대산학협력단, 세친구, 네이버비즈니스 플랫폼, 망고플레이트, 디스크라이, GS리테일, 소프트넷, 휴니버스글로벌, BC카드	개인의 건강진료와 카드 내역을 분석해 직장인 맞춤형, 정신 건강 관리, 장보기 서비스 제공

(계속)

07. 개인이 본인 정보를 수집·저장·관리 및 제3자에 제공할 수 있도록 지원하는 시스템으로 정보 주체에게 개인 데이터의 통제 권한을 부여하게 됨
08. 의료, 금융, 공공, 생활, 소상공인, 교통 등 6개 분야가 최종 선정. 당초 계획은 의료, 금융, 공공, 유통·물류, 문화, 통신·미디어, 교육, 기타(에너지 등) 등 8개 분야였음

분야	과제명	수행기관		서비스
		주관	참여	
소상공인	소상공인 신용평가 기반 상가 부동산 가치 정보 거래 플랫폼	소상공인연합회	신한카드, 한국간편결제진흥원, 코렉스플래닝, 한국금융솔루션, 케이원정보통신, 한국부동산사업협동조합	상가 부동산 데이터 수집·유통을 통한 대출연계 소상공인 평가모형, 점포 거래시장의 투명성 제고 및 상권분석 제공
교통	Micro-MaaS MyData를 활용한 도시문제 해결 데이터 에코 시스템 구축	데이터얼라이언스	레디우스랩, 나인투원, 이비카드, 아토리서치, 크로센트, 부천시	개인의 MaaS 데이터를 활용한 대중교통 사각지대 해소 방안 및 최적의 교통수단 재배치 방안 분석
복지	마이데이터 기반 교통약자 이동지원 서비스	한국과학기술 정보연구원	대전광역시, 대전복지재단, 대전도시철도공사, 하나카드, 메타빌드	교통데이터, 카드승인데이터 등을 연계하여 교통약자 대상 택시탑승 및 지하철 이용지원 서비스 제공

출처: 「의료·금융·공동 등 마이데이터 실증서비스 8개 선정(보도자료)」 과학기술정보통신부, 2020.6.11.
「대전시-KISTI '마이데이터 실증서비스사업' 최종 선정」 뉴스핌, 2020. 8. 11.

6. 데이터 보안

클라우드, 빅데이터, AI 등 디지털 신기술 활용이 증대되면서 정보보안(Cyber Security) 관점의 데이터 보안이 중요해지고 있다. 데이터에 대한 사이버 공격의 범위와 유형이 확대되고, 클라우드 등 아웃소싱으로 인한 제3자 리스크(Third Party Risk)도 심화되고 있다. 더욱이 오픈뱅킹 등 다양한 참여자가 공동망을 이용하는 사례가 늘어나면서 디지털 리스크의 상호연계성이 확대되고 있다.

정보 보안은 1995년부터 2003년까지 정보화가 급속히 진전되면서 해킹, 웜바이러스의 증가와 유무선 도청 문제에 따라 통신보안과 전자파보안을 포괄하는 개념으로 등장⁰⁹했다. 국내 정보화촉진기본법 제2조에서는 정보보안을 ‘정보의 수집, 가공, 저장, 검색, 송수신 도중에 정보의 훼손, 변조, 유출 등을 방지하기 위한 관리적, 기술적 수단을 강구하는 것’으로 정의하고 있다.

데이터 보안은 일반적으로 기밀성, 무결성, 가용성이라는 정보 보안의 세 가지 목표와 깊은 연관성을 갖는다. 먼저 기밀성(Confidentiality)은 기밀 정보를 권한이 없는 개인에게 노출되지 않도록 하는 것이다. 이는 어떤 정보를 수집, 저장하거나 누구와 관계 있는 정보인지 노출되지 않도록 하는 프라이버시 개념과 관련이 있다. 무결성(Integrity)은 정보와 프로그램이

09. 이연수 외 3명, 「주요국의 사이버안전관련 법, 조직체계 비교 및 발전방안 연구」, 국가정보연구 제1권 2호, 한국국가정보학회, 2008.

특정 인가된 방식으로만 변경되도록 하는 것으로, 시스템 기능을 고의나 부주의 등 허가되지 않은 조작에 의해 변경되지 않도록 하는 시스템 무결성과 관련이 있다. 가용성(Availability)은 시스템이 신속히 동작하고 인가된 서비스를 이용할 수 있도록 하는 것이다.

정부는 데이터 보안을 강화하기 위해 디지털 리스크를 사후 적발 중심에서 사전 예방 중심으로 관리할 수 있는 방안 강구와 보안 관련 민간 거버넌스 강화도 지원한다. 또한 데이터 보안 강화를 전제로 망 분리¹⁰ 규제의 합리화 방안을 계속 모색하고 있다. 망 분리 원칙의 전환은 사이버 위협의 수준과 네트워크의 연계성 등을 고려해 단계적으로 추진할 예정이다.

최근의 데이터 보안은 빅데이터 활용 과정에서 데이터 유출 문제에 대한 기술적 대책과 민감 데이터를 보호하는 기술 확보 쪽으로 방향을 잡고 있다. 따라서 암호화한 데이터를 복호화 없이 연산할 수 있는 동형암호(Homomorphic Encryption) 기술의 확보와 활용이 중요시되고 있다. 일례로 KCB 혁신금융서비스는 동형암호를 사용해 국민연금의 민감 데이터와 KCB의 신용 데이터를 결합·분석하는 방식의 서비스를 2020년 6월 15일부로 시작했다. 또 팬데믹 가드(Pandemic Guard)라 해, 코로나19 확진자와 의심자의 동선 경로를 암호화한 상태에서 계산함으로써 개인정보 노출 문제없이 감염자 동선을 파악하고 통보하는 시스템을 질병관리본부가 시행하기도 했다.

7. 마무리

데이터산업은 4차산업혁명 시대의 총아로 각광받고 있다. 새로운 서비스와 성장동력 확보, 신규 일자리 창출, 기존 산업들의 시장 융합 등 다양한 경제·사회적 효과를 불러올 산업이 될 것으로 기대된다. 반면 개인정보 보호를 위시한 데이터 보호와 데이터 주권 문제, 양자암호와 같은 데이터 관련 새로운 이슈에 대한 적시적인 대응과 활용, 새로운 융합서비스에 대한 법적 뒷받침 및 규제 샌드박스와 같은 제도적 정비, 경쟁력 있는 데이터 전문기업과 전문가 양성·확보 등 중요한 해결 과제도 안고 있다.

정부가 역점을 두고 있는 디지털 뉴딜사업의 성과 창출 또한 시대적으로 요구되는 과제다. 이들을 효과적으로 추진하고 조정하려면 데이터산업의 거버넌스를 확고히 하고, 관련 주체들에 대한 정책적 지원이 필요하다. 이런 집중과 대응이 따른다면, 데이터산업은 우리나라가 선도국가로 도약하는 데 중심 산업으로서 역할을 할 것이다.

10. 사이버 공격, 정보 유출 등 방지를 위해 업무용(내부망)과 인터넷용(외부망)을 논리적 또는 물리적으로 분리하는 것을 말함

제3장. 새로운 시장 생태계의 지각 변동

필자: 조현도
(넥스텔리전스 수석연구원)

최근 수년 간, 데이터 생태계의 기반인 데이터 유통 플랫폼 기술의 발달과 관련 제도 및 정책 개선이 활발하게 이뤄지고 있다. 데이터 생태계 기반 강화는 다양한 원천 소스들로 부터의 데이터 공급 활성화와 함께 데이터 활용의 수혜를 누리는 산업계와 정부 등의 데이터 수요자를 늘리는 효과를 가져온다. 이 효과는 데이터 생태계 전반에 활력과 역동성을 부여하면서 데이터 선순환 구조를 만들고 있다. 데이터 생태계는 경제적 측면에서 기존 제품·서비스의 생산성과 품질 향상은 물론, 융합형 신제품·서비스를 개발할 신규 플레이어의 시장 진입 촉진, 전통산업 질서의 ‘창조적 파괴’ 등 다양한 경제적 가치를 창출할 전망이다.

1. 데이터 생태계 형성과 성장

데이터 경제가 움트기 전의 초창기 데이터는 조직 내에 국한돼 폐쇄적이고 산발적인 형태로 존재했다. 그러나 데이터 경제의 데이터는 개별 조직을 넘어 공공 및 산업 부문 단위의 생태계는 물론, 생태계 간을 연계함으로써 데이터의 이동과 활용에 비약적인 자유도를 부여하고 있다. 이에 따른 데이터의 부가가치도 극대화하고 있다. 특히 데이터 생태계는 공생과 상생의 자연계처럼 원활한 데이터 흐름을 통해 조직 간 상호 영향을 미치는 관계로 성장하고 있다.

데이터 경제로의 전환과 데이터 생태계 조성의 이면에는 기술 및 정책 환경의 변화가 주요 동력으로 작용하고 있다. 기술 측면에서는 사물인터넷(Internet of Things, IoT) 및 소비자 기기의 급속한 확산에 따라, 산업(사물) 및 개인이 생성하는 데이터가 팽창⁰¹하고 있다. 이는 데이터 생산의 핵심 원천이다. 또한 발전하는 블록체인 기술은 민감한 정보 유통의 투명성과

신뢰성을 담보하며 데이터 연계와 축적을 촉진하고 있다. 빅데이터 분석과 인공지능 기술의 발달 역시 데이터의 활용성을 높이고 있다. 이렇듯 데이터 관련 기술들은 방대한 데이터의 분석을 가능하게 하고, 예측의 자동화와 정확성 향상에 크게 기여하고 있다.

[표 1-3-1] 데이터 생태계의 구성 요소와 촉진 역할

플레이어	데이터 활용 프로세스	활용 기술	생태계 촉진 역할
데이터 생산자	데이터 수집	IoT, 5G, LPWA ⁰² , 영상·음성인식기술, 스마트기기	• IoT로 수집한 실시간 데이터를 분석해 현실 세계의 물리적인 제품과 시스템을 디지털로 재현 • 디지털 혁신을 통해 데이터 기반 비즈니스 모델 전환 촉진 • 온라인 및 스마트기기 기반 실시간 개인정보 수집
데이터 서비스 제공자	외부 데이터와의 연계·축적	API ⁰³ 연계, 빅데이터 제어, 사이버 보안, 블록체인	• API 개방으로 다양한 서비스 간 융합 • 블록체인 기술을 이용함으로써 중앙 서버를 매개로하지 않고도 기업은 외부 서비스 연계 및 외부 데이터 공유를 실현 • 데이터 위변조가 원천적으로 불가능하므로 투명성과 신뢰성 높은 시스템을 구축
비즈니스 수요자	데이터 분석	인공지능, 엡지컴퓨팅 ⁰⁴ , 데이터 플랫폼	• 각종 디지털 기술로 수집된 데이터를 축적해 이용에 적합한 형태로 분석 • 기업은 고객 행동 로그 데이터와 속성 데이터 등을 결합·연계함으로써 고객관계관리(CRM) 등에 이용
	데이터 활용	로보틱스, 자율주행, 드론	• 수집 축적한 방대한 데이터를 사람이 처리하는 것은 불가능하기 때문에 AI를 활용함으로써 데이터 분석 및 예측을 자동화

출처: 디지털 생태계의 최전선 칼럼(デジタルエコシステムの最前線)コラム): 업종 제휴를 통한 디지털 생태계 형성(異業種連携によるデジタルエコシステム形成)』PwC⁰⁵ Japan, 2020.8.24. 재구성

정책 측면에서는 인공지능(AI) 학습용 데이터 구축과 정보주체 중심 데이터 활용 제도인 ‘마이데이터’에 이어 2020년 8월 데이터3법이 시행되며 데이터 경제 활성화를 위한 인프라 및 제도적 여건이 한층 성숙됐다. 특히 데이터3법은 개인정보 비식별화와 개인의

- 01. IoT 데이터 : 2020년 20ZB(Zettabyte=1조GB) 수준에서 2025년이면 79.4ZB(개인 데이터 포함 시 175ZB)로 증가 예고(IDC, *The Growth in Connected IoT Devices Is Expected to Generate 79.4ZB of Data in 2025, According to a New IDC Forecast*, 2019.6.)
- 02. Low-Power Wide-Area(저전력광역통신) : IoT를 위한 주요 통신 기술 범주의 하나로 근거리 저전력(배터리 수명 10년 이상) 무선통신 기술
- 03. Application Programming Interface : 프로그램들이 상호작용하도록 도와주는 매개체 역할을 하는 기술로 서버-데이터베이스, 애플리케이션-기기 간의 통신 및 연계를 지원
- 04. Edge Computing : 중앙 집중식 데이터 처리 방식이 아닌 데이터가 생성되는 네트워크의 종단(edge)과 가까운 곳에서 데이터를 처리하는 방식
- 05. 「디지털 생태계의 최전선 칼럼(デジタルエコシステムの最前線)コラム): 업종 제휴를 통한 디지털 생태계 형성(異業種連携によるデジタルエコシステム形成)』PwC Japan, 2020.8.24.

동의 없이도 가명정보와 익명정보의 산업적 연구·상업적 통계 목적 활용을 허용함으로써 다양한 분야에서 이중 빅데이터들 간의 안전한 결합 활용 및 분석을 가능하게 했다. 이로써 데이터3법은 데이터 생태계의 성장을 위한 중요한 제도적 전환점 역할을 했다.

2. 데이터 생태계의 정의 및 특징

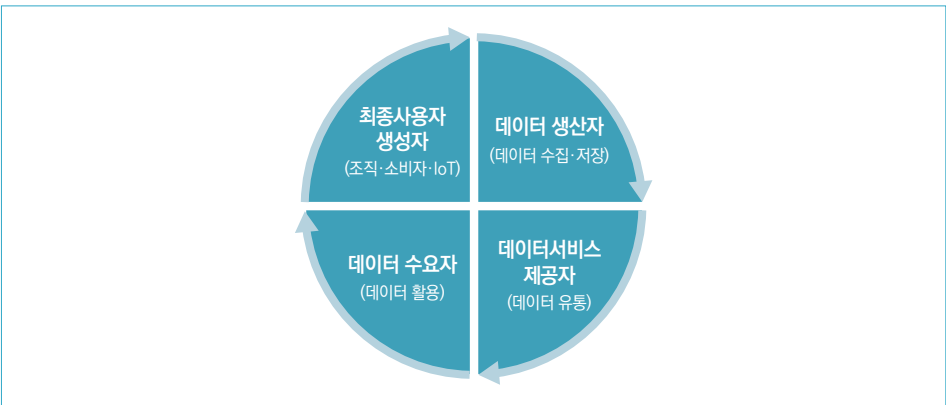
시장조사기관 IDC는 데이터 생태계를 ‘조직이 IoT 플랫폼을 통해 수집한 IoT 데이터와 인프라 시스템에 축적된 데이터 등 조직 내부의 다양한 데이터를 외부의 서드파티 데이터와 결합해 새로운 비즈니스 모델·수익 모델을 창출하기 위해 형성된 이해 당사자의 집합체’로 정의하고 있다. 이러한 데이터 생태계는 더 폭넓고 다양한 조직들 간에 데이터를 공유하고 유통하는 것을 주요 특징으로 삼고 있다.

데이터 생태계는 데이터를 생산·수집, 저장·처리, 유통, 활용하는 순환 과정에 참여하는 모든 이해 당사자들로 구성된다. 이때 데이터는 사물에 의해 생성되는 IoT와 사람이나 조직의 활동에서 생성되는 비IoT 데이터를 포괄한다.

IoT 데이터는 IoT 가전이나 공장 IoT 기기 등을 통해, 비IoT 데이터는 매장 POS(Point-of-Sales) 단말기나 교통정보 등 다양한 사람과 조직의 활동에서 각각 수집된다. 이렇게 생산·수집된 데이터는 데이터 전문 처리 기업이나 소비자(B2C) 또는 조직(B2B)들과 거래하는 기업에서 저장·처리된다. 데이터 유통은 데이터 신디케이터나 브로커, 빅데이터 플랫폼 운영 조직들에 의해 제공되고 있다. 최근 국내에서는 ‘마이데이터’를 통해 정보주체 중심의 유통 구조 확립이 본격화되고 있다.

비즈니스 수요자들은 확보한 다양한 데이터를 결합·분석해 마케팅이나 빅데이터 분석, 인공지능에 적용한다. 데이터 활용의 최종적인 대상인 고객은 자신이 생성한 다양한 쇼핑, 교통, 위치 등의 개인정보를, IoT 기기는 자체 생성 데이터를 다시 생산·수집 단계로 피드백함으로써 데이터 생태계의 순환 사이클을 반복한다.

[그림 1-3-1] 데이터 생태계 라이프 사이클



출처: 넥스텔리전스

3. 데이터 시장 생태계의 역동성과 지각 변동

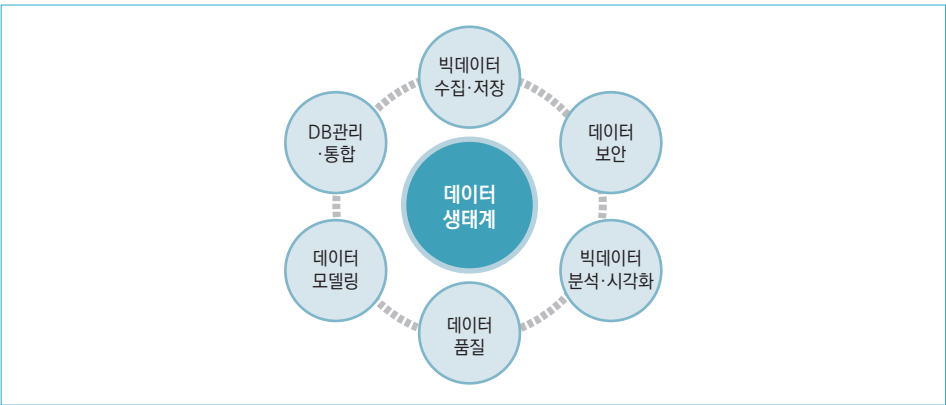
데이터 생태계의 활성화는 개인이나 개별 조직이 보유한 원천 데이터 공급원의 다변화와 함께 산업과 정부의 데이터 활용 니즈를 증폭시키고 있다. 또한 향후 데이터 생태계는 데이터 최종 사용자의 확장을 포함한 가치사슬 전후방의 확산을 야기하면서 데이터 시장의 파이를 확대시킬 것이다. 나아가 데이터 생태계의 역동성은 신제품·서비스 발굴 가능성을 높이고, 신규 기업의 시장 진입을 촉진하고, 전통적 산업 질서의 와해를 앞당길 것이다. 아울러 기존 제품·서비스 역시 생산성과 품질 향상 등 다양한 유무형의 경제적 가치를 창출할 것이다.

데이터 생태계의 역동성은 글로벌 거대 IT 기업들의 움직임과 벤처캐피탈의 데이터 관련 기술 기업들에 대한 투자 트렌드를 통해서도 확연하게 관찰된다. 스타트업 투자 분석 기업인 트랙스(Tracxn)에 따르면 데이터 애널리틱스 분야에 1700개 이상의 기업이 데이터 복구, 데이터 과학 및 첨단 예측 분석 기술을 개발하며 서비스를 제공 중인 것으로 조사됐다. 데이터 기술 기업으로 자본 유입도 매우 활발하게 이뤄지고 있음을 전하고 있다.

글로벌 데이터 분석 분야에서 현재까지 900개가 넘는 기업이 총 251억 미국 달러의 투자를 유치했으며, 이 중 절반 가까이가 2017년에서 2019년까지 3년 동안 이뤄진 것으로 집계하고 있다. 특히 글로벌 IT 기업들은 데이터 분야의 경쟁력을 확보한 스타트업을 적극 인수하고 있다. 구글(Google)은 2020년 2월 데이터 분석 및 비즈니스 인텔리전스 플랫폼 개발사인 루커(Looker)를 26억 미국 달러의 자금을 투입해 인수했다. 컨설팅 기업인 액센추어(Accenture) 역시 2019년과 2020년에 걸쳐 프래기스 바이둠(Pragis

Bidoo), 서터밀즈(Sutter Mills), 클래리티 인사이트(Clarity Insights), 바이트 프로페시(Byte Prophecy), 센텔리스(Sentelis) 등 금융 및 소매 분야에 특화된 데이터 분석 기업을 속속 인수했다. 이처럼 글로벌 IT 기업들의 역량 있는 데이터 기술 기업에 대한 적극적 투자와 인수 활동은 데이터 생태계 전반에 걸친 연계성을 확보하게 하고, IT 산업계 및 데이터 시장에서 경쟁력 우위를 확보할 수 있는 기반이기 때문이다.

[그림 1-3-2] 데이터 기술 관점에서 생태계 연계 및 통합성 강화



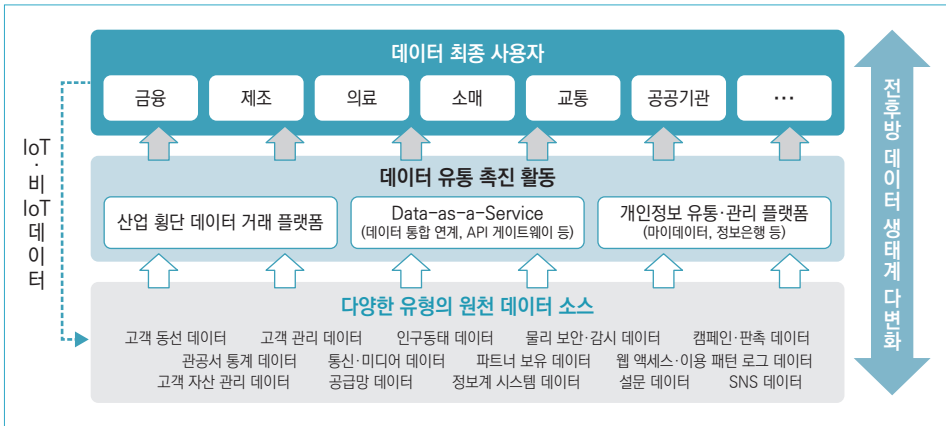
출처: 넥스텔리전스

아울러 포스트 코로나19 시대에 접어들면서 비대면 소비활동 관련 데이터가 급격히 늘어나고 있다. 고객 서비스 개선 방안 마련과 수요 변화 예측을 위한 소매 부문의 소비자 데이터 분석 기술에 대한 필요성도 크게 높아지고 있다. 또한 질병 역학 추적 및 모니터링을 위한 공공 및 의료 부문의 빅데이터 분석과 시각화 툴 역시 데이터 시장 생태계에서 더욱 주목받게 될 것으로 예상된다.

데이터 기술 발달과 함께 데이터 거래 유통을 위한 법적 근거와 제도적인 기반 마련 역시 중요하다. 이는 데이터 경제의 개별 가치사슬 전반에 걸친 역동성과 가치사슬 간 연계성 강화를 촉발하고, 데이터 생태계의 발전과 지각변동의 또 다른 동력을 제공하기 때문이다. 대표적으로 우리나라에서는 데이터3법을 토대로 데이터바우처, 데이터플래그십 사업 등 데이터 활용 지원사업 강화에 따라 데이터 순환과 관련된 시장의 부가가치 창출 환경이 크게 개선될 것으로 전망된다. 세계적 추세를 보면, 영국은 2020년 1월 금융 분야의 데이터 및 분석 기술 개발 촉진 전략(Data Strategy)을 발표하며 데이터 분석 기술 개발과 데이터 공유 관행의 현실적인 개선 방안을 내놓았다. 이후 9월에는 국가데이터전략(National Data Strategy) 차원에서 통신, 에너지, 소매, 금융 분야를 데이터에 기반해 고도화하는 사업에 민간의 의무적

참여를 규정한 법률 마련을 예고하기도 했다. 일본 역시 2020년 7월 빅데이터, IoT, AI 등 디지털 기술 기반 사회 혁신을 위한 법률 및 아키텍처 설계 재구축 방안에 관한 보고서⁰⁶를 내놓았다. 이 보고서는 데이터 관리, ID 기반 구축, AI 품질 평가 등을 위한 국가 자원의 신규 거버넌스 모델을 개발해 횡적인 결속을 강화하고, 수직적 연관산업 분야(자율주행, 금융, 인프라 등)를 활성화하기 위한 방향성을 담고 있다.

[그림 1-3-3] 데이터 생태계의 데이터 순환구조와 역동성



출처: 「国内「データエコシステム」事業者の調査結果を公表」, IDC Japan⁰⁷, 2020.8.20. 재구성

이처럼 데이터 자원의 시장 내 배분 메커니즘 작동을 위한 기술의 성숙과 데이터 유통을 위한 제도 기반 조성은, 금융·의료·소매·교통 등 수직산업 분야는 물론 공공 부문까지 데이터 활용을 촉진하고, 데이터 생태계의 선순환 구조에 더욱 활력을 제공하고 있다.

06. 「거버넌스 혁신: Society 5.0 실현을 위한 법과 아키텍처 재설계(GOVERNANCE INNOVATION : Society5.0の實現に向けた法とアーキテクチャのリ・デザイン)」, 일본 경제산업부, 2020.7.13.

07. 「国内「データエコシステム」事業者の調査結果を公表」, IDC Japan, 2020.8.20.

제 2 부

데이터산업 주요 정책 현황

제1장.
국내 데이터산업 정책 현황

제2장.
해외 데이터산업 정책 현황

제1장. 국내 데이터산업 정책 현황

필자: 김동일
(한국데이터산업진흥원 경영기획실 실장)

코로나19를 계기로 국내 전 분야에서 융합과 구조적 변화가 가속화되고 있다. 정부도 코로나19 위기 극복과 포스트 코로나 시대의 미래 경제 선도를 위해 범부처 차원의 계획인 한국판 뉴딜 종합계획을 발표했다. 이후 비대면화, 디지털 경제 전환 가속화를 위한 디지털 뉴딜 종합계획이 발표되었으며, 이를 소개한다. 특히 10대 중점 과제 중 5개의 대표 과제가 무엇인지를 중심으로 현 정책의 추진 방향을 짚어본다.

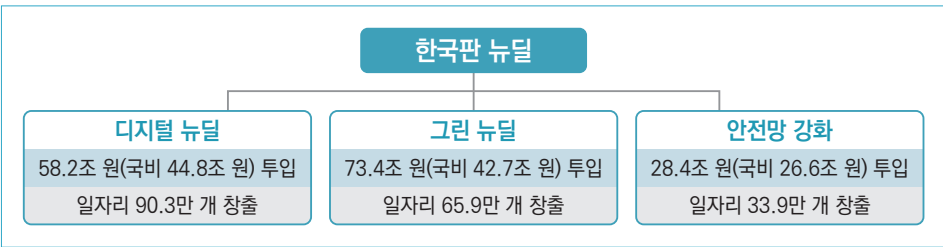
1. 한국판 뉴딜 정책과 방향

세계 각국이 코로나19 전염병 대응에 전력을 다하고 있다. 한국도 경기 침체와 일자리 충격 등에 직면한 상황을 타개하고 코로나19 이후 국가 경쟁력 제고를 위해 2020년 7월에 ‘한국판 뉴딜 종합계획’ 국가 발전전략을 발표했다. 이 전략은 포스트 코로나 시대에 유연하게 대응하고 세계적 흐름을 선도하겠다는 포부를 담고 있다.

한국판 뉴딜 정책은 크게 디지털(Digital) 뉴딜과 그린(Green) 뉴딜이라는 양대 축을 중심으로 안전망강화까지 3대 프로젝트와 10대 중점 추진과제로 구성됐다. 2025년까지 총 160조 원(국비 114.1조 원)을 투입해 총 190만 1000개의 일자리를 창출한다는 목표다.

전 산업 데이터와 디지털 혁신을 위해 DNA(Data-Network-AI) 생태계를 강화하고, 교육 인프라의 디지털 전환, 비대면 산업 육성 및 교통·수자원·도시·물류 등 사회기반시설을 디지털화하는 디지털 뉴딜 정책을 소개한다.

[그림 2-1-1] 한국판 뉴딜(2020-2025)

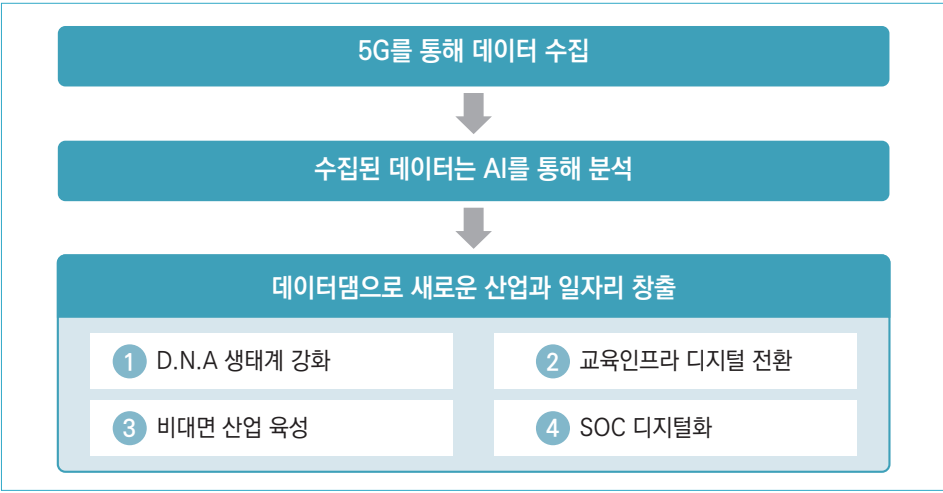


출처: 「한국판 뉴딜 종합계획 전체본」, 기획재정부, 2020.7.14.

2. 디지털 뉴딜 정책 배경과 주요 내용

디지털 전환(Digital Transformation)은 데이터, 네트워크, 인공지능(D.N.A) 등 디지털 신기술을 바탕으로 국가 경쟁력 및 시장 생산성 제고를 핵심 요소로 한다. 최근 코로나19 확산 방지 노력이 사회적·경제적 디지털화를 급속히 앞당기는 계기가 되고 있다. 온라인 소비, 원격근무 등의 비대면 사회활동, 이에 따른 디지털 기술과 데이터 활용의 증가 등 오프라인 사회가 온라인 세상으로 접어들며 비즈니스와 생활방식이 빠르게 바뀌고 있다. 디지털 뉴딜은 5G 이동통신 인프라를 기반으로, 전 산업 데이터를 수집·가공·결합해 AI 분석을 하여 의료, 교육, 제조 등 연관 분야에서 새로운 서비스와 일자리를 만드는 전략이다.

[그림 2-1-2] 디지털 뉴딜의 4대 분야와 목표



출처: 「한국판 뉴딜 종합계획 전체본」, 기획재정부, 2020.7.14.

디지털 뉴딜은 4대 분야 12개 추진과제로 구성된다. 4개 분야는 DNA 생태계 강화, 교육 인프라 디지털 전환, 비대면 산업 육성, SOC 디지털화다. 대표 과제로는 데이터 댐, 지능형(AI) 정부, 스마트 의료 인프라 등이다. 정부는 2020년 하반기 추경부터 2022년까지 총 23.4조 원(국비 18.6조 원), 2025년까지 58.2조 원(국비 44.8조 원)을 투입해 2022년까지 39만 개, 2025년까지 90만 3000개의 일자리를 각각 창출하고 디지털 대전환을 선도할 계획이다.

[표 2-1-1] 디지털 뉴딜 4대 분야 12개 추진과제와 목표

구분	분야	추진과제	투입예산(조 원)		일자리(만 개)
			~'22년	~'25년	~'25년
디지털 뉴딜	합 계		18.6	44.8	90.3
	1. D.N.A. 생태계 강화	소 계	12.5	31.9	56.7
		① 국민 생활과 밀접한 분야의 데이터 구축·개방·활용	3.1	6.4	29.5
		② 1·2·3차 전 산업으로 5G·AI 융합 확산	6.5	14.8	17.2
		③ 5G·AI 기반 지능형 정부	2.5	9.7	9.1
		④ K-사이버 방역체계 구축	0.4	1.0	0.9
	2. 교육 인프라 디지털 전환	소 계	0.6	0.8	0.9
		⑤ 모든 초중고에 디지털 교육 인프라 조성	0.3	0.3	0.4
		⑥ 전국 대학·직업훈련기관 온라인 교육 강화	0.3	0.5	0.5
	3. 비대면 산업 육성	소 계	1.1	2.1	13.4
		⑦ 스마트 의료 및 돌봄 인프라 구축	0.2	0.4	0.5
		⑧ 중소기업 원격근무 확산	0.6	0.7	0.9
		⑨ 소상공인 온라인 비즈니스 지원	0.3	1.0	12.0
	4. SOC 디지털화	소 계	4.4	10.0	19.3
		⑩ 4대 분야 핵심 인프라 디지털 관리체계 구축	3.7	8.5	12.4
		⑪ 도시·산단의 공간 디지털 혁신	0.6	1.2	1.4
		⑫ 스마트 물류체계 구축	0.1	0.3	5.5

출처: 「한국판 뉴딜 종합계획 전체본」 기획재정부, 2020.7.14.

3. 디지털 뉴딜 대표 과제와 주요 내용

2020년 7월 14일, 정부는 한국판 뉴딜 세부과제 중 경제활력 제고 등 파급력과 국민 체감도가 높은 사업을 선별해 10대 대표 과제를 선정했다. 10대 대표 과제는 데이터 댐, 지능형(AI) 정부, 스마트 의료 인프라, 그린 스마트 스쿨, 디지털 트윈, 국민 안전 SOC 디지털화, 스마트 그린 산단, 그린 리모델링, 그린 에너지, 친환경 미래 모빌리티 등이다. 그 중 데이터 활용 및 디지털 뉴딜 정책과 관련된 5개의 대표과제는 다음과 같다.

가. 데이터 댐

데이터 댐은 모든 산업에서 5G 이동통신망을 기반으로 데이터 수집·가공·결합·거래·활용하고, 많은 양의 데이터로 고도화한 인공지능 융합 서비스를 확산하는 사업이다. 이를 구체화하는 측면에서 공공데이터 개방, 데이터 수집 및 활용 확대, 데이터 거래 활성화, 인공지능 학습용 데이터 구축을 적극 추진한다. 과학기술정보통신부는 2020년 9월 디지털뉴딜 대표과제인 데이터 댐 프로젝트 7대 핵심사업에 4,991억원을 투입하여 본격 추진하기 시작했다.

1) 공공데이터 개방

정부는 민간 데이터산업 활성화와 데이터 기반 서비스·정책개발 촉진을 위해 공공데이터 개방을 강화할 계획이다. 공공 정밀 도로지도, 안전·취약 시설물 관리 데이터 등 2021년까지 14만 2000개의 공공데이터를 개방하고, 타기관 데이터와 연계·분석할 수 있는 공동 빅데이터 분석 시스템을 구축한다.

2) 데이터 수집 및 활용 확대

생산성 증대와 국민 생활편의 제고를 위해 제조·산업, 의료·바이오 등 생활 밀접분야의 빅데이터 구축과 활용을 확대한다. 2020년 하반기에는 공공·민간 등 국가 데이터 수집·연계·활용 정책을 총괄하는 민관 합동 컨트롤타워를 마련했다. 데이터3법 후속 시행령 개정, 가이드라인 제정 및 마이데이터 활용 활성화 등도 추진할 계획이다. 세부 분야별로 보면, 제조·산업 분야에서는 수요공급 예측, 공장 최적화 등을 위한 제조 데이터를 수집·분석하고, 소재 연구데이터 플랫폼을 구축해 신소재 탐색과 소재 설계 등에 활용할 계획이다. 의료·바이오 분야에서는 100만 명 분량의 임상정보·검체, 유전체 데이터 등의 바이오 빅데이터를 구축해 신약·의료기기 개발 등의 연구를 지원한다. 연구 목적의 보건·의료 공공데이터를 연계·분석할 수 있는 개방형 플랫폼도 확대 개편할 계획이다. 문화·관광 분야는

한국어-외국어·점자 말뭉치의 3300만 어절을 빅데이터화하고, 관광객 특성과 행태(소비, 이동경로, 관광지) 등 관광 빅데이터를 구축·분석해 지자체와 기업의 전략수립, 통번역과 AI 기술 개발을 지원한다. 그 외 축산업정보 연계 통합관리 플랫폼 구축과 소방정보 시스템 구축을 통해 국가 방역과 축산정책, 화재 예방에 적극 활용한다.

3) 데이터 거래

빅데이터 구축과 분석·활용도를 제고하기 위해 분야별 빅데이터 플랫폼을 기존 10개에서 30개로 확대한다. 데이터 가명처리 및 결합을 지원하며, 8400개 기업에게 데이터 구매·가공 바우처를 제공함으로써 데이터 거래·유통을 활성화한다. 더불어 데이터 거래 원칙과 기준을 마련해 공공데이터와 민간데이터 플랫폼 간 데이터·시스템 연동을 강화할 계획이다.

- 빅데이터 플랫폼, 데이터스토어, 공공데이터 포털 등의 개별 플랫폼을 통합해 데이터 지도에 연계
 - 가격산정, 품질 측정 등 데이터 거래 가이드라인 마련
- 예) 원가기준, 경쟁사 기준뿐만 아니라 품질, 사용성, 타당성 등을 고려한 이용자 관점의 데이터 가격을 정의하고 데이터 상품의 가치평가 모델 개발 등

4) 인공지능 학습용 데이터

제4차산업 선도를 위해 2025년까지 수요가 높고 기술적으로 구현 가능한(언어 말뭉치, 자율주행 영상 데이터 등) 인공지능 학습용 데이터 1300종을 추가 구축한다. 2025년까지 개발 리소스가 부족한 6700여 개 중소·스타트업에게 인공지능 기술개발과 적용을 위해 학습용 데이터 가공 바우처를 제공한다.

[표 2-1-2] 과학기술정보통신부의 데이터 댐 관련 7대 핵심 사업

No	사업명	2020년 예산	주요 내용
1	AI 데이터 가공바우처 지원사업	489억 원	인공지능 경쟁력 확보, 대규모 일자리 창출 및 중소기업 등 비즈니스 경쟁력 제고를 위해 인공지능 데이터 가공 지원 500여 건 확대
2	인공지능 바우처 사업	560억 원	중소기업 등에게 AI 솔루션 구매·적용을 지원해 기존 산업과 AI를 융합함으로써 신시장·가치 창출
3	인공지능 학습용 데이터 구축사업	2,925억 원	코로나19 이후 경기침체 극복 측면에서 일자리 창출효과가 높고 인공지능 모델 개발에 필수적인 인공지능 학습용 데이터 확대 구축·개발

(계속)

No	사업명	2020년 예산	주요 내용
4	AI 융합 프로젝트(AI+X)	282억 원	부처 간 협업으로 경제적 파급효과와 국민 체감도가 높은 전략 분야의 대규모 데이터를 AI로 분석해 신수요 창출
5	빅데이터 플랫폼 및 센터 구축	405억 원	각종 데이터를 수집·분석·유통하는 5대 분야별 빅데이터 플랫폼과 이와 연계된 센터를 기존 100개에서 150개로 증설
6	클라우드 플래그십 프로젝트	250억 원	제조, 물류, 헬스케어, 교육, 비대면복지 5개 분야에 혁신적이고 다양한 클라우드 서비스 공급
7	클라우드 이용 바우처 사업	80억 원	중소기업의 빅데이터 활용 방법 및 데이터 수집, 가공, 분석 역량 등 컨설팅·교육

출처: 「디지털 뉴딜의 핵심, '데이터 댐' 프로젝트 본격 착수」(보도자료), 과학기술정보통신부, 2020.9.2.

나. 지능형 정부 사업

지능형(AI) 정부 사업은 블록체인·AI 등 신기술과 5G·클라우드 등 디지털 기반을 활용해 비대면 맞춤형 정부 서비스 구현을 골자로 한다. 이 사업의 일환으로 모바일 신분증 도입, 지능형 공공서비스 제공, 스마트 업무환경 구현, 공공정보 시스템 클라우드 전환 등을 추진한다.

1) 모바일 신분증 도입 및 지능형 공공서비스 제공

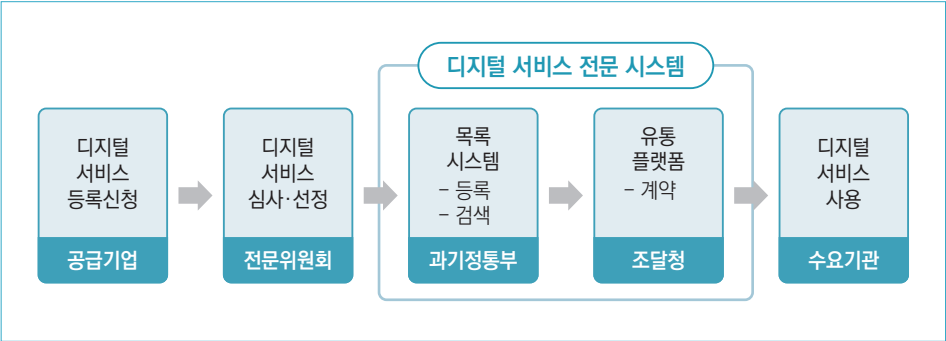
2020년 모바일 공무원증을 시작으로 블록체인 기반 모바일 신분증 도입, 2021년에는 모바일 운전면허증 도입을 추진할 계획이다. 이로써 비대면 디지털 신원증명체계를 구축하고, 납세·복지, 부동산 거래, 온라인 투표 등 체감도가 높은 분야에서 개인을 위한 데이터 활용 기반 맞춤형 서비스를 제공하게 된다.

2) 스마트 업무환경 구현 및 공공정보시스템 클라우드 전환

효과적이고 신속한 행정 업무처리를 위해 공공부문 유선망을 5G 이동통신 무선망으로 전환하고, 민간·공공 클라우드센터 간 데이터와 시스템을 연계해 사이버 위협 대응력을 강화한다. 기존의 입찰공고-입찰-낙찰자 선정-계약 체결로 이뤄지는 일반적인 조달 계약 시스템을 업그레이드하기 위해 디지털 서비스 전문계약 제도를 추진한다.⁰¹ 이를 위해 클라우드 컴퓨팅법 시행령 및 국가·지방계약법 시행령 등 연관 법령의 개정을 추진 중이다.

01. 디지털 서비스 전문위원회를 통해 사전에 엄선한 디지털 서비스 목록에서 필요한 서비스를 선택·구매하는 새로운 계약 제도. 디지털 서비스 등록·검색에서 계약까지 활용되는 전용 온라인 플랫폼도 구축하게 됨.

[그림 2-1-3] 디지털 서비스 전문계약 제도 운영방식



출처: 「한국판 뉴딜 종합계획 전제본」, 기획재정부, 2020.7.14.

다. 스마트 의료 인프라

디지털 기반 스마트 의료 인프라를 구축해 코로나19 등 감염병 위협에서 의료진과 환자를 보호하고, 의료 편의성을 제고한다. 2025년까지 18개 디지털 기반의 스마트병원을 구축해 신의료 모델을 도입하고, 감염병 대응을 위한 호흡기 전담 클리닉도 별도 신설 운영한다. 또한 의원급 의료기관에 비대면 영상진료 장비를 지원해 안전한 진료환경을 조성하고, 의료데이터 품질과 의료 서비스 질을 향상하기 위해 전자의무기록(Electronic Medical Record, EMR)도 표준화한다.

[표 2-1-3] 스마트 의료 인프라의 현재와 미래

불충분한 인프라로 비대면 의로서비스 활용 한계		스마트 의료 인프라 확충으로 비대면 의로서비스 기반 구축	
현재 상황		미래 모습	
성과지표	’20년	’22년	’25년
의료모델	스마트병원 기반 미흡	스마트병원 모델 9개	스마트병원 모델 18개
감염병 대응 인프라	호흡기전담 진료체계 미비	호흡기전담클리닉 1000개	호흡기전담클리닉 1000개
AI 기반 정밀의료	AI 진단 기반 미흡	8개 질환 AI 진단	20개 질환 AI 진단

출처: 「한국판 뉴딜 종합계획 전제본」, 기획재정부, 2020.7.14.

1) 감염병 안심 스마트 의료 인프라

입원환자의 실시간 모니터링, 의료기관 간 협진이 가능하도록 5G, IoT, 데이터를 활용한 디지털 스마트 병원을 구축한다. 격리병실과 집중치료실의 환자 영상정보와 데이터를 의료진에게 실시간 전송해 관리하고, 내외부 병원 간 협력체계도 구현한다. 2020년에는 코로나19 장기화 등에 대비해 감염병 대응 솔루션에 집중하고, 2025년까지 스마트병원을 매년 3개씩 구축해 2025년까지 총 18개로 확충할 계획이다.

2) 인공지능 정밀의료

간질환, 폐암, 당뇨 등 12개 질환별 AI 정밀 진단이 가능한 소프트웨어(SW) 실증사업인 닥터앤서 2.0 도입을 본격 추진한다. 닥터앤서 사업은 2018년부터 2020년까지 치매 등 8개 중증 질환별 인공지능 진단 소프트웨어를 개발했고, 향후 이를 확대 지원한다. 그 외 감염병 대응, 국민편의 제고를 위해 ICT기술기반 비대면 의료 제도화⁰²를 추진한다. 환자가 집에서도 필요한 의료 서비스를 받고 관리할 수 있도록 재택 의료 건강보험 수가 시범사업도 지속적으로 확대할 예정이다.

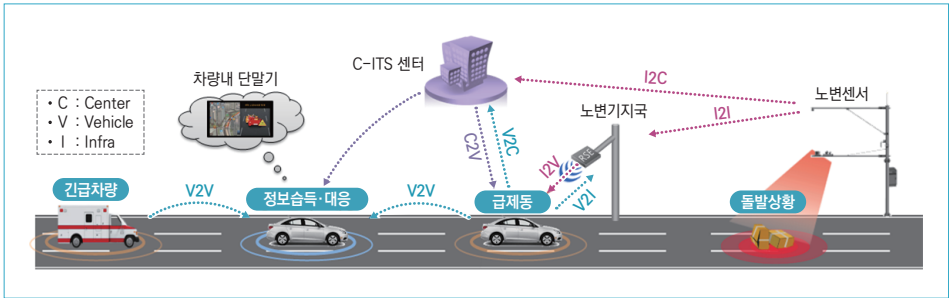
라. 국민안전 사회기반시설(SOC) 디지털화

도로·철도·공항 등 사회기반시설에 인공지능 기술과 디지털 관리 체계를 도입해 안전하고 효율적인 교통망을 구축한다. 자율주행 기준 마련과 규제 완화를 단계적으로 추진하고, 사물 인터넷 기반 조기 위험 경보시스템을 2022년까지 510개 설치해 디지털 안전체계를 확보한다.

자율주행차 등의 미래 기술을 적용한 도로환경을 제공하기 위해 차세대 지능형 교통 시스템(Cooperative Intelligent Transport System, C-ITS) 도입을 추진한다. 모든 철도에 사물 인터넷(IoT) 센서를 설치하고 ‘선로 안전점검 무인검측 시스템’을 도입해 실시간으로 상태를 모니터링하고, 다자간 대용량 데이터 통신이 가능한 4세대 철도무선망을 구축한다. 전국 15개 공항에 2022년까지 항공기 탑승권, 신분증, 지문, 얼굴정보 등 비대면 생체 인증 시스템을 구축한다. 전국 73개 국가 하천 배수시설에 자동·원격 제어시스템과 관리용 CCTV를 전 구간(3600km)에 설치한다.

02. 감염병으로 인한 안전, 의료사고, 상급병원 쏠림 등의 문제를 사전에 대비하기 위한 제도로 ICT 기반 원격 모니터링, 내원 안내 등 재택 의로서비스를 제공하고, 디지털치료제와 정밀의료의 조기 시장 진입도 지원한다.

[그림 2-1-4] 차세대 지능형 교통시스템(C-ITS) 체계도

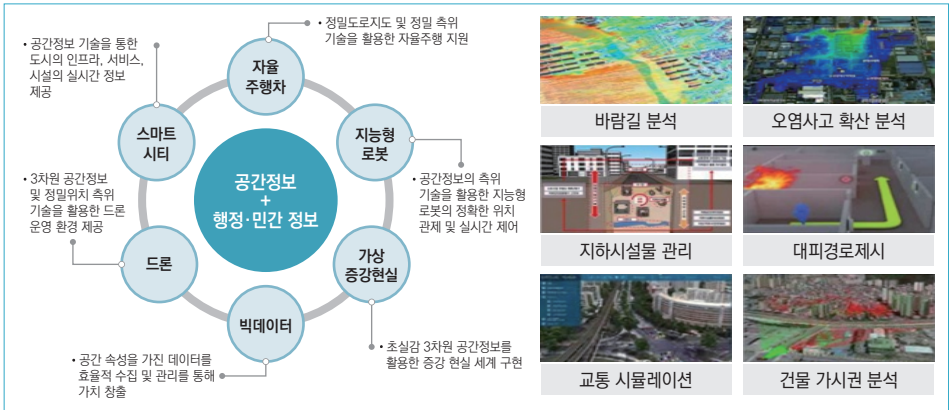


출처: 「한국판 뉴딜 종합계획 전제본」, 기획재정부, 2020.7.14.

마. 디지털 트윈

디지털 트윈은 스마트시티·자율주행차 등 신산업의 가상세계를 구축하기 위한 인프라 기술이다. 3차원 공간정보를 기반으로 행정·민간정보 등 각종 데이터를 결합해 국토·도시문제를 해결하고 안전관리를 강화하기 위해 등장했다. 주요 사업으로 전국 3차원 지도 구축, 지하 공간 통합지도·지하 공동구 지능형 관리 시스템, 정밀 도로지도 조기 구축 등이 있다.

[그림 2-1-5] 디지털 트윈의 구성



출처: 「한국판 뉴딜 종합계획 전제본」, 기획재정부, 2020.7.14.

도시지 주요 지형을 3차원으로 구축하고 12cm급 고해상도 영상지도를 작성한다. 또 국도와 4차로 이상 지방도로 중심의 정밀 도로지도를 확대해 자율주행차의 기반 환경을 조성한다. 현재 오프라인으로 제공되는 정밀도로 데이터를 온라인화하고, 산업계 의견도 반영해 디지털 인프라 조성을 확대할 계획이다.

[표 2-1-4] 디지털 뉴딜 4개 분야 12대 과제

과제명	소관 부처
1. D.N.A. 생태계 강화	
① 국민 생활과 밀접한 분야 데이터 구축·개방·활용	
· 공공데이터 개방, 공동 빅데이터 분석 시스템 구축	행정안전부
· 생활밀접 분야의 빅데이터 구축 및 활용 확대	과학기술정보통신부
【제조·산업】	
· 제조 데이터 수집·분석, AI 전문인력 네트워크 지원	중소기업벤처부
· 소재 연구데이터 플랫폼 구축	과학기술정보통신부
【의료·바이오】	
· 바이오 빅데이터 조기 구축	보건복지부
· 공공기관 의료데이터 활용 개방형 플랫폼 확대 개편	보건복지부
· 바이오 연구데이터 수집·제공, 바이오 데이터 관리 인력 양성	과학기술정보통신부
【문화·관광】	
· 한국어·외국어·점자 말뭉치 빅데이터 구축	문화체육관광부
· 관광 빅데이터 구축·분석	문화체육관광부
【기타】	
· 축산업 정보 연계·통합관리 플랫폼 구축	농림축산식품부
· 소방정보시스템 데이터 구축	행정안전부
· 마이데이터 활용 활성화	행정안전부
· 빅데이터 플랫폼 추가 구축	과학기술정보통신부
· 바우처 제공 등 데이터 거래·유통 활성화	과학기술정보통신부
· AI 학습용 데이터 구축	과학기술정보통신부
· AI 학습용 데이터 가공 바우처 제공	과학기술정보통신부
· 온라인 통합플랫폼 구축(디지털 집현전)	과학기술정보통신부
② 1·2·3차 전산업으로 5G·AI 융합 확산	
【5G 기반 융합서비스 확산】	
· 공연·관광 등 실감콘텐츠 제작 및 'XR 융합 프로젝트' 추진	과학기술정보통신부
· 스마트 박물관·미술관 구축, 예술·음악 등 융합서비스 개발지원	문화체육관광부
· 자율주행기술 개발, 테스트베드 고도화, 모빌리티서비스 실증	국토교통부
· 자율운항 핵심기술 개발·검증·실증 추진	해양수산부
· 5G 비대면 융합서비스 모델 개발 및 공공선도 적용	과학기술정보통신부

과제명	소관 부처
【AI 활용 확대를 통한 산업 고도화】	
• 스마트공장 보급 및 로봇설비 구축지원, 스마트화 도입 자금·보증	중소기업벤처부
• 지능형 건축설계 및 스마트건설기술 개발, AI 홈서비스 보급	국토교통부
• 첨단 제조로봇 실증장비 구축, 수요 맞춤형 AI로봇 개발·보급	산업통상자원부
• 융합 의료기기 개발, AI 기반 중환자 모니터링시스템 구축, 의료데이터 중심병원 지정	보건복지부
• 스포츠 AI코칭서비스 개발 및 온라인 미디어 예술활동 지원	문화체육관광부
• 농업 현장 실증·고도화, 차세대 융합·원천기술 개발, 첨단 무인자동화 농업 시범단지 조성	농림축산식품부
• 양식 클러스터 조성, 혁신기술 개발 및 청정어장 재생 추진	해양수산부
• ‘AI+X 7대 프로젝트’ 추진	과학기술정보통신부
• ICT 기반 환경영향평가 기술 개발 지원	환경부
• 실전형 과학화 훈련장 구축 및 가상 교육·훈련체계로 전환	국방부
【디지털 혁신기업 육성 및 산업의 디지털 전환】	
• 비대면 스타트업·벤처 발굴·지원 디지털 혁신기업 육성 추진	중소기업벤처부
• ‘스마트 대한민국 펀드’ 조성	중소기업벤처부
• 비대면 제품·서비스, 디지털 인프라 분야 우대보증	중소기업벤처부
• 데이터 기반 혁신창업 및 스타트업 맞춤형 지원 강화	중소기업벤처부
• 6G 통신 차세대 지능형 반도체 등 선도기술 개발 지원	과학기술정보통신부
• 차세대 AI 등 AI 원천 기술개발	과학기술정보통신부
• 차세대 지능형 반도체 기술개발	과학기술정보통신부
• 홀로그램 기술개발	과학기술정보통신부
• 비대면 및 MEC 기술개발	과학기술정보통신부
• AI 솔루션 구매 바우처 제공	과학기술정보통신부
• 데이터·AI기반 R&D 및 산업용 지능융합부품 개발 지원	과학기술정보통신부
• 스마트서비스 솔루션 구축 및 중소기업 혁신기술 개발 지원	중소기업벤처부
• 디지털 금융사업자 육성, 이용자 보호 강화 등 금융의 디지털 전환 가속화	금융위원회
• 금융회사 책임 강화 등 이용자 보호체계 구축	금융위원회
• 디지털 금융 법제화	금융위원회
• 디지털금융 리스크를 체계적 관리·감독	금융위원회
③ 5G·AI기반 지능형 정부	
• 모바일 신분증 등 All-Digital 민원처리 구현	행정안전부
• 부동산 종합공부시스템, 토지·건축 행정시스템, 건축물 생애이력 관리	국토교통부
• 국가보조금 맞춤형 안내, 지능형 연금복지 통합 플랫폼 등 시스템 구축·고도화	보건복지부

과제명	소관 부처
• 차세대 자동차관리 정보시스템 구축, 국토교통 데이터 활용 기반 마련, 교통안전 정보관리시스템 고도화	국토교통부
• 지진·해일·화산 정보 수집·분석·조기경보 제공, 빅데이터 기반 기상·기후 정보 제공, 산업 활성화를 위한 융합서비스개발	환경부
• DID 등 블록체인 시범사업 추진	과학기술정보통신부
• 온라인 투표시스템 등 블록체인 기술 도입	과학기술정보통신부
• 중앙부처 지자체 업무망 5G 전환 시범사업 추진(~’21)	행정안전부· 과학기술정보통신부
• 중앙부처 5G 국가망 전면도입(~’24), 지자체 확산(~’25~)	행정안전부· 과학기술정보통신부
• 대국민 공개용 홈페이지 등 민간 클라우드 이전	행정안전부
• 공공행정 업무 시스템 공공보안 클라우드센터 이전·통합	행정안전부
• 범용 클라우드 서비스 개발 지원	과학기술정보통신부
• 중소기업에 클라우드 바우처 제공	중소기업벤처부
• 국립중앙도서관, 국회도서관 소장물 디지털 DB 구축	문화체육관광부
• 디지털서비스 전문계약 제도 도입	과학기술정보통신부
④ K-사이버 방역체계 구축	
【기업의 사이버위협 예방·진단·대응 기반 강화】	
• 중소기업 보안 컨설팅 및 보안서비스 지원	과학기술정보통신부
• 보안 위협정보 탐지 및 빅데이터 분석 시스템 구축·운영	과학기술정보통신부
• 정보보호 관리체계(ISMS) 인증제 운영 내실화	과학기술정보통신부
• 비대면 서비스용 SW 보안진단 및 기술지원	과학기술정보통신부
• 개방형 보안취약점 분석플랫폼 구축	과학기술정보통신부
【국민 생활밀착형 사이버보안 강화】	
• 개인용 PC의 보안진단·점검 지원	과학기술정보통신부
• 공공·민간서비스 시스템(SW) 선제 점검	과학기술정보통신부
• 주요 웹사이트 침출 보안 모니터링 실시	과학기술정보통신부
【사이버보안 산업생태계 구축】	
• 4대 분야 보안기술 적용 시범사업 추진	과학기술정보통신부
• AI 보안 유망기업 발굴 및 사업화 지원	과학기술정보통신부
• 5대 분야별 보안모델 현장 배포 및 ‘보안리빙랩’ 운영 고도화	과학기술정보통신부
• 양자암호통신 시범망 구축	과학기술정보통신부
• IoT 제품 공공구매 촉진 및 보안인증 취득 지원	과학기술정보통신부

과제명	소관 부처
2. 교육인프라 디지털 전환	
① 모든 초중고에 디지털 기반 교육 인프라 조성	
• 초중고 교실 고성능 와이파이 구축	교육부
• 교원 노후 PC 교체 및 교육용 태블릿 지원	교육부
• 자율적 온라인 교과서 채택 단계적 허용	교육부
• ‘온라인 교육 통합플랫폼’ 구축	교육부
• 공공 온라인학습관리시스템(LMS) 고도화	교육부
• 학교의 저작물 이용범위 확대 기반 마련	교육부
• 온라인 공동교육과정의 정규수업시간 중 운영 및 교육청간 공동교육과정 개설 등 추진	교육부
② 전국 대학·직업훈련기관 온라인 교육 강화	
• 국립대 노후 서버 클라우드 전환 및 고속 전산망으로 교체	교육부
• 대학 공동 온라인학습 관리시스템(LMS) 구축, 권역별 원격교육 지원센터 설치	교육부
• 국공립 교원양성대학 미래교육센터 설치	교육부
• 온라인 공개강좌(K-MOOC) 우수 콘텐츠 확대 및 운영체계 혁신	교육부
• K-MOOC 플랫폼 강화 유료서비스 도입	교육부
• 공공 스마트 직업훈련 플랫폼(STEP) 고도화 및 콘텐츠 확대	고용노동부
• 민간 직업훈련기관 교육과정 온라인 전환 촉진	고용노동부
• 온라인교육 매치업(Match業) 교육과정 확대	고용노동부
3. 비대면 산업 육성	
① 스마트 의료 및 돌봄 인프라 구축	
• 디지털 기반 스마트 의료 인프라 구축	보건복지부
• 호흡기 전담클리닉 설치 및 화상진료장비 지원	보건복지부
• AI 정밀진단 SW 개발·실증	과학기술정보통신부
• 전자의무기록(EMR) 프로그램 표준화 지원	과학기술정보통신부
• 맞춤형 건강관리서비스 제공 보건소 모바일 헬스케어 확대	보건복지부
• 경증 만성질환자 대상 웨어러블기기 보급 및 질환 관리	보건복지부
• 건강취약계층 IoT·AI 활용 디지털 돌봄 시범사업 추진	보건복지부
• 어르신·장애인을 위한 돌봄로봇 4종 개발	보건복지부
• ICT 활용 재택의료 건강보험 수가 시범사업 확대	보건복지부
• 혁신의료기술 평가 대상 확대	보건복지부
• 재외국민 대상 비대면 진료특례 적용	보건복지부

과제명	소관 부처
② 중소기업 원격근무 확산	
• 중소기업 대상 원격근무 솔루션 바우처 지원	중소기업벤처부
• 중소벤처기업 밀집 주요거점에 공동활용 화상회의실 구축	중소기업벤처부
• ‘디지털 워크’ 기술개발 지원	과학기술정보통신부
③ 소상공인 온라인 비즈니스 지원	
• 소상공인 대상 온라인 기획전·쇼핑몰·라이브커머스 입점 및 상품개선 컨설팅 지원	중소기업벤처부
• 소상공인 대상 구독경제 시범사업 추진	중소기업벤처부
• 스마트기술 적용 소상공인 상점·공방 구축	중소기업벤처부
• 소상공인 대상 상권정보시스템 고도화	중소기업벤처부
4. SOC 디지털화	
① 4대 분야 핵심 인프라 디지털 관리체계 구축	
• 차세대 지능형 교통시스템(C-ITS)구축	국토교통부
• 자율차 제작·성능검증·보험 및 보안체계마련	국토교통부
• IoT 센서 설치 및 철도무선망 구축, ‘선로 안전점검 무인검측시스템’ 도입	국토교통부
• 공항에 비대면 생체인식시스템 조기 구축	국토교통부
• 원거리 도서지역 국가어항에 스마트 기술 유지관리 체계 시범 구축·운영	해양수산부
• 수치표고모형 구축 및 고해상도 영상지도 작성	국토교통부
• 차량 자율주행 핵심 인프라인 정밀도로지도 구축	국토교통부
• 지하구조물의 3D 통합지도 마련 등 지능형 관리시스템 구축	국토교통부· 과학기술정보통신부
• 항만 물류 자동화 테스트베드 등 디지털 관리체계 구축	해양수산부
• 국가하천 배수시설 자동·원격제어 시스템 및 실시간 모니터링 체계 구축	환경부
• 국가관리 저수지 원격수문제어 시스템 조기구축 및 계측기 설치	환경부
• 국가관리댐 스마트 안전관리체계 구축	환경부
• 재해위험지구 재난대응 조기경보시스템 구축	행정안전부
• 둔치주차장 침수위험 신속 알림시스템 구축	행정안전부
• 재난관리자원 통합관리시스템 구축	행정안전부
• 지능형 산림재해관리시스템 구축	산림청
② 도시·산단의 공간 디지털 혁신	
• 자자체 CCTV 통합플랫폼 구축	국토교통부
• 스마트 시티 솔루션 지자체 확산	국토교통부
• 스마트시티 국가시범도시 구축 지원	국토교통부
• 통합관제센터, 물류자원공유 플랫폼 확충	중소기업벤처부

과제명	소관 부처
• 제품생산 테스트 시뮬레이션센터 및 스마트 데모공장 구축	중소기업벤처부
• 설비관리·품질예측·공정최적화 등 솔루션 제공 혁신데이터센터 구축	중소기업벤처부
• 스마트산업 산학협력 프로그램 운영 및 스마트제조 혁신인력 양성사업 추진	중소기업벤처부
• 노후산업 대상 유해화학 물질 유·누출 원격 모니터링 체계 구축	중소기업벤처부
③ 스마트 물류체계 구축	
• 중소기업 공동 활용 스마트 공동 물류센터 조성	국토교통부
• E-Commerce 스마트 물류단지 조성 및 노후 물류기지 디지털화	국토교통부
• 스마트 물류센터 인증제 도입 및 물류시설 설비·투자 금융지원	국토교통부
• 스마트 물류기업 대상 정책금융 프로그램 지원	국토교통부
• 항만배후단지 스마트 공동물류센터 조성	해양수산부
• 항만 통합 블록체인 플랫폼 표준화 및 주요항만 확산	해양수산부
• 스마트항만 운용을 위한 재직자 교육 및 신규인력 양성체계 구축	해양수산부
• 농산물 등 소비촉진을 위한 공공급식 통합플랫폼 구축	농림축산식품부
• 축산물 온라인 경매플랫폼 구축	농림축산식품부
• 로봇, IoT 활용 스마트 물류기술 R&D 추진	국토교통부
• 디지털 물류시범도시 조성 및 디지털 물류서비스 실증	국토교통부

제2장. 해외 데이터산업 정책 현황

필자: 김희성
(넥스텔리전스 이사)

세계 여러 나라에서 데이터를 디지털 경제를 이끌어갈 주요 동력으로 인식하고, 혁신을 창출할 수 있도록 데이터 보호 및 규제 정책을 동시에 펼치고 있다. 미국, 영국, 일본은 공공데이터의 가용성을 높여 데이터 생태계를 구축하고 새로운 혁신을 창출하는 데 힘을 쏟고 있다. 유럽연합(EU)은 데이터 주권이라는 개념을 따라 프라이버시와 데이터 보호, 경쟁법 등의 충실한 준수를 강조하고 있다. 중국은 빅데이터 산업발전을 적극 추진하고 있는 가운데 최근 데이터보안법 초안을 발표해 보안과 경제적 편익 사이의 균형을 찾으려고 노력하는 모습이다. 동남아시아국가연합(ASEAN) 역시 역내 디지털 경제를 지원할 수 있는 데이터 거버넌스와 개인데이터 보호 정책을 펼치고 있다.

1. 미국의 정책

미국 연방정부는 일찍부터 빅데이터 기반의 생산성 향상과 혁신 창출 가능성을 확인하고 국가의 당면 과제 해결에 빅데이터를 적극 활용할 것을 견지해왔다. 오바마 정부에서는 다양한 부처가 참여하는 2억 달러 규모의 ‘빅데이터 연구개발 이니셔티브(Big Data Research and Development Initiative, 2012.03)⁰¹’를 시작으로 다양한 정책이 시행되었다. 우선 ‘빅데이터 연구개발 이니셔티브’에는 데이터에 대한 접근, 수집 및 관리에 요구되는 기술과 수단을 개선하는 데 목표를 두고, 빅데이터 기술 발전과 인력 확보, 혁신 프로세스 가속화 등을 추진한다는 내용을 담았다. 이와 함께 ‘네트워킹 정보기술 연구개발 프로그램’(Networking and Information Technology Research and Development, NITRD)의 일환으로 ‘빅데이터 고위 운영그룹(Big

01. *Big Data is a Big Deal*, Whitehouse, 2012.3; *Big Data R&D Initiative*, Federal Register, 2012.3.

Data Senior Steering Group)’을 설치하고 연방정부의 빅데이터 계획이 동 그룹을 통해 주도적으로 추진되도록 했다. 2016년 5월 NITRD는 ‘연방 빅데이터 연구개발 전략계획’(Federal Big Data Research & Development Strategic Plan)을 통해 빅데이터 핵심 기술 역량 확보, R&D 지원, 연구 인프라 구축, 데이터 공유 및 관리 촉진, 프라이버시 및 데이터 보안 강화, 빅데이터 인력 양성 및 혁신 생태계 강화 등을 시행한다고 밝힌 바 있다. 또한 오바마 정부는 취임 직후부터 ‘열린 정부(Open Government)’라는 기치 하에 정부 데이터를 민간이 적극 활용할 수 있도록 하는 오픈데이터(Open data) 정책을 적극 추진했다. 2009년 12월, ‘열린 정부지침(Open Government Directive)’에 따라 다음해 정부 공공데이터 포털 사이트인 Data.gov가 개설됐다. 이후에도 2013년에 발표한 ‘오픈데이터 정책(Open Data Policy)’, ‘정부 정보공개에 대한 행정명령(Making Open and Machine Readable The New Default For Government Information)’ 등을 통해 오픈데이터 정책을 공고히 했다.

트럼프 대통령은 당선 전후 오픈데이터와 관련한 어떠한 정책도 내놓지 않으면서 미국 오픈데이터 정책 후퇴 우려를 자아냈으나⁰² 이를 불식시키고 Data.gov를 유지하는 한편, 2019년 1월에는 ‘정부 데이터법(Open, Public, Electronic and Necessary Government Data(OPEN Act))’을 승인하는 등 미국 연방정부의 오픈데이터 정책 기초를 이어가고 있다. ‘정부 데이터법’, 일명 ‘오픈법(OPEN Act)’은 비민감 정보의 기계 판독성을 가능하도록 하여 공개성과 접근성을 높였다고 할 수 있다. 2014년 4월 ‘연방 데이터 전략’(Federal Data Strategy)’을 통해 동 법의 시행이 구체적으로 이뤄질 예정이다.⁰³ ‘연방 데이터 전략’은 미국 연방정부 데이터로부터 최적의 가치를 얻어내고 보안, 프라이버시, 기밀성 등을 보호하는 것을 목표로 한다. 데이터 관련 윤리, 설계, 문화 학습 등의 분야에서 10개 원칙과 40개 실천내용(practice)을 제시하고 있다. 이에 따라 향후 10년에 걸쳐 ‘연방 데이터 전략’을 실천할 수 있는 2020년도 실천계획(Federal Data Strategy 2020 Action Plan)과 업무도 제시했다.⁰⁴ 우선 정부기관은 2020년 내 △각 기관의 주요 과제에 대응하기 위한 데이터 니즈 규명 △다양한 데이터 거버넌스 기구 설립 △데이터 및 관련 인프라 성숙도 평가 △직원의 데이터 스킬 개선 기회 모색 △오픈데이터 계획을 위한 우선순위 데이터 자산 규명 △데이터 인벤토리 공표 및 업데이트 등을 달성하도록 계획됐다. 각 기관의 공통 과제와 관련된 특정 기관 등이 취해야 할 업무실천 커뮤니티(Community of Practice Actions)와 관련해서는 △연방 최고 데이터 책임자협의회(Federal

Chief Data Officer Council) 출범 △AI R&D를 위한 데이터 및 모델 리소스 개선 △재정관리 데이터기준 개선 △지리공간 데이터 업무의 연방 데이터 엔터프라이즈(Federal Data Enterprise)⁰⁵로의 통합 등이 포함됐다. 마지막으로 공유 솔루션 실천(Shared Solution Actions)에서는 △연방 엔터프라이즈 데이터 리소스 보관소 개발 △예산관리국(OMB) 연방데이터 정책위원회 신설 △큐레이트된 데이터 스킬 카탈로그 개발 △데이터 윤리 프레임워크 개발 △데이터 보호 톨킷 개발 △원스톱 표준 연구 애플리케이션 시범사업 실시 △데이터 인벤토리 신설 및 업데이트 지원 관련 정보수집 리뷰를 위한 증강툴 시범사업 실시 △연방기관을 위한 강화된 데이터 관리툴 시범사업 실시 △데이터 품질측정 및 보고 가이드스 개발 △데이터 기준 보관소 개발 등이 계획됐다.

한편 2019년 트럼프 정부는 ‘AI 이니셔티브(The American Artificial Intelligence Initiative)’와 ‘국가 AI R&D 전략(National Artificial Intelligence Research and Development Strategic Plan : 2019 Update)’ 등을 통해 AI 발전을 위한 데이터의 중요성을 강조하고, AI R&D 생태계 촉진을 위해 연방정부 데이터 등을 적극 개방할 것을 각 부처에 요청했다.

02. POLITICO, *What happened to Trump's war on data?*, 2017.7.25.; Engadget, *Trump administration is killing its open data portal*, 2017.4.14

03. ‘연방데이터전략’은 21세기 연방정부의 현대화를 비전으로 하는 트럼프 대통령의 ‘대통령경영관리 의제(President's Management Agenda)’의 일환으로 추진된다.

04. President's Management Agenda, Federal Data Strategy 2020 Action Plan, 2019.12

05. Federal Data Enterprise: 미국 각 연방 기관 및 기능을 공통의 엔터프라이즈 아키텍처 하에 통합하고자 하는 연방정부의 노력으로, 각 기관의 비즈니스, 서비스, 구성요소, 기술, 데이터 등을 하나의 공통된 접근 하에 통합함으로써 조직 설계나 성과 개선을 도모.

[그림 2-2-1] 미국 연방 데이터 전략의 2020년 실천계획 개요



출처: strategy.data.gov, FDS FRAMEWORK⁰⁶

06. <https://strategy.data.gov/assets/docs/2020-federal-data-strategy-framework.pdf>

2. 유럽의 정책

가. EU의 정책

EU는 2002년 공공부문 정보의 재사용과 관련한 공공부문 정보 전문가 그룹(Expert Group on Public Sector Information, PSI Group) 설치를 시작으로 데이터 정책에 시동을 걸기 시작했다. 2014년, 디지털 경제 활성화에 중점을 둔 ‘데이터 주도경제(Data-driven Economy)’, 2017년 ‘유럽 데이터 경제(European Data Economy)’를 거쳐 2020년 2월에는 유럽 데이터 전략(A European Strategy for Data)’을 발표했다. EU는 디지털 단일 시장(Digital Single Market Strategy for Europe) 구축을 목표로 AI 및 데이터 전략을 추진하고 있다. ‘데이터 전략’을 새로운 디지털 전략 중 한 축으로 삼고 향후 5년 간 유럽의 디지털 전환을 위한 프레임워크로 활용할 계획이다.

이번 전략은 산업 및 상업 데이터(industrial and commercial data)가 디지털 경제를 이끌어가는 주요 동인이라는 점을 강조하고, 데이터를 생산하는 주체에 대한 규제 적용과 동시에 경제·사회적 가용성 제고를 선언했다. 우선 데이터 단일 시장(single market for data) 창출을 위해 역내 및 부문 간 데이터 흐름이 ‘모두의 편익’을 위해야 하며, 데이터 규정, 특히 프라이버시 및 데이터 보호, 경쟁법 등을 충실히 준수해야 한다는 점을 강조하고 있다. 더불어 데이터 접근 및 사용 규정은 공정성, 실용성, 명확성을 담보해야 한다고 명시하고 있다.

이와 함께 데이터 단일 시장 구축을 위해 시행되는 새로운 조치들이 담아야 할 4가지 축(pillar)을 제시하고 있다. 구체적으로 살펴보면 다음과 같다.

첫째, 데이터 접근 및 재사용에 관한 분야 횡단적 거버넌스 프레임워크 구축으로 △유럽 공통의 데이터 스페이스(common European Data Space)의 거버넌스를 위한 법제 프레임워크 제안(‘20) △고가치 데이터세트 관련 실행규정(implementing act) 채택(‘21) △데이터 법안(Data Act) 제출(‘21) △디지털 경제에서 데이터의 중요성 분석 및 디지털 서비스법(Digital Service Act) 패키지와 관련해 기존 정책 프레임워크 검토(‘20) 등을 주요 실천사항으로 제시하고 있다.

둘째, 데이터에 대한 투자와 데이터 호스팅·처리·활용·상호운용성 역량 및 인프라 강화다. 이와 관련해서는 △데이터 공유 아키텍처, 거버넌스 메커니즘, 유럽 클라우드 인프라·서비스 연동(federation) 등을 포함하는 유럽 데이터 스페이스 관련 투자기금조성(‘22년) △클라우드 연동에 관한 회원국 양해각서 체결(‘20) △유럽 클라우드 서비스 시장(marketplace) 출범 △EU(자율)규제적 클라우드 규정집 작성(‘22) 등이 제시됐다.

셋째, 개인의 자기 데이터에 대한 권한 강화 및 스킬·중소기업 투자 등을 통한 역량 개선으로 일반 데이터 보호규정(GDPR) 20조 하에서 개인의 정보 이동권(portability right)을 강화해 기계가 생성하는 데이터에 대해 접근 및 사용하는 자에 대해 더 많은 통제가

가능하도록 조치할 예정이다. 이는 2021년 데이터법 일환으로 시행이 가능할 것으로 보고 있다.

마지막으로 전략 부문 및 공공 영역에서 유럽 공통의 데이터 스페이스 활성화다. 여기에서는 데이터의 흐름 및 경제적 가치를 측정할 수 있는 프레임워크 창출(21)을 계획하고 있다. 공공 영역으로 제시된 분야는 제조, 그린딜, 이동성, 건강, 금융, 에너지, 농업, 공공행정, 스킬 등이다.

[표 2-2-1] 유럽 데이터 전략의 4가지 축 개요

구분	주요 내용
데이터 접근 및 재사용에 관한 분야 횡단적 거버넌스 프레임워크 구축	• 유럽 공통의 데이터 스페이스(common European Data Space)의 거버넌스를 위한 법적 프레임워크 제안(20) • 고가치 데이터세트 관련 실행규정(implementing act) 채택(21) • 데이터 법안(Data Act) 제출(21) • 디지털 경제에서의 데이터의 중요성 분석 및 디지털 서비스 법 (Digital Service Act) 패키지와 관련해 기존 정책 프레임워크 검토(20)
데이터에 대한 투자와 데이터 호스팅·처리·활용·상호운용성 역량 및 인프라 강화	• 데이터 공유 아키텍처, 거버넌스 메커니즘, 유럽 클라우드 인프라·서비스 페더레이션(federation) 등을 포함하는 유럽 데이터 스페이스 관련 하이 임팩트 프로젝트 투자(22) • 클라우드 페더레이션에 관한 회원국 양해각서 체결(20) • 유럽 클라우드 서비스 시장(marketplace) 출범 • EU (자율)규제적 클라우드 규정집 작성(22)
개인의 자기 데이터에 대한 권한 강화 및 스킬·중소기업 투자 등을 통한 역량 개선	• 일반 데이터 보호규정(GDPR) 20조 하에서 개인의 정보 이동권(portability right)을 강화해 기계가 생성하는 데이터에 대해 접근 및 사용하는 자에 대해 더 많은 통제가 가능하도록 조치(2021년 데이터법 일환으로 시행이 가능할 것으로 보고 있음)
전략 부문 및 공공영역에서 유럽 공통의 데이터 스페이스 활성화	• 데이터의 흐름 및 경제적 가치를 측정할 수 있는 프레임워크 창출(21)을 계획 • 공공 영역으로 제시된 분야는 제조업, 그린딜, 이동성, 건강, 금융, 에너지, 농업, 공공행정, 스킬 등

출처: A European strategy for data, EC, 2020.2.

나. 영국의 정책

영국 정부는 2016년 6월, 국민투표를 통해 결정한 브렉시트를 계기로 산업 경쟁력 강화를 위한 정부의 적절한 개입 필요성에 대한 공론화의 발판을 마련했다. 이를 계기로 2017년 11월 ‘산업전략(Industrial Strategy)’을 마련하고 신기술·신산업에서 글로벌 리더로서의 지위를 유지할 수 있도록 영국의 강점을 활용하는 한편, 취약점에 적극 대응하겠다고 천명했다.

영국의 산업전략은 전 산업 부문을 관통하는 생산성 기반을 조성하고, 산업별로 민관협력체제인 섹터딜(Sector deal)⁰⁷을 추진함으로써 경제·사회적으로 직면한 4대 대과제를 해결한다는 내용을 골자로 한다. 4대 대과제 중 첫 번째로 AI·데이터 주도 경제가 선정됐다.

영국 정부는 이와 관련한 혁신을 촉진하고, 영국이 글로벌 센터로 자리매김할 수 있도록 AI 섹터딜(AI Sector Deal) 등 다양한 지원을 시행하고 있다. 최근에는 ‘영국 연구개발 로드맵(UK Research and Development Roadmap, 2020.07)’을 통해 혁신 과정에서 신뢰할 수 있는 데이터 사용을 최대화할 필요가 있다는 점을 강조한 바 있다.

한편 2017년 3월에는 세계 최고 수준의 디지털 경제로의 도약을 위한 국가 차원의 디지털 전략인 ‘영국 디지털 전략(UK Digital Strategy 2017)’을 발표했다. 이 전략에는 디지털 인프라, 역량, 사업 환경, 기업의 디지털 전환, 사이버 안전, 디지털 정부 등의 내용과 함께 데이터 개방 및 활용 기반 강화 전략이 포함됐다.

데이터 개방 및 활용 기반 강화와 관련해 민관 데이터 연계 확대 및 데이터 활용을 위한 인프라 구축, 2016년 6월 브렉시트 결정⁰⁸ 이후 EU와의 데이터 공유를 위한 1988년의 ‘데이터 보호법(Data Protection Act)’ 개정, 정부·기업·시민 간의 데이터 사용에 대한 신뢰관계 구축 및 개인정보보호 문제에 대한 강력한 대응방안 수립 등을 주요 내용으로 담고 있다.⁰⁹

이러한 일련의 정책을 기반으로 지난 9월, 영국의 디지털·문화·미디어·스포츠부는 데이터 기반(foundation), 스킬, 가용성 및 책임성 등을 골자로 하는 ‘국가 데이터 전략(National Data Strategy)’을 발표했다. 이번 전략은 데이터의 사용성, 접근성 및 가용성을 둘러싼 환경을 개선함으로써 영국 경제 전반에 데이터의 가치가 널리 실현될 수 있도록 하는 한편, 성장 친화적이며 신뢰성 있는 데이터 제도 확충, 정부 데이터 사용의 효율성 및 공공서비스 개선, 데이터 인프라의 보안성 및 복원성 확보, 국제적 데이터 흐름 보장 등을 목표로 하고 있다. 이 목표 달성을 위해 데이터 기반, 스킬, 가용성 및 책임성과 관련해 다음과 같은 구체적인 방안도 제시하고 있다.

07. 섹터딜(sector deal)은 이번 ‘산업전략’을 통해 새롭게 도입된 정책 개념으로 해당 산업의 생산성, 고용 및 혁신 제고를 목적으로 특정 이슈를 논의하는 정부-산업계 파트너십

08. 2016년 6월 23일에 EU 탈퇴에 대한 국민투표를 거쳐 2020년 1월 31일에 공식 탈퇴

09. Department for Digital, Culture, Media & Sport, UK Digital Strategy 2017, 2017.3.1.; 『2016년 다포스포럼 4차산업혁명 발표 전후 주요국 국가정보화 전략 분석 및 시사점』 한국정보화진흥원, 2017.10.31.

‘국가 데이터 전략’ 아래 체계적으로 접근

첫째, 데이터 기반은 경제사회의 광범위한 데이터 기반 조성, 정부 및 공공부문의 데이터 기반 조성, 국제적 데이터 기반 지원을 주요 내용으로 한다. 경제사회의 광범위한 데이터 기반 조성과 관련해 △디지털트윈 관련 정보관리 프레임워크(Information Management Framework) 개발 등을 비롯해 위치, 인공 구조물, 자연환경, 운송, 기타 인프라 등에 관한 데이터 표준화를 통한 물리적 환경 조성 △‘영국 연구개발 로드맵’에 따라 국가 및 지역 차원의 혁신에서 신뢰할 수 있는 데이터 사용 최대화 △소비자 데이터가 소비자를 위해 활용될 수 있도록 노력 등의 내용을 포함하고 있다.

정부 및 공공부문의 데이터 기반 조성과 관련해서는 △데이터 품질 프레임워크(Data Quality Framework)와 정부 데이터 성숙도 모델(Data Maturity Model for government) 등을 마련 △정부의 통합 데이터 플랫폼 개발 △지방정부의 데이터 편익 지원 △범부처 데이터 리더십을 발휘할 수 있는 고위급 CDO(Chief Data Officer) 채용 등을 제시하고 있다. 이와 함께 글로벌 데이터 상호운용성 확보 지원 등 국제적 데이터 기반 지원도 할 예정이다.

공교육과 직업교육에 데이터 스킬 추가

둘째, 데이터 스킬은 데이터 스킬의 명확성 확보 및 협력 추동, 공공교육 및 직업교육 상에서의 데이터 스킬 대응, 공공부문 전반의 데이터 스킬 확충을 통한 역량, 리더십 및 문화 확충 등을 골자로 하고 있다. 데이터 스킬의 명확성 확보 및 협력 추동에서는 △경제 전반에서 사용될 수 있는 데이터 스킬을 정리하고 데이터 스킬, 디지털 스킬, AI 스킬 간 차이를 제시 △앨런튜링연구소, 국가데이터혁신센터, 공공데이터연구원, 데이터스킬 태스크포스, AI위원회, 영국 사이버보안위원회 등의 역할 명확화 등이 제시됐다.

공공교육 및 직업교육에서의 데이터 스킬 대응에서는 교육 수준에 따라 대응이 구분되고 있다. △우선 학교에서는 디지털 전략의 일환으로 국가 스킬 기금(National Skills Fund)를 통해 데이터 스킬 교육 방안 마련 △대학 학부생 대상 스킬 교육은 AI, 사이버 보안, 디지털스킬 등과 같은 폭넓은 주제를 포함하는 모듈을 제공하거나 데이터 스킬을 다른 과목에 통합하는 등 두 가지로 접근하고 있다. △연구 분야 엔지니어와 전문가의 고급 데이터 스킬 확충을 위한 방안 마련 △국가 스킬 기금을 통해 기업의 데이터 스킬 훈련을 지원할 수 있는 온라인 포털 출범 등이 명시됐다.

공공부문 전반의 데이터 스킬 확충을 통한 역량, 리더십 및 문화 확충과 관련해서는 △정부 전반의 올바른 스킬 구축을 우선 실시 △데이터 과학 캠퍼스(Data Science Campus)를 통해 공공부문에서 데이터 과학 관련 500명 규모의 분석가 훈련 △지방정부의 데이터 관리·사용·전파 역량과 관련된 니즈 검토 등이 포함됐다.

지속 가능한 데이터의 사용성 및 책임성 강조

셋째, 데이터 가용성으로 여기에서는 경제사회, 정부 및 공공부문 국제사회에서의 데이터 가용성 확보가 논의되고 있다. 경제사회의 데이터 가용성 확보에서는 △통신, 금융, 에너지, 연금 분야에서 추진되던 ‘스마트 데이터(Smart Data)’ 관련 이니셔티브들을 가속화하기 위해 분야 횡단적 스마트 데이터 워킹그룹 신설 △스마트데이터 관련 이니셔티브들에 대한 참여 지시 역량을 개선하기 위한 1차 입법 실시 △오픈데이터와 관련해 데이터의 일관성을 보장하기 위해 오픈데이터 발간물 및 의사결정 과정 검토 △에너지 데이터 태스크포스의 권고사항 시행 및 에너지 데이터 접근 현대화 프로그램(Modernizing Energy Data Access) 촉진 △아동 성학대, 혐오발언, 자해 및 자살 생각 등 온라인 유해물 모니터링 및 보고 관련 인프라의 검토·개선 등이 포함됐다.

정부 및 공공부문의 데이터 가용성 확보에서는 2017년 ‘디지털 경제법(Digital Economy Act)’에 따라 데이터 공유를 둘러싼 장애요인에 대비한다는 입장이다. 한편 국제 데이터 가용성 확보에서는 △EU, 미국, 일본, 호주, 뉴질랜드 등과 국경 간 데이터 흐름의 불필요한 장애요인 제거 협력 추진 △신설 ‘글로벌 인공지능 파트너십(Global Partnership on AI)’에서 영국이 공동의장을 맡고 있는 데이터 거버넌스 워킹그룹(Data Governance Working Group)을 활용해 국제 데이터 접근 및 공유 추진 △영국으로부터 이뤄지는 개인데이터 이전과 관련해 영국 자체의 데이터 타당성 평가를 시행할 수 있는 정부의 독립적 역량 구축 △국제 데이터 이전과 관련된 과도기적 방식, 영국 외에서 개인 데이터 이전이 적절하게 보호될 수 있는 대안적 이전 메커니즘 활용 등을 검토한다는 계획이다.

마지막으로 데이터 책임성에서는 성장 친화적 데이터 권리 제도 마련과 안전하고 지속 가능한 데이터의 사용을 제시하고 있다. 성장 친화적 데이터 권리 제도 마련과 관련해서는 △혼란과 관성을 초래해온 패스트트래킹 가이드선(Fast-Tracking Guidance), 공동 규정틀 사용 등 기존 데이터 제도의 명확성 담보 △기업, 특히 중소기업의 규제준수 부담 경감 △규제 샌드박스과 같은 개입을 통한 혁신자(innovator) 지원 △공공부문 내에서 알고리즘 지원 의사결정 활용과 관련한 투명성을 담보할 수 있는 적절하고 효과적인 메커니즘 활용 △소비자 통제 권한 및 신뢰 강화를 위한 프라이버시 강화 기술 역할 고찰 △민간 및 제3부문의 데이터 사용과 관련한 공정성, 투명성 및 신뢰성 등을 보장할 수 있는 추가적 조치 모색 △‘글로벌 인공지능 파트너십’의 창립 멤버로서 영국의 포지션 강화 등을 주요 내용으로 하고 있다. 안전하고 지속 가능한 데이터의 사용과 관련해서는 기후위기에 대비하는 ‘정부 녹색화 전략(Greening Government: ICT & Digital Services Strategy 2020-2025)’에 발맞춘 ‘데이터 지속 가능성 헌장(Data Sustainability Charter)’ 등의 마련을 강조했다.

[표 2-2-2] 영국의 국가 데이터 전략의 주요 내용

구분	주요 내용
데이터 기반	• 경제사회의 광범위한 데이터 기반 조성 - 디지털 트윈 관련 정보관리 프레임워크(Information Management Framework) 개발 등을 비롯해 위치, 인공 구조물, 자연환경, 운송, 기타 인프라 등에 관한 데이터 표준화 실시를 통한 물리적 환경 조성 - ‘영국 연구개발 로드맵’에 따라 국가 및 지역 차원의 혁신에서 신뢰할 수 있는 데이터 사용 최대화 - 소비자 데이터가 소비자를 위해 활용될 수 있도록 노력 • 정부 및 공공부문의 데이터 기반 조성 - 데이터 공유 관련 법제도 해석의 일관성을 확보할 수 있는 전문가팀 구성 - 데이터 품질 프레임워크(Data Quality Framework), 정부 데이터 성숙도 모델(Data Maturity Model for government) 등 마련 - 정부의 통합 데이터 플랫폼 개발 - 지방정부가 데이터 편익을 누릴 수 있도록 지원 - 범부처 데이터 리더십을 발휘할 수 있는 고위급 CDO(Chief Data Officer) 채용 • 국제적 데이터 기반 지원 - 글로벌 데이터 상호운용성 확보 지원
데이터 스킬	• 명확성 및 협력 추동 - 경제 전반에서 사용될 수 있는 데이터 스킬을 정리하고 데이터 스킬, 디지털 스킬, AI 스킬 간 차이를 제시 - 앨런튜링연구소, 국가데이터혁신센터, 공공데이터연구원, 데이터스킬 태스크포스, AI위원회, 영국사이버보안위원회 등의 역할 명확화 • 공공교육 및 직업교육 상에서의 데이터 스킬 대응 - 디지털 전략의 일환으로 국가 스킬 기금(National Skills Fund)을 통해 학교에서의 데이터 스킬 교육 방안 마련 - 학부생 대상 스킬 교육은 AI, 사이버 보안, 디지털 스킬 등과 같은 폭넓은 주제를 포함하는 모듈을 제공하거나 데이터 스킬을 다른 과목에 통합하는 등 두 가지로 접근 - 연구 분야 엔지니어 및 전문가의 고급 데이터 스킬 확충을 위한 방안 마련 - 국가 스킬 기금을 통해 기업의 데이터 스킬 훈련을 지원할 수 있는 온라인 포털 출범 • 공공부문 전반의 데이터 스킬 확충을 통한 역량, 리더십 및 문화 확충 - 정부 전반의 올바른 스킬 구축을 우선 실시 - 데이터 과학 캠퍼스(Data Science Campus)를 통해 공공부문에서 데이터 과학과 관련된 500명 규모의 분석가 훈련 - 지방정부의 데이터 관리·사용·전파 역량과 관련된 니즈 검토
데이터 가용성	• 경제사회의 데이터 가용성 확보 • 정부 및 공공부문의 데이터 가용성 확보 • 국제 데이터 가용성 확보
데이터 책임성	• 성장 친화적 데이터 권리 제도 마련 - 혼란과 관성을 초래해온 패스트 트레이킹 가이드نس, 공동 규정률 사용 등 기존 데이터 제도의 명확성을 담보 - 기업, 특히 중소기업의 규제준수 부담 경감 - 규제 샌드박스와 같은 개입을 통해 혁신자(innovator)를 지원 - 공공부문 내에서 알고리즘 지원 의사결정 활용과 관련한 투명성을 담보할 수 있는 적절하고 효과적인 메커니즘 활용 - 소비자 통제 권한 및 신뢰 강화를 위한 프라이버시 강화 기술 역할 고찰 - 민간 및 제3부문의 데이터 사용과 관련한 공정성, 투명성 및 신뢰성 등을 보장할 수 있는 추가적 조치 모색 - 신설 ‘글로벌 인공지능 파트너십(Global Partnership on AI)’의 창립 멤버로서 영국의 포지션 강화 • 안전하고 지속 가능한 데이터의 사용 - 기후위기에 대응하는 ‘정부 녹색화 전략(Greening Government: ICT & Digital Services Strategy 2020~2025)’에 발맞춘 ‘데이터 지속 가능성 헌장(Data Sustainability Charter)’ 마련

출처: Department for Digital, Culture, Media & Sport, National Data Strategy (2020.9.9.)를 재구성

다. 독일의 정책

독일 정부는 2010년 ‘디지털 독일 2015(Deutschland Digital 2015)’를 시작으로 본격적인 디지털 전략을 마련했다. 2014년 ‘디지털 어젠다(Digitale Agenda)’를 거쳐 2016년 독일의 디지털 경제로의 이행 가속화를 목표로 하는 ‘디지털 전략 2025(Digitale Strategie 2025)’를 발표했다. 데이터가 디지털 전략에 본격 등장하는 것은 ‘디지털 전략 2025’다. 이 전략에서는 정보보안 및 정보의 자기 결정권 강화라는 측면에서 데이터를 바라보고 있다. 특히 정보의 자기 결정권은 4차산업혁명과 함께 미국 플랫폼형 비즈니스의 급속한 확산에 대한 독일 및 EU 차원의 대응 필요성과 맞물리면서 데이터 주권(Data Sovereignty)의 개념으로 발전하게 됐다. 이 개념은 현재까지도 독일 연방정부 데이터 정책의 골자를 이루고 있다.

공정 경쟁과 데이터 주권 확보에 역점

데이터 주권이란 데이터의 사용자가 데이터 사용의 조건을 특정할 수 있는 권리를 의미한다. 이 개념은 독일의 디지털 기술, 서비스, 플랫폼의 창출과 평가, 활용, 보호가 국제적으로 상당히 뒤처졌다는 위기감에서 비롯됐다고 볼 수 있다.¹⁰ 독일 연방경제에너지부는 2017년 3월 「디지털 플랫폼 백서」에 디지털 자유방임(Digital laissez-faire)이나 국가가 조직한 현대화 프로그램이 아닌 제3의 방식으로 디지털화를 추진할 것임과 함께 디지털 플랫폼의 역동적 성장과 이에 대한 참여 활성화 의지를 보였다. 이와 더불어 공정 경쟁 및 데이터 주권을 확보할 수 있도록 관련 조치를 취한다는 입장을 밝혔다.¹¹ 2019년 10월, 독일 연방정부가 발표한 GAIA-X 프로젝트(Project GAIA-X)¹² 역시 독일과 유럽의 데이터 주권 및 관련 기업의 경쟁력 확보 등을 목적으로 하고 있다는 점에서 이러한 맥락을 엿볼 수 있다. 이 밖에도 지난 2019년 11월, 독일 연방정부는 「데이터 전략 초석(Eckpunkte einer Datenstrategie der Bundesregierung)」이라는 보고서에서 데이터가 혁신 및 경쟁력뿐만 아니라 디지털 주권을 확보하는 데 중요한 역할을 한다는 점을 강조했다. △데이터 제공 개선 및 안전한 데이터 접근 △책임 있는 데이터 사용 촉진 및 혁신 잠재력 제고 △데이터 문해력 증진 및 데이터

10. Fraunhofer, Industrial Data Space-Digital Sovereignty over Data, 2016; Polyas, “What Do We Know about Digital Sovereignty?”, 2016.6.

11. BMWi, White Paper: Digital Platforms-Digital regulatory policy for growth, innovation, competition and participation, 2017.3.

12. GAIA-X 프로젝트는 유럽 데이터 인프라의 공통요건을 개발하는 것으로 공개성, 투명성 및 유럽 국가간 연결성을 핵심으로 함

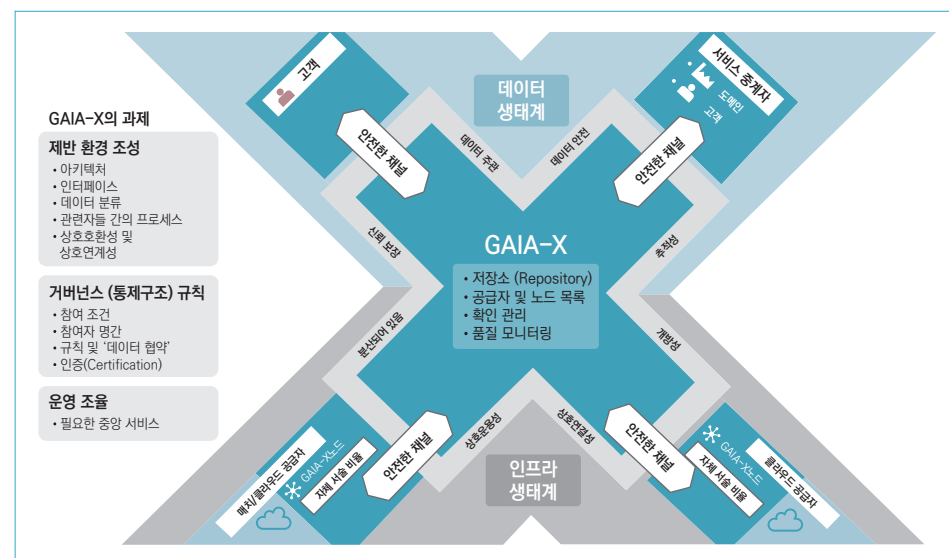
문화 확립 △데이터 사용과 제공의 개척자로서의 정부 역할 확립 등 4가지 분야에서 전략을 개발할 것이라고 명시했다.

‘산업 데이터 스페이스 이니셔티브’에 주변국 합류

독일 연방정부의 데이터 관련 정책은 앞서 서술한 데이터 주권과 함께 산업의 경쟁력을 강화하기 위한 데이터 기반 혁신에 초점을 두고 있다. 특히 '인더스트리 4.0'과 여기서 파생되는 스마트 서비스 세계(Smart Service Welt) 등과 관련한 데이터 기반 혁신을 위해서는 우선적으로 데이터의 안정적인 교환·공유가 이루어져야 하는데, 프라운호퍼연구소(Fraunhofer Institute)는 2016년부터 연방교육연구부의 지원을 받아 비즈니스 생태계에서 데이터의 안정적 교환과 용이한 연계를 목적으로 표준 및 공통 거버넌스 모델을 사용한 데이터 플랫폼인 산업 데이터 스페이스(Industrial Data Space) 및 국제 데이터 스페이스(International Data Spaces) 이니셔티브를 운영하고 있다.

산업 데이터 스페이스 이니셔티브는 연방교육연구부 자금이 지원되는 경쟁 전 연구 프로젝트(pre-competitive)의 성격을 가지고 있다. 산업 데이터 공간의 레퍼런스 아키텍처 개발 및 시범 테스트를 수행하고 있다.¹³ GAIA-X 프로젝트¹⁴도 독일과 유럽의 데이터 제공자와 사용자가 참여하는 데이터 생태계를 창출하기 위한 인프라 구축 사업이다. 현재는 프랑스 정부가 합류했으며, 다양한 국가에서 300개 이상의 기관들이 참여하고 있다.¹⁵

[그림 2-2-2] 독일 GAIA-X 프레임워크



출처: BMWi & BMBF, Project GAIA-X, 2019.10.; 김은, 「독일의 제조 분야 디지털 트랜스포메이션」 34부 - GAIA-X 프로젝트, 2020.5.11., ZDNet Korea에서 재인용

3. 일본의 정책

일본 정부는 2011년 동일본 대지진을 계기로 사회 전반의 협력을 위한 오픈 데이터의 중요성을 깨달으면서 ‘전자행정 오픈 데이터 전략(電子行政オープンデータ戦略)’을 수립하는 등 본격적인 오픈 데이터 정책을 펼치고 나섰다. 2016년을 전후로 4차산업혁명이 전 세계적인 주목을 받게 되면서 데이터 이용·활용을 성장전략 내로 적극 포함했다.

우선 오픈 데이터 정책을 살펴보면¹⁶, ‘전자행정 오픈 데이터 전략’에 이어 이듬해인 2013년 6월 백서, 방재·감재정보, 지리공간정보, 사람 이동정보, 예산·결산·조달정보 등 중점 5개 분야의 우선적 정책 시행을 골자로 하는 ‘전자행정 오픈 데이터 추진 로드맵’을 발표했다.

공개에서 활용과 과제 해결로 초점 이동

일본 정부는 2015년 6월 ‘새로운 오픈 데이터의 전개를 향해(新たなオープンデータの展開に向けて)’를 발표하며 공개에 맞춰졌던 오픈 데이터 정책의 초점을 과제 해결형으로 옮겼다. 이 차원에서 공공데이터를 포함한 빅데이터의 분석·활용 효과를 사회적 과제 해결로 이어가야 한다는 점을 강조했다. 2016년에는 ‘오픈 데이터 2.0(オープンデータ2.0)’을 발표하고 2020년까지 데이터 공개 중심의 활동에서 데이터 활용을 통한 과제 해결을 집중 추진한다는 입장을 재천명했다. 같은 해인 2016년 12월에는 ‘민관 데이터 활용 추진 기본법(官民データ活用推進基本法)’을 제정해 민관 데이터 활용과 관련된 국가, 지방 공공단체 및 사업자의 책무를 명확히 했다.

데이터 활용이 강조된 오픈 데이터 정책은 한 걸음 더 나아가, 모든 국민이 IT 및 데이터 활용을 통해 그 혜택을 온전히 누릴 수 있도록 하는 민간 데이터 활용 사회를 목표로 환경 정비 등의 내용을 담은 ‘세계 최첨단 IT 국가 창조 선언·민간 데이터 활용 추진 기본계획(世界最先端IT國家創造宣言·官民データ活用推進基本計画)’, 2017.05’으로 구체화됐다. 이 계획은 2013년 발표한 ‘세계 최첨단 IT 국가 창조 선언(世界最先端IT國家創造宣言)’의 4차 개정이다. 공공부문을 포함한 사회 전 분야에서 IT 활용 촉진, IT를 통한 사회적 과제 해결, 데이터 공개 및 유통 등과 같은 기존의 정책 방향에서 IT 전략을 데이터 활용으로 대전환하고 있는

13. 「산업데이터플랫폼 확산 및 정책방향」 KIAT, 2018.10

14. BMWi & BMBF, Project GAIA-X, 2019.10

15. GAIA-X, <https://www.data-infrastructure.eu/GAIA-X/Navigation/EN/Home/home.html>; “미국 클라우드 의존 낮추자”...독일 프랑스 ‘가이아X’ 프로젝트 추진, 디지털투데이, 2020.6.5.

16. 「オープンデータの推進(政府全体の取組)」総務省, 2012.7.
https://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/opendata/seihu_od_torikumi.html

모습이다. 2020년 개정된 이 계획에서는 코로나19 사태와 맞물려 국민의 생명 보호 및 경제 재생을 위한 데이터 활용이 강조됐다. 구체적으로는 안전하고 공정한 디지털 시장의 규제 형성, 원활한 데이터 활용 환경 정비, 오픈 데이터 심화, 공유경제의 추진 확대 등을 명시하고 있다.¹⁷

‘리얼 데이터 플랫폼 분야’ 선점 의지

다음으로는 성장 전략 측면에서의 데이터 이용·활용이다. 4차산업혁명이 본격적으로 등장한 이후 일본 정부는 기술 혁신과 데이터를 활용한 새로운 수요 창출이 불가피함을 인식하고 생산 현장의 리얼 데이터(real data)¹⁸를 최대한 활용해 제조기술과 데이터를 결합하는 방식으로 새로운 경쟁 우위를 확보해 나간다는 입장을 표명했다. 이는 미국의 구글, 애플, 페이스북, 아마존 등 대형 플랫폼과의 데이터 경쟁에서 우위를 점하기는 사실 어렵다는 점을 자각해, 자국이 잠재적 경쟁력을 갖는 리얼 데이터 플랫폼 분야를 선점하겠다는 인식이 작용했기 때문이다.¹⁹ ‘4차산업혁명을 향하여(第4次産業革命に向けて)’라는 부제가 붙은 2016년도 성장 전략인 ‘일본재흥전략(日本再興戦略 2016)’에서는 기업·조직의 틀을 넘는 데이터 이용·활용 플랫폼, 공유경제 추진, 사이버 보안 강화 등의 실시를 명시했다. 이후 2018년도 성장전략인 ‘미래 투자전략(未来投資戦略)’에서는 소사이어티 5.0과 데이터 구동형 사회로의 전환을 목표로 경제구조 혁신기반 조성을 위해 데이터 기반 사회의 공통 인프라 정비를 추진한다는 내용을 명시했다.

구체적으로는 산업 데이터의 연계·활용, 개인 데이터의 활용, 민간기업 분야에서의 디지털 전환 촉진 등이 포함됐다. 특히 산업 데이터의 연계·활용과 관련해 2018년 6월 ‘생산성 향상 특별 조치법(生産性向上特別措置法)’에 근거한 데이터 활용 지원 시책을 마련하는 한편, 자동주행·이동성 서비스, 제조·로보틱스, 바이오·소재 등 커넥티드 인더스트리(Connected Industries)의 중점 분야를 중심으로 산업 데이터 공유·연계 사례를 확대하는 동시에 구현 지원을 강화한다고 명시했다.

한편 일본 경제산업부는 정부의 성장 전략과 발맞추어 4차산업혁명에 대응하기 위해 ‘신산업 구조 비전(産業構造ビジョン, 2016.4.)을 발표한 바 있다. 여기에서도 데이터

이용·활용 촉진을 위한 환경 정비를 강조하고 있다.²⁰ 이러한 정책 방향에 의거해 일본 경제산업부는 ‘생산성 향상 특별조치법’을 근거로 2018년 6월부터 산업 데이터 공유 사업 인정제도(産業データ共有事業の認定制度)를 시행하고 있다. 이는 사물인터넷 발전으로 데이터의 유통량이 폭발적으로 증가하는 가운데, 산업 경쟁력 강화와 사회적 문제 해결을 위한 데이터 사용 촉진 차원에서 데이터 공유 사업을 수행하는 민간 사업자를 인증·지원하는 제도다. 주무대신으로부터 민간 사업자의 사업계획이 인정을 받으면, 공공기관에 특정 데이터 제공을 요청할 수 있고, 사이버 보안이나 세제 지원 등을 받을 수 있다.²¹

포스트 코로나 시대 대비해 AI 데이터의 중요성 강조

최근에는 총무성 산하 정보통신정책연구소가 「2020년도 AI 네트워크 사회 추진회의(AIネットワーク社会推進会議) 보고서(2020.7.)」를 발표했다. 이 보고서에서는 포스트 코로나 시대의 AI 활용을 전제로 한 경제·사회로의 이행 가속화 가능성이 높다는 점을 강조하고 있다. 또한 AI 시대의 데이터 경제 구축을 위해서는 △중소기업 및 지방 기업에 대한 AI 데이터 활용 활성화를 위한 민관의 지원 △AI 등의 신기술에 대한 올바른 이해를 위한 민관의 활동 촉진 △포용적 AI 경제·사회 이미지 공유 △AI 데이터 활용을 촉진할 수 있는 제도의 재검토 △포스트 코로나 사회를 염두에 둔 AI 데이터 활용의 추진을 구체적 방안으로 제시하고 있다. 이는 향후 일본 정부의 데이터 정책이 AI 경제·사회 구현과 적극 결합될 것임을 보여주는 바라 할 수 있다.

17. 「世界最先端デジタル国家創造宣言・官民データ活用推進基本計画」, 高度情報通信ネットワーク社会推進戦略本部(IT 総合戦略本部), 2020.7.17.

18. 리얼데이터란 건강 정보, 주행 데이터, 공장설비 가동 데이터 등과 같이 실제 세계의 활동(개개인의 생활정보 및 제품의 가동상황 등을 네트워크에 접속된 센서를 통해 직접적으로 수집하는 데이터를 의미

19. 「日本再興戦略 2016」, 日本経済再生本部, 2016.6.2.

20. 「新産業構造ビジョン」中間整理, 経済産業省, 2016.4.

21. 「産業データ共有事業について」, 経済産業省, 2018.12.11.

4. 중국의 정책

중국 정부는 2013년부터 경제발전 모델 전환을 위해 적극적인 노력을 기울여왔다. 이 측면에서 2014년 3월 정부업무보고(政府工作报告)에서 빅데이터를 정책영역으로 포함하면서 빅데이터 정책 출범을 알렸다. 이후 빅데이터는 정부 정책 전반으로 확대돼 적극적인 지원 정책과 함께 일정 정도의 규제 환경을 정비했다.²² 다음해인 2015년 8월 중국 국무원은 「빅데이터 발전 촉진을 위한 행동요강(促进大数据发展行动纲要)」에서 빅데이터라는 기초적인 전략자원을 활용해 ‘대중창업·만중혁신’ 촉진, 기존 산업의 질적 전환, 정부 관리·감독 및 사회관리 수준 제고 등을 추진할 것이라고 명했다. 이러한 중국의 빅데이터 정책은 ‘국민경제 사회발전 제13차 5개년 계획(国民经济和社会发展第十三个五年规划, 2016~2020)’의 핵심 이니셔티브인 ‘인터넷 플러스(互联网+)’, ‘중국제조 2025(中国制造 2025)’ 등 정보통신기술을 전면에 내세워 산업 구조를 근본적으로 개혁하려는 시도와 맞물려 더욱 탄력을 받았다. 중국 공신부는 2016년부터 2020년까지 ‘빅데이터 산업발전계획(大数据产业发展规划)’을 통해 일상생활에서 광범위하게 활용되는 빅데이터를 제조 분야로 확대하고, 특히 생산 모델의 고도화 및 스마트 공장을 추진하는 과정에 빅데이터 활용 필요성을 강조했다.²³

구체적으로 추진할 중점 프로젝트로는 △빅데이터 기술제품 개발 강화 △산업 빅데이터 혁신 및 응용 △산업계의 빅데이터 애플리케이션 개발 촉진 △빅데이터 산업 주체 육성 가속화 △빅데이터 표준 시스템 구축 촉진 △빅데이터 산업 지원 시스템 개선 △빅데이터 보안 역량 제고 등이다. 한편 과학기술혁신 측면에서도 빅데이터는 중요한 과제로 등장했다. 중국 국가과학기술혁신 13차 5개년 계획(十三五国家科技创新规划)에서 2030년을 지향하는 과기혁신 2030 중대 프로젝트 중 하나로 빅데이터 분야가 명시됐다.²⁴ 2017년 10월에 개최된 제19차 중국 공산당 전국대표대회(中国共产党全国代表大会)에서는 빅데이터와 실물경제의 융합 추진이 제안되면서 향후 빅데이터 산업 발전에 대한 방향성을 제시했다. 2019년 3월, 정부 업무보고에서 빅데이터가 6번이나 언급되는 등 빅데이터 관련 정책을 강조했다.

22. Big Data White Paper(2019), CAICT, 2019.12.

23. 「工业和信息化部关于印发大数据产业发展规划(2016 - 2020年)的通知」, 工业和信息化部, 2016.12.; 유영신, 「4차산업혁명을 대비하는 중국의 ICT 산업 및 정책 동향」, 정보통신기술진흥센터, 2017.5.

24. 「院关于印发促进大数据发展行动纲要的通知」, 国务院, 2015.8.

데이터센터 설립에 박차 가하는 중국 지방정부

중국은 중앙정부의 정책에 따라 지방정부 역시 빅데이터 정책을 적극적으로 추진하고 있다. 국무원이 「빅데이터 발전 촉진을 위한 행동요강」을 발표한 직후 랴오닝성, 칭하이성을 필두로 대부분의 지방정부에서 빅데이터와 관련된 정책을 도입했다. 최근 일부 지방정부에서 빅데이터 산업을 포함한 빅데이터 발전을 담당할 조직을 신설하고 있다. 예를 들어 광둥성은 2014년 중국에서는 최초로 광둥빅데이터청(Guangdong Big Data Administration)을 설립하는 등 2019년 5월까지 총 14개 지방정부가 관련 조직을 신설한 것으로 파악되고 있다. 이와 함께 지방정부에서도 빅데이터 센터 설립을 가속화하여 상하이 빅데이터센터, 텐진 빅데이터관리센터 등이 설립됐다. 장시성이 설립한 장시성 정보센터도 빅데이터 관련 업무 일부를 담당하고 있다. 2018년 10월 통계에 따르면, 성급 이하 행정구 중 79개 행정구가 빅데이터관리 특별조직을 설립한 것으로 나타났다.

최근 중국 정부의 데이터 관련 정책은 다양한 분야에 걸쳐 강화되고 있다. 신인프라 건설, 데이터 보안, 산업 빅데이터 발전 등이 대표적이다. 첫째, 신인프라(新基建)는 2018년 12월 중앙경제업무회의에서 처음으로 등장한 개념이다. 전통적인 ‘철도기반 인프라(铁公基)’에 상응해 새로운 과학기술 혁명과 산업 변화의 특징을 결합하고, 국가의 전략적 수요에 보조를 맞추면서 경제사회의 혁신, 조화, 녹색, 개방, 공유발전 기반을 조성하기 위한 신개념이라고 할 수 있다.²⁵ 신인프라에는 5G 통신망, 특고압 설비, 도시 간 고속도로 및 도시철도, 새로운 에너지 자동차 충전장(充电桩), 빅데이터 센터, 인공지능, 산업 인터넷 등 7대 분야가 포함돼 있다.

국가 안보 일변도에서 탈피한 데이터 보안정책

중국 정부는 7대 분야의 중점 산업을 대상으로 디지털화와 스마트화를 통해 산업의 데이터 활용을 촉진하고, 초고도화, 스마트 제조, 스마트 커넥티드 자동차, 신에너지 자동차 등 관련 선도 산업 분야에서 응용 환경을 개선·보완해 산업 발전의 교두보를 확보한다는 입장이다. 신인프라 개념이 발표된 이후 2020년 중반까지 신인프라 건설과 관련된 중점 프로젝트 투자 계획을 발표한 지방정부는 13개에 달하는 것으로 나타났다.²⁶

둘째, 데이터 보안 측면에서 중국 정부는 2020년 7월 제13차 전국인민대표회의에서 ‘데이터보안법(数据安全法)’ 초안을 공개했다. 이 법의 초안은 그동안 보안에만 중점을 뒀던 중국의 데이터 전략이 보안과 경제적 편익 사이에서 균형을 찾는 것으로 선회했음을

25. 「新基建”发展白皮书」, 赛迪智库电子所, 2020.3.19

26. 「中 신 인프라 투자 확대를 통한 미래 성장동력 가동」, KOTRA 중국 상하이무역관, 2020.6.9.

보여줬다.³⁸ 중국 내에서는 새로운 상품으로서의 데이터 시장을 창출할 수 있도록 규제를 적용하고, 국제적으로는 무역전략에 데이터를 포함시키는 노력을 기울이기 시작한 것이다.

셋째, 산업용 빅데이터의 발전이다. 2020년 5월 중국 공신부(工业和信息化部)는 빅데이터 발전 전략 시행, 산업 디지털 전환 촉진, 산업 데이터 자원의 잠재력 추동, 빅데이터 산업 발전 가속화 등을 위한 「산업 빅데이터 발전에 관한 지도의견(关于工业大数据发展的指导意见)」을 발표했다.³⁹ 이 지도의견에서 산업 데이터 발전을 위해 △데이터 통합 가속화 △데이터 공유 촉진 △데이터 융합·혁신 심화 △데이터 거버넌스 역량 제고 △데이터 보안 강화 △산업발전 촉진 △조직지원 강화 등 7대 분야에서 산업 빅데이터 플랫폼의 통합 구축, 산업 데이터 공유 및 개방 촉진, 산업 데이터 시장 활성화, 산업 데이터 응용 생태계 조성 등 21개 지침을 제시했다.

5. 동남아시아의 정책

아세안은 역내 디지털화를 통해 지속적인 경제사회 발전을 이끌어갈 수 있도록 2010년부터 5년 단위의 정보통신기술 마스터플랜(ASEAN ICT Masterplan)을 발표해왔다. 2020년까지의 디지털 경제로의 전환 목표를 담은 2015년 마스터플랜에서는 정보보안 및 보장이 새로운 핵심 전략 중 하나로 추가됐다. 이 전략 방향을 배경으로 2018년 12월 ASEAN 정보통신장관회의에서는 회원국 간 데이터 관리 강화 및 데이터 규제 조화 촉진, 역내 데이터 이동 활성화 등을 골자로 하는 ‘ASEAN 디지털 데이터 거버넌스 프레임워크(ASEAN Framework on Digital Data Governance)’를 승인했다.⁴⁰

이는 ASEAN이 하나의 통합 공동체로서 연결성을 강화하기 위해 마련한 ASEAN 연결성 마스터플랜 2025(ASEAN Master Plan on ASEAN Connectivity 2025, 2016.8.)에서도 이미 명시됐던 사항이다. ‘디지털 데이터 거버넌스 프레임워크’는 ASEAN 디지털 경제를 지원할 수 있도록 △데이터 라이프사이클 및 생태계 △국경 간 데이터 흐름 △디지털화 및 신기술 △법규제 및 정책 등 4개 분야를 디지털 데이터 거버넌스 전략 우선 분야로 선정하고, 이러한 전략 우선

분야의 시행을 위해 4개 이니셔티브의 추진을 명시했다.

4개 이니셔티브는 △데이터 분류 프레임워크(Data Classification Framework) △국경 간 데이터 흐름 메커니즘(Cross-Border Data Flows Mechanism) △디지털 혁신 포럼(Digital Innovation Forum) △데이터 보호 및 프라이버시 포럼(Data Protection and Privacy Forum) 등이다.

한편 ASEAN은 2016년 11월 ASEAN 회원국과 지역 차원에서의 데이터 보호 조치 시행 원칙을 제시한 ‘개인데이터 보호 프레임워크(Framework on Personal Data Protection)³⁰’도 승인한 바 있다. 이 프레임워크는 개인정보 사용에 대한 고지 및 사용 목적에 대한 동의, 개인데이터의 정확성, 보안, 접근성 및 정보 수정, 타국으로의 데이터 이전, 보유, 책임성 등에 관한 원칙을 명시하고 있다. 여기에는 개인정보보호 강화와 역내의 무역 및 정보 흐름 촉진 등의 지원 계획이 담겨 있다.³¹

27. Technode, China sets the rules for its new data economy, 2020.8.4.

28. 「工业和信息化部关于工业大数据发展的指导意见」, 工业和信息化部, 2020.5.13.

29. ASEAN TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY MINISTERS MEETING (TELMIN), FRAMEWORK ON DIGITAL DATA GOVERNANCE, 2018.12.

30. ASEAN TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY MINISTERS MEETING (TELMIN), FRAMEWORK ON PERSONAL DATA PROTECTION, 2016.11.

31. GSMA, Regional Privacy Frameworks and Cross-Border Data Flows, 2018.9.

제 3 부

데이터산업
시장 현황

제1장.
국내 데이터산업 시장 현황

제2장.
해외 데이터산업 시장 현황

제1장. 국내 데이터산업 시장 현황⁰¹

필자: 유효라
(한국데이터산업진흥원 산업지원실 선임)

국내 데이터산업 현황은 2019년 데이터산업 현황조사 보고서⁰² (과학기술정보통신부, 한국데이터산업진흥원, 2020.3.)를 요약하여 작성했다. 2019년 데이터산업 분류에 따라 데이터 처리 및 관리 솔루션 개발·공급업(이하 ‘데이터 솔루션’), 데이터 구축 및 컨설팅 서비스업(이하 ‘데이터 구축·컨설팅’), 데이터 판매 및 제공 서비스업(이하 ‘데이터 서비스’)으로 구분했다. 본 백서에 수록된 시장규모는 데이터 비즈니스를 영위하는 사업체의 데이터 관련 매출기반의 추정치이다. 2018년 이전까지 시장규모는 확정치인 반면, 2019년 시장규모는 잠정치이므로 향후 공표예정인 확정치 통계 결과와 상이할 수 있다.

1. 국내 데이터산업 시장규모

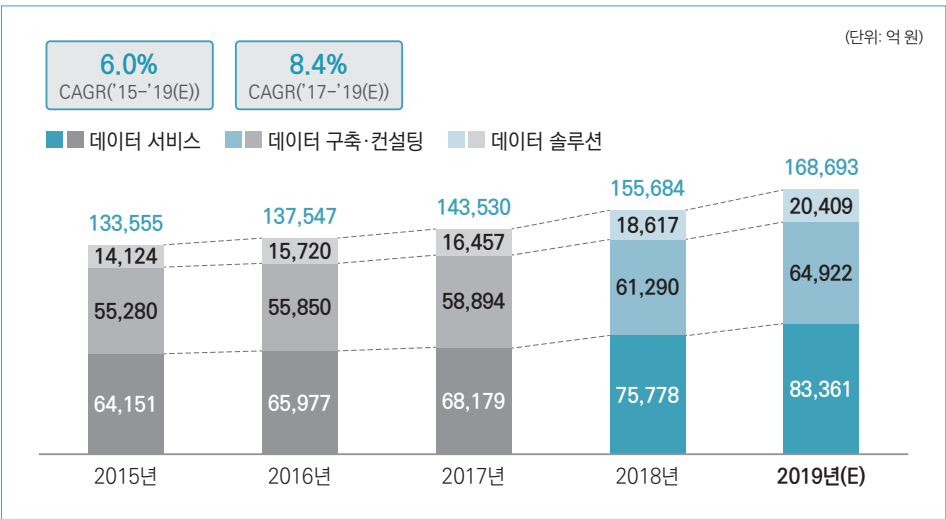
2019년 데이터산업 전체 시장규모 잠정치는 16조 8,693억 원으로, 2018년 15조 5,684억 원에 비해 8.4% 성장했다. 2015년부터 2019년 잠정치까지의 5년간 연평균성장률 (Compound Annual Growth Rate, CAGR)은 6.0%이고 2017년부터 3년간의 CAGR은 8.4%로 나타났다.

2019년 데이터산업은 총 3가지 대분류와 11개 중분류로 구성된다. 부문별 시장규모 잠정치는 데이터 솔루션이 2조 409억 원, 데이터 구축·컨설팅이 6조 4,922억 원, 데이터

01. 본 장에서 데이터산업이란, 데이터의 생산·수집·처리·분석·유통·활용 등을 통해 가치를 창출하는 상품 및 서비스 생산·제공 산업으로 정의

02. 2019년 데이터산업 현황조사는 국가승인통계 127004호로 연간 발간되며, 해당 보고서는 2019년 10~11월에 조사된 결과를 수록해 2020년 3월에 발표했다음

[그림 3-1-1] 최근 5개년(2015~2019(E)) 데이터산업 시장규모



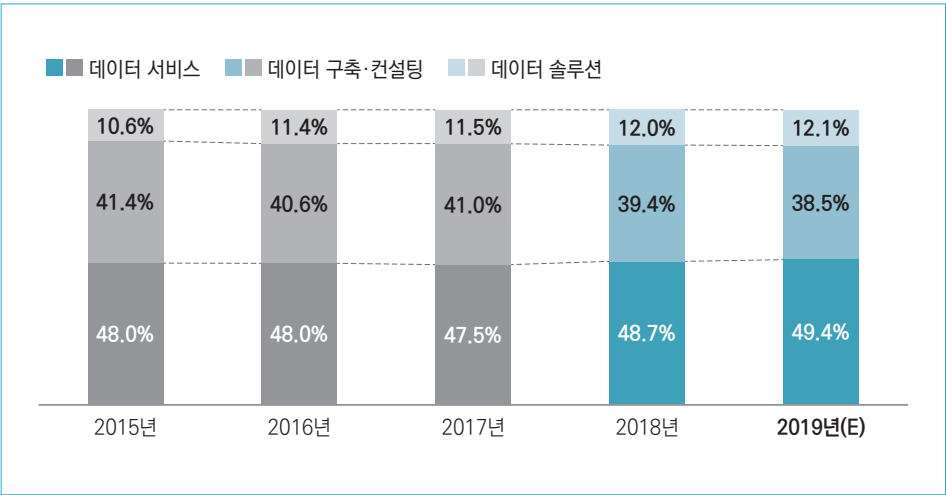
출처: 「2019 데이터산업 현황조사」 보고서, 과학기술정보통신부·한국데이터산업진흥원, 2020.3.(이하 표·그림 모두 동일)

[표 3-1-1] 최근 5개년(2015~2019(E)) 데이터산업 시장규모 총괄

구 분	2015년	2016년	2017년	2018년	2019년(E)	증감률 '18~'19(E)	CAGR	
							'15~'19(E)	'17~'19(E)
전 체	133,555	137,547	143,530	155,684	168,693	8.4%	6.0%	8.4%
데이터 솔루션	14,124	15,720	16,457	18,617	20,409	9.6%	9.6%	11.4%
데이터 구축·컨설팅	55,280	55,850	58,894	61,290	64,922	5.9%	4.1%	5.0%
데이터 서비스	64,151	65,977	68,179	75,778	83,361	10.0%	6.8%	10.6%

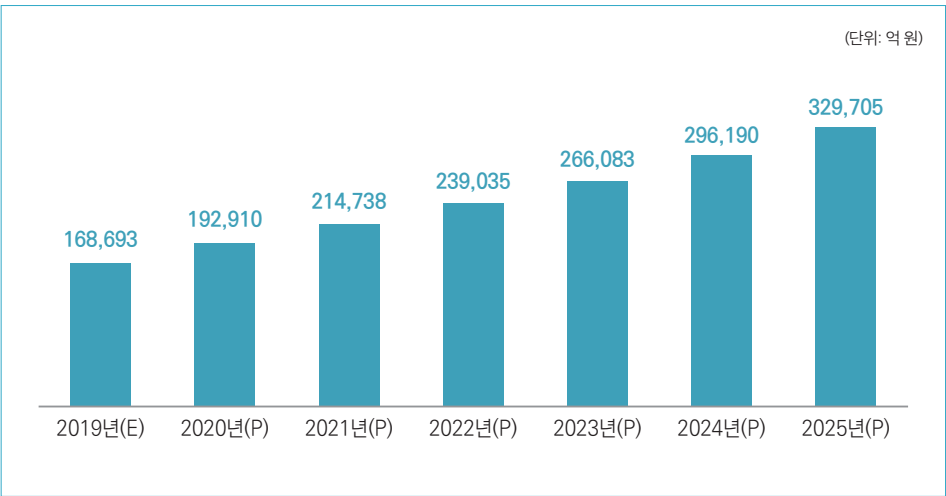
서비스가 8조 3,361억 원으로 예상되고, 그 중 데이터 서비스가 49.4%로 가장 큰 비중을 차지한 것으로 나타났다.

[그림 3-1-2] 최근 5개년(2015~2019(E)) 데이터산업 부문별 시장규모 비중



향후 국내 데이터산업 시장 규모는 2019년부터 2025년까지 연평균 11.3% 지속 성장한다고 가정⁰³한다면, 2025년에 32조 원을 넘어설 것으로 예측된다.

[그림 3-1-3] 2019(E)~2025(P) 데이터산업 시장 전망



* E: 잠정치, P: 추정치

03. 2018년 조사 잠정치 대비 확정치는 11.3% 증가한 확정치를 반영함. 동일한 성장률로 2019년 예측치를 조정했으며, 향후 2025년까지 동일하게 확대된다고 가정함

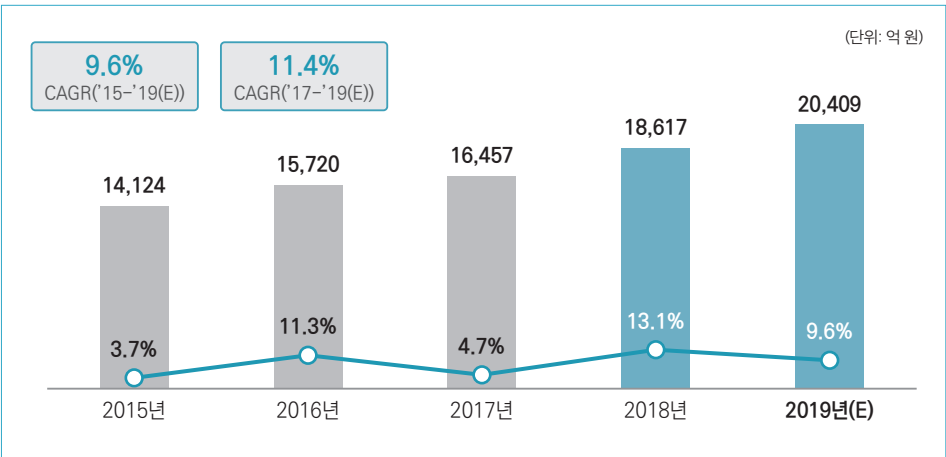
2. 국내 데이터산업 부문별 시장규모

가. 데이터 솔루션 시장

데이터 솔루션 분류는 데이터 처리 및 관리 솔루션 개발 공급업을 의미한다. 데이터 수집 솔루션, 데이터베이스관리시스템, 데이터 분석 솔루션, 데이터 관리 솔루션, 데이터 보안 솔루션, 빅데이터 통합 솔루션을 포함해 6개 중분류로 구분된다. 이들은 주로 솔루션 제품을 판매하는 비즈니스를 의미하고 라이선스, 개발 및 커스터마이징, 유지보수 등에서 매출이 발생한다.

2019년 데이터 솔루션 부문의 시장규모 잠정치는 전년 대비 9.6% 성장한 2조 409억 원으로 조사됐다. 2019년도 데이터 솔루션 시장의 부문별 매출액은 DBMS 솔루션이 약 7,500억 원, 데이터 관리 솔루션이 약 5,000억 원으로 가장 높은 비중을 차지하는 것으로 나타났다. 2017년부터 3년간 연평균성장률(CAGR)은 11.4%로 조사됐다.

[그림 3-1-4] 최근 5개년(2015~2019(E)) 데이터 솔루션 시장규모

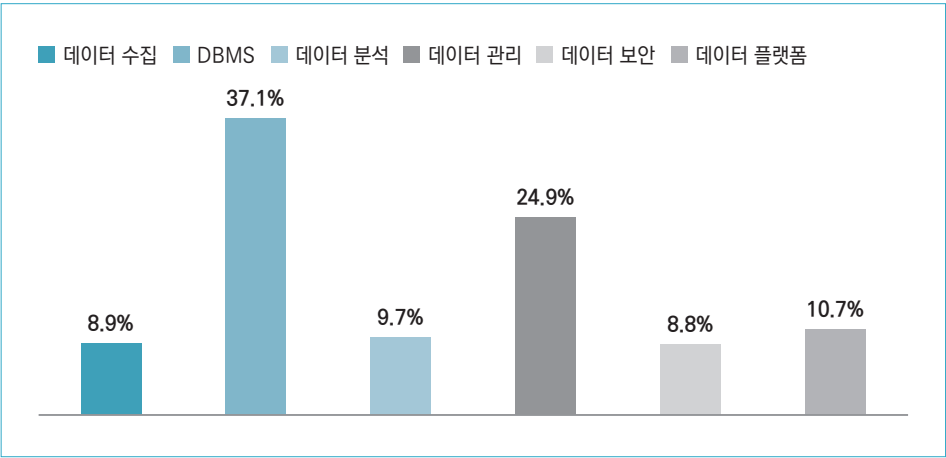


[표 3-1-2] 최근 5개년(2015~2019(E)) 데이터 솔루션 중분류별 시장규모 (단위: 억 원)

구 분	2015년	2016년	2017년	2018년	2019년(E)	증감률 '18~'19(E)	CAGR '15~'19(E)	
데이터 수집	1,115	1,345	1,393	1,622	1,821	12.3%	13.0%	14.3%
DBMS	5,727	6,148	6,121	6,775	7,565	11.7%	7.2%	11.2%
데이터 분석	1,157	1,249	1,325	1,782	1,981	11.2%	14.4%	22.3%
데이터 관리	4,699	5,367	4,628	4,972	5,074	2.1%	-	4.7%
데이터 보안			1,213	1,517	1,794	18.3%	-	21.6%
데이터 플랫폼	1,426	1,611	1,776	1,949	2,174	11.5%	11.1%	10.6%
전체	14,124	15,720	16,457	18,617	20,409	9.6%	9.6%	11.4%

※ 데이터 관리 및 데이터 보안의 경우 2018년 변경된 데이터산업분류체계에 따라 각각 분리됨

[그림 3-1-5] 2019년(E) 데이터 솔루션 중분류별 시장규모 비중



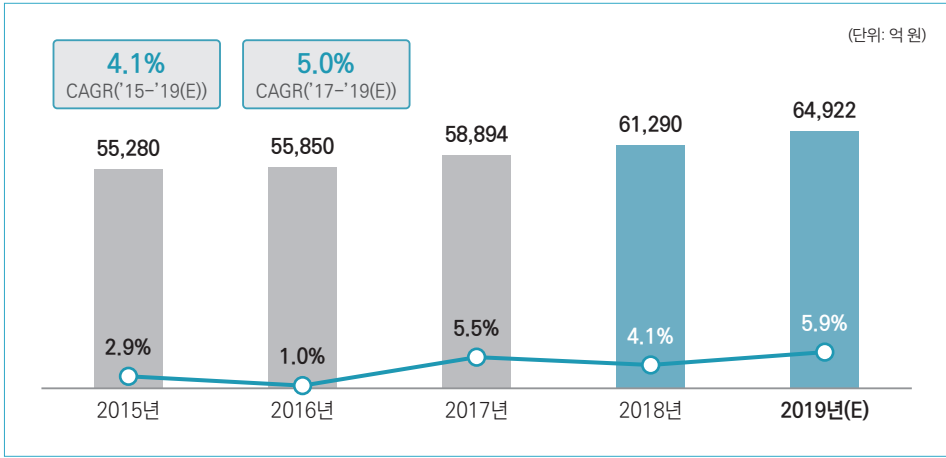
나. 데이터 구축 및 컨설팅 시장

2019년 잠정적인 데이터 구축·컨설팅 시장 규모는 전체 6조 4,922억 원으로 지난해 대비 5.9% 성장할 것으로 나타났다. 이중 데이터 구축 시장 규모는 6조 2,576억 원으로 전체 데이터 구축·컨설팅 시장의 96.4%를 차지했다. 데이터 컨설팅은 2,346억 원인 3.6% 수준이었다. 2017년부터 3년간 연평균성장률(CAGR)은 5.0%로 조사됐다.

데이터 구축 및 컨설팅은 데이터 구축 서비스업과 데이터 컨설팅 서비스업으로 구분된다. 데이터 구축 서비스는 물리적 DB 설계와 구축, 데이터 이행, 데이터 구축 및 가공

서비스 등을 포함하고, 데이터 컨설팅 서비스는 데이터 설계·품질, DB 성능개선, 데이터 거버넌스, 데이터 분석 및 활용 등의 컨설팅 서비스를 의미한다. 비즈니스 매출은 주로 제품·서비스 제공 매출, 용역 및 운영비, 컨설팅 매출 등으로 발생한다.

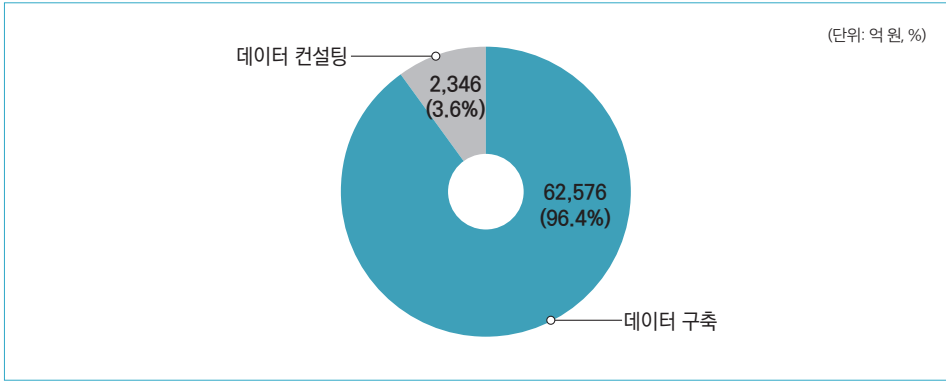
[그림 3-1-6] 최근 5개년(2015~2019(E)) 데이터 구축·컨설팅 시장규모



[표 3-1-3] 최근 5개년(2015~2019(E)) 데이터 구축·컨설팅 중분류별 시장규모 (단위: 억 원)

구 분	2015년	2016년	2017년	2018년	2019년(E)	증감률 '18~'19(E)	CAGR '15~'19(E)	
데이터 구축	54,142	54,571	57,207	58,993	62,576	6.1%	3.7%	4.6%
데이터 컨설팅	1,138	1,279	1,687	2,297	2,346	2.1%	19.8%	17.9%
전체	55,280	55,850	58,894	61,290	64,922	5.9%	4.1%	5.0%

[그림 3-1-7] 2019년(E) 데이터 구축·컨설팅 중분류별 시장규모 비중



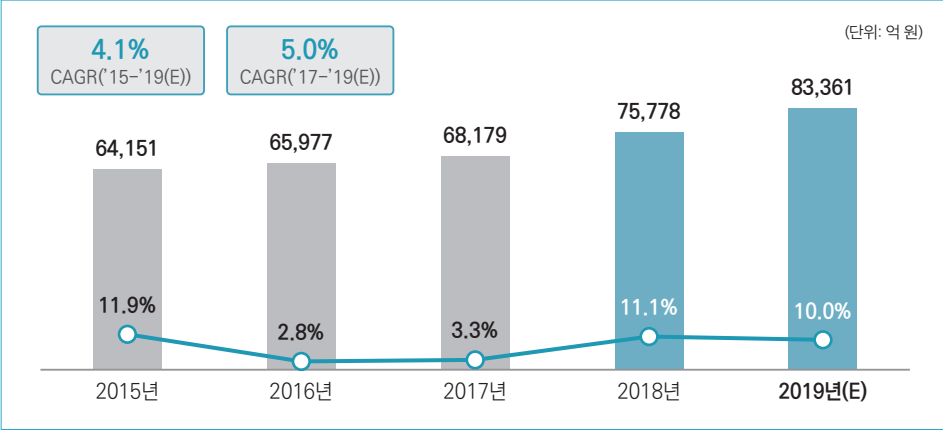
다. 데이터 서비스 시장

2019년 데이터 서비스 시장 규모 잠정치는 전년 대비 10.0% 성장한 8조 3,361억 원으로 조사됐다. 2017년부터 3년간 연평균성장률(CAGR)은 10.6%이다.

데이터 서비스는 데이터 판매 및 제공 서비스업을 의미한다. 원천 데이터와 DB를 수요자에게 가공·활용·분석해 제공하는 비즈니스다. 주로 데이터 이용료·수수료 등의 직접매출과 광고료 등의 간접매출로 수익을 창출한다. 세부 분류로 데이터 중개와 판매, 데이터 신디케이션을 포함한 데이터 거래 서비스업, 포털매개 서비스와 목적별 정보검색을 포함한 정보제공 서비스업, 소셜·마케팅·리스크·기타 데이터 분석결과를 제공하는 데이터 분석제공 서비스업을 포함한다.

2019년 세부 부문별로는 데이터 거래 서비스 시장은 4,315억 원, 데이터 분석제공 시장은 4,382억 원, 정보제공 서비스 시장은 7조 4,664억 원으로 나타났다. 이 중 정보제공 서비스의 비중이 데이터 서비스 내 89.6%로 가장 큰 부분을 차지하고 있으나, 최근 3년간 연평균성장률은 데이터 거래 시장이 21.6%로 성장폭이 가장 크게 나타났다.

[그림 3-1-8] 최근 5개년(2015년~2019년(E)) 데이터 서비스 시장규모



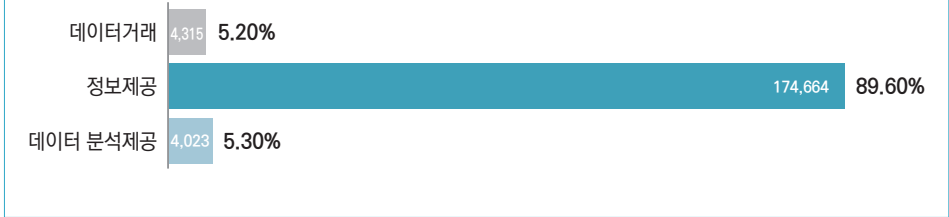
[표 3-1-4] 최근 5개년(2015~2019(E)) 데이터 서비스 중분류별 시장규모

(단위: 억 원)

구 분	2015년	2016년	2017년	2018년	2019년(E)	증감률 '18~'19(E)	CAGR '15~'19(E)	CAGR '17~'19(E)
데이터 구축	2,379	2,594	2,918	4,175	4,315	3.4%	16.1%	21.6%
정보제공	58,171	59,854	61,570	67,580	74,664	10.5%	6.4%	10.1%
데이터 분석제공	3,601	3,529	3,690	4,023	4,382	8.9%	5.0%	9.0%
전체	64,151	65,977	68,179	75,778	83,361	10.0%	6.8%	10.6%

[그림 3-1-9] 2019년(E) 데이터 서비스 중분류별 시장규모 비중

(단위: 억 원)



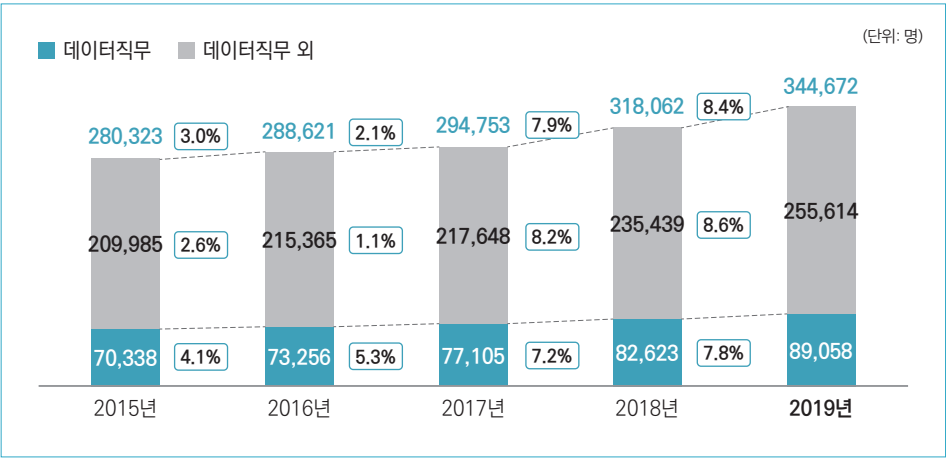
3. 국내 데이터직무 인력 현황

가. 데이터산업 내 데이터직무 인력 현황

2019년 데이터산업에 종사하고 있는 전체 종사자 수는 2018년 전년 대비 약 8.4% 증가한 34만 4,672명이었다. 이 중 데이터 직무인력⁰⁴은 8만 9,058명 수준으로, 전체 데이터산업 종사자의 25.8%를 차지한다. 데이터산업 내 데이터 직무인력은 전년대비 7.8% 증가했다.

04. 데이터개발자, 데이터엔지니어, 데이터분석가, 데이터베이스관리자, 데이터과학자, 데이터컨설턴트, 데이터아키텍트, 데이터기획자 8개 직무를 기준으로 함

[그림 3-1-10] 최근 5개년(2015~2019년) 데이터산업 전체 인력 현황



[표 3-1-5] 최근 5개년(2015~2019년) 데이터산업 전체 인력 현황

구 분	2015년	2016년	2017년	2018년	2019년	증감률 '18~'19	CAGR '15~'19
합계	280,323	288,621	294,753	318,062	344,672	8.4%	5.3%
데이터직무	70,338	73,256	77,105	82,623	89,058	7.8%	6.1%
데이터직무 외	209,985	215,365	217,648	235,439	255,614	8.6%	5.0%

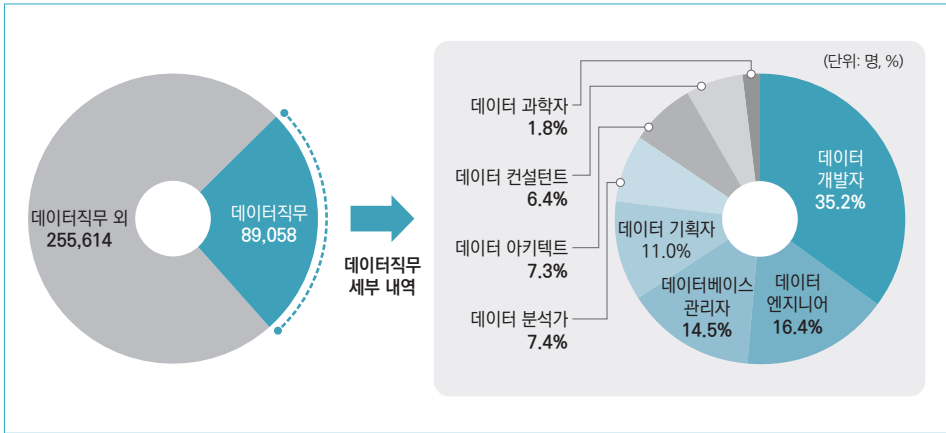
2019년에 부문별로 데이터 직무인력 현황과 추이를 살펴보면, 데이터 구축 및 컨설팅 분야의 데이터직무 인력 수가 4만 2979명으로 가장 많았다. 반면 전년 대비 데이터직무 인력 비율이 가장 크게 증가한 분야는 데이터 솔루션으로, 2018년 대비 16.7% 증가했다.

[표 3-1-6] 최근 5개년(2015~2019년) 데이터산업 부문별 데이터직무 인력 현황

구 분	2015년		2016년		2017년		2018년		2019년		증감률 '18~'19	CAGR '15~'19
	인력수	비중	인력수	비중	인력수	비중	인력수	비중	인력수	비중		
전체	70,338	100.0%	73,256	100.0%	77,105	100.0%	82,623	100.0%	89,058	100.0%	7.8%	6.1%
데이터 솔루션	8,886	12.6%	9,272	12.7%	10,291	13.3%	11,541	14.0%	13,467	15.1%	16.7%	11.0%
데이터 구축·컨설팅	34,323	48.8%	35,404	48.3%	37,516	48.7%	40,197	48.7%	42,979	48.3%	6.9%	5.8%
데이터 서비스	27,129	38.6%	28,580	39.0%	29,298	38.0%	30,885	37.4%	32,611	36.6%	5.6%	4.7%

데이터산업 내에서 직무별 비율을 살펴보면, 데이터 개발자가 3만 1330명(35.2%)으로 가장 많은 비중을 차지하였고, 데이터엔지니어 1만 4592명(16.4%), 데이터베이스관리자 1만 2928명(14.5%) 순으로 나타났다.

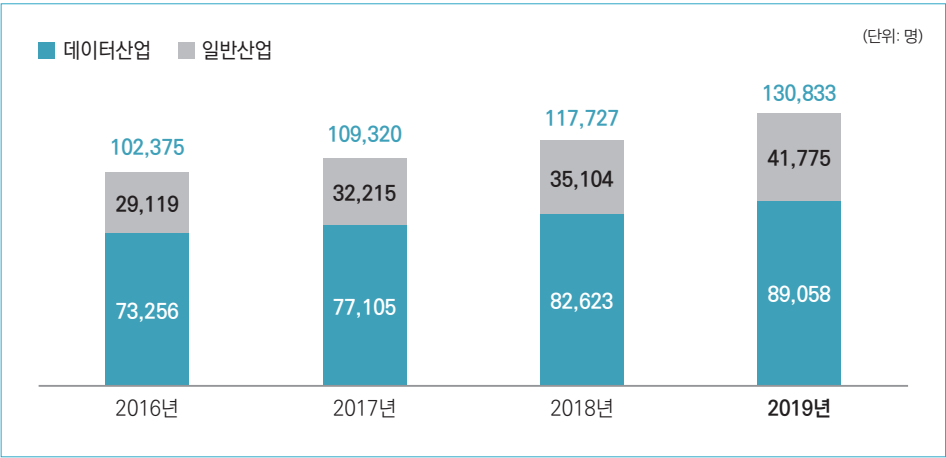
[그림 3-1-11] 2019년 데이터산업 인력 및 데이터직무별 인력 비중



나. 전 산업 내 데이터직무 인력 현황

2019년 일반산업⁰⁵까지 포함한 전 산업의 데이터직무 인력은 총 13만 833명으로, 전년 대비 11.1% 증가했다. 데이터산업을 제외한 일반산업의 데이터직무 인력은 4만 1775명으로 전년도 대비 19.0%로 크게 증가한 것으로 나타났다.

[그림 3-1-12] 2019년 전 산업의 데이터직무 인력 현황



[표 3-1-7] 최근 4개년(2016~2019년) 전 산업의 데이터직무 인력 현황

구 분	2016년		2017년		2018년		2019년		증감률 '18~'19
	인력수	비중	인력수	비중	인력수	비중	인력수	비중	
전 산업	102,375	100.0%	109,320	100.0%	117,727	100.0%	130,833	100.0%	11.1%
데이터산업	73,256	71.6%	77,105	70.5%	82,623	70.2%	89,058	68.1%	7.8%
일반산업	29,119	28.4%	32,215	29.5%	35,104	29.8%	41,775	31.9%	19.0%

전 산업의 데이터직무별 인력은 데이터 개발자가 3만 7007명(28.3%)으로 가장 큰 비중을 차지했고, 데이터베이스관리자 3만 4644명(26.5%), 데이터엔지니어 1만 6165명(12.4%) 순으로 그 뒤를 이었다. 데이터산업과 일반산업의 데이터인력 직무별 비중을 살펴보면, 데이터산업은 데이터 개발자(35.2%)의 비중이 큰 반면, 일반산업은 데이터베이스관리자(52.0%)의 비중이 가장 크다.

[표 3-1-8] 2019년 전 산업의 데이터직무별 인력 현황

구 분	데이터산업		일반산업		전 산업	
	인력수	비중	인력수	비중	인력수	비중
합계	89,058	100.0%	41,775	100.0%	130,833	100.0%
데이터아키텍트	6,508	7.3%	3,852	9.2%	10,360	7.9%
데이터개발자	31,330	35.2%	5,677	13.6%	37,007	28.3%
데이터엔지니어	14,592	16.4%	1,574	3.8%	16,165	12.4%
데이터분석가	6,571	7.4%	5,588	13.4%	12,159	9.3%
데이터베이스관리자	12,928	14.5%	21,716	52.0%	34,644	26.5%
데이터과학자	1,608	1.8%	194	0.5%	1,802	1.4%
데이터컨설턴트	5,686	6.4%	272	0.7%	5,958	4.6%
데이터기획자	9,836	11.0%	2,902	6.9%	12,738	9.7%

05. 일반산업은 데이터산업 외 12개 분야(금융, 제조, 유통·서비스, 교육, 공공, 통신·미디어, 의료, 건설, 물류, 농림·축산광업, 숙박·음식점, 유틸리티(전기, 수도, 가스 등))의 100인 이상 사업체가 포함된 산업을 의미함

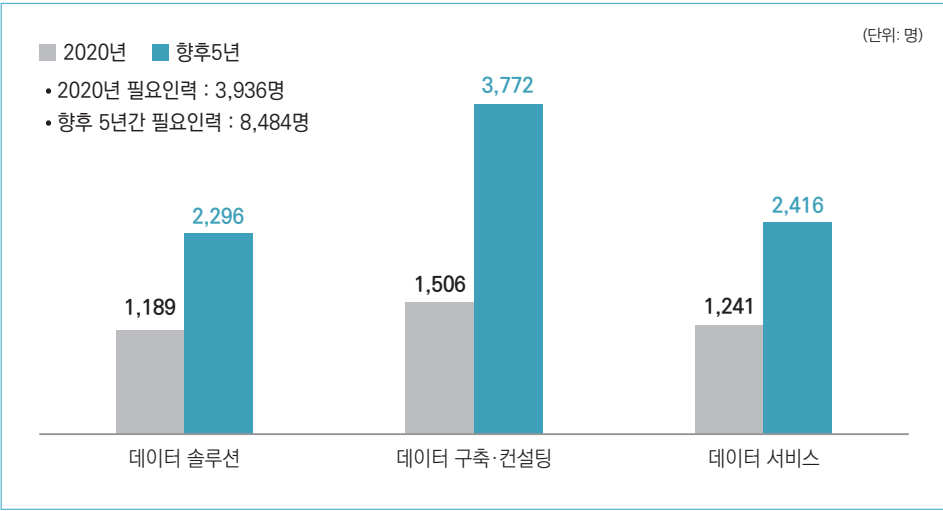
4. 데이터직무 인력 수요

가. 데이터산업 내 데이터직무 인력 수요

데이터산업에서 데이터직무 필요 인력⁰⁶은 2020년까지는 3936명, 향후 5년 이내 총 8484명으로 조사됐다. 직무별 수요는 데이터개발자 수요가 4870명(57.4%)으로 가장 높았고, 데이터분석가 1145명(13.5%), 데이터과학자 836명(9.9%), 데이터엔지니어 785명(9.2%) 순으로 나타났다.

데이터산업 중에서도 데이터 구축·컨설팅 분야가 향후 5년간 3772명으로 가장 많은 인력이 필요할 것으로 나타났고, 데이터 서비스 분야가 2416명, 데이터 솔루션이 2296명의 인력이 필요한 것으로 조사됐다.

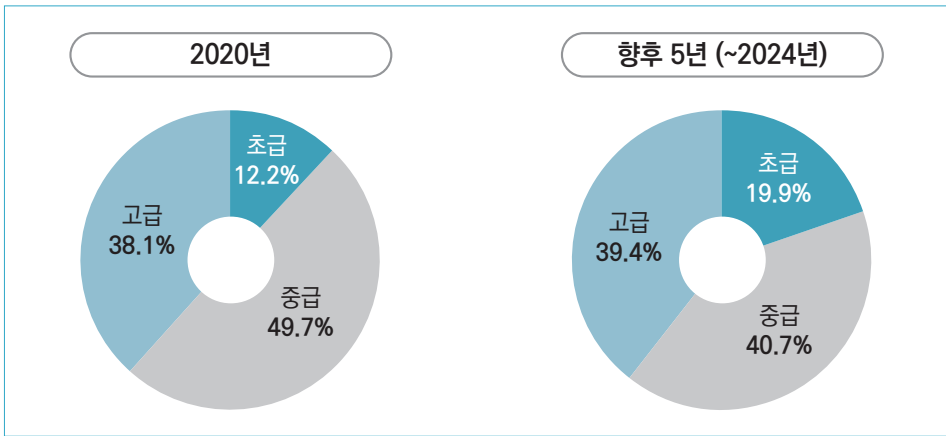
[그림 3-1-13] 향후 5년 내 데이터산업 부문별 데이터직무 필요 인력



향후 5년 내 데이터직무의 기술등급별 필요 인력 수를 살펴보면 중급이 3455명(40.7%)으로 가장 높게 나타났고, 고급 3341명(39.4%), 초급 1689명(19.9%) 순으로 그 뒤를 이었다.

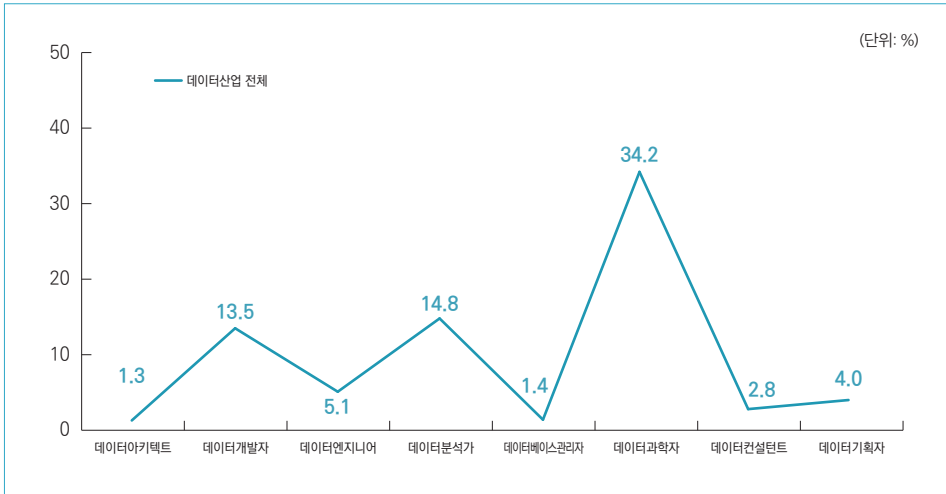
06. 필요 인력은 기업에서 현재 인력보다 추가로 더 필요로 하는 인력 수를 의미함. 즉 현재 부족한 인력 수를 의미하며, 채용할 인력수는 아님

[그림 3-1-14] 향후 5년 내 데이터산업의 기술등급별 필요 인력 비중



데이터산업 내 데이터직무 평균 부족률⁰⁷은 8.7%였다. 부족률은 현재 인력과 필요 인력을 합한 것 대비 필요 인력이 차지하는 비중을 의미한다. 가장 부족한 직무는 데이터과학자(34.2%)이며, 이어서 데이터분석가(14.8%), 데이터개발자(13.5%) 순으로 나타났다.

[그림 3-1-15] 향후 5년 내 데이터산업 부문별 데이터직무 인력 부족률

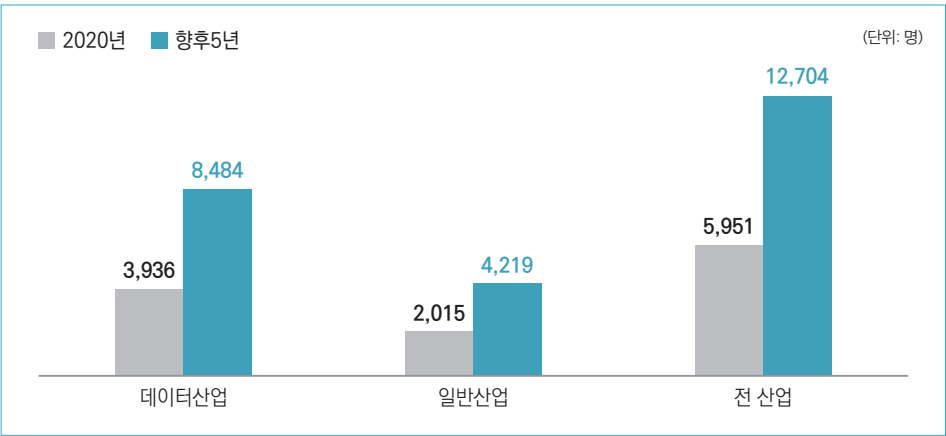


07. 부족률(%) = {필요인력수/(현재인력 수 + 필요인력 수)} × 100

나. 전 산업 내 데이터직무 인력 수요

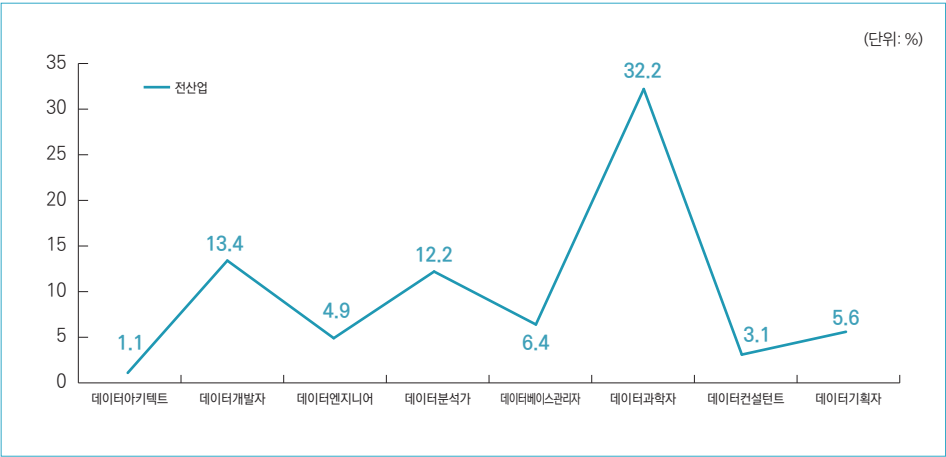
향후 5년 내 일반산업을 포함한 전 산업에서 추가적으로 필요한 데이터직무 인력은 총 1만 2704명이다. 이 중 데이터 개발자가 5717명(45.0%)으로 가장 높게 나타났고, 데이터베이스관리자 2358명(18.6%), 데이터분석가 1694명(13.3%) 순으로 조사됐다.

[그림 3-1-16] 향후 5년 내 데이터산업·일반산업·전 산업 데이터직무 필요 인력



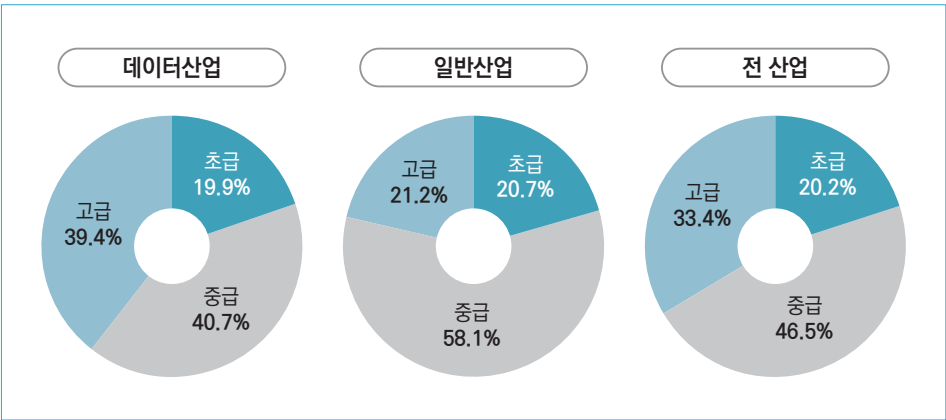
향후 5년 내 일반산업을 포함한 전 산업 내 데이터직무별 인력 부족률은 평균 8.9% 수준이며, 데이터과학자 직무의 부족률이 32.2%로 가장 높게 나타났다. 평균 이상의 부족률이 나타난 직무는 데이터개발자 13.4%, 데이터분석가 12.2%로 조사됐다.

[그림 3-1-17] 향후 5년 내 전 산업 내 데이터직무 인력 부족률



향후 5년 내 일반산업을 포함한 전 산업에서 추가로 필요한 기술등급별 인력은 중급이 5906명(46.5%)으로 가장 높게 나타났고 고급 4237명(33.4%), 초급 2561명(20.2%) 순으로 나타났다.

[그림 3-1-18] 향후 5년 내 전 산업의 필요 인력 기술등급별 비중



5. 마무리

한국은행 기업경영분석 통계에 따르면 국내 전 산업의 매출 증감률은 2015년부터 2019년까지 연평균 3.7%씩 성장하는 것(CAGR)으로 나타났다. 같은 기간에 정보통신업은 5.6%, 정보통신기술산업은 4.1%, 서비스업은 연평균 5.3% 수치를 보였다. 데이터산업은 2015년부터 2019년 잠정치까지로 살펴보면 6.0%로 비교적 높은 증가세를 보였다.

향후 데이터산업은 디지털 경제를 이끄는 핵심 동력으로 강조되고 AI 기술 및 인프라 환경과 결합돼 지속적으로 성장할 것으로 전망된다. 전 산업에서 데이터와 융합된 디지털 전환과 이를 위한 데이터산업 생태계와 거버넌스 수립이 강조됨에 따라 데이터직무 인력의 수요도 올라가고 있다. 반면 앞 조사 결과와 같이 중고급 이상의 기술을 보유한 데이터 직무인력 수급은 여전히 부족한 실정으로 이에 대한 지원이 필요하다. 2020년 디지털 뉴딜 정책을 기점으로 더 체계적이고 실무에 맞는 데이터 전문인력 양성과 투자가 지속되기를 기대한다.

【표 3-1-9】 최근 5개년(2015~2019년) 국내 주요 산업별 시장규모 추이

(단위: 억 원)

구 분	2015년	2016년	2017년	2018년	2019년	CAGR '15~'19
전 산업	3,578,571,726	3,660,045,282	3,991,487,136	4,151,608,790	4,141,161,549	3.7%
제조업	1,678,047,032	1,670,375,610	1,821,878,190	1,900,756,205	1,854,703,239	2.5%
서비스업	1,379,873,826	1,446,948,438	158,807,3621	1,652,628,121	1,699,584,879	5.3%
정보통신업	147,335,151	154,555,444	165,118,449	170,210,276	183,077,830	5.6%
정보통신기술산업	530,769,320	533,044,960	606,000,669	647,134,154	624,327,928	4.1%
데이터산업*	133,555	137,547	143,530	155,684	168,693	6.0%

출처: 한국은행 경제통계시스템, 기업경영분석(2015-2019) 및 2019년 데이터산업현황조사 자료를 정리

* 데이터산업은 2019년 잠정치를 반영하고, 그 외 산업은 모두 확정치를 반영

제2장.
해외 데이터산업
시장 현황

필자: 나영민

(날리지리서치그룹 이사)

AI와 디지털 기반 혁신이 전 세계적으로 확대됨에 따라 데이터의 중요성은 더욱 강조되고 있다. 이에 따라 국내뿐 아니라 해외에서도 데이터 관리, 데이터 분석, 머신러닝 등 데이터 및 데이터 관련 산업도 지속적인 성장세를 이어가고 있다. 데이터를 중심으로 하는 신생기업의 수가 증가하고 있으며, 파생 시장 또한 더욱 성장하고 있다. 여러 글로벌 시장 자료 기관의 데이터 산업 관련 통계를 소개한다.

1. 데이터산업의 시장규모

데이터 시장은 각 나라와 기관마다 그 범위와 정의가 다양해 시장의 정의와 분류를 어떻게 하느냐에 따라 시장 규모가 다르게 나타난다. 이 글에서 해외 시장은 글로벌 시장 자료를 제공하는 기관의 시장 정의에 따른 시장 통계를 소개한다.

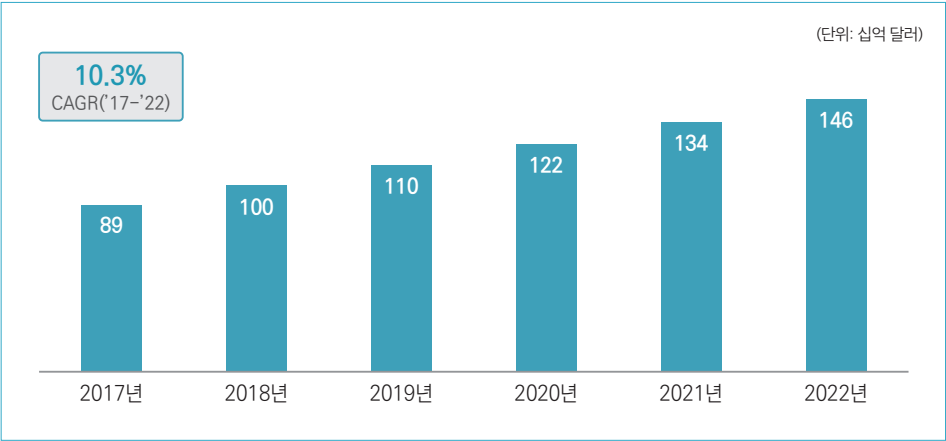
가. 데이터 기반 솔루션 시장

미국 ‘451Research’는 데이터 시장을 8개의 세부 솔루션 시장⁰¹으로 분류하고 2020년 1,223억 달러 규모에 이를 것으로 전망했다. 전체 데이터 시장은 2017년 894억 달러 규모를 형성했으며, 2017년부터 2022년까지 연평균성장률(CAGR) 10.3%를 기록할

01. 8개 세부 시장 : 운영 데이터베이스(Operational databases), 분석 데이터 플랫폼(Analytic data platforms), 리포팅 및 분석(Reporting and analytics), 데이터 관리(Data management), 기업 성과 관리(Corporate performance management), 이벤트/스트림 프로세싱(Event/stream processing), 분산데이터 그리드/캐시(Distributed data grid/cache), 검색기반 데이터 플랫폼 및 분석(Search-based data platforms and analytics)

것으로 예측했다. 데이터 시장은 2022년에 1,461억 달러 규모로 크게 성장할 것으로 전망된다. 8개 세부 시장 중 가장 규모가 큰 시장은 운영 데이터베이스 시장이지만 성장세가 가장 높은 부문은 분산 데이터 처리 프레임워크 분야로, 2017년부터 2022년 사이의 연평균성장률이 29.0%에 달하는 것으로 나타났다. 그 외에도 분산데이터 그리드/캐시 부문의 연평균성장률은 26.5%, 리포팅 및 분석의 하위 분류인 고급분석 및 머신러닝 부문이 24.0%, 검색기반 데이터 플랫폼 및 분석 부문 23.2%, 이벤트/스트림 프로세싱 부문이 19.7%로 높은 연평균성장률을 보였다.

[그림 3-2-1] 2017~2022년 데이터 기반 솔루션 전체 시장규모



출처: 451Research, 2018. 바탕으로 재구성

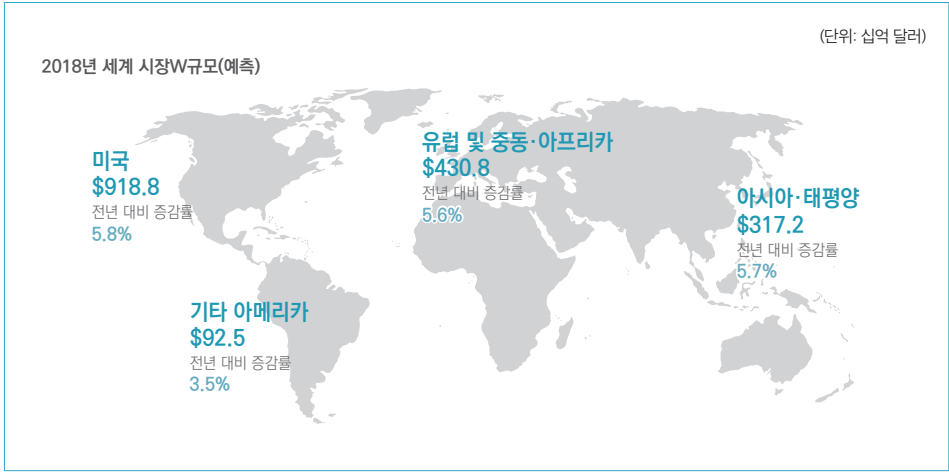
나. 정보서비스 시장

정보서비스 시장은 데이터를 활용하여 온라인/오프라인 정보서비스를 제공하는 영역⁰²으로 데이터 솔루션 시장과 구분된다. 2018년 기준 전체 시장 규모는 약 1조 8,000억 달러로

02. 아웃셀(OUTSELL)에서는 데이터 기반 정보서비스를 'Information Industry'로 정의하고 총 27개 영역으로 구분함. 27개 정보서비스 영역 : Marketing Services, Adtech, News, B2B Media & Business Information, Marketing Research, Yellow Pages, CRM Solutions, Martech, Supply Chain Automation & Procurement Solutions, Company, Contact & Personal Information, IT Research, Financial Information & Solutions, Tax & Accounting Solutions, Governance, Risk & Compliance Solutions, Legal & Regulatory Solutions, Credit Information & Solutions, Health Information & Health IT, Scientific & Technical Information & Solutions, Medical Information, Geophysical & Geomapping, Pharma Information & Solutions, HR Services & Solutions, Teaching & Learning, Corporate Training, Education Management Solutoins, Consumer Entertainment, Content Technology & Distribution Services

집계됐으며, 대륙별 시장 분포로 보면 미국이 전체 정보서비스 시장규모의 절반을 넘는 약 52.2%를 차지하고 있는 것으로 나타났다. 미국의 2018년 정보서비스 시장 규모는 전년 대비 5.8% 성장한 9,188억 달러 수준이었다. 미국의 뒤를 이어 유럽 및 중동·아프리카는 5.6% 성장한 4,308억 달러로 세계 시장의 약 24.5%를 차지하고 있으며, 아시아·태평양은 5.7% 성장한 3,172억 달러로 세계 시장의 약 18.0%를 차지하고 있다. 미국과 유럽 및 중동·아프리카, 아시아·태평양은 전체 시장의 약 94.7%를 차지하고 있는 셈이다.

[그림 3-2-2] 전 세계 대륙별 데이터 기반 정보서비스 시장 비교

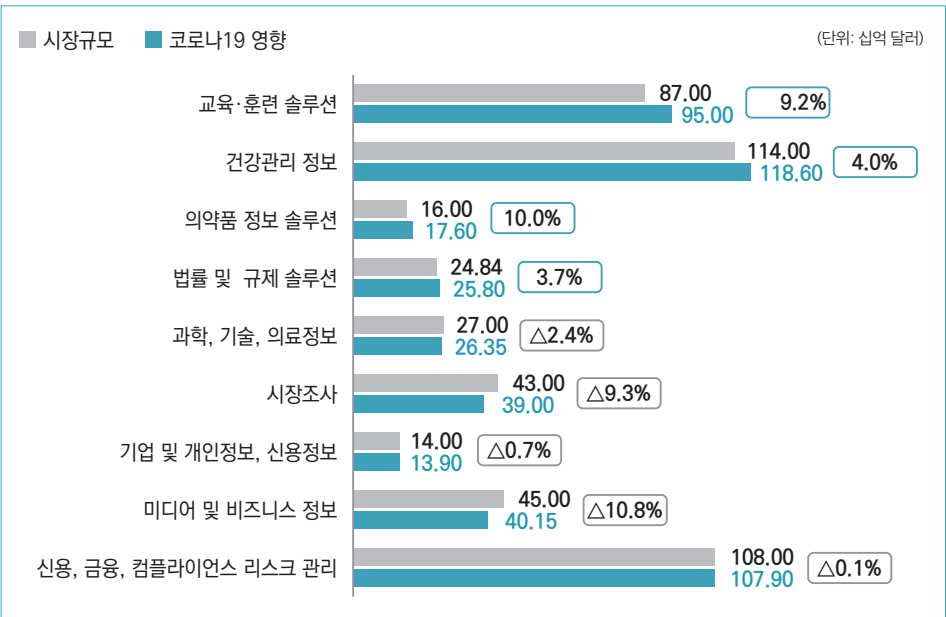


출처: The outsell information industry outlook 2020, Outsell

2020년 초 창궐한 코로나19의 장기화가 예상됨에 따라 데이터 기반의 정보서비스 시장에도 적지 않은 영향을 미칠 것으로 전망된다. 코로나19 전염병에 따라 시장이 확대되는 영역은 교육 및 트레이닝 솔루션 영역과 헬스케어 정보영역, 제약 정보 솔루션 영역, 법률 및 규제 솔루션 영역이다. 기존에 예상된 2020년 시장 규모보다 각각 9.2%, 4.0%, 10.0%, 3.7% 확대될 것으로 예상됐다.

반면 코로나19에 따라 악영향을 받을 수밖에 없는 영역은 과학, 기술과 의료 정보영역, 시장 조사 영역, B2B 미디어 및 기업정보 영역이다. 기존의 예상 시장 규모 대비 각각 2.4%, 9.3%, 10.8% 감소할 것으로 예상됐다. 기업 및 개인정보, 신용정보(Company, Contact and Personal Information) 영역과 신용, 금융, 컴플라이언스 리스크 관리 영역에서는 코로나19의 영향이 크지 않아 시장 규모별 변화는 매우 미미할 것으로 나타났다.

[그림 3-2-3] 코로나19에 따른 2020년 데이터 기반 정보서비스 시장규모 변화⁰³



출처: Outsell, https://outsell-sandbox.zapnito.com/posts/covid-19-s-impact-on-the-data-information-and-analytics-industry?room_id=1013-covid-19-council

다. 디지털 데이터 시장

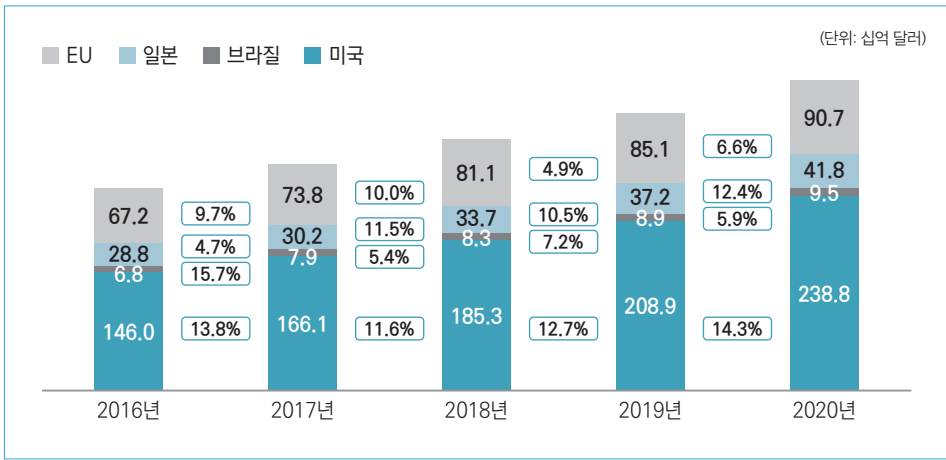
유럽 집행위원회에서 발간한 The European Data Market Monitoring Tool Report(2020)에서는 ‘데이터를 가공함으로써 제품 및 서비스로 재생산되는 것’으로 디지털 데이터 시장범위를 정의했다. 여기에는 이미지 데이터처럼 디지털 기술을 통해 수집, 저장, 가공, 전송되는 멀티미디어 데이터도 포함됐다.

디지털 데이터 시장규모는 데이터 관련 기업의 매출액을 기반으로 산출됐으며, 데이터 기업의 매출은 데이터 관련 제품의 총 가치에 해당한다. 수출을 포함해 해당 국가에 기반을 둔 기업이 생산한 서비스 매출을 기반으로 하고 있다.

미국의 디지털 데이터 시장은 세계에서 가장 큰 규모이며, 시장 성장세 또한 2020년 14.3%로 높게 나타났다. 2018년 1,853억 달러, 2019년에 2,089억 달러 수준을 기록했으며, 2020년에는 약 2,388억 달러 규모로 나타나 지속적인 두 자릿수 성장세를 보이고 있다.

03. 시장규모는 2020년 초 예측한 규모이며, 코로나19 영향은 코로나19 이슈로 바뀐 시장 규모를 의미

[그림 3-2-4] 2016~2020년 디지털 데이터 시장규모



출처: The European Data Market Monitoring Tool Report 21st June 2020 (이하 표·그림 동일)

미국의 이러한 빠른 성장 속도는 유럽연합⁰⁴보다 두 배 이상의 시장규모를 형성하는데 영향을 주고 있으며, 세계에서 가장 빠른 성장 속도다. 유럽연합의 데이터 시장 성장세 또한 미국과 견줄 정도로 빠른 편이었으나, 최근에는 5% 내외로 성장세가 완화되고 있다. 2018년에 811억 달러 수준을 기록했으며, 2019년에는 4.9% 성장한 851억 달러, 2020년에는 907억 달러까지 성장했다.

일본의 데이터 시장규모는 2018년에 약 337억 달러를 기록했으며, 2019년에는 372억 달러로 성장했다. 2020년에는 12.4% 성장한 418억 달러 규모로 추정돼 높은 성장세를 이어가고 있는 것으로 나타났다.

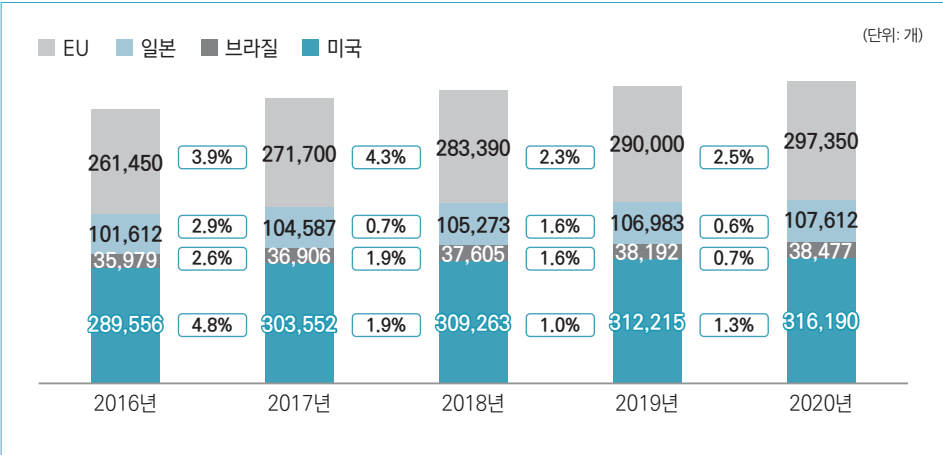
한편 브라질의 데이터 시장규모는 2018년에 약 83억 달러 수준을 기록했으며, 2019년에 7.2% 성장한 89억 달러, 2020년에는 95억 달러까지 성장했다.

04. 유럽연합(EU)은 영국의 탈퇴 이전인 다음의 28개국을 의미함. 오스트리아, 벨기에, 불가리아, 크로아티아, 키프로스, 체코, 덴마크, 에스토니아, 핀란드, 프랑스, 독일, 그리스, 헝가리, 아일랜드, 이탈리아, 라트비아, 리투아니아, 룩셈부르크, 몰타, 네덜란드, 폴란드, 포르투갈, 루마니아, 슬로바키아, 슬로베니아, 스페인, 스웨덴, 영국

2. 데이터 기업 수

데이터 기업이란 시장에 데이터를 공급하는 기업으로서 데이터의 생산과 전달을 담당하는 곳으로 정의된다. 즉 데이터 관련 제품, 서비스 및 기술에 대한 데이터를 공급하는 기업을 의미한다. 두 자릿수 내외를 기록한 시장 규모의 성장세만큼은 아니지만, 데이터 시장이 성장함에 따라 데이터 관련 기업 수도 증가추세를 보인다.

[그림 3-2-5] 2016~2020년 디지털 데이터 기업 수



미국, EU, 일본, 브라질의 디지털 데이터 기업 수는 데이터 종사자 수와 마찬가지로 매년 꾸준히 증가하고 있다. 미국의 디지털 데이터 기업 수는 2019년 31만 2215개에서 2020년에는 31만 6190개로 1.3% 증가했다. EU의 디지털 데이터 기업 수는 2019년 약 29만 개였으나 이후 꾸준히 증가해 2020년에는 약 29만 7350개에 이르렀다.

미국과 EU 데이터 시장의 규모 차이는 크지만, 데이터 기업의 수는 큰 차이를 보이지 않았다. 이것은 EU의 디지털 데이터 기업이 상대적으로 다양한 데다 중소기업 중심으로 형성돼 있어 매우 건전한 상태라고 할 수 있다. 미국의 경우 몇몇 소수의 글로벌 대기업이 전체 데이터 시장을 주도하는 양상이며, 대체로 대규모 데이터 기업들로 시장이 형성돼 있다.

일본의 데이터 기업 수는 2019년 약 10만 6983개였으나 2020년에는 0.6% 증가한 10만 7612개인 것으로 집계됐다. 장기적인 경제 침체를 겪는 중에도 데이터 기업 수는 꾸준히 증가했다. 이는 일본 기업들의 데이터에 대한 관심과 시장성을 보여주는 수치다.

한편 브라질의 데이터 기업 수는 2019년 약 3만 8192개였는데, 2020년에는 3만 8477개로 늘었다. 브라질 경제에서 ICT 시장이 상대적으로 작고, ICT의 보급률 또한 매우

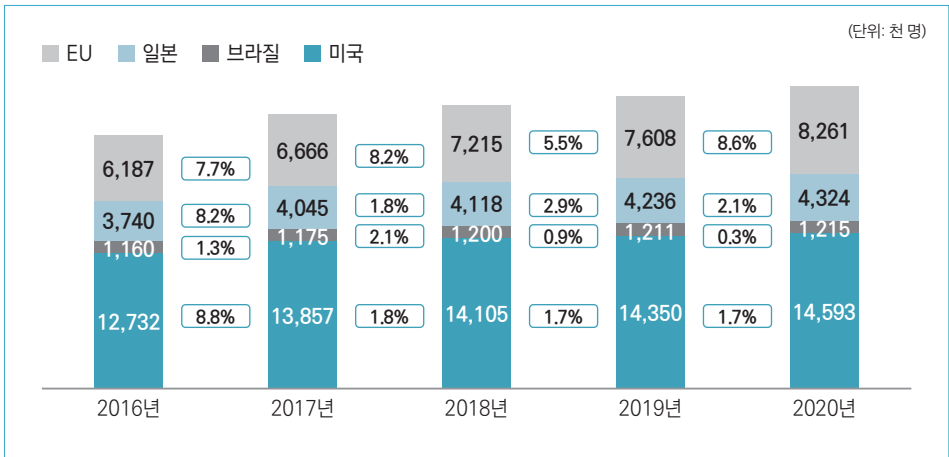
낮기 때문에 기업 수는 적지만 향후 지속적 성장이 기대되고 있다.

3. 데이터 전문인력

데이터 전문인력은 대용량 데이터를 수집, 저장, 관리, 분석, 해석, 시각화할 수 있는 인력으로 정형/비정형 데이터의 사용에 능숙해야 한다. 막대한 양의 데이터를 다루고 새로운 데이터 분야 기술에 능통한 자로 정의된다. EU의 데이터 전문인력은 2019년에 761만 명 수준이었으나 2020년에는 826만 명으로 가장 높은 증가율을 보였다. 미국 또한 144만 명이었던 2019년에 비해 2020년에는 145만 명 수준으로 지속 증가했다.

데이터 전문 인력이 가장 폭발적으로 성장했던 시기는 2016년과 2017년 사이였다. 이 기간에 미국은 8.8%, 유럽연합은 7.7%, 일본은 8.2%가 증가했다.

[그림 3-2-6] 2016~2020년 데이터 전문인력 수



EU의 데이터 전문인력의 활동 분야별 구성 비중은 전문 서비스(Professional Services)와 도소매업(Retail and Wholesale) 분야에서의 비중이 높게 나타나고 있다. 인력 확대의 속도를 의미하는 성장률 측면에서는 정보 및 커뮤니케이션(Information & Communication) 분야, 전문 서비스(Professional services)와 건강(Health) 분야가 비교적 높게 나타났다.

[표 3-2-1] 2016~2020년 유럽연합의 데이터 전문인력 수 (단위: 천 명)

구 분	2016	2017	2018	2019	2020	'18~'19 증감률	'19~'20 증감률
건설	126	138	152	156	168	2.1%	8.0%
교육	459	489	530	566	616	6.8%	8.9%
금융	615	651	704	728	781	3.4%	7.3%
건강	488	529	573	610	666	6.5%	9.1%
정보 및 커뮤니케이션	680	750	835	880	967	5.4%	9.9%
채광, 제조	725	779	836	883	956	5.6%	8.3%
전문 서비스	1,281	1,392	1,511	1,621	1,777	7.2%	9.6%
공공 행정	388	417	447	469	507	4.9%	8.1%
도소매업	1,141	1,215	1,299	1,352	1,452	4.1%	7.4%
운송	192	207	221	233	252	5.6%	8.2%
전력·가스·수도	93	99	106	110	118	4.4%	7.2%
총계	6,187	6,666	7,215	7,608	8,261	5.5%	8.6%

4. 데이터산업의 경제적 효과

가. 직접 효과

데이터산업의 경제적 효과는 데이터 시장이 경제 생태계 전체에 미치는 전반적인 영향이 어느 정도인가를 평가하는 분석 방법론이다. 디지털 기술이 발달함에 따라 데이터의 수집, 저장, 처리, 배포, 분석, 정교화, 전달 및 활용이 쉬워졌다. 이렇게 데이터의 활용이 쉬워짐에 따라 데이터 산업과 시장이 형성됐고, 이 시장은 직접 또는 간접적으로 경제 전체에 영향을 미쳤다. 데이터의 경제적 효과는 직접 효과와 간접 효과로 나뉜다. 직접 효과란 데이터 산업 그 자체에 의해 형성된 효과를 의미한다. 즉 데이터 생산과 관련한 모든 비즈니스 활동으로 형성된 효과다. 정량적인 직접 효과는 판매된 데이터 제품 및 서비스의 수익을 기준으로 측정된다.

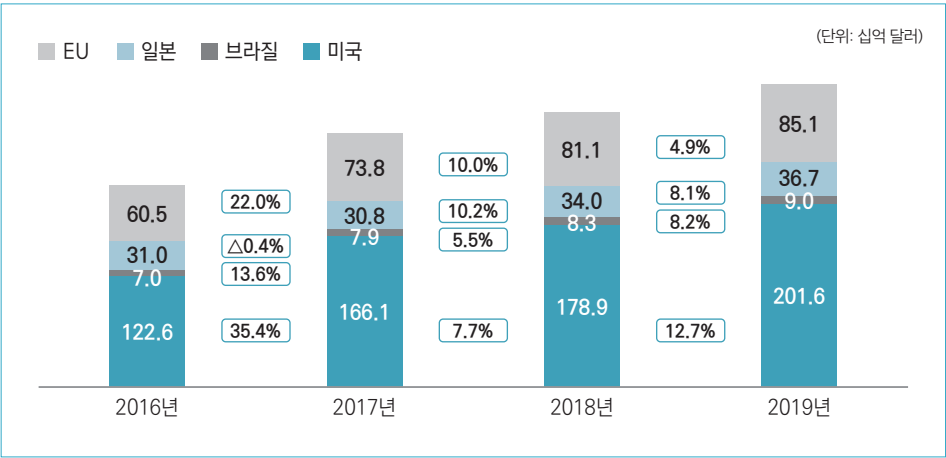
미국의 데이터 경제 효과 중 직접 효과는 2016년에 약 1,226억 달러를 기록한 이후 지속적으로 증가해 2019년에는 2,000억 달러를 돌파했다. 즉 데이터에 기반을 둔 제품 및 서비스의 생산에서 파생된 직접적 영향은 EU를 비롯한 일본과 브라질보다 높게 나타났다. 이는 데이터 시장의 규모와 GDP에서의 비중이 상대적으로 높기 때문이다.

‘직접 효과는 데이터 사업 그 자체에 의해 형성된다’는 정의를 따르면, 미국은 데이터 생산에 매우 적극적임을 알 수 있다. 또한 미국의 데이터 직접 효과는 유럽연합의 직접 효과의 거의 두 배에 해당하는 수치다. 이는 미국의 데이터산업이 좀 더 발전되어 있고, 경제적인 효과의 확산 속도가 더 빠름을 의미한다.

EU의 데이터 경제적 효과 중 직접 효과는 2016년에 약 605억 달러를 기록한 이후 점차 증가해 2019년에는 약 851억 달러로 집계됐다. 특히 2016년에서 2017년의 성장세가 22.0%로 매우 두드러졌다. 이는 시장규모의 증가보다 더 높은 성장세를 보임으로써 EU의 데이터 활용이 다른 국가들보다 더욱 긍정적인 영향을 주고 있는 것으로 판단된다. EU의 데이터산업 규모가 지속해서 증가한 것처럼 그 직접 효과 또한 점진적으로 영향력을 높여 나갈 것으로 보인다.

일본의 데이터 경제적 효과 중 직접 효과는 2016년에서 2017년 사이 소폭 하락세를 보였다. 하지만 2018년에는 반등에 성공했으며, 2019년에는 8.1% 증가한 367억 달러를 기록했다. 브라질의 데이터 경제적 효과 중 직접 효과는 2016년에 약 70억 달러를 기록한 이후 꾸준히 증가해 2018년 90억 달러를 기록했다. 빠르게 발전하는 브라질의 ICT 산업 덕분에 데이터산업 또한 빠르게 성장하는 것으로 분석된다. 브라질은 경제 위기를 심하게 겪었음에도 데이터에 기반을 둔 제품 및 서비스 확산세가 두드러져 데이터산업의 성장 가능성을 보여준다.

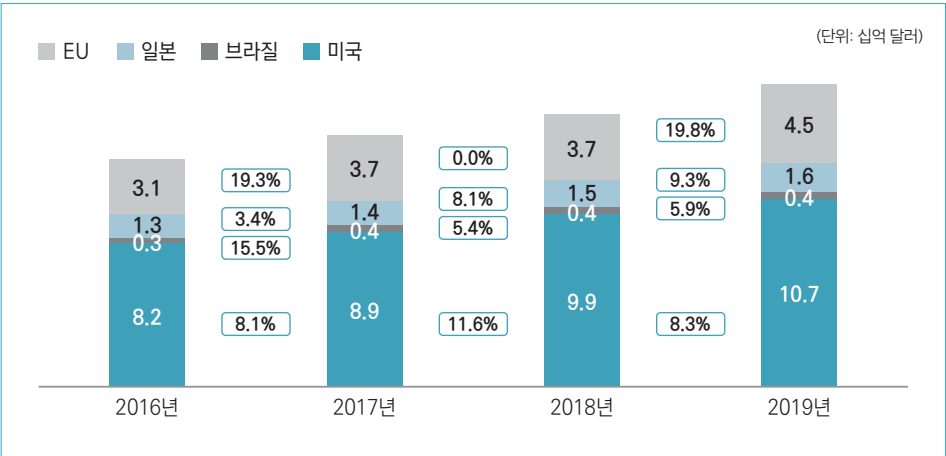
[그림 3-2-7] 2016~2019년 경제적 효과 중 직접 효과



나. 간접 후방 효과

데이터의 경제적 효과 중 직접 효과가 데이터 생산 활동이 데이터 산업 내에 미치는 영향을 설명하는 것이라면, 간접 효과는 타 산업에 미치는 영향을 의미한다. 간접 효과는 전방 효과와 후방 효과로 구분된다. 후방 효과란 데이터 산업에 속한 기업이 데이터 기반의 제품 및 서비스를 생산활동 과정에서, 타 산업으로부터 자원을 받을 때 자원을 제공해준 산업에서 발생하는 효과를 말한다.

[그림 3-2-8] 2014~2018년 경제적 효과: 간접 후방 효과



반면 전방 효과는 데이터산업에서 생산된 데이터 기반의 제품 및 서비스를 이용해 타 산업에서 다른 형태의 제품 및 서비스로 가공되는 효과를 말한다. 즉 타 산업에서 데이터를 활용해 제품 및 서비스를 제공함으로써 얻는 수익을 일컫는다.

전방 효과는 세 가지로 구분할 수 있는데, 첫째는 생산 및 공급 프로세스 최적화를 들 수 있다. 이는 데이터에 기반을 두고 프로세스를 통해 생산을 효율화하는 것이다. 둘째는 데이터를 활용해 표적 광고 및 개인 맞춤형 광고를 제공함으로써 마케팅 기술을 향상하는 것이다. 즉 타 산업에서 데이터 분석 결과를 토대로 마케팅함으로써 수익 상승 효과가 발생한다는 점이다. 마지막으로 데이터 기반의 조직을 구성할 수 있는 점이다. 즉 데이터를 활용해 기존 조직의 관리 방식을 개선함으로써 불필요한 자원 소모를 최소화해 조직의 생산성을 높이는 것도 전방 효과라 할 수 있다.

데이터산업으로 인한 후방 간접 효과는 매우 빠르게 증가하고 있다. 특히 미국의 경우 2016년 약 82억 달러 규모였으며, 2019년에는 높은 증가세를 보이면서 107억 달러를 기록했다. 2016년 이후 매년 8%가 넘는 증가율을 보이고 있다.

EU의 후방 효과는 2016년에 31억 달러 규모였으나 2017년 약 20%의 성장률을 보였다. 반면 2018년에는 규모의 변동이 없는 것으로 나타났다. 하지만 2019년에는 19.8% 성장해 45억 달러를 기록했다. EU는 공공데이터 개방 정책을 통해 민간데이터보다 공공데이터 중심의 데이터산업이 상당한 비중을 차지하고 있다. 산업 내에 1인 기업 및 소규모의 기업들이 많아 건전한 시장 구조의 성과가 본격적으로 나타나기 시작한 것으로 분석된다.

일본의 후방 효과도 빠르게 증가하는 추세다. 2016년에 약 13억 달러 규모였던 후방 효과는 증가세를 지속하면서 2019년에 약 16억 달러 수준에 도달했다.

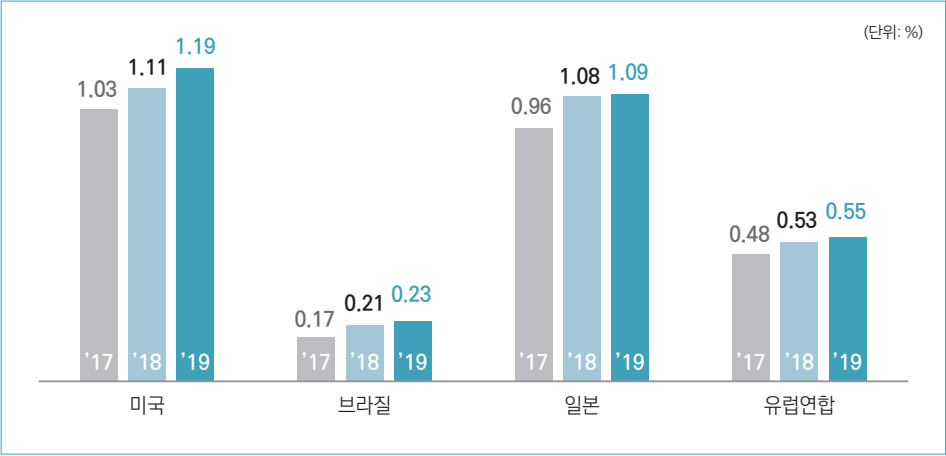
5. GDP 대비 데이터 경제적 효과의 비율

앞서 데이터산업의 경제적 효과를 직접 효과와 간접 후방 효과를 중심으로 살펴보았다. 직접 효과와 간접 후방 효과 모두 미국이 가장 앞선 것으로 나타났다. 두 효과를 합했을 때 2019년 기준 약 2,124억 달러 수준이었다. 시장 규모와 마찬가지로 경제적 효과 또한 미국은 EU의 두 배 가까운 규모였으나 성장률은 EU가 더 높은 것으로 나타났다.

GDP에서 경제적 효과가 차지하고 있는 비율을 살펴보면 미국은 2017년 1.03%에서 2019년 1.19%로 나타나 약 0.16%p가 증가했다. EU는 2017년 0.48%에서 2019년

0.55%로 상승해 0.07%p의 성장세만 기록했다. 일본의 경우 GDP 대비 비중이 2017년 기준 0.96%였으나 2019년 1.09%로 약 0.13%p 증가해 미국에 견줄 정도의 비율을 보였다. 브라질의 경우 0.23% 수준에 그쳐 GDP 대비 데이터 활용 및 경제적 효과는 다소 처지는 것으로 나타났다.

[그림 3-2-9] 2017~2019년 GDP 대비 경제적 효과의 비율(직접 효과 + 간접 후방 효과)

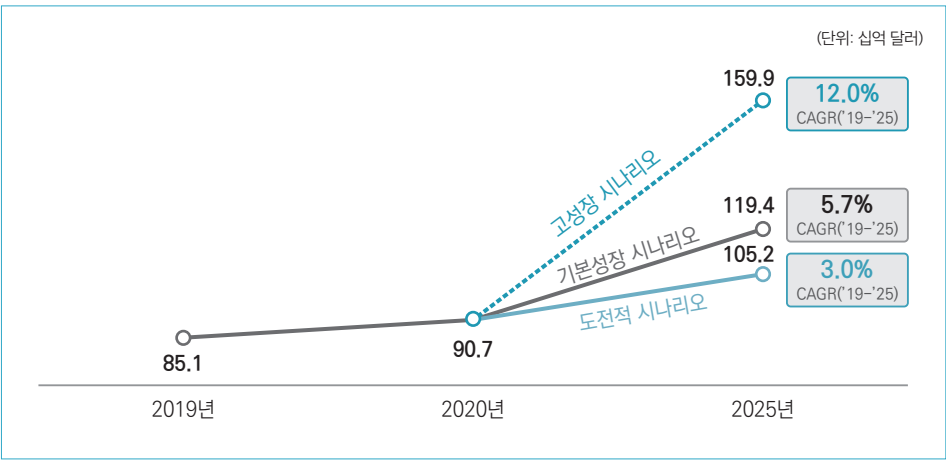


6. 중장기 데이터 시장 전망

향후 데이터 시장은 지속적인 성장세를 기록할 것으로 전망된다. 유럽 집행위원회의 2020 European Data Market Monitoring Tool 보고서에 따르면 기본 시나리오 기준 2020년에서 2025년까지 연평균성장률은 5.7%로 예측된다. 2025년 EU의 데이터 시장 규모는 1,194억 달러까지 성장할 것으로 전망되며, 만약 고성장 시나리오가 전개된다면 1,599억 달러까지 내다볼 수도 있다는 전망이다.

데이터 기업의 매출은 같은 기간 전체 ICT 시장의 성장률을 훌쩍 뛰어넘을 것이라는 전망이 지배적이다. 이러한 데이터 기반 시장은 경제 규모가 큰 국가일수록 데이터 경제 확대에 크게 이바지할 것으로 예측됐다.

[그림 3-2-10] 2019~2025년 EU의 중장기 데이터 시장 전망⁰⁵⁾



데이터 관련 인력 또한 데이터 관련 기업의 향후 성장세를 견인하는 중요한 자원으로 평가받고 있어 지속적인 성장세를 보일 것으로 예측될 뿐만 아니라, 높은 고용률과 고액 연봉 체제를 구축해 나갈 것으로 전망된다.

05. 3가지 시나리오를 산출하기 위해 활용된 데이터는 다음과 같다. (1)IDC의 최근 전 세계 시장 예측 추정 및 BlackBook(표준판), (2)IDC의 빅데이터 분석, Digital Transformation, Innovation Accelerators 및 2019년 시장 동향, (3)EU GDP 및 ICT 지출의 2025 예측 업데이트, (4)2013~2018년 EDM 모니터링 Tool에서 나타난 추세, EDM 지표를 모델링하고 예측에 사용된 시나리오 참조

제 4 부

산업별 데이터 활용 비즈니스 현황

제1장.
금융 분야 데이터 활용 비즈니스 현황

제2장.
헬스케어 분야 데이터 활용 비즈니스 현황

제3장.
제조 분야 데이터 활용 비즈니스 현황

제4장.
통신 분야 데이터 활용 비즈니스 현황

제5장.
광고 분야 데이터 활용 비즈니스 현황

제6장.
교육 분야 데이터 활용 비즈니스 현황

제7장.
유통·물류 분야 데이터 활용 비즈니스 현황

제1장. 금융 분야 데이터 활용 비즈니스 현황

필자: 장혜리
(에이젠글로벌 이사)

최근 금융서비스는 데이터 기반의 의사결정과 고객 중심의 맞춤형 디지털 전환을 가속화하고 있다. 특히 금융 비즈니스 영역은 핀테크, 빅테크 기업이 등장하고 이중 산업간의 융합이 확대되면서 간편결제, 오픈뱅킹, API, 인공지능, 클라우드 등 신기술과의 결합이 활발하게 이루어지고 있다. 금융 데이터 오픈, 혁신금융서비스 같은 정부 정책을 근간으로 해 오픈된 금융 데이터 기반의 새로운 금융서비스가 등장했다. 또한 데이터 기반의 AI 금융은 AI 뱅킹서비스, 카드 뱃터닝 이상금융거래탐지시스템(FDS), 보험영역을 망라해 인공지능 플랫폼 도입 및 서비스형 AI 모듈을 활용해 디지털 전환이 활발하게 진행되고 있다.

1. 금융 데이터 시장과 정책 변화

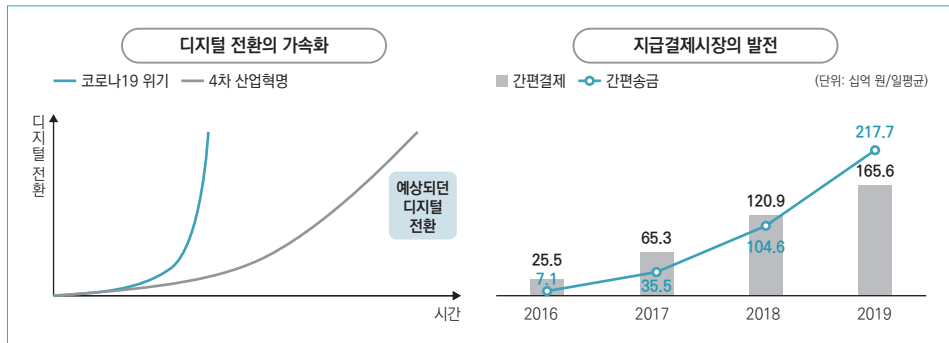
가. 디지털 금융의 발전

4차산업혁명의 흐름과 더불어 코로나19의 예기치 못한 확산은 금융 분야의 디지털 전환에 속도를 내는 데 큰역할을 했다. 디지털 금융은 데이터와 기술, 실물 경제의 융합을 가속화하는 핵심 산업으로 부각되고 있다. 최근 모바일 환경과 ICT를 기반으로 전자결제, 송금, 중개, 대출 등의 서비스를 통합한 새로운 금융 플랫폼이 등장하고 있다.

디지털 금융 혁신은 기술을 기반으로 한 핀테크 기업 및 빅테크 기업이 금융 산업에 적극 진출하면서 기존 금융기관 중심의 산업 생태계를 변화시키고 있다. 국내 핀테크 기업은 2019년 기준으로 지급결제, 자산관리, 보안 등 다양한 분야에서 300개가 넘게 존재한다. 그 중 유니콘 기업은 토스(Toss) 1곳으로 나타났고, 대부분은 초기 단계의 스타트업이다.⁰¹

01. 조영은, 「금융 산업 구조 측면에서의 디지털 금융 혁신 동향과 향후과제」, 국회입법조사처, 2020.5.30.

[그림 4-1-1] 디지털 금융시장의 성장속도



출처: 「4차 산업혁명 시대의 디지털금융 종합혁신방안」(보도자료), 금융위원회, 2020.7.

반면 네이버, 카카오 같은 빅테크 기업은 기존의 브랜드 인지도와 고객층, 기술력을 바탕으로 간편결제, 인터넷전문은행 등을 시작으로 빠르게 고객층을 확보하며 성장성과 확장성을 보여주고 있다.

나. 활성화를 위한 정부 정책

글로벌 디지털 금융 흐름에 따라 국내에서도 경쟁과 혁신을 촉진하기 위해 법·제도를 정비하고 있다. 올해는 국내 데이터3법 시행과 마이데이터 사업 추진으로 데이터 개방과 신규 산업을 위한 육성 환경이 마련됐고, 디지털 금융거래와 데이터 개방이 강화되고 있다.

1) 데이터3법 시행 및 마이데이터 사업 본격화

2020년 8월 5일 데이터3법⁰²이 본격 시행됐다. 또한 부처별로 나뉘어 있던 개인정보보호에 관한 법률을 통합해 불필요한 중복 규제를 없앴으며, 개인은 정보 주체로서 이동권을 행사하고, 기업은 필요한 데이터를 확보해 새로운 서비스를 개발할 수 있는 마이데이터 사업⁰³ 환경이 갖춰졌다.

이로 인해 각 금융기관 내부에 있던 정보를 고객 동의하에 타 기관으로부터 수집·통합해 하나의 플랫폼에서 제공하는 것이 가능해졌고, 금융권 및 핀테크 기업의 플랫폼 경쟁이

02. 데이터3법이란 '개인정보보호법', '신용정보법', '정보통신망법'의 3개 법률을 총칭하는 것으로 4차산업 시대의 도래에 따른 세계 각국의 데이터 활성화 추진 전략에 대응하고 국내 데이터의 효율적이고 안전한 활용을 도모하고자 개정됨

03. 마이데이터 산업(본인 신용정보 관리업)이란 신용정보주체의 권리행사에 따라 개인신용정보 등을 수집하고, 수집된 정보를 신용정보주체가 조회·열람 등 제공하는 행위를 영업으로 하는 산업을 말함

심화되고 있다. 해당 사업을 위해서는 금융위원회의 본인신용정보 관리업 허가⁰⁴가 필요하다. 운영 중인 서비스에 대해서는 2021년 2월 전까지 허가를 마쳐야 하며 금융 당국은 사전 신청서를 제출한 63개 기업 중에서 2021년 상반기까지 40개 기업을 우선적으로 심사할 계획이다.⁰⁵

2) 금융데이터거래소 출범

2020년 5월 금융보안원이 주도해 출범한 금융데이터거래소는 금융 데이터 활성화를 위해 공급자와 수요자를 연결해주는 중개시스템을 구축했다. 해당 거래소에 참여 중인 84개 기업은 금융 데이터 및 API를 등록하고, 수요자 요청에 따라 맞춤 데이터를 제공한다. 데이터 거래 체결 이후엔 거래소 내 분석 환경을 제공해 공급자의 데이터 권리를 보호하고, 수요자에게는 합리적인 가격으로 데이터를 제공한다. 또한 안전한 데이터 유통환경을 위해 이중 보안체제로 운영되고 있다.

해당 거래소의 출범 이후 참여 기관 및 등록 데이터가 빠르게 증가하며 높은 관심을 받고 있다. 최근에는 유통업의 씨유⁰⁶와 네이버⁰⁷가 참여하면서 금융 데이터뿐만 아니라

비금융 데이터 중개가 이루어짐으로써 금융 데이터와 결합한 고부가가치 데이터 생산이 기대된다.

3) 금융 빅데이터 개방시스템 활성화

2019년 6월부터 한국신용정보원은 금융 빅데이터 개방시스템을 개시해 일반신용정보, 기술신용정보, 보험신용정보를 3차에 걸쳐 차례로 공개했다. 해당 플랫폼을 활용해 이용자에게 금융신용정보 DB와 원격분석 환경을 제공함으로써 금융업권 간의 데이터 격차를 해소하고자 했다. 분석용 DB는 신용정보원에 등록된 전체 정보 중 표본 추출 후 비식별 조치를 거쳐 표본 DB 형태로 제공되는데, 세 번에 걸쳐 금융 빅데이터를 활용할 핀테크 기업을 선정한 뒤 DB를 제공하며 금융서비스 고도화를 지원하고 있다.

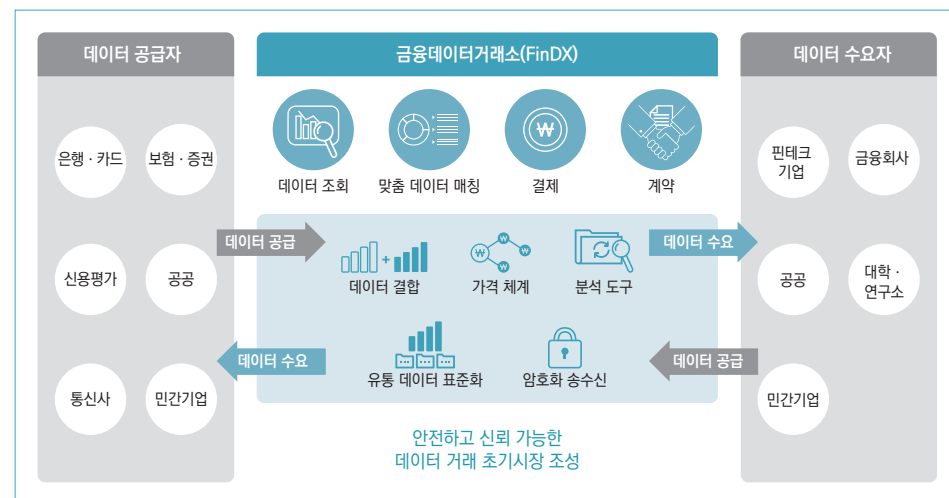
2. 금융 데이터 융합을 위한 기술환경 변화

가. 표준 API와 클라우드 컴퓨팅

마이데이터 사업으로 기업 간 데이터 이동이 활발해지면서 표준 API(Application Programming Interface)를 이용한 데이터 송수신이 주목받고 있다. 기존의 스크린 스크래핑 방식은 기업이 사용자의 인증 정보를 제공받아 화면에 보여지는 정보를 추출하는 형식이었다. 표준 API의 경우 기업이 직접 데이터에 접근해 필요한 정보만을 수집함으로써 과도한 개인정보 수집과 보안 위험에 대한 부담이 줄어든다. 표준 API 방식은 표준화된 형식으로 데이터를 관리하고 수집해 상호 운용성을 증대시키고 프로그램 개발을 용이하게 한다. 다수의 기업과 공공기관이 데이터를 공유하기 위해 표준 API를 개방하고 금융서비스에 표준 API 방식을 이용해 데이터 가공 결과에 대한 응답을 빠르게 적용하면서 AI 금융 모델에 대한 실시간 예측모델 도입이 쉬워졌다.

기업이 소유하는 빅데이터가 점차 방대해지면서 이를 저장·처리·분석하기 위한 비용이 증가했으며, 이를 해결할 대안으로 클라우드로의 전환이 가속화되고 있다. 클라우드 컴퓨팅은 서버 등의 IT인프라, 애플리케이션, 개발도구를 직접 소유하지 않고 인터넷 상의 서버를 이용해 필요한 만큼의 IT 자원에 대해 합리적 비용을 지불하고 이용하는 서비스다.

[그림 4-1-2] 금융데이터거래소 소개



출처: 금융데이터거래소(FinDX)

04. 2020년 8월 5일부터 시행되는 개정 신용정보법에 따라 본인신용정보 관리업이 허가제로 운영됨. 본인신용정보 관리업의 고유 업무를 영위하고자 하는 모든 회사는 금융위원회의 허가를 받아야 함

05. 박소정 기자, 「마이데이터 허가 경쟁 과열에…당국'40곳 한bane 심사」, 조선비즈, 2020.8.19. https://biz.chosun.com/site/data/html_dir/2020/08/19/2020081903964.html

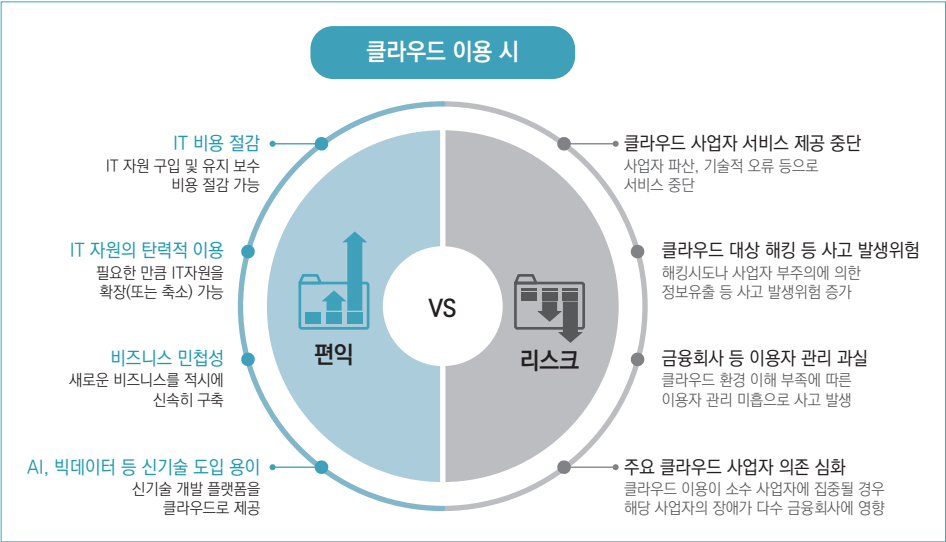
06. 윤미혜, 「금융데이터거래소 '비금융'도 중개한다」, IT조선, 2020.8.13. http://it.chosun.com/site/data/html_dir/2020/08/12/2020081204011.html

07. 이효석 기자, 「네이버, 온라인 쇼핑 트렌드 데이터 금융데이터 거래소에 등록」, 연합뉴스, 2020.9.18. <https://www.yna.co.kr/view/AKR20200918039400017?input=1195m>

이를 이용해 IT 기반 시설 구축과 유지비용 감소 그리고 신기술을 개발하는데 필요한 IT 자원의 유동적 이용이 가능하다. 이에 따라 KB금융그룹은 하이브리드-멀티 클라우드⁰⁸ 전략을 세워 2024년까지 그룹 공동 클라우드 구축하는 계획을 발표했고, 농협은행은 2017년부터 클라우드를 도입해 점진적으로 기존 전산시스템을 클라우드로 전환하고 있다. 하나금융그룹은 2018년 금융권 최초로 그룹 공용 프라이빗 클라우드 시스템을 오픈한 뒤 지속적인 고도화 작업을 진행 중이다.⁰⁹

최근 전자금융감독 규정 개정에 따라 클라우드 도입이 가능한 정보가 비중요 정보에서 개인신용정보, 고유식별정보를 포함한 중요 정보로 확대되었으나 중요 정보를 클라우드로서 다룰 경우 금융감독원에 보고해야 하는 의무가 따른다.¹⁰ 클라우드 이용 시 발생할 수 있는 장애나 재해 시 복구를 위해 클라우드 제공 기업은 서버 분산과 백업 등을 통해 이에 대비해야 한다.

[그림 4-1-3] 클라우드 도입의 편익과 리스크



출처: 2020년 디지털 금융 이슈 전망 보고서, 금융보안원

08. 클라우드는 배포 유형에 따라 프라이빗(온프레미스), 퍼블릭 클라우드와 둘의 장점을 통합한 하이브리드 클라우드로 나뉜다. 하이브리드 클라우드는 이미 구축된 환경에서 클라우드의 장점을 활용할 수 있음. 멀티 클라우드는 복수의 벤더(클라우드 공급 업체)가 제공하는 동일한 유형의 클라우드를 2개 이상 배포하는 것을 말하며 특정 벤더에 종속되지 않는 장점이 있음

09. 박기록·이상일 기자, 「더 거세진 금융권 '클라우드'바람…은행·2금융권, U2L가속화」, 디지털데일리, 2020.7.17. <http://www.ddaily.co.kr/news/article/?no=198585>

10. 전자금융감독규정 제3장 14조의 2

나. 인공지능을 활용한 데이터 융합

최근 쇼핑, 유통, IT처럼 실생활과 밀접한 분야에서 소비 및 행동 트렌드 파악을 기반으로 한 고객별 세분화한 맞춤 서비스가 각광받고 있다. 금융 분야의 AI 서비스 도입은 영업활동 전반에서 변화를 가져오고 금융의 본질적인 신용 및 운용에 대한 비용 구조를 개선한다. 고객을 상대하는 앞단에서는 챗봇을 도입하고 상담 자동화를 실시해 업무를 경감시키고, 맞춤형 상품 추천이 가능해 고객 편의성 및 만족도를 높일 수 있다. AI를 활용한 신용평가, 여신심사, 보험 인수 과정의 고도화는 비즈니스 모델 혁신을 가져오고 있다. 이뿐만 아니라 내부의 단순, 반복적인 업무를 자동화함으로써 업무 효율성을 개선시킨다.

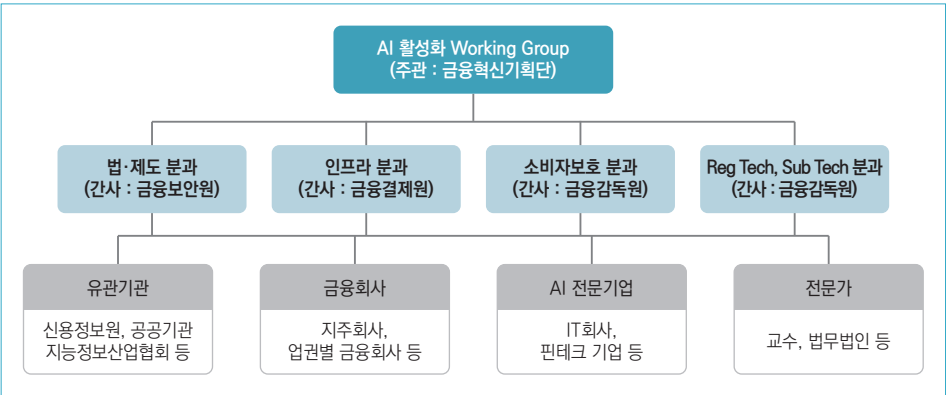
AI 신기술을 앞세운 빅테크 기업들은 기존 서비스와 간편결제 서비스를 접목해 금융시장에 진출하기 시작했다. 대표적으로 아마존 페이, 애플 페이, 구글 페이, 알리 페이 등이 있다. 통신과 금융 데이터를 융합해 소상공인에게 양질의 금융서비스를 제공하는 방식도 나타나고 있다.

[표 4-1-1] AI 기반 금융서비스 분류

구분	활용 사례
프론트 오피스 서비스	챗봇, AI 스피커, 로보어드바이저, 스마트 ATM 등
미들 오피스 서비스	렉테크(Regtech), 자금세탁방지(AML), 이상금융거래탐지시스템(FDS), 약관 자동 비교 분석 등
백 오피스 서비스	AI 기반 신용평가, 로보틱 프로세스 자동화(RPA), 기업위험관리, 여신서비스 자동화 등

2020년에 금융 데이터에 AI를 적용하는 사례가 늘어나고 있다. 특히 금융의 프론트 서비스에 해당하는 챗봇, 로보어드바이저에 적용되던 AI가 미들·백 서비스로 확대되며 실질적인 AI 금융을 도입한 성과가 나타날 전망이다. 최근엔 이상거래탐지, 자금세탁방지, 머신러닝 로보틱 프로세스 자동화 등 업무탐지 부분과 자동화 부분으로 확대됐다. 나아가 금융의 핵심 업무인 신용평가, 기업위험도 예측, 여신서비스 자동화, 디지털 전략까지 확대되며 금융AI 활성화가 진행되고 있다. 이 같은 상황과 더불어 금융위원회는 인공지능 워킹그룹을 운영해 ‘금융분야 인공지능(AI) 활성화’ 방안을 2020년 말까지 마련할 계획이라고 밝혔다.

[그림 4-1-4] 금융분야 AI 활성화 워킹그룹 구성안



출처: 「코로나 이후 시대의 디지털 대전환을 선도하기 위해 금융분야 인공지능을 활성화하겠습니다」(보도자료), 금융위원회, 2020.7.16.

3. 데이터 활용과 금융업 비즈니스 사례

적극적인 데이터 개방과 정책 변화로 금융업계는 초개인화된 금융상품 및 서비스 제공이 가능해지고 있다. 금융과 비금융 데이터의 결합이 활발해지는 추세로 이종 산업 간의 경계가 불분명해지는 ‘빅블러(Big Blur)’ 현상이 가속화되고 있다.

가. 은행업

은행의 오픈뱅킹 서비스는 은행의 핵심 금융기능을 표준화해 다른 사업자에 개방하는 은행권 공동 인프라를 구축해 고객이 하나의 앱에서 금융정보 조회와 은행 업무를 하는 것을 가능하게 한다. 오픈뱅킹 서비스는 2019년 10월 말 시범 도입된 뒤 동해 12월 전면 시행됐으며 각 은행은 차별화된 서비스와 편리한 UI를 내세워 고객을 선점하기 위해 노력하며 대부분의 시중은행에서 우리원(WON)뱅킹(우리은행), 솔(SOL)뱅킹(신한은행), 스타뱅킹(KB국민은행), 하나원큐뱅킹(하나은행) 등의 오픈뱅킹 앱을 선보였다.

나. 카드사

카드사는 고객의 구매 이력이라는 활용성이 높은 정보를 바탕으로 데이터 상품 등록과 이종 산업 간의 데이터 거래가 활발한 업종이다. 카드사는 금융정보 이력이 부족한 씬파일러(Thin Filer)의 공과금, 세금, 통신비 납부 등 다양한 카드 사용 내역을 추가적인 신용평가 데이터로 활용해 대출 상품의 문턱을 낮추고 있다. 신한카드와 KB국민카드는 개인사업자 신용평가

서비스 ‘마이크레딧’과 ‘크레딧트리’를 차례로 출시했다.

카드사는 이상거래탐지(FDS) 업무에 머신러닝을 활용하는 수준까지 끌어올렸다. 이는 방대한 양의 고객의 카드 이용 패턴을 학습하고 부정사용 징후를 탐지하는 것으로, 카드사와 핀테크 기업 간에 활발한 협업이 이뤄지고 있다. 그 예로 우리카드는 에이젠글로벌과 함께 딥러닝 기반 이상금융거래탐지시스템(FDS) 구축 사업을 추진하고 자동화한 AI 솔루션 아바커스를 활용해 이상거래 탐지율을 높이고 새로운 사고 패턴의 재학습이 가능한 모델을 구축했다.

다. 금융투자업

기존의 금융투자업은 고액 자산가들을 대상으로 투자자문과 일임 등의 서비스를 제공했다. 하지만 고객의 금융 정보는 물론 투자 성향에 대한 정보를 수집하고 분석하는 것이 가능해지면서 일반 고객을 대상으로 한 자산관리 및 투자 상품 추천 등이 가능해졌다. 실제 신한금융투자는 신한카드 등 계열사 간 정보를 교류해 해외주식구매와 추천 모델을 개발했다.¹¹ KB증권은 엔씨소프트, 디셈버앤컴퍼니 자산운용과 간편투자 증권사 출범을 계획하고, 로보어드바이저와 인공지능 프라이빗뱅크(PB)를 활용해 자산운용 및 자산관리 자문 서비스를 제공할 계획이다.¹²

라. 보험업

2019년 7월 금융당국은 보험사의 건강관리 서비스 부수업무를 허용했다. 헬스케어 서비스는 고객의 건강 데이터를 분석해 차별화된 건강관리 서비스를 제공하고, 고객의 목표달성 여부에 따라 정해진 보상 혹은 보험료 할인 혜택을 제공한다. 그 예로 교보생명엔 앱 ‘케어(Kare)’를 선보여 고객의 신체정보에 따른 맞춤형 건강증진 솔루션을 제공한다. 10여 개 질환의 위험도 예측과 건강관리 방안, 건강검진 데이터 트래킹 등의 서비스가 이에 해당한다.¹³ 인슈어테크는 기존 보험산업에 데이터 분석과 인공지능 기술을 활용해 혁신적인 서비스를 제공하는 것으로, 캐롯손해보험은 주행거리만큼 보험료를 계산하는 퍼마일 자동차보험을 출시해 저렴한 보험료를 제공하고 있다.

11. 유근일 기자, 「계열사 간 빅데이터 교류 물꼬…금융투자업계 데이터 기반 신상품 준비 한창」, 전자신문, 2019.7.31.
12. 윤민혁 기자, 「NC-KB증권·디셈버앤컴퍼니, ‘인공지능 PB’만들기 손잡았다」, 조선비즈, 2020.10.7.
13. 이종현 기자, 「교보생명, 통합 고객서비스 앱 ‘케어’ 출시」, 조선비즈, 2020.8.10.

마. 핀테크 기업

에이젠글로벌은 오픈된 개인신용정보를 활용해 ‘Credit’을 시각적으로 지도화한 크레딧투맵(Credit to Map)을 완성했다. 또한 기업의 신용데이터를 지역, 업종, 등급 등 차주의 속성별 대출규모와 AI 기반의 분석을 통해 기업의 빅데이터 분석에 활용했다. 이같은 데이터 분석 및 AI 적용은 금융서비스 초개인화의 근거를 마련할 수 있다. 향후 보험신용정보를 활용해 생애주기별 보험의 평균 가입율과 종목 등을 분석해 보험 상품 추천에 활용될 것이다.

통신업의 금융업 진출도 활발히 이루어지고 있다. SK텔레콤은 하나금융그룹과 손잡고 핀테크 업체 ‘핀크(Finnq)’를 선보였다. 핀크 앱을 이용해 오픈뱅킹 기반의 편리한 금융서비스를 제공하고, 고객의 휴대폰 이용 정보를 바탕으로 한 비금융정보 신용평가서비스인 ‘T스코어’를 제공한다. 이 같은 T스코어는 기존에 신용평가에서 불이익을

[표 4-1-2] 데이터 활용과 금융서비스 예시

업권	활용 데이터 분석	주요 서비스 목적	서비스 예시
은행	계좌거래 내역, 대출잔액, 금리·이자 등 다양한 금융자산 현황 분석	저축, 재테크 방안 안내 등을 통한 자산 형성 지원	• 오픈뱅킹: 타 금융사 계좌 정보와 자사 은행의 정보를 함께 제공 대부분의 은행에서 제공 중
카드	카드사용 일시, 결제내역, 카드대출 이용 등 소비패턴 분석	다양한 카드사용 혜택 제공 및 합리적인 소비습관 형성 지원	• 자산조회 및 소비 컨설팅 서비스 - 현대카드: 소비케어서비스 운영 - 우리카드: 핀테크 앱 ‘브로콜리’와 제휴해 자산관리서비스 제공 - 신한카드: 토털 소비관리 신규 추가 - KB국민카드: 리브메이트를 기반으로 개인사업자 특화 신용평가 서비스인 크레딧 트리(Credit Tree) 운영 • 그 외: 고객구매 이력의 활용도 높은 정보를 바탕으로 데이터 상품화, 이종산업 간 데이터 거래 추진
금융 투자	투자종목, 투자금액, 자산규모 등의 투자정보를 분석해 투자패턴 제공	세제 혜택, 투자습관 개선 등 다양한 포트폴리오 제공	• 투자자문 및 일임업: 고객 자산가에서 일반 고객으로 대상 확대
보험	보험료 납입내역, 보험기간, 보장내역 등 보험정보를 이용해 노후예측 및 건강 분석	연금관리를 통한 노후설계와 저비용 건강관리 서비스 제공	• 헬스케어: 앱을 이용해 맞춤형 건강습관을 제시하고, 목표달성 시 다양한 혜택이나 포인트 제공 - SK텔레콤: AIA vitality와 제휴해 걷기 목표 달성 혹은 헬스푸드 구매시 통신비 할인 등의 혜택 제공
핀테크	은행, 카드, 증권, 보험 등 여러 금융 업권의 금융상품 및 정보를 종합 비교·분석	맞춤형 금융상품 추천 및 정보주체의 정보권리 행사 대행	•뱅크샐러드: 카드, 은행, 투자, 대출, 연금 등 종합적 자산조회 및 소비컨설팅, 금융상품 추천 앱을 운영
IT/융합	금융과 통신, 유통 등 데이터와 융·복합	통신·유통 정보기반 금융상품 등 고부가가치 혁신서비스 제공	• 금융정보이력이 부족한 이용자의 공과금, 세금, 통신비, 카드이력 등을 조회한 신용평가

받은 씬파일러의 신용평가 등급을 개선시킴으로써 금융 소외계층을 포용하는 데 기여했다.

4. 향후 기대 및 전망

전 세계적으로 전통적인 금융과 빅테크의 금융업 진출, 핀테크의 등장으로 금융 데이터 활용이 경쟁력의 핵심이 되고 있다. 국내에서는 카카오페이가 2014년 간편결제 서비스를 도입해 송금, 대출, 신용조회, 전자인증 등 다양한 금융서비스 포트폴리오를 완성했고, 2020년 3월 자산관리 서비스를 출시해 현재 약 350만 명이 사용하고 있다.

네이버파이낸셜은 자동차, 개인, 보험료 정보를 조합한 맞춤형 자동차보험 조회, 자산, 소득 수준에 맞는 부동산 매물 추천, 통합데이터 기반의 새로운 대출 심사모형을 이용한 대출 등의 서비스를 준비 중이다. 신한은행은 신한SI 설립, 우리은행과 KB국민은행은 업무 가치사슬에 적극적으로 인공지능 기술을 도입해 발 빠르게 디지털 전환을 구축하고 있다. 통신업계도 클라우드 서비스를 이용한 금융서비스 제공 및 통신 데이터를 활용한 핀테크 결합 등의 서비스를 새롭게 설계하고 있다.

앞으로 ‘모든 기업은 금융서비스를 제공하기 원할 것이다’라고 예측해도 과언이 아닐 만큼, 데이터 플랫폼은 데이터 활용과 인공지능을 이용해 데이터를 융합해 가치 있는 금융을 ‘서비스(as-a-Service)’ 형태로 기획할 것이다. 금융 빅데이터 활용 및 데이터 경제 활성화를 토대로 앞으로 금융시장은 얼마나 혁신적인 고객서비스를 제공하느냐가 관건이 될 것이다. 앞서 다룬 개방된 데이터 시장, 마이데이터 산업, 오픈 API, 클라우드, 금융 및 비금융 데이터 융합, 인공지능 금융이 그 바탕이 될 것이며 데이터 비즈니스의 고도화가 디지털 금융 혁신으로 더욱 발전될 것이라 기대한다.

제2장. 헬스케어 분야 데이터 활용 비즈니스 현황

필자: 김준연
(인바이츠헬스케어 대표)

코로나19 전염병에서 촉발된 비대면 사회로의 진입은 데이터를 활용한 헬스케어 산업에 대한 관심을 증폭시켰다. 다양한 분석 도구를 바탕으로 헬스케어 데이터에 대한 수집·이해·활용 방식이 확대되고 있으며, 이종 산업 간 데이터를 융합해 개인화된 맞춤형 헬스케어 사업이 시도되고 있다. 주요 헬스케어 데이터 종류에 따라 추진 중인 주요 사업화 사례들과 이종 산업 데이터 간의 융합 사례를 같이 살펴보고자 한다.

1. 디지털 헬스케어 시장규모와 성장

디지털 헬스케어는 2020년 전 세계를 강타한 코로나19를 해결하는 데 중요한 기술로 지목되고 있다. 전 세계에서 코로나19 예측과 유입 방지, 현장대응과 확산 방지를 위한 실시간 모니터링과 원격의료, 신약개발까지 사회 전반에서 디지털 헬스케어 발전을 체감하고 있다.

글로벌 시장조사기관 그랜드뷰리서치(Grand View Research, 2019)의 전망에 따르면, 글로벌 디지털 헬스케어 시장은 2019년 1,145억 달러(약 140조 원)에서 2025년 5,092억 달러(약 610조 원)로 폭발적인 성장이 예상된다. 이 중 2019년 기준 전체 시장의 절반 이상을 북미 지역이 차지하며, 국내 시장은 전체의 1.4%인 2조 원 수준으로 추정된다.

2. 디지털 헬스케어 분야의 데이터 활용 확대

과거 헬스케어 데이터는 병원에서 환자를 진료하고 치료하는 과정에서 생성돼 의료기관이 보관하는 의료 정보 및 이에 수반되는 약 처방과 의료보험에 반영되는 정보들로 보는 협의의 개념이었다. 그러나 최근에는 제약, 의료기기, 보험, 금융, 생활 습관 등 건강활동에 수반되거나 사회경제, 주변환경 등 건강에 영향을 미치는 여러 분야의 데이터를 포괄한다. 의료기관 외부에서 생성되는 데이터는 의료 및 유전체 등 서비스 제공 영역의 내부 데이터보다 매우 클 뿐만 아니라 건강에 미치는 영향도 크다는 분석이 있다.

비즈니스에서 많이 활용되는 데이터는 [표 4-2-1]과 같다. 현재는 인공지능(Artificial Intelligence, AI) 학습용 빅데이터, 임상시험 데이터를 적극적으로 수집, 정리, 관리하고 사업에 활용하는 추세다. 인공지능은 방대한 데이터를 활용하는 데 있어 자체 알고리즘을 이용해 학습한다. 따라서 헬스케어 분야에 활용할 수 있는 인공지능은 헬스케어 빅데이터를 통해 학습되고 있다.

[표 4-2-1] 비즈니스에서 많이 활용되는 데이터 원천

데이터 구분	종류	Source
임상 연구 및 진료 데이터	전자의무기록, 전자건강기록, 처방 정보, 의료 영상기록, 시술 기록 등 의약품·의료기기 임상시험 데이터를 포함한 연구 데이터	병의원 의무기록관리(EMR), 병원정보시스템(HIS), 의료영상저장전송장치(PACS), 병원 및 임상시험수탁기관의 임상시험, 생물학적 지식 DB
유전체를 포함한 오믹스 데이터	유전체, 전사체, 단백체, 후성 유전체 등 다양한 분자 수준에서 생성된 데이터	의료기관, 연구기관, 검사 기관, 분석기관
개인 건강 데이터 및 라이프로그	웨어러블기기, 모니터링 시스템, 사물인터넷, 모바일 애플리케이션에서 수집된 데이터로, 개인의 일상활동에 관한 모든 데이터	개인건강기록(PHR), 웨어러블기기, 스마트폰 등
공공기관에서 발생하는 청구데이터	자격 및 보험료 관련 데이터, 진료내역, 건강검진 결과, 사망 정보 등 공공기관에서 수집, 보관, 관리하는 데이터	건강보험공단, 통계청, 건강보험심사평가원 등

출처: 관련 자료를 종합 분석해 저자 재구성

가. 임상 연구 및 진료 데이터

임상데이터는 의사, 간호사, 약사, 물리치료사, 의무기록사, 행정 담당자 등 의료기관 관련자가 사람을 대상으로 한 진료 과정 중에 생성하는 인구 통계학적 정보, 진단, 치료, 처방, 실험실 검사, 생체 모니터링, 입원 기록, 간호기록 및 영상, 진료비 등의 데이터를 의미한다. 진단 기술이 발전함에 따라 인간이 입력하지 않고 의료장비로부터 직접 생산돼 저장되는

데이터의 양이 점점 늘어나고 있다. 전자문서 형태인 전자의무기록과 처방전달 시스템에 기록되며 엑스레이나 MRI 같은 영상 데이터는 영상저장전송시스템(PACS)에 저장돼 관리된다.

[표 4-2-2] 임상진료 데이터 활용 사례

기업명	사업·프로젝트	내용
구글헬스와 자회사 베릴리(Varily)	의료영상 빅데이터로 학습시킨 인공지능을 이용한 질병 진단 ⁰¹	- 구글헬스는 2016년에 당뇨병의 증상인 망막증과 황반부종의 영상 빅데이터를 시에게 학습시켜 이를 토대로 당뇨병 망막병증을 정확하게 진단할 수 있는 딥러닝 기반의 인공지능을 발표함. 이후 발전을 거듭해 구글헬스의 인공지능은 망막 안저사진(Retinal Fundus Photographs)을 판독해 최고 수준의 인간 안과의사와 대등한 수준의 진단 정확성을 확보 - 구글헬스는 2019년에 인도의 아라빈드 안과 병원에서 본격적인 임상 실험 개시. 인도는 안과 의사가 10만 명이 부족하고 당뇨병을 앓는 7200만 명 가운데 600만 명 이상이 당뇨병 안과 질환으로 판정 - 구글헬스의 이 알고리즘은 유럽연합 의료기기 표준 인증을 획득
ICON plc, Medpace Inc, Syneos Health 등 글로벌 CRO 기업	빅데이터를 활용한 임상주기 단축과 비용절감 ⁰²	- 글로벌 CRO 업체들과 디지털 솔루션 제공 업체들이 함께 실사용증거(Real World Evidence)의 임상 빅데이터를 효율적으로 사용해 임상 비용을 줄이고, 임상 주기를 단축하고자 'Align Clinical CRO'를 구축 - 축적된 정형 및 비정형 데이터를 활용해 각 임상에 최적화된 환자를 빠르게 모집하고, 후보약물을 빠르게 발굴하며 이를 통해 임상이 진행되지 못한 희귀질병에 대한 임상과 신약개발이 가능할 것으로 기대

출처: 관련 자료를 종합 분석해 저자 재구성

나. 유전체를 포함한 오믹스 데이터

오믹스(Omics) 데이터는 인체로부터 수집된 인체구성물의 DNA/RNA 시퀀스 정보 및 미세결실검사(microarray) 분석 정보를 의미한다. 최근에는 국내 병원에서도 암유전체를 분석한 맞춤형 의료 서비스가 본격적으로 시작되고 유전체분석 전문 기업들이 등장해 사업화가 활발하게 진행되고 있다⁰³. 차세대 유전체분석기술(Next Generation Sequencing, NGS)에 기반을 둔 암유전체 패널 검사 서비스는 지금까지 알려진 유전자 변이 수십 개를 하나의 패널로 구성해 유전성 질환과 비유전성 암의 필수 유전자를 대상으로 각 질병을 일으키는 유전자 돌연변이를 찾아내고, 돌연변이에 따른 표적 치료제를 추천해준다. 유전체분석 전문 기관에서 수탁하고 여기에서 제시된 유전체분석 리포트를 기반으로 환자 치료를 시행하는 방식이다.

01. Google Health, <https://health.google/>
02. Align Clinical CRO, <https://www.alignclinicalcro.org/>
03. The Cancer Genome Atlas Program, <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>

[표 4-2-3] 유전체 포함 오믹스 데이터 활용 사례

기업명	사업·프로젝트	내용
닥터노아 바이오테크	유전체 데이터와 인공지능을 활용한 희귀질환 타겟 약물 재창출	- 희귀질환 치료제 중 복합체를 예측할 수 있는 플랫폼 'ARK' 개발 - 의약학 문헌 정보 및 유전체 데이터, 인공지능을 이용해 환자의 특성에 맞는 최적의 약물 조합을 예측해주는 시스템 - 기존에 승인된 약물 혹은 임상에서 실패한 약물을 2~3개 조합해(약물 재창출 기법) 새로운 기능을 갖는 복합체로 재탄생
신टे카바이오	차세대 유전체분석기술(NGS)로 개인 유전체 데이터 분석을 통한 정밀의학	- 개인 유전체 데이터 분석으로 컴퓨터 기반 가상 임상시험, 환자 맞춤형 치료 및 약물 재창출 등을 개발하는 유전체 헬스케어 빅데이터 기업 - 핵심기술은 특정 질환을 앓는 환자들의 유전체 빅데이터를 이용해 특정 약물에 효과가 있는 군과 그렇지 않은 군, 또는 부작용이 있는 군을 찾아내는 '개인유전체맵 플랫폼' 임
메디젠 휴먼케어	유전체 패널검사를 통한 질병 위험도 분석	- 자체보유한 질병패널검사 시스템으로 남성 암 17종, 여성 암 19종 등 다양한 질병에 대한 위험도를 분석할 수 있는 유전체 검사 진행 - 유전체분석 결과 및 생체지표 검사를 바탕으로 전문의, 전문 영양사 및 유전학 전문가의 자문을 이용해 개인 맞춤 솔루션 제공, 정기적으로 검사

출처: 관련 자료를 종합 분석해 저자 재구성

유전 데이터는 정밀의학 분야에서도 활발하게 사용된다. 한 개인의 유전체 서열을 명확히 파악해 개인별로 특정 질병으로 이환될 확률, 약물 효과, 건강 관련 데이터(생활습관, 식사, 운동, 거주 환경 등)와의 연계성 등을 예측하고 면밀하게 분석해 개인 맞춤형 최적의 의료 서비스를 제공한다. 이외에도 신약 개발, 약물 재창출, DTC(Direct-to-Consumer, 소비자 직접 의뢰 방식) 유전자 검사에 대한 사업화 영역에서 유전체 빅데이터가 활발하게 활용되고 있다.

다. 개인 건강 데이터 및 라이프로그

개인 건강 데이터(Personal Health Record, PHR)는 병원에서 진료가 이뤄지는 동안 지속적으로 수집·제공되는 개인의 일상 헬스케어 데이터이며 개인의 기록, 웨어러블기기, 스마트폰 등을 통해 수집된다. 이는 유전 데이터 및 의료 데이터, 건강검진 데이터와 합쳐져 상호 상관 관계 분석을 통해 개인 맞춤 의료(Personalized Medicine), 정밀 의료(Precision Medicine)의 중요한 한 축을 이룬다.

[표 4-2-4] 개인 건강 데이터 활용 사례

기업명	사업·프로젝트	내용
모바일닥터	‘열나요’ 앱 서비스	• 부모가 입력하는 아이들의 체중, 증상, 해열제 복용, 진단명, 예방접종 여부 및 체온 등의 데이터를 분석해 영유아 체온 관리를 위한 새로운 의학지식과 다양한 서비스를 제공 • 체온기록 400만 건, 해열제 입력 100만 건, 진단명 5만 건 및 예방접종기록 5만 건에 해당하는 헬스케어 빅데이터 보유 • 최근에는 이와 같은 데이터를 이용해 머신러닝 기반의 독감 스크리닝 프로그램을 개발중
웰트	벨트형 헬스케어 디바이스로부터 측정된 PHR 빅데이터를 활용한 헬스케어 솔루션 및 서비스 개발	• 환자에 의해 생산되는 건강 데이터 중 주로 디바이스에서 생산되는 라이프로그를 기반으로 허리둘레, 활동량, 식습관 등을 모니터링해 독창적인 맞춤형 비만관리를 위한 웰니스 솔루션을 확보
메디블록	블록체인 기술을 이용한 개인건강기록 플랫폼 개발	• 핵심 기술은 블록체인 기술을 이용해 이를 바탕으로 개인건강기록 플랫폼을 만들고 분권화된 저장소에 데이터를 저장하며 플랫폼 참여자에게 가상화폐로 보상을 주는 것 • 플랫폼을 이용해 개인 사용자는 헬스케어 빅데이터를 만들어내고 연구자나 기업은 사용자로부터 헬스케어 빅데이터를 안정적으로 획득

출처: 관련 자료를 종합 분석해 저자 재구성

라. 공공기관에서 발생하는 청구 데이터

청구 데이터는 국민건강보험공단, 건강보험심사평가원, 보험사 등에서 발생하는 데이터를 말한다. 보험급여가 책정된 항목의 질병에서는 시계열별로 처치, 시술, 검사 등의 진료 내역, 진단명, 보험자 지급 비용, 환자부담금, 환자 인구통계학적 특성, 요양기관 정보 등 다양한 정보를 얻을 수 있다. 그러나 비급여 항목에 대한 데이터⁰⁴는 확보하기 어렵다는 단점이 있다.

[표 4-2-5] 청구 데이터 활용 사례

기업명	사업·프로젝트	내용
코아제타	청구 데이터를 분석해 제약사에게 제공	• 청구 데이터인 건강보험심사평가원의 빅데이터를 기반으로 의약품의 사용 패턴과 특징을 분석하고 시각화해 제약사에게 서비스 제공 • 이를 위해 PBD(PharmaBigData)라는 의약품 분석 및 시각화를 위한 서비스 개발
미국 누나(NUNA) 헬스케어 ⁰⁵	메디케이드 데이터 관리 분석 플랫폼 제공	• 미국 정부 의료 프로그램인 메디케이드와 연관된 데이터 분석 표준화 플랫폼 구축 • 미국 전역에 걸쳐 7300만 명이 넘는 빈곤층, 장애인 및 어린이 데이터를 관리하는 메디케이드에 대한 최초의 표준화된 데이터 플랫폼 제공 • 메디케이드 대상자들 치료 등의 청구 데이터 정보를 이용해 클라우드 컴퓨팅 데이터베이스를 구축하고 고객들에게 자료 제공

출처: 관련 자료를 종합 분석해 저자 재구성

04. 보건의료빅데이터개방시스템, <https://opendata.hira.or.kr/home.do>

3. 디지털 헬스케어 데이터와 다른 데이터의 융합 비즈니스

최근의 트렌드는 비대면 사회로의 급격한 변환을 통해 이종 산업 간의 데이터 융합이 활발하게 일어나고 있다는 것이다. 국내 기업의 주목할 만한 몇 가지 사례를 소개한다.⁰⁶

[표 4-2-6] 디지털 헬스케어 데이터와 타산업 융합 비즈니스 사례

기업명	융합분야	내용
KB손해보험	금융·건강정보	• 보험업계 처음으로 ‘빅데이터 자문 및 판매 서비스’ 부수 업무 자격 취득. 고객 건강정보를 바탕으로 종합건강관리 서비스를 제공하는 헬스케어 사업과 여러 기관에 산재한 금융정보를 통합 관리하게 해주는 마이데이터 사업 추진
삼성생명	금융·건강정보	• 업계 두 번째로 ‘빅데이터 자문 및 판매 서비스’ 부수 업무 자격 취득. 삼성생명은 고객 건강정보를 기반으로 종합 건강관리서비스를 제공하는 헬스케어 사업을 강화하고 출어진 고객 금융거래 정보 등을 일괄수집해 제공하는 마이데이터 사업 등을 추진 예정
GC녹십자헬스케어·에이블애널리틱스	빅데이터·건강정보	• GC녹십자헬스케어는 누적된 빅데이터를 기반으로 최근 인수한 빅데이터 분석 전문 컨설팅 기업인 에이블애널리틱스의 데이터 분석 역량을 활용해 차별화된 개인 맞춤형 건강관리 솔루션 제공 계획 • 2020년 하반기 ‘비만도 및 대사증후군 고객 세분화 모델’, ‘검진센터 추천 모델’, ‘건강검진 맞춤 및 검사항목 추천 모델’ 등을 추진
SK건설	건설·헬스케어	• 체성분 분석 세계 1위 기업인 인바디와 ‘홈 IoT 랩(Home Internet of Health Things LAB)’을 오픈 • ‘홈 IoT 랩’은 기존 세대별 조명, 가스밸브, 가전제품 등을 원격으로 제어하는 IoT 기술에 건강(Health) 기능을 결합한 IoT 기술이 구현됐으며 실내 공기질을 측정해 최적으로 유지하는 SK건설의 지능형 환기시스템도 적용 • 고객은 앱을 통해 체성분, 운동 능력 분석 등 현재 건강상태를 진단받고, 개인별 맞춤 운동 프로그램과 식단 정보를 제공 받음. 생체 징후 원격 모니터링과 각종 건강 교육 서비스가 가능한 인프라도 구축돼 원격진료 시대를 대비한 토털 헬스케어를 체험하는 공간으로 구성
유비케어	제약사·보험·헬스케어	• 헬스케어 데이터 분석·유통 플랫폼 ‘UBIST Data Bank’ 출시(2019.8.19) • 보유한 의료·헬스케어 빅데이터를 기반으로 광범위한 의약품정보, 처방정보, 환자 데이터 등과 융복합하고 다각적으로 분석해 다양한 콘텐츠를 생산 및 공급 • 의약품 원외처방 시장 전망 및 마케팅 분석 콘텐츠 • 의약품 처방 패턴 분석 콘텐츠 • 의약품 이상사례 분석 콘텐츠 • 보험업권 대상의 질환 패턴 변화 분석과 의약품 특허 만료에 따른 시장 전망 등의 신규 콘텐츠 출시 계획

(계속)

05. NUNA Healthcare, <https://www.nuna.com/>

06. 주요 기사 및 보도자료 분석, 정리해 저자 재구성

기업명	융합분야	내용
아리아케어· 폴무원 푸드머스	식품·유통·건강정보	• 시니어 플랫폼 기업 아리아케어가 폴무원푸드머스와 시니어 세대 만성질환 개선을 위한 맞춤형 케어테크 솔루션인 ‘푸드케어’ 서비스 제공 • ‘푸드케어’ 서비스는 2020년 10월부터 노인장기요양보험 수급자를 대상으로 푸드케어 꾸러미를 무료로 공급하는 사회공헌형 시범 사업으로 시작 • 시범사업 방식은 아리아케어 플랫폼 고객 중 100여 가구를 선정해 어르신들의 건강 상태와 음식 기호, 식생활 환경 변화 추이 등을 조사하고 이를 기반으로 폴무원 연구진들에 의해 개발된 맞춤형 고령친화식단이 제공될 예정 • 기간내 수집된 사용자들의 빅데이터는 가톨릭의대 빅데이터센터와 스마트헬스케어센터 연구진들과 협업해 향후 시니어를 위한 맞춤형 고령친화식품 연구 개발에 활용될 예정 • 두 기업은 ‘방문요양 서비스’로 점차 확대하고 빅데이터 분석을 통한 시니어 맞춤형 케어 제품 공동 연구, 푸드케어 제품 공급 및 유통 협력 등 계획
바디프렌드· 메디플렉스 세종병원· 메디컬에이아이	제약사·보험·헬스케어	• 심전도 빅데이터를 이용한 AI 분석을 통해 빈혈 진단 성공 • 7만 건 이상의 심전도 데이터와 나이, 성별 등의 데이터를 AI 기술로 분석해 높은 수준의 정확도로 빈혈 진단과 수치까지 파악할 수 있음을 밝혀냈으며 빈혈 외 중대한 심장 관련 질환 등 다른 질환에 대해서도 예측과 진단을 할 수 있는 연구를 진행중

출처: 관련 자료를 종합 분석해 저자 재구성

산업이 활성화하는 한편, 안전한 데이터 활용을 위한 제도적 장치(데이터3법 등)에 대한 끊임 없는 연구와 개선이 함께 이루어지길 기대한다.

4. 향후 전망과 기대

헬스케어 데이터의 활용가치는 무궁무진하다. 향후 더 많은 이종 산업의 데이터를 헬스케어 데이터와 접목하는 시도가 활발하게 진행될 것으로 예상된다. 특히 올해는 비대면 사회로의 급속한 진화에 따라 ICT를 활용한 융합 기반 헬스케어 데이터 사업 모델들의 필요성이 공론화되고, 디지털 헬스케어 산업은 더 가속화 될 것으로 보인다.

2020년 5월에 발표된 4차산업혁명위원회의 설문조사⁰⁷⁾에 따르면, 국민의 87%가 의료보건기술을 위해 개인정보를 제공할 의향이 있으며, 전문가 집단은 올해 개정된 데이터3법으로 빅데이터를 활용한 금융업과 의료보건업이 가장 크게 성장할 것이라고 내다봤다. 하지만 의료법과 국민건강보험법에서 개인정보를 별도로 보호한다는 점은 개인정보보호법 개정안과 충돌된다는 지적이 있었다.

헬스케어 산업의 중요한 요소 중 하나인 ‘공공성’이라는 측면에서 이러한 변화에 부응하되 실제적인 사회경제적 가치를 만들어내기 위한 사회적 공감대 형성이 필요하다. 또한 헬스케어 데이터를 활용하는 사업자에게는 강한 윤리적 잣대가 요구된다. 헬스케어

07. 「4차위 데이터3법 관련 공동 설문조사 결과」 보도자료, 4차산업혁명위원회·대한상공회의소·한국인터넷기업협회, 2020.5.18.

제3장. 제조 분야 데이터 활용 비즈니스 현황

필자: 이재혁
(마키나락스 공동 대표)

제조기업은 사물인터넷과 빅데이터 프로젝트를 통해 많은 데이터를 축적했다. 축적한 데이터를 분석하기 위해 AI 기술에 대한 관심이 높아졌다. 아직은 시장 초기이므로 제조기업은 프로젝트 위주로 기술을 시험·적용하고 있다. 제조데이터는 직관적 해석이 어려워 기술적인 제약이 많다. 또한 제조업은 수율과 안전이 중요해 AI 기술 도입과 운영이 쉽지 않다. 하지만 최근에 이를 도입하는 제조기업이 늘면서 적용 분야가 다양해지고 급속도로 기술 개발이 이뤄지면서 시장규모가 커지고 있다.

1. AI와 접목한 제조업 시장 동향

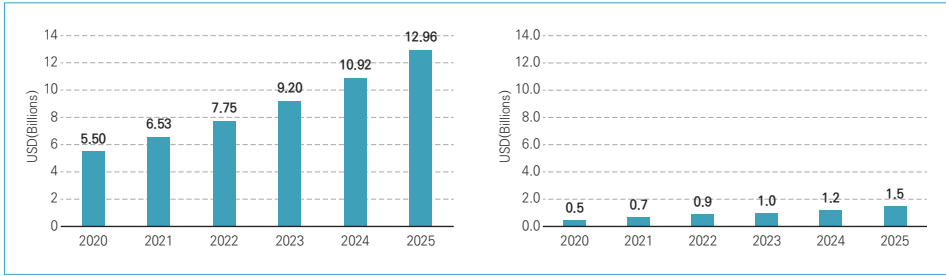
인공지능(AI) 기술을 소비재가 아닌 제조업 같은 산업에 적용하는 것을 산업 인공지능(Industrial AI)이라고 한다. 2015~2017년 사이에 제조업에서 사물 인터넷(IoT), 빅데이터 기술을 통해 광대한 데이터가 축적되면서 AI 기술을 도입하려는 시도가 본격화됐다. 24시간, 365일 쉬지 않고 생산이 이뤄지는 공정에서는 산업 분야에 따라 하루에도 테라바이트 이상의 방대한 데이터가 생산장비, 설비, 공정, 제품에서 만들어진다. 제조기업은 이런 빅데이터를 머신러닝(Machine learning), 딥러닝(Deep learning) 기술로 분석하려는 다양한 시도를 하고 있다.

최근 2년 동안 제조업에 AI를 도입하려는 수요는 계속 증가하고 있다. 아직은 시장 초기 단계라 정확한 시장 크기를 추정하기는 어렵다. 하지만 AI 도입 프로젝트를 수행하는 기업이 예전에는 매출 10조 원 규모의 제조 대기업 위주였다면, 지금은 매출 1조 원 미만의 중소기업까지 제조 공정에 AI를 도입하려는 시도가 늘고 있다. 시장조사 자료를 따르면 글로벌 제조 분석 시장은 2020년 기준 550만 달러로 추정되며, 2027년까지 연평균성장률(CAGR) 18.7%로 성장해 1,820만 달러 규모에 이를 것으로 전망된다. 한국

시장은 2020년 50만 달러에서 2025년까지 150만 달러 규모로 성장할 것으로 추정된다.

초기 AI를 도입하려는 수요는 반도체, 자동차, 철강 관련 기업에서 많았다. 정유 산업이나 화학 산업과 달리 이들 산업은 공정, 원재료, 제품이 자주 바뀌고 복잡도가 높아 제조 이슈를 해결하는 데 AI 도입을 검토했다. 반도체 산업은 3~5년마다 신규 공장을 건설하고, 3년마다 신규 장비를 도입해 신규 라인을 증설하고 있다. 이렇게 변화가 많은 경우에 아무리 뛰어난 엔지니어라도 새로운 변화를 지속적으로 수용해 최적의 수율을 달성하기 어렵다. 이런 부분에서 빅데이터, AI 기술 도입에 대한 수요가 발생한다.

[그림 4-3-1] 제조기업 시장 규모



출처: Artificial Intelligence In Manufacturing Market - Global Forecast To 2023, Markets and Markets, July 2017.7.
<https://www.researchandmarkets.com/reports/4805237/manufacturing-analytics-global-market#pos-0>

2. 주요 기업 동향

아직 시장 초기이기 때문에 다양한 시도를 하는 스타트업이 꾸준히 등장하고 있다. 시장 지배 기술이 나타나거나 성숙된 단계가 아니므로 다양한 연구개발 노력이 필요하다. 마이크로소프트, 액센추어, 아마존웹서비스, 휴렛팩커드엔터프라이즈 같은 IT 서비스

[표 4-3-1] IT 서비스 대기업의 스타트업 인수 현황

IT 서비스 대기업	마이크로소프트	액센추어	아마존웹서비스	휴렛팩커드엔터프라이즈
인수된 스타트업	Bonds.ai (강화학습)	Clarity Insights (AI 컨설팅 서비스)	Angel.ai (자연어 처리)	MapR (데이터 플랫폼)
목적·내용	마이크로소프트의 클라우드 서비스 애저 경쟁력 강화	액센추어 인공지능 컨설팅 강화	아마존 웹 서비스의 자연어 처리 인력 확보	휴렛팩커드 엔터프라이즈의 데이터처리분석 플랫폼 강화

출처: 마키나락스 자체 분석

대기업은 AI 기술 개발 외에도 인수합병(M&A)을 통해 기술력 높은 스타트업을 계속 인수하고 있다.

구글은 B2C 비즈니스에서 시장 지배력이 높아 AI 기술을 기존 사업에 적용하는 데 집중하고 있다. 직접 산업 인공지능 시장에 진입하고 있지는 않지만, 구글 클라우드와 안드로이드 플랫폼에 AI 적용을 시도하고 있다. 아직 초기 단계지만, 로봇팔 제조사에 구글 클라우드를 SaaS, PaaS 플랫폼으로 제공하며, 로봇팔에 이상이 생기면 안드로이드 폰에서 모니터링하고 오류를 파악하게 하는 시도를 하고 있다. AI 기술 개발기업은 고복잡도 산업 vs. 상품(Commodity) 산업, 범용 제품 vs. 맞춤화(Customization) AI 서비스로 차별화해 자리를 잡아가고 있다.

예를 들어 C3.ai와 업테이크사는 사물인터넷(IoT) 시기부터 개발해온 솔루션과 자체 데이터 플랫폼을 이용해 발전, 화학, 철도 같은 인프라 산업에서 사업을 전개하고 있다. 반면 액센츄어는 제품보다 데이터 분석 컨설팅으로 사업을 하고 있다. 일본 AI 기업인 Preferred Networks(PFN)는 제품 개발보다 일본 내 대기업과 장기 R&D 계약을 맺고 데이터 분석 서비스를 제공하고 있다.

3. 제조업에서 AI 도입의 도전

AI 기술의 성공사례는 B2C 산업에 AI를 도입한 예가 대표적이다. 구글 포토는 AI의 딥러닝 기술을 이용해 사진 분류를 사용자 지정 없이 스스로 할 수 있다. 네이버 파파고는 딥러닝 기술을 도입해 번역 품질을 크게 향상시켰다. 하지만 제조업에 AI가 실제 현장까지 적용돼 성공한 사례는 많지 않다. 이유는 제조업에 AI를 도입하는 것이 B2C 산업에 도입하는 일보다 어렵기 때문이다. 맥킨지 컨설팅이 실시한 조사 자료에 따르면, AI를 도입하는 3개 기업 중 2개 기업이 AI 기술을 테스트하는 파일럿 단계까지 진행하고 실제 생산라인에 도입하는 것을 이루는 것으로 나타났다.⁰¹

그 이유는 제조업과 B2C가 사용하는 데이터와 AI 기술 도입 목적, 응용분야가 다르기 때문이다. B2C에서는 AI를 인간이 하는 작업을 대체하는 용도로 많이 활용한다. 제조업에서는 사람이 하기 어려운 일, 예를 들면 오랜 경력의 엔지니어라도 하기 어려운 공정

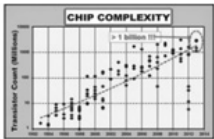
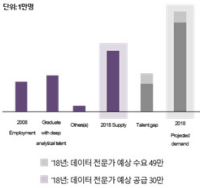
최적화, 장비 고장 예측 분야 등에 AI를 도입하려고 한다.

[표 4-3-2] B2C AI와 B2B AI 비교

구분	소비자 기반 AI	제조업 특화 AI
목적	인간의 지능을 모방해 인간이 할 수 있는 일을 대체	인간의 인지·분석 능력 범위보다 복잡한 업무를 대체
데이터	직관적 해석이 가능한 데이터 - 이미지, 동영상 - 텍스트, 음성 - 고객 프로필 & 행동	제조 공정 중 생성되는 데이터(비직관적 데이터) - 센서에 의한 측정 값 - 작업자에 의한 설비 설정 값 - 품질 계측 값 - 고장, 유지·보수, 수리 등의 업무일지
응용	개인화 광고, 개인화 콘텐츠·상품 추천 영상처리 기반 의료 진단 기계번역, 비디오 감시 시스템, 자율주행	제조업에서의 - 장비고장 예측, 원인인자 분석 - 공정 최적화, 제품 품질 관리 - 전기 설계 해석(Electrical Design Analysis)

출처: 마키나락스 자체 분석

[표 4-3-3] 제조 데이터 분석의 어려움

직관적 분석 어려움	문제 난이도 증가	데이터 분석가 부족
- AI에서 흔히 활용되는 이미지 데이터와 달리, 산업 센서 데이터는 인간의 직관적 해석이 어려움 - 해석을 위해서는 복잡한 데이터 전처리 필요	- 공정 수의 지속 증가(예: 반도체 '08년 300개 → '18년 800개) - 칩 집적도 증가	고급 데이터 분석 전문가의 제조업 기피 현상 발생. 데이터 분석 전문가 수급불균형 심화
		

출처: 마키나락스 IR 자료

01. Mckinsey Consulting report, Preparing for the next normal via digital manufacturing's scaling potential, 2020.4.

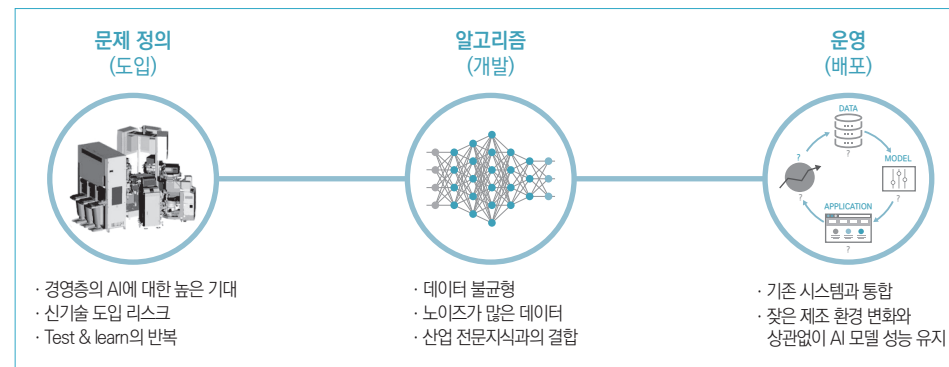
[표 4-3-3]은 제조업 데이터의 직관적 분석이 어렵다는 것을 설명하고 있다. 펄킨 그림을 보면, 대다수 펄킨과 다른 하나의 펄킨을 가운데 부분에서 쉽게 찾을 수 있다. 펄킨 그림 옆의 그래프는 제조업에서 흔히 사용되는 온도, 압력, 전류 센서에서 나오는 센서값 그래프다.

일반 사람은 그래프에서 데이터 정상패턴과 다른 이상값을 찾기 어렵다. 왜냐하면 이 그래프를 해석하기 위해서는 전문지식이 필요하기 때문이다. 이런 데이터를 분석하는 능력을 얻기 위해서는 시간과 노력이 많이 소요된다. 산업 전문지식을 갖춰야 하고 데이터를 전처리해 분석하기 쉬운 형태로 변환시킬 줄 알아야 하기 때문이다.

공정과 제품의 복잡도가 지속적으로 증가하기 때문에 분석하고자 하는 문제의 난이도도 높아지고 있다. 무엇보다도 아직도 제조업 데이터를 분석할 데이터분석가가 부족하다. AI 스타트업이 많은 미국 실리콘밸리에서조차 데이터분석가들은 상대적으로 재미있고 빠른 해석을 얻을 수 있는 B2C 분야에 분포해 있다.

[그림 4-3-2]은 AI를 도입하고 구축하는 과정에서 극복해야 할 어려움을 정리한 것이다.

[그림 4-3-2] 제조업의 AI 도입·구축 과정에서 어려움



가. AI 도입 단계

경영진은 알파고 같은 B2C에서의 성공사례를 기반으로 AI 도입 효과에 대해 기대가 크다. AI 기술은 계속 발전하고 있고, AI가 적용된 제조업은 시장 초기이기 때문에 신기술 도입 위험이 존재한다. 대부분 초기 도입 고객은 Test & learn을 목적으로 소규모 R&D 프로젝트만 테스트하려는 경향을 보인다.

나. AI 개발 단계

1) 데이터 불균형

데이터 분석을 위해서는 정상 데이터와 비정상 데이터가 모두 필요하지만, 제조업에서는 비정상 데이터가 드물다. 비정상 데이터가 많다는 것은 생산장비의 셧다운(Shut down)이 많거나 제품 이상이 많다는 의미다. 그런 기업은 생존할 수 없으므로 현재 잘 운영되는 기업에서 비정상 데이터는 적을 수밖에 없다. 이렇듯 데이터 불균형은 분석을 어렵게 만드는 원인 중 하나다.

2) 데이터 노이즈

제조업 데이터는 대부분 센서에서 생성된다. 예를 들면 200억 원 대의 복잡한 반도체 장비에는 온도, 진공, 진동, 전압 등 각종 센서가 100개 이상 달려있다. 이런 다양한 센서에서 하루에도 수 기가바이트 분량의 막대한 데이터가 만들어진다. 문제는 이런 데이터가 항상 정확하게 생성되지 않는다는 점이다. 통신 네트워크의 오류, 외부 환경(센서에 영향을 줄 만큼 높거나 낮은 온도 등), 센서 자체 오류 등으로 데이터가 정확하게 수집돼 저장되지 않는다. 그래서 부정확한 데이터가 섞여 있고 이것을 데이터 노이즈라고 한다. 이런 데이터 노이즈는 분석의 어려움을 가중시킨다.

다. AI 운영 단계

1) 기존 시스템과 통합

새로운 AI 시스템을 기존 제조 관련 시스템과 통합하는 것은 늘 위험을 동반한다. 제조 시스템은 수율, 환경 안전과 직결돼 있으므로 가장 보수적이면서 안전한 방법으로 통합해야 한다. 아마존웹서비스, 마이크로소프트 애저 같은 클라우드 서비스나 시스템과 결합하면 시스템 운영이 수월해진다. 하지만 제조업체는 B2C 수준보다 높은 보안 요구 때문에 클라우드 도입에 적극적이지 않다.

2) 잦은 제조 환경 변화와 모델 성능 저하

제조업의 제조 환경은 계속 바뀐다. 예를 들어 자동차 생산라인은 자동차 모델이 부분 변경되거나 신규 모델이 출시되면 변경된다. 차체가 달라지기 때문에 로봇팔의 동작이 변화하고 공정이 달라진다. 이렇게 환경이 바뀌면 데이터가 달라지고 모델이 바뀌므로 모델 성능이 지속적으로 유지되기 어렵다.

4. 제조업에서 AI 적용 분야

제조업에서 AI 적용 분야는 제조 관련 분야뿐 아니라 회계, 인사관리 등 일반 관리 분야에도 다양하게 도입되고 있지만, 제조 관련 분야를 집중적으로 기술하려고 한다.

[표 4-3-4] 제조업에서 AI 기술 주요 적용 분야

적용 분야	설명	적용 AI 기술(예시)
주요 생산장비 예지 정비	장비에 부착된 센서에서 나오는 데이터나 공정 등 생산 환경에서 나오는 데이터를 분석해 장비의 예기치 못한 셧다운(Shut down) 혹은 이상징후의 발생을 예측하고 계획되지 않은 셧다운(Unplanned Shut down)을 최소화함	딥러닝, 평생학습
제품 표면 불량 검사	생산 과정 중에 있는 반제품·주요 부품·제품의 품질 검사를 공정 장비와 테스트 장비의 측정 데이터를 분석해 과거의 무작위 샘플링(Random sampling) 검사 기법을 대체함	머신비전
공정 최적화	품질 최적화를 위해, 공정 수행에 필요한 원·부재료 투입량 최적화, 생산설비 최적 제어값 등을 추천함	강화학습
생산 안전환경	공장 작업 환경, 외부 환경에서 나오는 데이터를 분석해 안전사고를 예방함	머신러닝
일반 업무관리	챗봇, RPA(Robotics Process Automation) 등을 도입해 정형화된 업무를 AI가 수행하고, 인간은 비정형화한 업무를 수행함	자연어 처리

출처: 마키나락스 자체 분석

가. 제품 표면 불량 검사 분야

우리나라에서 다른 분야보다 먼저 도입돼 의미 있는 성과를 보여주는 분야는 제품 불량 검사다. 머신비전(Machine Vision) 기술이 사용되며, 주로 제품 표면을 검사해 사람이 놓치기 쉬운 미세 스크래치 등 표면 불량을 찾아낸다. 2016년부터 본격 도입됐고, 현재는 자동차부품·패션·가구 업계 등 제품 표면 검사가 필요한 기업에서 도입하고 있다.

현재 머신비전협회에 등록된 기술·서비스 기업은 400여 개가 넘는다. 이 분야 주요 기업인 수아랩은 설립한 지 6년 만에 글로벌 비전검사 기업인 코그넥스(Cognex)에 약 2,300억 원에 인수돼 스타트업 엑시트(Exit)의 새로운 기록을 남겼다.

나. 생산장비 예지정비 분야

PdM(Predictive Maintenance)으로도 불린다. 이런 발상은 오래전부터 존재했고 GE, 지멘스 등 장비 제조사와 장비를 사용하는 대기업이 많이 시도하는 영역이지만 괄목할 만한 성공사례가 나오지 않는 영역이기도 하다. 주요 원인 중 하나는 장비 제조사와 장비 사용기업 간 데이터

협업이 필요한데, 제조업 특성상 생산 데이터의 기밀성 때문에 협업이 쉽지 않은 탓이다.

생산장비 데이터는 장비에 부착된 센서에서 생성되고, 해당 센서 체계를 잘 아는 주체는 장비제조 기업이다. 이중 어떤 데이터는 암호화돼 있어 장비 제조사만 해석할 수 있다. 반면 데이터 분석에 필요한 장비 정상 운영 데이터와 비정상 운영 데이터는 장비사용 기업이 갖고 있다. PdM 개발을 위해서는 장비 제조사와 사용기업의 데이터와 역량이 모두 필요하다.

하지만 장비 사용기업과 장비 제조사 간에 협업이 잘 이뤄지지 않는다. 장비 사용기업은 생산수율의 노출 등을 이유로 생산 데이터를 장비 제조사에 공개하는 것을 반대하는 경우가 발생하고, 장비 제조사도 장비의 센서 노하우 공개를 꺼리는 경우가 발생한다.

다. 공정최적화 분야

알파고에 적용된 강화학습 같은 계열의 기술을 활용해 생산장비를 운영할 때, 최적 운영값을 추천하거나 조립 공정에 사용되는 로봇팔의 최적의 움직임을 설정하는데 적용된다. 아직 초기 단계이며 실험적으로 테스트하고 있다.

라. 생산안전환경 분야

IoT 기술 도입 때부터 다양한 시도가 있었던 영역이며, AI 기술과 결합돼 더욱 다양한 분야에 적용되고 있다.

5. 제조업에 AI 주요 도입 사례

제조업은 생산 관련해 신기술 적용에 적극적인 초기 수용자(Early adopter)이지만, 생산 공정에 실제 적용하는 데에는 가장 보수적이다. 즉 생산성 향상 가능성이 있는 기술에 대해서는 R&D 프로젝트를 통해 다양하게 시험하지만, 실제 생산 공정에 적용할 때에는 검증은 반복하며 완벽한 기술을 적용하려고 한다. 왜냐하면 새로운 기술 적용이 생산수율과 안전에 영향을 주기 때문이다.

시장은 대부분 테스트 및 학습(Test & learn) 단계이고, 대부분 개념 정립(Proof of Concept, PoC), 파일럿(Pilot) 단계로 기술 검증 성격의 프로젝트가 대부분이다. 하지만 디지털 전환이 지속되면서 제조업의 다양한 분야에서 기술 채택이 일어나고 있다.

[표 4-3-5] 제조업 AI 적용사례 요약

적용분야	요구사항	이슈	사용 데이터	기대효과
설비 예지 정비	웨이퍼 생산장비의 예기치 못한 고장 발생 예측	장비가 복잡하고 다양한 레시피 때문에, 모든 고장유형을 사전에 알 수 없음	장비에 부착된 100개 이상의 센서 데이터	고장 전 12~48시간 전에 알람
	로봇팔의 예기치 못한 고장 예측	도장 작업에 사용되는 로봇팔이 제품 모델에 따라 동작하면서, 예방정비로도 잡지 못하는 예기치 못한 고장 발생	로봇팔의 각 축마다 부착된 모터의 전류값, 인코더 값(각도 측정), 진동센서값 등	로봇팔 고장 발생 4~6일 전 예측
공정 최적화	로봇팔을 활용한 용접 공정에서 불량 발생	다양한 자동차 모델을 단일 용접 설비에서 생산하기 때문에, 모델별 초기 설정값이 안 맞는 경우 발생	로봇팔 제어운영 값, 용접품질 정상, 레이블링 값	과다/과소 용접 사이의 최적 용접 설비 제어값 추천
	로봇팔 모션 최적화	제품 모델이 변경될 때마다 로봇팔 동선이 변경돼야 하는데, 옆의 로봇팔과의 동선을 고려해 최적의 모션을 설정해야 함	로봇팔 움직임 값, 조립 대상 제품 데이터(크기, 부품, 조립 포인트 등)	기존의 시행착오 (Trial&Error) 방식보다 모션 설정 속도가 빨라지고 정확도가 향상됨. 비용과 시간 절감
	화학산업 보일러 클로킹(막힘) 예측	Reboiler 공정에서 제품 생산과정에서 발생한 찌꺼기 때문에 클로킹 현상 발생	압력, 유량, 온도 등	클로킹 시점을 사전에 예측해 생산설비 운전 방식을 변경하고 클로킹 발생 시점을 늦춤
	전기자동차 배터리룸 최적 온도 유지	배터리 성능에 큰 영향을 주는 배터리룸 온도 최적유지를 위해 냉각팬, 에어컨로우의 자동 조정 필요	쿨링팬 속도, 외부 온도	배터리 성능 최적화, 배터리 수명 향상
	정밀화학 반응기 자동 운전	정해진 표준운전 조건이 있지만, 외부 환경에 따라 수동 조작이 필요함. 이 부분에서 인간에 의한 에러 발생함	반응기(Reactor)에서 발생하는 센서값 (용존산소량, 압력값 등), 외부 온도, 습도 등	휴먼에러(Human Error) 최소화로 수율 향상 및 안정화
제품 표면 검사	자동차 부품 표면에 발생한 미세 스크래치, 자국 등 검사	제품 검수 과정에서 몇 mm 수준의 미세한 스크래치는 발견되지 못하는 경우 발생	제품 표면 촬영한 이미지	불량제품 출하율 감소, 검사 인건비 감소

(계속)

적용분야	요구사항	이슈	사용 데이터	기대효과
신재생 에너지 분야	태양광 발전량 예측	태양광 발전량의 정확한 예측이 가능해지면, 전력계통 운영비용 절감(태양광 발전량이 부족하면, 전력 계통 운영을 위해 현물가격으로 다른 발전원 전력을 구매해야 함)	기상 예보, 관측데이터, 실제 발전량 값	발전회사는 정확한 예측에 대한 추가 정산금 확보 가능, 계통 운영기업은 비용 절감
	ESS 이상탐지	최근 발생한 ESS 화재 사고 같은 심각한 상황에 대한 예측 필요	ESS 배터리에서 발생한 전류, 전압값, 배터리 온도 값, ESS 운영데이터	사고 예방을 통한 ESS 운영 효율 증가

출처: 마키나릭스 자체 분석

6. 향후 제조업의 AI 적용 전망

가. 제조업과 AI기술, 클라우드의 결합

현재까지 제조기업은 퍼블릭 클라우드(Public Cloud)를 많이 도입하지 않았다. 관련 사고가 일어날 경우 수율과 안전이 큰 영향을 받기 때문에 초기 수용자 역할을 하고 있지 않다. 하지만 AWS, MS 애저 등 클라우드 사업 강자들은 외부에 있는 서버에 데이터를 저장하는 것을 실행하기 꺼리는 제조기업의 요구에 맞춰 프라이빗 클라우드(Private Cloud) 등을 앞세워 적극적으로 문을 두드리고 있다.

제조기업과 오랫동안 비즈니스 관계를 맺어 시스템 안정성, 장애 발생 시 빠른 복구 등 서비스 신뢰 관계가 굳건한 기존 HPE, 델 같은 서버 기업은 퍼블릭 클라우드 기업과 제휴하거나 직접 인수해 프라이빗 클라우드 방식으로 서비스를 제공하기 시작했다. 제조기업도 다른 산업 기업의 안정적인 퍼블릭 클라우드 도입 사례를 검토하면서 당장은 제조 데이터가 아닌, 판매 및 일반 관리 데이터를 대상으로 도입을 시도하고 있다. 이런 초기 테스트 및 학습(Test & Learn) 도입 사례가 성공으로 이어지면 제조 데이터까지 확대될 수 있다.

클라우드 환경이 구축되면, 제조 AI 도입과 관리가 온프레미스(on-premise)보다 용이해져 제조 시의 확산에 긍정적인 요인으로 작용할 수 있다. 예를 들어 제조 AI 솔루션 기업은 특정 업종 및 특정 사용목적(애플리케이션, 로봇팔 이상탐지 등)의 AI 솔루션을 클라우드 기업의 앱을 활용해 판매하고 유통시킬 수 있다.

일부 맞춤형 서비스를 별도 판매해 고객의 수요를 충족시킬 수 있고, 모델 관리를

원격으로 서비스할 수 있는 표준 환경이 클라우드를 통해 구축되기 때문에 서비스 공급기업의 고민을 줄여준다. 클라우드라는 표준 유통, 관리 환경이 구축되면 서비스 공급 원가가 감소한다. 이는 수요기업이 부담하는 가격 감소로 이어져 수요기업의 범위가 현재의 대기업 및 일부 중견기업 위주 시장에서 중소기업으로 확산될 수 있는 계기가 될 수 있다.

나. 운영 AI 중요성 증대

시장 초기 단계라 기술 도입 자체가 공급기업과 수요기업 모두에게 화두이지만, 도입이 어느 정도 이뤄진 시장 성장기에 접어들면 도입한 AI 운영이 점점 더 중요해진다.

AI 모델은 과거 데이터로 개발되므로 제조 공정에서 나오는 운영 데이터가 과거와 많이 다르면 모델 성능이 떨어진다. 문제는 대부분의 제조기업이 원재료, 공정, 제품 등에서 변화가 생기기 때문에 모델 성능을 유지하기 쉽지 않다. 예를 들어 완성차 기업은 신규 자동차 모델이 매년 나오고 출시 주기가 점점 더 짧아지고 있다. 이렇게 제품이 달라지면 해당 제품을 제조하는 공정이 달라진다. 차체 크기가 달라지면 로봇팔 움직임이 달라져야 하고, 제품 표면 검사 AI 모델이 달라져야 한다.

또 다른 예로 반도체 제조 공정을 설명하면, 반도체 장비는 복잡성이 높아 예기치 못한 고장이나 고장은 아니더라도 장비의 이상동작이 발생하는 경우가 종종 있다. 장비를 유지보수하게 되면 장비에서 나오는 센서 데이터 단위가 유지보수 전 데이터와 비교할 때 많이 달라질 수 있다. 유지보수 전 데이터로 개발된 모델은 이 데이터가 입력되면 이상 알람을 울리게 된다.

제조 환경의 변화로 모델 변경이 필요할 때마다 모델을 처음 도입한 때와 비슷한 수준의 노력, 시간, 비용을 들여 개발해야 한다면 AI를 도입하려는 기업에게는 상당한 부담이 될 것이다. 개발 노하우가 축적되면 처음 도입한 때보다는 부담이 줄 수 있지만, 환경 적응 기술이 필요하다.

지속적 학습, 평생학습 AI 기술이 최근에 화두가 된 데는 이런 배경에서 비롯된 것이다. 이 기술의 주요 목적은 새로운 데이터 패턴이 등장했을 때 빠르게 적응하는 것이다. 예를 들어, 새로운 데이터 패턴마다 작은 경량 모델을 만들어 데이터베이스화하고, 데이터베이스에 있는 모델에 정상, 비정상 여부를 확인한다. 이 과정에서 새로운 패턴을 갖는 최신 데이터뿐만 아니라 적절히 정규화된 과거 데이터를 함께 사용해 모델을 학습해 새로운 패턴에 대한 대응 시간을 최소화할 수 있다.

모델을 재교육해 업데이트하는 등 모델 관리를 손쉽게 할 수 있는 기능과 지속적 학습 기술까지 포함된 운영 AI 플랫폼의 중요성이 높아질 전망이다. 현재 일부 AI 기업들이 운영 AI 플랫폼 개발에 박차를 가하고 있다.

제4장. 통신 분야 데이터 활용 비즈니스 현황

필자: 김정선
(SK텔레콤 T3K/박사)

통신 비즈니스 영역은 개인(Person)에서 사물(IoT)로, 사물에서 사업을 이어주는 플랫폼으로 확장하고 있다. 기술 혁신을 통해 기업과 사회의 혁신이 시도되는 실제 사례가 코로나19가 가져온 삶의 방식 변화와 맞물려 구현됐다. 이러한 삶의 변화는 통신 데이터가 가진 다양한 적용성을 증명하는 기반을 제공했다. 비록 통신사의 데이터를 활용하기 위한 여러 가지 사회적·제도적 제약이 존재하지만, 데이터 중심의 기회와 비전은 매우 클 것으로 보인다. 통신 사업자의 기존의 비즈니스 핵심이 네트워크였다면 미래 시장의 핵심은 데이터다. 특히 5G 환경으로 변화될 서비스에서 발생되고 수집되는 데이터는 통신 회사의 미래 가치일 것이다.

1. 국내외 통신시장 현황

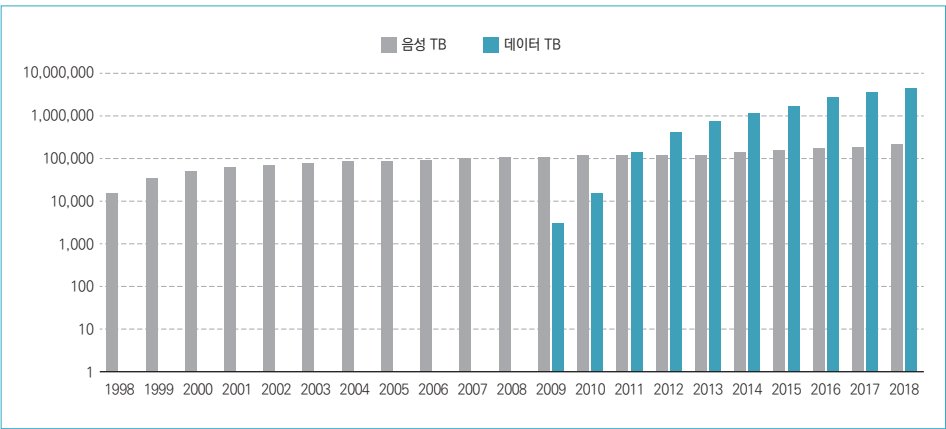
국내외 통신시장은 기술 세대변화에 따른 진화를 지속하는 중이다. 2020년에 2G 서비스는 역사 속으로 사라졌고 세계 최초로 5G 서비스를 시작했다. 이렇게 통신 서비스에 대한 수용도가 높아지면서 가입자와 트래픽이 급격히 증가하고 있다.

통신 분야에서 트래픽이란 통신망에서 전송되는 데이터 양을 의미한다. 모바일 트래픽(Mobile traffic)은 무선통신에서 전파를 이용해 전송되는 정보량이라 할 수 있으며⁰¹ 이러한 트래픽은 개인화된 데이터의 흐름이라고 할 수 있다.

Omdia에 따르면 국내 통신가입자는 2025년까지 89%의 가입율(subscription rate)을 보일 것으로 전망된다. 현재 3만 원대를 유지하는 유저당 평균매출(Average Revenue Per User, ARPU)는

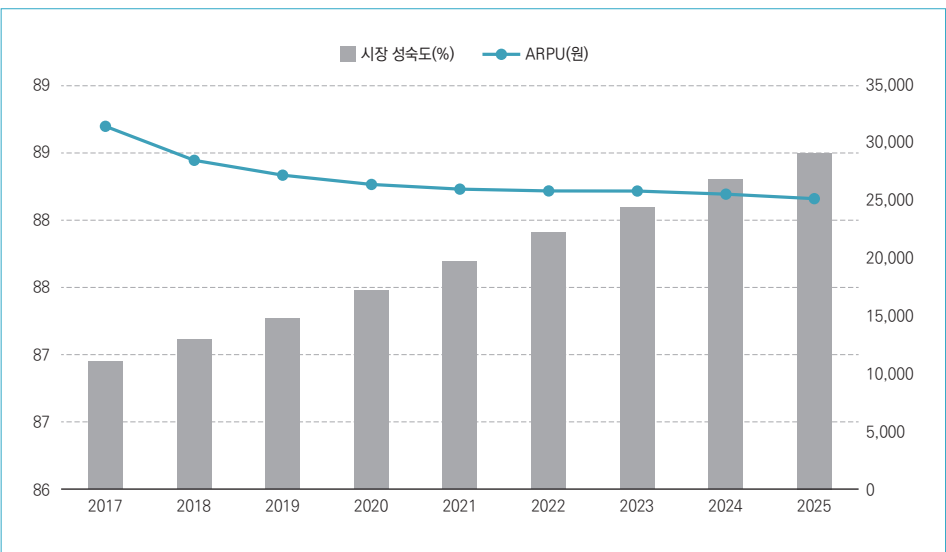
01. 장재혁·박승근, 「모바일 트래픽 동향」, 2019.

[그림 4-4-1] 모바일 트래픽 추이 (단위: TB/년도, 로그 스케일)



출처: 「모바일 트래픽 동향」, 2019.

[그림 4-4-2] 국내 이동통신 가입자 및 ARPU 전망



출처: Omdia

2만 5,000원대로 감소할 것으로 예측하고 있다.⁰² 현재 통신시장은 새로운 비즈니스의 전환과 모색이 필요한 시점이며, 전환을 위한 핵심 기제로 데이터의 가치가 새롭게 부각되고 있다.

2020년 통신시장에 큰 변화를 가져온 주요 변인을 네 가지 꼽을 수 있다. 우선 국제적 관점에서의 두 가지는 전 세계적으로 코로나19 확산과 디지털 전환의 가속화다. 글로벌

02. Mobile Subscription and Revenue Forecast: 2020-25, Omdia, 2020.7.

통신사업자들은 포스트 코로나 시대 5G(5세대) 이동통신의 역할을 강조하고 나서고 있다. 이는 막대한 인프라 투자를 새로운 수익원으로 연결시키기 위한 설득으로 보인다.⁰³ 2020년 7월 온라인으로 진행된 GSMA 쓰라이브(GSMA Thrive) 행사에서 귀 핑 화웨이 순환 회장은 “5G가 팬데믹 이후 디지털 전환을 촉발시킬 것”이라고 강조했다. 코로나19 사태 당시 5G, AI, 클라우드, 빅데이터 기반으로 개발된 애플리케이션 덕분에 사람들은 가정에서 ‘사회적 거리두기’를 할 수 있었다는 것이다.

이처럼 디지털 전환은 2020년에 실제적인 전환기를 맞았다. 기술 혁신을 통해 기업과 사회의 혁신이 비롯되는 실제 사례가 코로나19가 가져온 삶의 방식 변화를 통해 구현됐다. 온라인 수업과 원격진료, 재택근무와 비대면 멀티 스트리밍 공연 등 개인과 집단의 삶의 방식이 완전히 바뀌고 있다. 그 바탕에는 통신과 통신활동으로부터 생성된 트래픽(데이터의 흐름)이 있다.

국내 상황으로 관점을 돌려보면 통신 데이터 비즈니스 영역에 가장 큰 변화를 가져온 사건은 ‘데이터3법’의 통과다. 개정된 법안에 따르면, 통계 작성, 과학적 연구, 공익적 기록 보존 등을 위해 정보 주체의 동의 없이 가명정보 사용이 가능해졌다. 또한 제3자에게 가명정보의 제공 및 이종 데이터 간의 결합이 가능해졌다.⁰⁴ 아울러 신용정보보호법에 따라 본인신용정보 관리업 및 비금융전문 CB, 개인사업자 CB 등 새로운 사업 영역이 등장하게 됐다.

마지막으로는 정부의 디지털뉴딜 정책이다. 디지털뉴딜 정책은 코로나19로 최악의 경기 침체와 일자리 충격 등 어려움에 직면한 상황에서 위기를 극복하고 코로나19 이후 글로벌 경제를 선도하기 위해 마련된 국가발전 전략으로, 정부는 3대 프로젝트와 10대 중점 추진과제를 담아 그 추진방향을 발표했다.⁰⁵ 한국판 뉴딜의 핵심은 ‘데이터 댐’으로 꼽을 수 있다. 인공지능 학습용 데이터 수집·가공에서부터 의료, 교육, 제조 분야까지 새로운 비즈니스 기회를 데이터 분야에서 찾겠다는 것이다. 아울러 인공지능의 고도화도 데이터 댐으로 달성한다는 계획이다.

이러한 국내 동향에 맞추어 통신 사업자들은 디지털 뉴딜 TF를 발족하거나, O2O 마케팅 플랫폼과 언택트 보안 솔루션 등 비대면 사업 강화 계획을 발표하는 등 정부 시책에 부응하기 위한 다양한 전략을 수립, 실행하고 있다.

03. 류은주 기자, 「포스트 코로나=5G 시대」 역설하는 글로벌 통신업계, IT조선, 2020.7.3. http://it.chosun.com/site/data/html_dir/2020/07/02/2020070203746.html

04. 김정선, 「가명 데이터 활용연구-기술적 처리방법 및 기업의 활용방향을 중심으로」 정보보호학회 논문지 30권 2호, 253-261, 2020.

05. 「한국판 뉴딜 종합계획」, 관계부처합동, 2021.7.14.

2. 통신분야 데이터 비즈니스 동향

가. 통신분야 데이터 비즈니스 현황

현재 통신사가 처한 환경은 그리 녹록치 않다. 전 세계 이동통신 보급률이 104%에 달해 정체기에 진입했고, 경쟁이 심화되면서 유저당 평균매출액(ARPU)도 하락 추세에 직면하고 있다. 이동통신 기술이 보편화됨에 따라 통신품질의 격차가 해소되고 있으며, 경쟁기업들의 모방이 가능함에 따라 서비스 차별화도 불가능한 형편이다.

그러므로 각국의 통신사들은 비즈니스 차별화를 위해 노력을 다하고 있다. 미국 4위 이동통신 사업자인 T 모바일은 기존과 차별화되는 ‘Un-carrier’ 캠페인 및 탈통신사 전략 등을 통해 3위 사업자인 스프린트를 추월했다. 기존 사업자들이 약정제도를 통해 보조금을 지급하며 가입자를 유치한 반면에 T 모바일은 저렴한 요금제 출시, 단말기 업그레이드 프로그램, 무제한 동영상과 음악 스트리밍 서비스 등을 제공했다. 이후 선발 사업자인 AT&T, 버라이존, 스프린트가 유사한 전략을 도입하면서 소모적인 보조금 경쟁보다 파격적인 서비스를 통한 경쟁이 격화되고 있는 상황이다. 또한 타 산업과의 인수합병을 통해 시너지 창출을 도모하는 움직임도 발생하고 있다. AT&T는 미디어업체 인수를 통해 콘텐츠 사업을 강화했으며, 버라이존도 디지털 미디어, 광고, 사물인터넷 기업의 인수를 확대하는 중이다. AT&T는 DirecTV(디지털 위성방송) 및 타임워너(CNN, HBO, 워너브라더스 등 보유) 등을 인수하며 방송 콘텐츠를 확장(비중 37%)하고 있다.

스페인의 텔레포니카(Telefonica)는 위성방송사업자인 Canal Plus를 인수했으며, 프랑스 케이블사업자 Numeri-cable은 통신기업인 SFR을 인수했다. 프랑스의 Orange는 금융기관 Groupama Banque를 인수하는 등 비즈니스 간 융복합이 다양하게 일어나고 있다.

국내에서도 KT스카이라이프가 현대 HCN을 인수했고, SKT와 LG유플러스의 경우도 유료방송 시장의 인수합병을 지속적으로 모색 중이다. 또한 SKT는 한화손해보험, 현대자동차 등과 캐롯손해보험을 설립하고 데이터 기반의 특화 보험을 출시 중이다. SKT는 모바일 광고대행사 인크로스 지분을 인수하고 고객특화 혜택을 제공하는 ‘T-deal’ 서비스를 출시했다.

이렇듯 이동통신사, 미디어, 방송, 금융서비스 간 융합은 최근 통신 비즈니스 영역의 보편적 움직임이다. 그런데 이런 합종연횡의 움직임 근간에는 데이터가 있다. 다양한 사업 영역에서 발생된 가입자의 데이터를 수집하고 개인화된 관점의 분석을 통해 새로운 인사이트를 도출할 수 있다. 지금까지 데이터 이용과 활용이 분절된 해당 산업의 카테고리 안에서 이루어졌다면, 이제부터는 개인을 중심으로 한 데이터의 횡적 연결을 통해 말 그대로 360도의 라이프로그 데이터 획득이 가능한 것으로 변모하게 됐다.

[그림 4-4-3] 2019년 디지털 광고 지출이 최소 50%에 달하는 국가⁰²

Segment	2017	2018	2019	2020	2021	2022	2023	2024	2025
Total mobile connectivity subscriptions	63568	66130	68712	71896	74169	76035	77564	79046	80388
<i>Analog</i>	-	-	-	-	-	-	-	-	-
<i>1xRTT</i>	2435	1699	1019	537	219	162	110	84	67
<i>GSM</i>	-	-	-	-	-	-	-	-	-
<i>iDEN</i>	175	-	-	-	-	-	-	-	-
<i>PHS</i>	-	-	-	-	-	-	-	-	-
<i>US TDMA</i>	-	-	-	-	-	-	-	-	-
<i>1xEV-DO</i>	-	-	-	-	-	-	-	-	-
<i>TD-SCDMA</i>	-	-	-	-	-	-	-	-	-
<i>W-CDMA</i>	10493	9327	7429	6006	5073	4526	3912	3533	3080
<i>LTE</i>	50465	55103	55596	52496	49342	43539	37152	32069	28269
<i>5G</i>	-	-	4668	12858	19535	27808	36390	43360	48973
<i>1G</i>	-	-	-	-	-	-	-	-	-
<i>2G</i>	2609	1699	1019	537	219	162	110	84	67
<i>3G</i>	10493	9327	7429	6006	5073	4526	3912	3533	3080
<i>4G</i>	50465	55103	55596	52496	49342	43539	37152	32069	28269
<i>5G</i>	-	-	4668	12858	19535	27808	36390	43360	48973
<i>1G % of total subscriptions</i>	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
<i>2G % of total subscriptions</i>	4.1%	2.6%	1.5%	0.7%	0.3%	0.2%	0.1%	0.1%	0.1%
<i>3G % of total subscriptions</i>	16.5%	14.1%	10.8%	8.4%	6.8%	6.0%	5.0%	4.5%	3.8%
<i>4G % of total subscriptions</i>	79.4%	83.3%	80.9%	73.0%	66.5%	57.3%	47.9%	40.6%	35.2%
<i>5G % of total subscriptions</i>	0.0%	0.0%	6.8%	17.9%	26.3%	36.6%	46.9%	54.9%	60.9%
Prepaid	3529	3925	3820	3917	3959	3975	3970	3960	3939
Postpaid	60039	62200	64892	67978	70210	72059	73594	75086	76449
Prepaid % of su bscriptions	5.6%	5.9%	5.6%	5.4%	5.3%	5.2%	5.1%	5.0%	4.9%
Postpaid % of su bscriptions	94.4%	94.1%	94.4%	94.6%	94.7%	94.8%	94.9%	95.0%	95.1%
Unique users	44088	44347	44582	44841	45091	45332	45564	45790	46008
Unique users - Population penetration	87%	87%	87%	87%	88%	88%	88%	88%	89%
Service revenues	19964	18872	18661	18978	19346	19749	20053	20893	20504
<i>Non-SMS</i>	10227	11202	11859	12954	13820	14777	15606	16326	16965
<i>Mobile voice</i>	7459	6066	5390	5204	4640	4297	3948	3606	3281
<i>SMS</i>	2278	1062	1412	1146	886	674	498	361	258
ARPU	26.66	24.25	23.06	22.50	22.07	21.91	21.76	21.60	21.43
<i>Non-SMS</i>	13.66	14.40	14.66	15.36	15.77	16.40	16.93	17.37	17.73
<i>Mobile voice</i>	9.96	7.80	6.66	6.17	5.29	4.77	4.28	3.84	3.43
<i>SMS</i>	3.04	2.06	1.75	1.36	1.01	0.75	0.54	0.38	0.27

출처: Omdia

통신 사업자들의 기존 비즈니스 핵심이 네트워크였다면 앞으로 다가올 미래 시장의 핵심은 데이터다. 특히 5G 환경으로 변화될 서비스에서 발생되고 수집되는 데이터는 통신회사가 바라보는 미래 가치다.

개정 발의된 '독점 규제 및 공정거래에 관한 법률 개정안'에 따르면 현재는 금융업으로 분류돼 설립할 수 없는 기업형 벤처캐피탈(CVC)을 허용하게 함으로써 대기업의 스타트업 기업에 대한 투자와 그에 따른 지분 소유가 가능해진다.⁰⁸

국내 CVC 설립허가를 전제로 앞으로 미래가치가 높은 스타트업의 발굴은 데이터를 얼마나 잘 활용해 혁신적인 서비스를 발굴 가능한가의 역량으로 평가받을 것이다. 특히나 통신사가 제공하는 다양한 데이터를 기반으로 서비스를 제공한다면 기존 더미 파이프(dummy pipe)화를 방지하면서도 가치 있는 기업을 발굴해 동반 성장을 위한 비즈니스 생태계 모델 설계가 가능하다. 일본의 소프트뱅크는 로봇, AI 등 미래 산업에 대해 장기적 관점으로 투자 중이며, 기업형 벤처캐피탈(CVC)을 통한 스타트업 기업 투자를 진행 중이다. 또한 버라이즌이 설립한 CVC의 투자 포트폴리오를 보면, 기존 통신사업에 25%, 통신 연관사업에 약 40%, 그 외 새로운 사업에 약 35%를 투자 중인 것으로 알려져 있다.

스타트업은 자본 확보를 통해 기업 성장에 대한 동력을 확보할 수 있으며, 글로벌 기업으로부터의 투자유치 등으로 시장에 신뢰감을 줄 수 있는 장점이 있다. 반면 통신사는 스타트업 기업의 혁신성을 활용함으로써 급변하는 환경에서 빠른 적응이 필요한 데이터 기반 신규 사업으로의 진출이 가능해질 것이다.

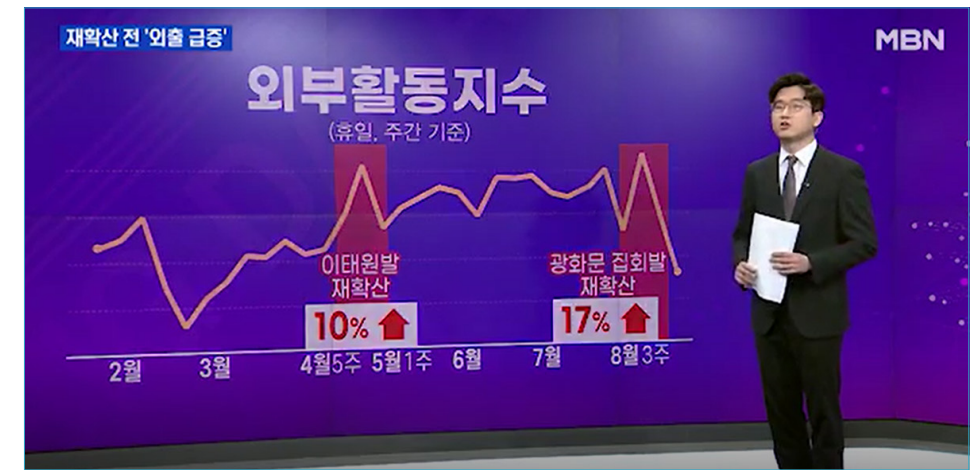
나. 통신분야 데이터 활용 비즈니스 사례

통신사의 데이터 기반 적용 사례는 매우 다양한데, 특히 올해는 코로나19 상황에서 통신사의 데이터를 활용한 다양한 서비스들이 선보여졌다.

KT는 과학기술정보통신부와 함께 코로나19 확산예측 모델을 개발하는 연구기관과 '코로나19 확산예측 연구 얼라이언스'를 구축했다. 해당 연구는 '감염병 해외유입 위험도 예측'과 '국내 확산지역 예측'의 두 가지 주제로 진행됐다. 이번 연구는 여러 연구기관이 참여해 국내 감염병 확산 연구에 기여했다는 평가를 받았다.

SKT는 경북경찰청과 함께 SKT의 공간통계 데이터를 활용해 인구 밀집도가 높은 지역을 실시간으로 확인하고 위험도가 높은 지역을 대상으로 실시간 '핀 포인트' 순찰을 실시함으로써 코로나19 확산을 방지하고 주민의 체감안전도를 높이고 있다. 그리고 유통소비 빅데이터 플랫폼을 통해 전국의 주요 시군구 단위 재택지수, 여행지수를 제공해

[그림 4-4-4] 통신사 데이터를 활용한 코로나19 확산의 징후 방송사례



출처: 「코로나19, 재확산 직전에는 '외출 급증' 있었다」, MBN, 2020.9.1.

코로나19 상황에 대한 모니터링에 기여하고 있다. 또한 수도권 지하철 객실 한 칸 단위별 혼잡도를 제공해 시민들이 혼잡한 시간과 노선을 피해 코로나19 감염 위험도를 낮출 수 있도록 도움을 제공하고 있다.

이처럼 코로나19를 계기로 통신사의 빅데이터 연구와 활용이 탄력을 받게 됐다. 확진자 추적 과정에서 일부 확인한 빅데이터의 힘을 전염병 예측과 확산 방지까지 활용할 수 있는 사례가 만들어진 것이다. 코로나19가 그간 온갖 규제로 억눌린 한국 데이터산업과 경제를 활성화하는 계기가 됐다. 향후 다양한 감염병 발생 시 현재의 연구 및 서비스를 바탕으로 더욱 고도화된 서비스가 가능해질 것이다.

통신사의 데이터 활용방식은 아직까지는 실질적 재무성과 창출보다는 기업 운영 프로세스 향상 등 '비금전적 효과(soft benefits)'에 주력하는 모습이다. 인공지능이나 빅데이터 분석이 기업의 재정적 수익 향상을 위한 도구로 널리 활용되고 있으나, 실제 효과에 대해서는 여러 가지 의견이 상충되는 상황이다. 특히 통신사는 대규모 가입자 기반의 비즈니스 특성상 개인정보 등 데이터의 대외 사업화에 따라 발생 가능한 위험에 대해 자유롭지 못한 상황이며, 기간통신 사업자로서의 데이터 활용이 마치 공공재처럼 시장 내 이해되는 분위기가 있어 데이터 비즈니스화에 적극적이기 어려운 상황이다. 기업들의 투자는 일반적으로 기술 도입이 가져오는 직접적인 수익 효과를 기반으로 이루어지는데, 아직까지 빅데이터 분석 관련 기술은 수익에 직접적인 영향을 미치는 정도는 아니므로 해당 사업 및 기술에의 투자가 적극적인 상황은 아니다.

이러한 가운데에서도 빅데이터 및 인공지능 기반 도구들이 성과 향상 효과를 창출하고

08. 장서연 기자, 「기업형 벤처 캐피탈 설립허가, 어떤 효과가 있을까?」, THE DAILYPOST, 2020.7.2. 참조

환경 변화에 민첩하게 적응할 수 있도록 지원하는 등 금전적 이점만으로는 평가할 수 없는 다양한 이점을 제공한다는 평가가 제기돼 주목할 만하다.

이와 같은 사례는 단순히 수익 증대나 비용 절감 등과 같은 재정적 효과보다는 고객 만족 및 인재 채용 프로세스를 향상시키고, 기업에서 운영 중인 장비의 고장 시기나 공급망을 예측하는 등의 부수적인 '비금전적 효과'를 가져오는 것으로 평가되고 있다.

특히 자연어 처리(Natural Language Processing, NLP) 기술을 활용해 고객이 서비스 담당자와 상담하는 동안 실시간으로 고객의 감정을 감지하도록 하거나 더 많은 고객 정보를 획득하기 위한 데이터 활용 등의 노력은 고객 측면에서는 서비스 만족도 향상을 가능하게 할 수 있다.

이러한 데이터 활용 노력은 수익으로 정량화돼 평가할 수는 없지만 바이럴 마케팅이나 소셜 미디어를 통한 브랜드 가치나 기업 평판 향상에 도움을 줄 수 있다. 또한 인재 채용 프로세스에서도 인공지능 및 자연어 처리를 통해 이력서의 키워드를 식별하고 소셜 미디어 분석을 통해 후보자를 선별하는 방식으로 구직자 채용 과정을 개선한 사례가 있다. 이러한 과정을 통해 혁신과 신규 수익 창출이 가능한 제품에 의존하는 기술 산업 분야에서는 역량 있는 인재의 채용 과정 역시 차별화 요소로 작용할 수 있다. 기지국 장비의 수명 주기 예측을 위해 센서와 같은 장비의 고장 시기를 예측하기도 하는데 이를 통해 고장 전에 유지 보수 작업이 가능하므로 관리주기에 효율성을 꾀할 수도 있을 것이다.

다. 통신분야 데이터 비즈니스 활성화의 제약점

통신사의 데이터는 방대하다. 전 국민이 통신사의 비즈니스 대상이다. 이에 따라 기업 비즈니스의 변화 속도에 맞게 데이터를 운영하는 것이 쉽지 않다. 이러한 측면에서도 그렇고 활용을 위한 정제나, 관리의 측면에서도 데이터 운영은 쉬운 문제가 아니다. 빅데이터와 관련된 기술은 매우 빨리 발전하고, 이를 이용하는 사람들의 요구는 다양하기 때문에 해당 구조를 반영하는 문제도 쉽지 않다. 더불어 점점 더 첨예해지는 보안 이슈와 정부 규제가 증가하고 있는 어려움이 존재한다.

우리나라는 세계적 수준의 5G 네트워크와 ICT 인프라를 보유하고 있는데도 데이터 비즈니스가 선진국에 비해 활성화되지 않고 있다. 주요 요인은 가입자 기반의 서비스이므로 데이터를 활용해 얻는 득보다 실이 클 수 있다는 경영진의 우려와 데이터 활용을 위해 살펴봐야 하는 과도한 규제 요인들 때문이다. 데이터3법 개정을 통해 일부 데이터 활용을 위한 물꼬가 볼 수 있다. 하지만 실제 데이터를 사업적으로 활용하는 데 있어 통신사가 살펴야 하는 규제요소가 너무나 많다. 관련 법만 하더라도 데이터3법 외 전기통신사업자법, 위치정보 보호법(위치정보의 보호 및 이용 등에 관한 법률), 통신비밀보호법, 지적재산권보호법, 공정거래법(독점규제 및 공정거래에 관한 법률) 등 유관한 법제도간의 해석적 충돌을 살펴야 한다. 특히 신규

융복합 산업영역에서는 해당 유관 법령까지도 살펴보아야 하는 것뿐만 아니라 규제기관의 관리감독을 감당해야 한다.

기업용 데이터 클라우드 기업인 클라우데라의 앤드루 살티스(Andrew Psaltis) 최고 기술책임자(CTO)는 “코로나19 사태를 계기로 기업 비즈니스 영역에서 데이터의 중요성이 이전보다 더욱 부각되기 시작했다는 것은 흥미롭다”며 “코로나19 여파가 비즈니스에 미칠 악영향을 최소화하려면 데이터 활용 방식과 영역을 적극적으로 확장할 필요가 있다”고 강조했다.⁰⁹

팬데믹을 맞아 공공 분야와 비즈니스 영역에서 통신 데이터를 활용하려는 노력은 급격히 늘어났다. 하지만 아직 그 수요가 불충분하거나 수익 측면의 접근 방법이 적절하지 않은 것도 사실이다.

데이터의 가치를 공공재로 규정하는 사회적 분위기도 개선돼야 한다. 데이터의 가치를 정당하게 산정하고 이를 기반으로 한 부가가치의 가치사슬이 정상적으로 작동해야 데이터 비즈니스가 활성화될 수 있을 것이다.

3. 향후 전망

데이터 기반의 혁신을 추구하는 이유는 불가능할 것 같던 일을 가능하게 만들기 때문이다. ‘측정할 수 없으면 관리할 수 없고, 관리할 수 없으면 개선할 수 없다’는 피터 드러커의 말처럼 데이터는 기존에 개선할 수 없었던 영역이나 새로운 아이디어를 가능하게 만든다.

전 국민 규모의 통신 데이터를 더욱 정교하게 만들어 새로운 상품과 서비스를 출시하고, 시장을 선점하기 위해 이종 간 데이터의 결합 활용은 더욱 가속화될 전망이다. 이종 산업의 데이터와의 결합사유는 데이터가 정교해지고 연결될수록 활용도가 확장되고 가치가 증가하기 때문이다. 이를 기반으로 새로운 상품 서비스를 선보일 수 있고, 현재의 포화한 통신시장에서 벗어나 비로소 데이터 기반의 신규 시장으로의 진입이 가능해질 것이다. 기존까지는 통신사의 비즈니스가 네트워크를 통한 개인의 연결성을 제공하는 것이었다면 향후 통신사의 비즈니스는 이를 뛰어넘어 개인, 사물을 연결하면서 발생된 데이터를 통해 비즈니스와 비즈니스를 연결하는 구조로 변모될 것이다. 그 구조를 실현하기 위해서는 클라우드와 모바일 엣지 컴퓨팅(MEC) 등 통신사가 지닌 역량이 경쟁요소가 될 것으로

09. <https://www.mk.co.kr/news/business/view/2020/05/519003/> 참조

기대한다.

데이터는 기업의 ‘라이프 사이클’ 안에서 무한히 생산되고 확장되며 데이터 기반 디지털 전환의 원동력이 될 것이다. 다가올 미래 시장의 핵심은 데이터다. 특히 5세대 통신 환경으로 변화될 서비스에서 발생되고 수집되는 데이터와 이를 연결해주는 플랫폼은 통신회사가 바라보는 미래 가치다.

제5장. 광고 분야 데이터 활용 비즈니스 현황

필자: 안익진
(물로코 대표)

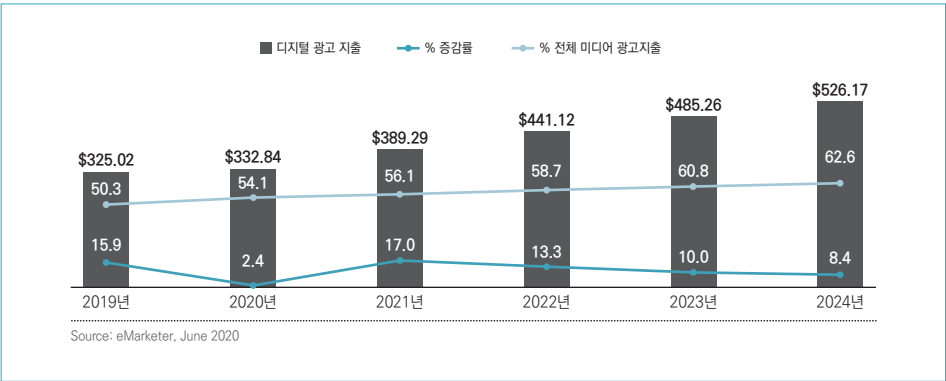
광고는 개인데이터의 사업 가치를 가장 먼저, 그리고 광범위하게 찾은 영역이다. 구글과 페이스북 모두 수익의 80% 이상이 광고 비즈니스에서 나오고 있고, 아마존 역시 10조 원 이상의 광고 매출을 목전에 두고 있다. 구글, 페이스북 등이 이러한 사업적인 성공을 바탕으로 유튜브, 인스타그램과 같은 추가 미디어를 확보해 정보와 미디어의 쏠림 현상은 가속화되고 있다. 광고 집행 방식 역시 전통적인 지면 위주 미디어 구매에서 데이터 기반 타기팅 위주의 퍼포먼스 구매로 빠르게 진화하고 있다. 한편 시장 확대의 반대 급부로 일반 사용자들 사이에선 개인데이터 사용과 프라이버시 침해에 대한 우려가 높아지고 있다. 이에 EU의 GDPR을 시작으로 미국 캘리포니아의 CCPA 등 각 국은 광고 데이터 사용에 사용자들의 권리를 보장하는 방법을 강구하고 있다.

1. 개인데이터를 기반으로 한 광고 사업의 성장

가. 디지털 광고 시장

디지털 광고는 인터넷이나 온라인을 기반으로 PC나 모바일 등의 기기를 통해 상품이나 서비스를 홍보하기 위해 송출되는 모든 광고를 의미한다. 기술 발전에 맞춰 디지털 광고 시장은 최근 몇 년간 기하급수적으로 성장했다. 2015년에만 해도 새롭게 시도해볼 만한 광고 기법으로 거론되던 디지털 마케팅은 2019년 약 376조 원(3,250억 2,000만 달러)의 자본을 끌어들이는 시장이 되었고, 2020년에는 약 385조 원(3,328억 4,000만 달러) 규모로 성장할 것으로 예측된다. 중국, 영국, 노르웨이, 미국, 호주, 캐나다를 포함한 많은 주요 경제국들은 꽤 오래전부터 전체 미디어 광고의 절반 이상을 디지털 광고에 지출하고 있다.

[그림 4-5-1] 2019~2024년 전 세계 디지털 광고비(단위: 10억 달러)⁰¹



[그림 4-5-2] 2019년 디지털 광고 지출이 최소 50%에 달하는 국가⁰²

	2018	2019
중국	65.3%	69.5%
영국	63.8%	66.4%
노르웨이	61.7%	65.5%
아일랜드	58.8%	62.6%
덴마크	57.8%	61.1%
스웨덴	56.3%	60.3%
호주	55.6%	57.1%
미국	48.6%	54.2%
뉴질랜드	51.8%	54.0%
캐나다	50.2%	53.5%
네덜란드	47.8%	52.6%
러시아	45.1%	50.0%

Note: includes advertising that appears on desktop and laptop computers as well as mobile phones, tablets and other internet-connected devices, and includes all various formats of advertising on those platforms; excludes SMS, MMS and messaging on those platforms; excludes SMS, MMS and P2P messaging-based advertising; *excludes Hong Kong
Source: eMarketer, February 2019

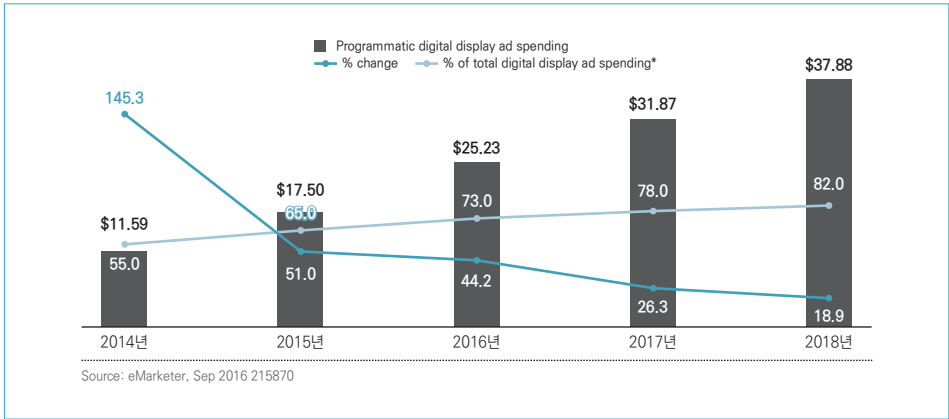
이런 흐름이 유지된다면 디지털 광고는 2023년에 전체 광고 시장의 60% 이상을 차지할 것으로 전망된다. 대부분의 유럽 국가들이 이 흐름에 함께할 것이며 아직 비중이 적은 중남미와 동남아시아 일부 지역의 개발도상국이 가까운 미래에 합세할 것으로 예상된다.

01. Global Digital Ad Spending Update Q2 2020, eMarketer, 2020.7.6.
02. Global digital Ad spending 2019, eMarketer, 2019.5.28.

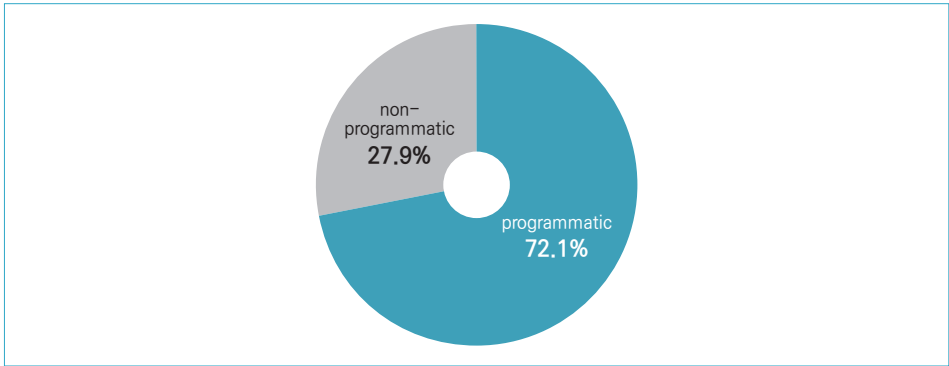
나. 프로그래매틱 광고 시장

디지털 광고 중에서도 프로그래매틱 광고(Programmatic Advertising) 분야의 성장세가 도드라진다. 프로그래매틱 광고는 빅데이터와 머신러닝 등 디지털 기술을 활용해 사용자 데이터를 분석하고 이를 통해 파악한 관심사를 기반으로 광고를 집행하는 기법으로, 데이터의 사업 가치를 극대화할 수 있는 마케팅 방식이다. 프로그래매틱 광고는 미국 전체 디지털 광고 시장의 82%, 유럽의 경우 약 75%의 점유율을 차지한다.

[그림 4-5-3] 2014~2018년 미국 프로그래매틱 디지털 광고비(단위: 10억 달러)⁰³



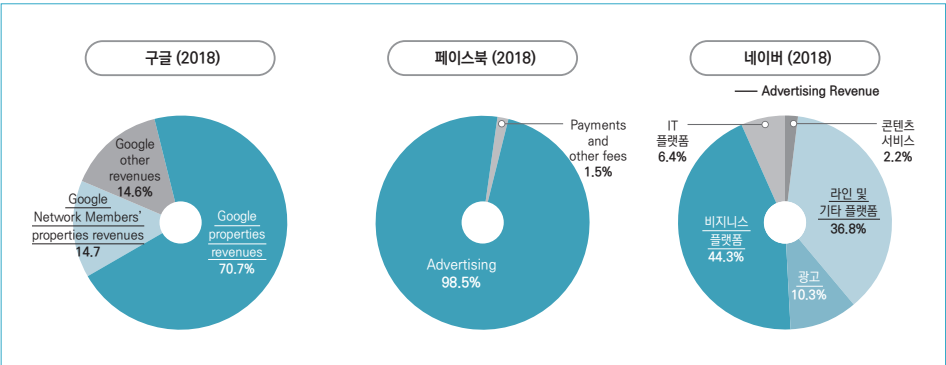
[그림 4-5-4] 2018년 유럽 전체 디지털 광고 대비 프로그래매틱 광고비 비중⁰⁴



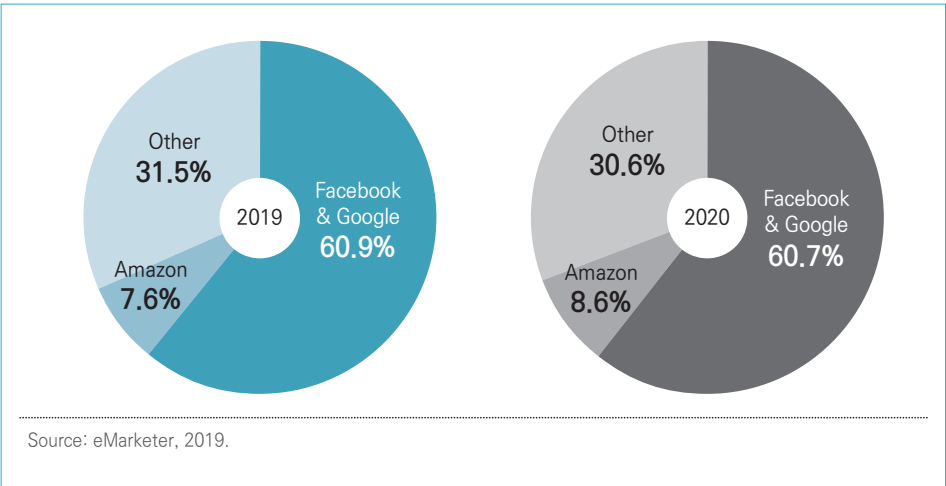
03. US Programmatic Digital Display Ad Spending 2014-2018, eMarketer, 2018.
04. Programmatic Ad spend in Europe, IAB, 2019.9.

현재 전체 프로그래매틱 광고 시장의 절반 이상을 구글과 페이스북이 장악하고 있다. 구글 전체 매출의 80% 이상이 광고 사업에서 나오고 있으며, 페이스북의 경우 전체 매출의 98.5%가 광고 사업이다. 국내의 경우, 네이버의 전체 매출의 50% 이상이 광고를 통해 들어오고 있다. 최근에는 세계 최대 유통 플랫폼인 아마존이 합세해 디지털 시장의 양대 산맥에 도전장을 내밀었다.

[그림 4-5-5] 구글, 페이스북, 네이버의 기업별 전체 매출 대비 광고 매출의 비중⁰⁵



[그림 4-5-6] 2019, 2020 구글과 페이스북 vs. 아마존의 디지털 광고 점유율⁰⁶



05. Google, Facebook, Naver 2018 IR 자료, 물로코 재구성

06. Facebook-Google Duopoly Won't Crack this Year, eMarketer, 2019.11.4. <https://www.emarketer.com/content/facebook-google-duopoly-won-t-crack-this-year>

다. 실시간 입찰(RTB) 기술 시장

프로그래매틱 광고 중에서도 RTB(Real Time Bidding)는 인간의 개입을 최소화하고 더 큰 규모의 데이터 처리와 광고 거래를 가능하게 하는 기술 기반의 광고 거래 방식이다. RTB는 실시간으로 광고 지면에 접근하는 사용자 데이터를 분석해 광고 지면의 값을 지정하고, 경매를 통해 광고를 거래한다.

지난 5년간 RTB는 디지털 마케팅 환경을 근본적으로 변화시켜 왔다. 무엇보다 RTB의 알고리즘 기반 마케팅 덕분에 광고주는 적절한 가격으로 적절한 잠재고객을 더 쉽게 타기팅할 수 있게 됐다. 미국에서 RTB를 활용한 광고비는 2014년과 2018년 사이 3배 증가했으며 2025년 말까지는 약 35조 5,000억 원(290억 달러)에 이를 것으로 예측된다. RTB에 대한 자세한 설명은 다음 장에서 이어가도록 하겠다.

2. 데이터를 기반으로 한 광고 시장의 진화

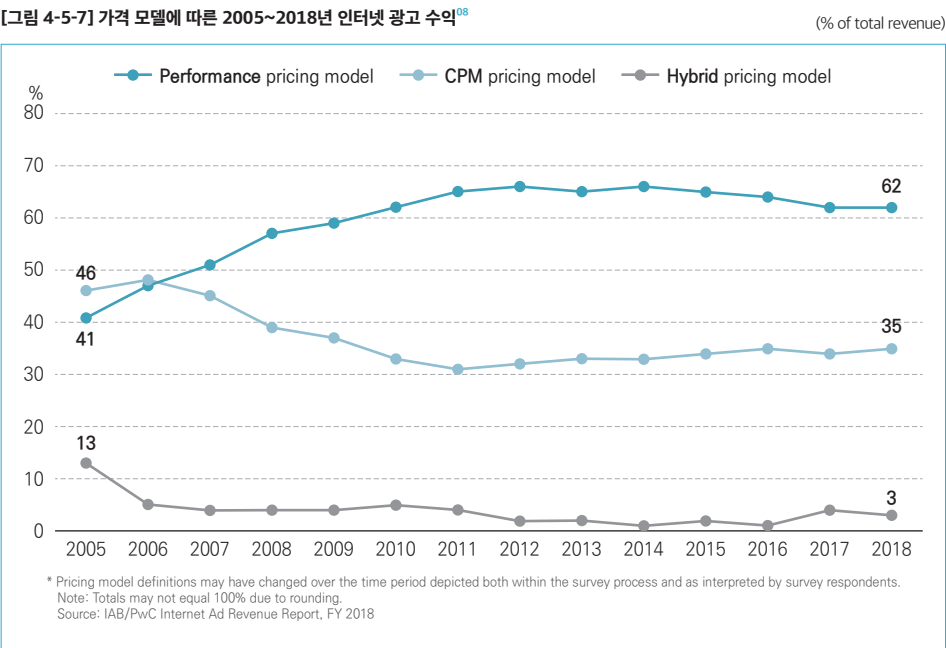
가. 표준화 시장: 프로그래매틱

인공지능, 머신러닝, 클라우드 기반의 서비스는 많은 것을 자동화시켰고, 이 중 RTB로 대변되는 프로그래매틱 마케팅은 짧은 시간에 전 세계적으로 괄목할만한 성장을 거두었다. 2000년 이후 전 세계 중 아시아 중산층이 수천만 명 증가하며 자동차, 여행, 식품, 금융 등 다양한 상품군의 구매력이 크게 증가했다. 때맞춰 이뤄진 스마트폰 보급의 증가는 곧 데이터 사용량의 폭증으로 이어졌다. 더 많은 동영상 시청과 더 많은 앱 사용, 그리고 상호작용이 가능한 미디어와 광고 형식의 등장을 촉진시켰다.

성장에 목마른 글로벌 기업들은 아시아 고객의 마음을 얻기 위해 다양한 마케팅 전략을 전개하기 시작했다. 데이터를 기반으로 가장 적절한 고객에게 접근하는 데에 적합한 프로그래매틱 광고 수요 또한 자연스럽게 증가했다. 프로그래매틱 광고는 수천만 개의 앱과 웹페이지에 동시다발적으로, 그러나 광고주가 필요로 하는 사람에게 광고를 노출할 수 있는 표준화된 규격을 제시했다. 이는 과거 유명한 혹은 사용자가 많은 광고 지면을 예약하는 방식에서, 적은 수의 광고 소재만으로 더 많은 잠재고객과 지면에 실시간으로 도달할 수 있는 효과적인 광고 방법으로 주목받았다.

프로그래매틱 광고의 주된 과금 방식은 사용자가 광고 지면에 등장했을 때마다 데이터에 근거해 가격을 매기는 광고 노출 당 비용, CPM(Cost Per Mille)이다. CPM 광고를 게재하는 광고주는 광고가 1000회 게재될 때 지급할 금액을 설정하고 이에 맞춰 비용을

정산하는데, 최근 프로그래매틱 광고에서는 1000회가 아닌 모든 사용자마다 잠재 성과를 예측해(예. 모바일 앱을 설치할 확률) 최적의 광고 비용을 매기는 광고 1회 노출당 비용(CPM)으로 입찰하는 성과 기반 방식(Performance Pricing Model)이 일반적이다.



앞서 설명한 RTB 방식에 따라 CPM 광고는 텍스트, 이미지 또는 동영상 광고일 수 있으며 복합적인 데이터에 기반한 타기팅(targeting)을 사용한다. 결국 한 명의 사용자에게 경쟁사가 아닌 자사의 광고 소재를 선택해 노출하기 위해서는 표준화된 광고 입찰 시스템에 기반해 빠른 시간에 데이터를 분석해 최적의 입찰가격을 제시해야 한다. 이 과정을 탁월하게 처리할 수 있는 것이 광고주를 대리해 적절한 광고 지면을 자동화해 구매해 주는 DSP(Demand Side Platform)이다.

프로그래매틱 광고에서는 사용자 데이터를 실시간으로 처리, 분석, 예측하는 DSP의 머신러닝 혹은 인공지능 엔진의 성능에 따라 광고 성과가 달라진다. 특히 1초에 수백만 건의 데이터가 쏟아지는 인터넷, 모바일 환경에서의 데이터 수집과 분석 기술은 광고주에게도 다양한 데이터 기반의 타기팅 선택지를 제공할 수 있다.

나. 타기팅

RTB로 대표되는 프로그래매틱 광고에서 실시간으로 광고를 노출하기 위해서는 0.1초 안에 모든 의사결정이 이루어져야 한다. 이는 사용자가 제공에 동의한, 개인식별이 불가능한 익명의 데이터에 기반하는데, 모바일 환경에서의 광고 ID(Advertising ID)가 가장 대표적이다. 광고 ID는 사용자가 모바일 내 광고를 노출할 수 있는 지면(예. 메신저 사업자가 설정한 광고 지면)에 접근할 때 여러 정보를 전송하고, 이를 통해 DSP와 같은 플랫폼은 해당 사용자가 광고주의 광고에 어떻게 반응할지를 예측해 가장 적합한 광고를 송출하게 된다. RTB 알고리즘 내에서 광고주는 어떤 사용자에게 자신의 광고를 노출할지 결정할 수 있다. 모바일 앱을 기준으로 타기팅이 가능한 대표적인 데이터를 분류하면 아래와 같다.

1. 광고 지면(Publisher): 광고가 노출될 수 있는 지면, 즉 앱을 의미한다.
2. 앱 분류(Category): 모바일 앱은 대표적으로 안드로이드, iOS 정의에 따라 범주가 나뉜다. 예를 들어 카카오톡은 구글 내 ‘커뮤니케이션’에 속하고, 모바일 게임은 ‘게임’이라는 대분류 내 ‘액션’, ‘전략’ 등으로 나뉜다.
3. 지역(Geo): 국가 혹은 시, 도로 나뉘 타기팅이 가능하다.
4. 모바일 기기 운영체제 버전(OS Version): 운영체제에 대한 정보가 수집되므로, 최신 운영체제 버전에서 제공하는 모바일 게임 광고 등에 활용할 수 있다.
5. 인터넷 연결 상태: 와이파이 혹은 모바일 인터넷 사용자를 구분할 수 있다.

실제로 사용자가 전송하는 데이터는 수백 개이므로 사람이 잠재 사용자를 모두 구분해 타기팅(Targeting)하는 것은 물리적으로 불가능하고 효율적이지도 못하다. 다양한 사용자 데이터를 보유한 대형 플랫폼에서는 광고주가 희망하는 고객 군을 구분해 인구학적 통계 기반(Demographic Targeting)이나 관심사 기반의 타기팅(Interest Targeting)을 제공하고 있다. 이를 통해 “강아지가 있는 20~30대 남성, 여성”과 같은 세부적인 타기팅이 사실상 가능해졌다. 그러나 인간이 인간의 모든 관심사를 손수 분류할 수 없고 모든 모바일 사용자가 특정 미디어에 존재하는 것은 아니다. 그러므로 정확한 타기팅을 위해서는 결국 기술의 도움이 필요하다.

머신러닝 엔진이 탁월한 DSP의 경우, 이러한 사용자의 데이터 신호를 실시간으로 분석해, 모든 사용자별로 잠재 전환 확률(예. 모바일 앱 설치, 상품 구매 등)을 계산하는 잠재 고객 타기팅(Audience Targeting)을 제공하고 있다. 이는 광고주 입장에서 어떠한 타기팅도 제공하고 있지 않는 것으로 보이지만, 실제로는 인간의 인지로 분류할 수 없는 수많은 데이터를 분석하고 학습함으로써 최적의 결과를 이끌어내는 방식이다. 예를 들어 구글의 광고는

특별한 타기팅 설정 없이도, 머신러닝 기술을 활용해 몇 번의 클릭만으로 성과 기반의 광고 실행이 가능하다.

다. 성과 측정

과거에는 ‘잠재 사용자가 하루에 n만 명 등장하는 지면’이라는 방식으로 광고 가격을 설정하고, 노출, 클릭 수준의 데이터만을 수집했다면 앞선 오디언스 타기팅 기반의 프로그래매틱 광고는 ‘타기팅의 정의’보다는 광고가 노출되는 순간부터 클릭, 그리고 그 이후의 실제 성과까지 측정하는 완벽한 투명성을 보장한다. 특히 노출 단위당 가격을 정하는 RTB 광고는 광고 1회의 노출을 1원 단위까지 측정해 잠재 사용자가 어떤 광고 지면에서, 어떤 광고 소재를, 언제 보았고, 언제 클릭해 결과적으로 언제 목표하는 전환이 일어났는지를 측정할 수 있다. 특히 몇 초라도 사용자의 눈을 잡아두려는 모바일 시장에서는 아래와 같은 지표를 주요 목표로 측정한다.

- CPM(Cost Per Mille): 광고의 1000번 노출 당 평균 가격
- IPM(Install Per Mille): 광고 1000번 노출당 발생한 앱 설치 수
- CPI(Cost Per Install): 앱 설치 1번당 평균 광고비
- 잔존율(Retention Rate): 앱을 설치한 후 경과일에 따른 앱 사용자 수
- CPA(Cost Per Action): 회원 가입, 구매 등의 액션으로의 전환 당 광고비

경영 관리의 구루 피터 드러커가 말했듯, 모든 성과는 ‘측정할 수 없다면 관리할 수 없고, 관리할 수 없다면 개선시킬 수도 없다.’ 프로그래매틱 광고는 모든 것을 계량화할 수 있다는 장점이 있다. 그 덕분에 데이터를 통한 효과적인 성과 기반의 광고가 가능해졌다.

프로그래매틱 시대의 마케터는 데이터에 기반해 더 효과적인 의사결정을 내리고 마케팅 성과를 입증할 수 있는 무기를 갖추었다. 광고 성과의 개선을 위한 소모적인 커뮤니케이션이 아닌 자동으로 최적화된 결과를 보고 필요한 행동에 집중할 수 있게 됐다. 그러나 이러한 자동화 광고 시장의 환경을 악용해 가짜 노출과 클릭, 심지어는 모바일 앱 설치나 상품 구매까지 일어났다고 속이는 사기 행위도 등장했다. 이를 광고 사기(Ad Fraud)라고 하는데, 이는 광고 없이 순수 유입된 사용자의 데이터뿐만 아니라, 정상적으로 광고를 송출한 광고 플랫폼의 광고 기여를 납치해 자신의 광고 성과인 양 제시한다. 따라서 데이터 기반의 광고에서 투명성을 제공하지 못하는 광고 서비스를 주의하고 걸러내는 것도 중요하다.

3. 데이터 기반 광고 시장의 미래

가. 규제 환경 변화

전 세계적으로 데이터 기반의 디지털 광고 시장이 성장하며 반대 급부로 일반 사용자들 사이에선 개인데이터 사용과 프라이버시 침해에 대한 우려가 높아지고 있다. 이에 유럽연합(EU)의 GDPR을 시작으로 미국 캘리포니아의 CCPA 등 각 국은 광고 데이터 사용에 사용자들의 권리를 보장하는 방법을 강구하고 있다.

1) 유럽연합 개인정보보호 일반규정

EU(유럽연합)는 개인의 개인정보 보호권을 강화하고 EU 회원국 간의 자유로운 데이터 이동을 보장하기 위해 개인정보보호 일반규정(General Data Protection Regulation, GDPR)을 제정했다. 기업이 보유한 개인데이터를 새로운 사업자에게 이전시킬 수 있는 데이터 이동권(Right to data portability)을 명시해 기술적으로 가능한 경우, 개인은 본인 또는 다른 기업에게 직접 정보를 전자적으로 전송하도록 데이터 보유 기관에 요청할 수 있도록 했다.

GDPR은 개인정보보호와 관련해 EU 회원국 소재 기업들은 물론 일정한 요건을 충족하는 EU 이외 국가에 소재하는 기업들에 대해서도 직접적으로 효력을 가지는 첫 입법으로 해석된다. 일반 위반은 위반 기업의 전 세계 연간 매출액 2%나 약 128억 원(1,000만 유로)의 금액을 과징금으로 내도록 하고, 중대한 위반은 위반 기업의 전 세계 연간 매출액 4%나 약 256억 원(2,000만 유로)의 높은 과징금을 내도록 한다. 정보 업계는 GDPR을 ‘인터넷이 생긴 이후 가장 강력한 개인정보보호규정’이라고 평가한다.

2) 미국 캘리포니아주 소비자프라이버시법

EU의 GDPR에 이어 미국에서도 글로벌 수준의 개인정보보호를 위해 새로운 법이 입법, 올해 초 시행됐다. CCPA는 캘리포니아 소비자 프라이버시 보호법(California Consumer Privacy Act 2019)의 준말로, GDPR이 유럽 내 국가별로 상이한 개인정보보호 체계를 단일화하고자 하는 시도였다면, CCPA는 연방법이 아닌 캘리포니아 주법으로 제정됐다.

CCPA는 캘리포니아 주민을 대상으로 서비스를 제공하는 서비스에도 적용된다. 따라서 캘리포니아 내에 사업장이 없어도 사실상 미국의 거주자를 대상으로 서비스가 제공되는 모든 사업은 CCPA의 영향권 안에 있다고 할 수 있다.

이로 인해 기업은 개인정보 수집 시 일정한 사항을 소비자에게 고지해야 한다. 또한 개인정보를 제3자에게 판매하는 경우 ‘Do not Sell my info’ 문구와 링크를 웹사이트를 통해 제공해야 하고, 정보주체에게 사후 거부 의사를 밝힐 수 있도록 해야 한다. 미성년자의 정보

처리와 관련해서도 제3자에게 이를 판매할 경우 나이에 따라 미성년자 본인 또는 보호자의 동의를 얻도록 의무화하고 있다.

CCPA는 데이터 주체에게 통지를 요구할 수 있는 권리, 데이터에 대한 접근을 허용할 권리 및 개인정보에 대한 자신의 권리를 행사하는 소비자에 대한 차별을 금지하는 권리를 부여하고 있다. 특히 CCPA는 개인정보처리방침에 캘리포니아의 개인정보에 대한 권리에 대한 설명을 수립하고 이를 공개해야 하며, 개인정보의 판매 여부 및 목적의 고지 및 공개 의무뿐만 아니라 이러한 사항에 대한 연 1회 이상의 업데이트 의무도 부과하고 있다.

3) GDPR과 CCPA의 비교

GDPR과 CCPA는 개인데이터에 대한 강력한 보호를 보장하고 온라인 또는 오프라인 모두에서 소비자 데이터를 수집, 사용 또는 공유하는 기업에 적용하는 것을 목표로 한다. CCPA는 대부분의 조항을 GDPR을 참고해 미국 국경 내에 투명성, 사용자 통제 및 책임 향상을 위해 유럽의 추진력을 가져왔다고 해석된다. 그러나 CCPA는 GDPR보다 범위와 엄격성이 축소돼 ‘GDPR Lite’라는 별명으로 불리기도 한다.

구체적으로, 글로벌 매출액의 4%라는 엄청난 과징금 체계를 갖춘 GDPR과는 다르게 CCPA는 고의 위반에 대해 소비자 당 최대 7,500달러까지의 벌금이 부과될 수 있고 소비자 한 명 당 750달러까지의 배상 및 집단소송을 보장하고 있다. 또한 GDPR은 데이터 보호 책임자의 임명을 의무화하고 이에 대한 처리 활동 기록의 유지 및 특정 상황에서 데이터 보호 영향 평가의 필요성과 관련된 의무를 규정하고 있지만, CCPA는 기업이 소비자의 요청을 처리하는 직원들을 훈련시켜야 하는 의무와 같은 책임 관련 의무에는 특별히 초점을 맞추지 않는다. CCPA는 GDPR보다 범위와 엄격성이 축소돼 일부 제약이 있으나 개인정보보호에 관한 일반법이 없던 미국에서 사실상 기업에 대한 일반법으로 작동할 것으로 기대되고 있다.

나. 규제 환경 변화로 인한 광고산업의 미래

앞서 살펴본 바와 같이 타기팅에 응용 가능한 개인데이터는 폭발적으로 증가하고 있다. 이에 따라 개인데이터에 대한 규제 환경은 개인데이터 사용 과정에서 발생할 수 있는 프라이버시 침해 가능성을 방지하고, 투명한 데이터 사용승인 절차를 거쳐 데이터 주권을 강화하는 방향으로 고도화되고 있다. 변화에 따라 거대 플랫폼 사업자들은 개인정보 보호를 명목으로 개인데이터의 광고 활용을 위해 필수적으로 필요한 쿠키 혹은 광고 ID(Advertising ID)와 같은 식별자(Identifier)의 활용을 제한하는 방향으로 움직이고 있다.

전 세계 웹 브라우저 점유율 1위인 구글의 크롬은 향후 2년 내 쿠키 지원을 중단한다고 발표했으며, 애플은 iOS에서 광고 ID 기반의 광고 추적을 제한하는 기능을 OS의 기본

설정값으로 바꿀 예정이라고 공표했다. 데이터 기반의 타기팅 광고 사업자는 이렇게 식별자의 활용이 불가능해지는 환경에서라면 광고 캠페인의 성과와 성과분석에 어려움을 겪을 수밖에 없다.

이에 따라 식별자가 없는 익명의 데이터 속에서 동일한 사용자를 식별할 수 있는 기술이 광고 업계에서 광고 캠페인의 성과를 좌우하는 새로운 기준이 될 것으로 예상된다. 즉 과거에는 머신러닝 기술이 광고에 있어 잠재 고객 타기팅을 위해 주로 활용돼 왔다면, 이제는 식별자가 없는 방대한 데이터들 사이에서 특정 데이터가 동일한 사용자로부터의 데이터임을 효과적으로 인식할 수 있도록 하는 데에도 필수적으로 활용되는 것이다. 이렇게 광고 산업은 개인 차원에서 더 적절한 수준의 프라이버시를 보장하면서도 더 효과적인 성능의 타기팅 광고 서비스를 제공하는 방향으로 흘러가고 있다. 그렇다면 향후 광고 사업자의 경쟁력은 얼마나 다양한 개인데이터를 프라이버시 문제없이 확보하고 있는지와 함께, 고도화한 머신러닝 기술로 이 데이터들을 얼마나 잘 식별해 타기팅 성능을 높일 수 있는지에 달려 있다고 할 수 있다.

제6장. 교육 분야 데이터 활용 비즈니스 현황

필자: 김진숙
(한국교육학술정보원 수석연구위원)

교육 데이터 비즈니스를 종합해 보면 한마디로 에듀테크라 지칭할 수 있다. 에듀테크는 교육과 기술의 융합으로, 빅데이터·인공지능·가상현실·클라우드 등의 최신 정보기술이 교육에 접목된 형태다. 코로나19에 대응하기 위해 원격수업을 속속 도입함에 따라 에듀테크의 성장 가능성에 전 세계가 주목하고 정책 지원과 투자를 확대하고 있다. 에듀테크 적용 범위 역시 학습 콘텐츠 활용을 넘어 데이터 분석에 기반한 개인 맞춤형 서비스로 확대되고 있다. 교육의 목적 가운데 하나가 미래사회의 경쟁력 있는 인재 양성인 만큼, 인간이 중심이 되는 에듀테크 서비스 활용, 특히 데이터로부터 통찰을 이끌어내는 능력을 키우는 방향으로 발전돼야 한다.

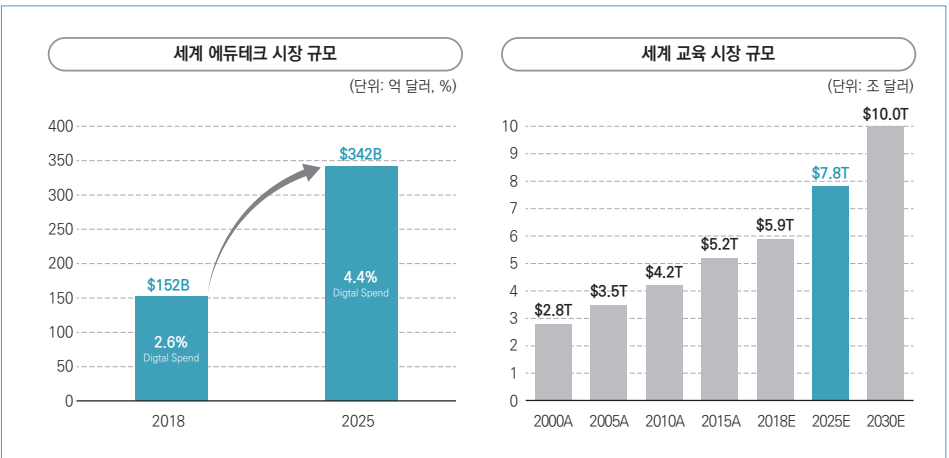
1. 세계 에듀테크 산업 및 기업 동향

가. 국내외 에듀테크 산업 동향

세계 에듀테크 산업 시장은 연평균 12% 성장을 유지하고 있다. 2018년 1,530억 달러(183조 원)에서 2025년에는 3,420억 달러(약 409조 원) 시장으로 성장할 것으로 전망된다. 주로 미국, 중국, 영국을 중심으로 시장이 형성되고 있으며, 코로나19로 인해 전 세계적으로 원격수업이 실시되면서 공교육에서의 에듀테크 도입 사례도 늘어나고 있다. 세계 교육 시장의 지속적인 성장과 함께 교육 시장에서 에듀테크의 비중도 확대되는 추세다. 성장세인 세계 교육 시장에서 에듀테크의 비중은 2018년 2.6%에서 4.4%로 확대될 것으로 전망된다.

국내 에듀테크 산업 시장은 2018년 3조 8,450억 원에서 2019년 3조 9,515억 원으로 성장하면서 2.8%의 성장세를 유지하고 있다. 이는 세계 교육 시장 성장률에 비해 낮은 수준이다. 그동안 공교육보다는 사교육 시장을 중심으로, 국내 시장보다는 해외 시장에서

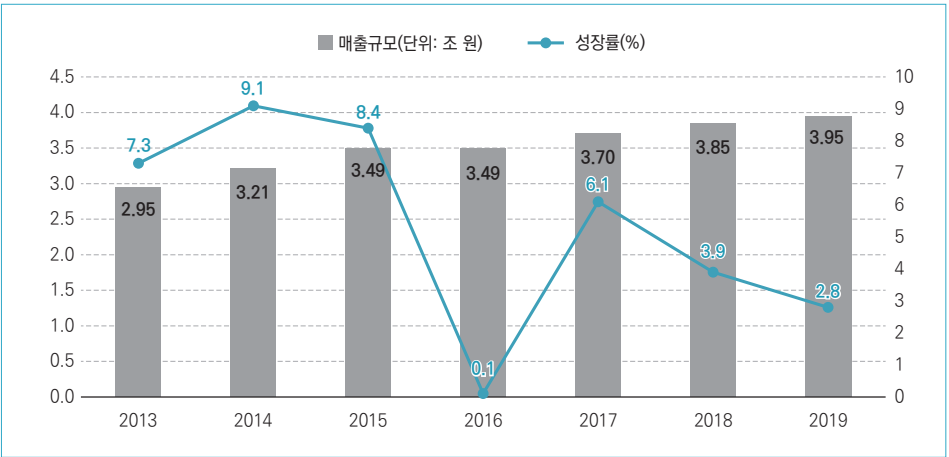
[그림 4-6-1] 세계 에듀테크 시장 전망과 교육 시장 규모



출처: Holon IQ(2019). Smart Estimates™

가시적 성과를 거두고 있다. 현재 국내외 산업 규모를 파악할 수 있는 데이터가 코로나19 이전의 2019년 것이므로, 코로나 이후 에듀테크 산업 현황에 대한 면밀한 데이터 분석이 필요하다. 결론적으로 세계 에듀테크 시장 규모나 성장률에 비해 국내 에듀테크 시장은 성장률이 3% 이내에 머무르고 있다.

[그림 4-6-2] 국내 에듀테크 시장 규모 및 성장률⁰¹



출처: 「2019년 이러닝산업 실태조사」(정보통신산업진흥원, 2019.) 재구성

01. 에듀테크 산업에 대한 조사가 별도로 이뤄지지 않고 있어, 매년 이뤄지는 이러닝산업 실태조사를 기반으로 저자 재구성

나. 국내외 에듀테크 서비스 동향

신성장산업으로의 에듀테크 발전 가능성이 주목되면서, 에듀테크의 적용 범위도 확대되고 있다. 빅데이터, 인공지능 기술을 활용한 학습 진단 및 분석 서비스 개발이 본격화되고, 학습자들의 커뮤니티를 지원하며, 콘텐츠 제공 범위가 기존의 동영상 위주에서 가상 현실(AR) / 증강 현실(VR) 기술을 활용한 실감형 콘텐츠로 확대되고 있다. 크게 학습 콘텐츠, 학급 관리 및 커뮤니티, 학습 진단 및 분석 등으로 분류해 대표적인 국내외 서비스를 살펴보면 다음의 [표 4-6-1], [표 4-6- 2]⁰²와 같다.

[표 4-6-1] 국외 에듀테크 서비스 동향

구분	서비스	주요 특징
학습 콘텐츠	Google Expedition	VR과 AR을 활용한 콘텐츠 중 가장 빠르게 성장하는 플레이어다. 해외 유적지, 박물관, 우주 등을 직간접적으로 체험할 수 있는 체험형 플랫폼. 사진을 기반으로 한 가상현실 앱으로 전 세계 주요 장소, 박물관, 우주 등을 탐험할 수 있다. 학습 설명, 주제, 특징, 질문 등 다양한 콘텐츠를 포함하고 있다.
	Amazon Inspire	아마존 킨들을 앞세워 전자책 형태의 퍼블리싱 플랫폼을 성공시킨 경험으로 콘텐츠 서비스를 확대하고 있다. 아마존의 ‘인스파이어(Amazon Inspire)’는 학습 콘텐츠 공유 서비스로서 관심도 수준, 이용자 평가, 인기도, 등급 등의 조건에 따라 자료를 검색할 수 있다.
	Discovery Education	증강현실과 가상현실 기술을 학습 경험과 접목해 가상 실험실, 영상 자료 추천 등 탐험 환경을 제공한다.
	Dreambox Learning	초·중학교의 수학 학습 과정을 게임처럼 구성하여 1800개가 넘는 콘텐츠를 제공하고 있다. 학습경험 분석을 통해 사용자에게 맞춤 영상을 추천해 주는 넷플릭스 기술을 채택하고 있다.
학급 및 커뮤니티 관리	Google Classroom, G Suite for Education	Google Classroom은 클라우드 기반으로 교사와 학생의 학습 자료 및 데이터를 공유하고, 공동 작업을 지원한다. G Suite for Education은 구글의 메일, 캘린더, 드라이브, 문서 등 생산성 관련 도구를 학습 상황에 맞춰 제공한다.
	MS Office 365 Education	학급이 구성되고, 구성원들간에 수업 자료와 과제를 공유하고, 협업할 수 있는 플랫폼이다. OneDrive, Teams, Skype 등의 전문 학습 커뮤니티와 연결해 원활한 소통을 지원한다.
빅데이터를 통한 학습 진단 및 분석	Knewton	누적한 학습자의 학습 데이터를 관리, 분석하여 개인별 맞춤 학습 정보를 제공한다. 특히 수학과 독해, 기타 교과목간의 학습 개념 연계성을 기반으로 적합한 학습법 추천과 콘텐츠를 제공한다.
	KidAdaptive	적응형 학습플랫폼(Adaptive Learning Platform, ALP)으로 학습자의 흥미를 유발해 참여도를 높여 교육 효과를 높이는 서비스를 제공한다. 클라우드 기반의 다양한 학습 활동으로부터 데이터 분석 결과에 의미를 부여하는 플랫폼이다.

(계속)

구분	서비스	주요 특징
빅데이터를 통한 학습 진단 및 분석	Duolingo	외국어 학습에 대한 진단, 평가 서비스다. 학습자의 언어 수준을 진단해 단계적으로 학습할 수 있는 카드형 학습 콘텐츠를 지원하고, 평가 결과를 분석·진단해 반복 학습을 유도한다.
	Carnegie Learning	수학, 컴퓨터과학에 특화된 서비스다. 대표적인 서비스인 MATHlab는 학생들의 개별적 학습 진단과 학습 결과에 대한 종합 성취 리포트를 제공한다. LIVELAB 기능에서 학생들의 집중도, 완수 현황 등을 교사가 즉각적으로 확인할 수 있다.
	HUBERT AI	인공지능 자연어처리(Natural Language Processing) 기술을 적용해 대화형 피드백을 제공할 수 있다. 학습자들의 대화를 통해 수집한 데이터를 분석해 학생별로 강점 등을 교사에게 제공한다.

국외 에듀테크 서비스는 교육 기업이 아니었던 기업들이 비즈니스 등의 서비스를 교육 영역으로 확대하거나, 교육 시장에서 강세를 보였던 기업들이 빅데이터와 인공지능 기술 등을 접목해 학습자 맞춤형 서비스를 강화하는 모습이다. 특히 구글, 마이크로소프트, 아마존 등 글로벌 기업들의 교육 진출이 돋보인다. 이들 기업들은 자체 보유한 전문적이고 안정적인 기술을 토대로 서비스에 대한 연구 및 개발 투자를 확대해 에듀테크 영역에서도 경쟁력을 높일 것으로 전망된다. 또한 학습 진단 및 분석의 기반이 되는 인공지능과 빅데이터 분석 기술 적용을 확대하면서, 학습 수준뿐 아니라 학습 집중도와 학습 습관, 행동 등에 대한 정서 및 감성까지 분석하는 등 관련 기술이 날로 발전하고 있다.

[표 4-6-2] 국내 에듀테크 서비스 동향

구분	서비스	주요 특징
학습 콘텐츠	타임교육 팩토schule (factoschule)	수업 교구 제조 경험을 살려 AR, 사물인식(OR) 기술 등을 접목한 교구, 교재 등을 제공한다. 타임교육이 개발한 ‘팩토schule’는 오프라인 교실 수업과 온라인 수업을 지원하기 위해 모바일 앱과 교구, 교재 등을 통합적으로 제공한다.
	럭스 로보 (luxrobo)	코딩교육 콘텐츠에 특화된 하드웨어 모듈을 제작해 로봇을 만들면서 소프트웨어 개념을 배울 수 있는 서비스다. 하드웨어 모듈 ‘모디’는 각기 다른 사물인터넷(IoT) 기술을 내장해 코딩 기초 개념 형성을 돕는다.
학급 및 커뮤니티 관리	클래스팅 (Classting)	학급 구성, 수업 자료 온라인 공유, 학교 간 교류를 지원한다. 커뮤니티 기능 보강 및 국내 출판사들과 제휴를 통해 인공지능 기반의 학습 진단 및 분석을 강화하고 있다.
	위두랑 (Wedorang)	공공에서 만든 교육용 커뮤니티 서비스다. 학급 개설, 과제 제출, 토의토론, 학습 포트폴리오 기능을 지원한다. 디지털교과서와 연계한 학습 과제 관리를 할 수 있다.
	엘리스 (elice)	실습 중심의 온라인 코딩에 특화된 학습 관리 서비스다. 대학 및 기업의 사내 교육 및 개인용으로 사용되며, 개인별 코드 리뷰와 멘토링이 가능하다. 강의 영상 업로드와 실습 과제 생성 및 자동 채점 등 학습 평가 관리도 지원한다.

(계속)

02. 관련 에듀테크 산업 동향 보고서 및 언론보도 기사 등을 종합 분석해 저자 재구성

구분	서비스	주요 특징
빅데이터를 통한 학습 진단 및 분석	산타토익 (Santa Toeic)	TOEIC 대비를 위한 지능형 튜터 서비스다. 학생 수준을 미리 진단하고 예측해 취약 분야를 집중 보강할 수 있도록 관련 학습 자료를 추천한다. 최근 공인중개사 시험 대비 서비스를 내놓았으며, SAT와 GRE 등의 학습 평가 서비스 분야로 해외 진출을 추진중이다.
	노리 (KnowRe)	온라인 개인교사를 표방하는 수학 분야에 특화한 서비스다. 학생들이 반복적으로 틀리는 문제의 풀이 과정을 분석해 부족한 영역의 수학 개념을 찾아내고, 핵심 개념에 대한 이해를 지원한다.
	웅진 (웅진생크빅 AI학습)	국의 사례에서 소개된 ALP 플랫폼을 활용해 웅진생크빅 AI 학습 서비스를 출시했다. 학습시 행동패턴을 분석해 잘못된 학습 습관을 교정하도록 지원한다.
	마타수학	수학 개념에 대한 개인별 취약점 분석 및 문제를 추천하는 서비스다, 수학의 개념별 연계도를 제시하고, 누적된 진단 결과를 생성해 개인별 맞춤 문제를 제시한다.
	천재교육 (밀크T)	학습자의 개별 성향을 진단하고 분석해 맞춤 시간표를 제공한다. 학습 유형을 분석해 개인화한 학습법을 추천하고, 유명 강사진들의 강의 및 전문가들의 1:1 학습 관리 서비스를 제공한다.

국내 에듀테크 시장은 상대적으로 협소한 시장 규모와 공교육에서의 활용률 저조에 따라 사교육의 대체제 수단으로 성장해왔다. 따라서 내용도 영어, 수학 등 전통적인 사교육 영역에 집중돼 있다. 이에 따라 학습지 출판사와 교육 기업에서 주도하는 모습이다. 다만 원격수업시행에 발맞춰 온라인 학습 이력과 사용자 참여도 분석 및 학습 관리 등에 초점을 둔 서비스 출시가 잇따르고 있다. 인공지능 교육, SW 교육 확대 정책에 대한 기대감으로 관련 서비스를 출시하는 모습이다. 국내 교육 시장은 국가 정책에 영향을 받는 구조이므로 향후 전개될 정책 기조나 규제 개선 등에 주목할 필요가 있다.

2. 에듀테크 정책 추진 동향

가. 국가 차원의 정책 추진

코로나19 사태가 국가 차원의 미래사회 대응 정책인 한국판뉴딜(2020년 7월 발표)을 불러왔다. 한국판뉴딜 정책⁰³에서 에듀테크 활성화와 직접적으로 연계되는 분야는 디지털뉴딜에 속해 있다. 교육 분야 디지털 뉴딜은 전국 초·중고·대학·직업훈련기관의 온/오프라인 융합 학습

환경 조성을 위해 디지털 인프라 기반 구축 및 교육 콘텐츠 확충을 추진하는 것이다(기획재정부, 2020). 주요 내용을 살펴보면, 2022년까지 초·중고 전체 38만 실에 고성능 와이파이 무선 인터넷 구축, 교원의 노후 PC 교체, 온라인 플랫폼 구축 등이 이뤄진다. 대학의 온라인 강의를 지원하기 위한 국립대의 원격교육지원센터나 교대와 사대에 미래교육센터 설치 등이 추진된다. 이외에 평생교육 수요에 대응하기 위한 K-MOOC(Massive Open Online Course, 한국형 온라인공개강좌)의 확대나 공공, 민간의 직업훈련 기반 구축 등이 포함돼 있다. 뉴딜이 갖는 성장 기반 마련 기조에 맞춘 인프라 구축이 주된 내용이지만, 활성화를 위한 토대를 마련한다는 데 의미를 둘 수 있다.

이외에도, 대통령 직속 4차산업혁명위원회에서는 미래교육을 위한 에듀테크 활성화 권고안을 발표했다⁰⁴. 권고안 자체가 에듀테크 활성화에 초점을 두고 있으며, ‘에듀테크 활용 활성화를 통한 미래역량 기반 디지털 교육 체제로의 대전환’을 비전으로 3가지 권고 원칙을 제시했다.

3가지 권고는 맞춤형 역량 기반 교육을 실현하기 위한 에듀테크 기반 통합 교육 환경 구축, 학교 내 에듀테크 활용 환경 개선 및 교원의 디지털 교육 역량 강화 지원을 통한 교육 수준 향상, 교육 혁신을 위한 에듀테크 산업 육성 및 활용 활성화 기반 구축 등이다. 이 중에서 에듀테크 산업 육성 권고 원칙에 따른 내용이 눈길을 끈다. 학교 자율에 맞춘 에듀테크 서비스 활용 환경 조성을 위해 바우처 제도 도입·운영을 구체적으로 제안하고 있다. 에듀테크 활용 활성화의 대상 영역을 초·중등 공교육에 두고 있다는 것을 감안하면, 바우처 제도를 도입해 학교에서의 에듀테크 민간 서비스의 자유로운 선택과 활용, 이를 지원하기 위한 예산을 지원하는 데 주안점을 둔 것으로 판단된다. 이를 위한 자원 확보 노력과 후속적으로 관련 부처의 과제 개발이 이뤄져야 실효성 있게 추진될 것으로 판단된다.

나. 교육부 차원의 정책 추진

국가 차원의 정책과 연계해 교육부의 미래교육 10대 과제⁰⁵는 교육의 비전과 목표를 담아 유·초등부터 고등·평생교육에 걸친 혁신 과제를 제시하고 있다(그림 4-6-3 참조). 이 중에서 디지털 기반 구축과 관련한 9번 과제를 분석해 보면, 향후 에듀테크의 교육 적용 범위를 예측해 볼 수 있다.

디지털 전환에 대응하기 위한 교육 기반 마련에서 가장 대표적인 과제는 K-에듀 통합 플랫폼 구축이다. 이는 흩어져 있는 학습 콘텐츠와 학습관리 시스템(Learning Management

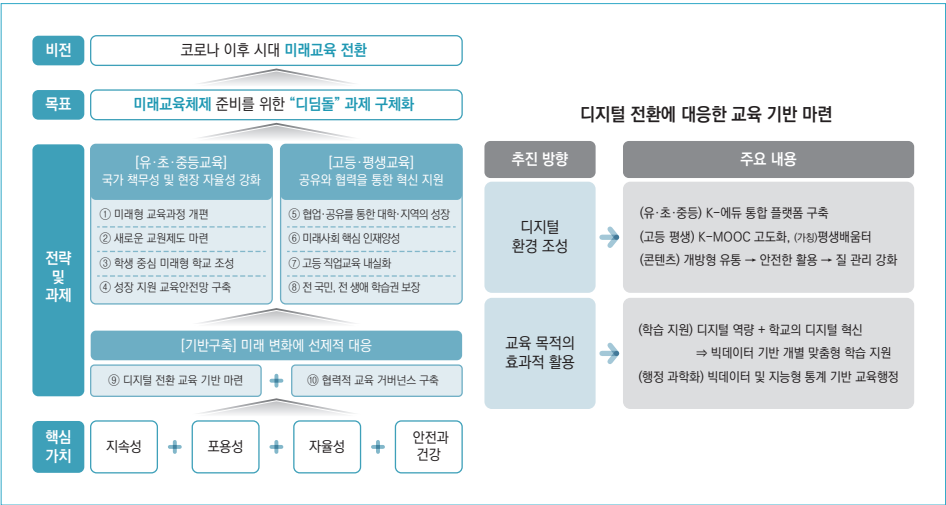
03. 「한국판 뉴딜 종합계획」, 기획재정부, 2020.7.14.

04. 「미래교육을 위한 에듀테크 활성화 권고안」, 4차산업혁명위원회, 2020.11.

05. 「코로나 이후, 미래교육 전환을 위한 10대 정책과제」, 교육부, 2020.10.15.

System, LMS) 등을 하나로 연결해, 유초중고등학교에서 사용 가능하도록 서비스를 통합해 나간다는 내용이다. 향후 이용자 관점에서의 서비스 접근성, 콘텐츠 활용의 용이성을 높이고, 민간과 공공이 협력해 한국형 에듀테크 서비스를 발전시키는 데 기본 목적과 방향을 두고 있다. 이외에도 데이터 및 기술 활용 역량 차이로 인한 교육 격차를 해소하기 위한 미디어 리터러시 교육 강화, 공공 데이터의 품질 제고를 통한 교육 분야에서 빅데이터를 활용해 교육 정책을 과학화하는 과제 등은 원격수업의 현안 해결을 넘어선 발전적 방향을 담고 있다.

[그림 4-6-3] 미래교육 전환을 위한 데이터 활용 정책 방향



출처: 「코로나 이후, 미래교육 전환을 위한 10대 정책과제(안)」, 교육부, 2020.10.5.

현재 발표되는 미래교육 정책은 거시적으로 국가 인재 양성을 위한 교육 혁신에 에듀테크 적용을 확대하는 것에 중점을 두고 있다. 한국판뉴딜이 디지털 인프라 구축에 초점을 맞췄다면, 4차산업혁명위원회의 권고안은 에듀테크 산업 활성화를 염두에 둔 구체적인 정책 과제의 제시라고 볼 수 있다. 교육부의 미래교육 10대 과제가 교육 정책 전반에 걸친 중장기 계획이라고 보기는 어렵다. 하지만 범위를 평생교육까지 확대했고 지속성과 포용성, 자율성 등을 핵심 가치로 단기적인 실천 과제들을 제시한 점은 눈길을 끈다. 다만 정책 추진의 기반이 되는 추진체제 정비와 관련 법·제도 정비, 자원 확보에 대한 구체적인 계획이 뒷받침돼야 실효성을 확보할 수 있을 것이다.

3. 에듀테크 발전을 위한 제언

코로나19가 촉발한 원격수업 경험은 에듀테크 산업 발전의 기폭제가 될 것으로 전망된다. 데이터 비즈니스 관점에서 보면, 국내외 모두 공급자 중심의 일방적인 콘텐츠 서비스에서 벗어나는 모습이다. 즉 인공지능 기술을 적용해 학습자의 활동 데이터를 수집·분석해 학습력을 개선하는 데 주력하는 것이다. 새로운 기술이 에듀테크 산업의 글로벌 경쟁력 제고로 이어지기 위해서는 몇 가지 과제가 선결되어야 한다.

먼저 국가 차원의 정책 측면의 해결 과제다. 현재로서는 가장 대표적인 한국판뉴딜과 교육부에서 제시하는 미래교육 정책 간 연계성 확보가 필요하다. 에듀테크가 최신의 데이터 활용 기술을 기반으로 한다면, 디지털뉴딜은 새로운 에듀테크가 교육 현장에 투입될 수 있도록 추진 과제 간 연결성을 확보할 수 있어야 한다. 예컨대 가상현실 기술은 교육 현장에서 학습자의 학습 몰입 환경을 제공하는 기반이므로, 교육 영역에 우선 적용하도록 제안하고 조율하는 기능을 하는 부처 간 협의기구가 필요하다.

둘째, 국내 에듀테크 산업은 기존의 사교육 중심의 시장 구조로 형성된 특징이 있다. 정부가 진로, 직업, 평생교육 차원에서 에듀테크 서비스의 활용을 권장하고, 이에 대한 재정 투자와 연구 개발을 지원하는 등 공교육에서의 시장 조성 노력이 필요하다. 4차산업혁명위원회에서 제안한 바우처 제도가 그 첫발이 될 것이다. 산업 육성 측면이 아니라, 데이터와 관련 기술의 교육 분야 적용 사례를 확보해 기술발전과 교육 환경 개선이라는 선순환 구조를 조성해야 한다.

셋째, 에듀테크가 교육에 대한 이해와 고도화한 기술 적용이 전제가 된다면, 산업 분야에서는 교육 현장에 대한 이해가, 교육 분야에서는 기술 활용에 대한 이해가 선행돼야 한다. 이 또한 정부의 역할이다. 이 분야를 연계할 수 있는 비영리 성격의 관련 연구소를 정부 출자로 설립해 연구·개발을 지원할 필요가 있다. 이는 기획, 설계, 적용과 실행 단계를 넘어 구체적인 교육의 효과와 효율성을 검증해 내는 모든 단계에서의 지원을 말한다. 이는 세계 시장으로 진출할 수 있는 에듀테크 비즈니스 모델의 레퍼런스가 될 것이다.

마지막으로, 인공지능으로 대표되는 학습자 개인 맞춤형 에듀테크 서비스의 성공 여부는 학습 데이터의 확보에 있다. 나이스(교육행정정보시스템)의 데이터를 공개하고, 최근 교육 분야 공공 데이터 공개에 초점을 두고 있는 EDS(Educational Data System)의 기능과 역할을 확대할 필요가 있다. 이는 연구자 중심의 한정된 활용 대상이 아닌, 에듀테크 산업계에서도 활용할 수 있는 기반이 될 것이다. 또한 개인정보보호법 등 관련 법안의 지속적인 수정, 보완과 함께 데이터 활용으로 얻게 된 교육 성과에 대한 성공 사례를 확보함으로써 데이터 활용의 중요성에 대한 인식을 지속적으로 개선하는 것도 필요하다.

제7장.

유통·물류 분야

데이터 활용 비즈니스 현황

필자: 신광섭
(인천대학교 동북아물류대학원 교수)

우리는 당장 필요한 물건이 있으면 스마트폰을 꺼내 상품을 주문하고 바로 다음날 받을 수 있는 시대에 살고 있다. 오프라인 중심 유통 시장의 온라인 전환은 4차산업혁명의 시작과 코로나19 확산으로 더욱 가속화되고 있다. 온라인 전환이 가능했던 것은 4차산업혁명 기반의 물류 서비스 혁신 덕분이라고 할 수 있으며, 쿠팡과 마켓컬리, SSG닷컴과 같은 기업들이 다양한 상품을 빠른 시간에 배송하고 있다. 구체적으로는 사용자로부터 생성되고 과거 서비스 제공 과정에서 확보된 엄청난 양의 데이터를 실시간으로 축적하고 분석해 정확하고 효율적인 방식으로 물류 서비스를 운영할 수 있게 되었기 때문이라고 할 수 있다. 최근의 유통 물류 분야의 시장 변화 과정을 통해 데이터 기반의 서비스 혁신에 대해 알아본다.

1. 유통 물류 서비스의 혁신

내일 아침 당장 필요한 물건이 있으면 예전에는 급하게 가까운 대형마트를 방문하거나, 동네 인근 상품을 판매하는 상점을 찾아 헤매야 했다. 지금은 스마트폰을 열고 내일 새벽 배송이 가능한 품목을 찾아 주문만 하면 되는 시대다. 모든 것은 빨라졌고 편리해졌으며, 말그대로 모든 것이 풍요로운 시대라고 할 수 있다. 4차산업혁명이 갑작스럽게 우리 일상을 바꾸기 시작했고, 코로나19 전염병은 우리 일상을 ‘언택트(Untact)’라는 단어로 송두리째 바꿔놓았다. 언택트는 당일 혹은 새벽 배송에 대한 요구를 증폭시켰으나, 사람은 만나지 않더라도 물건은 반드시 목적지를 향해 실제로 이동을 해야만 한다. 배송이 필요한 물건의 양은 폭발적으로 증가했고, 배송에 필요한 시간은 더 줄어들었다. 무엇이 먼저인지 인지하기도 전에 이미 세상은 변했다. 이 모든 것이 최근 2년 이내에 일어난 일이다.

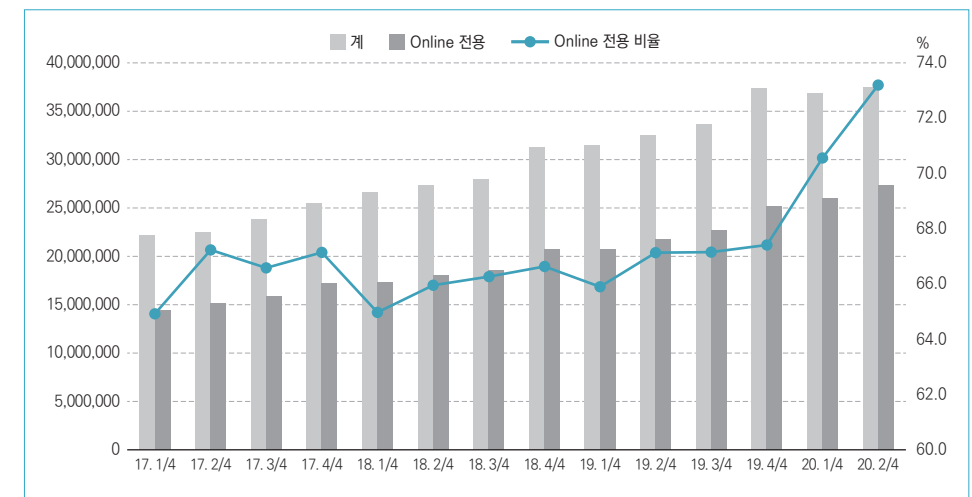
앞서 이야기한 큰 변화의 과정을 유통과 물류산업 관점에서 조금 더 구체적으로

살펴보고, 유통과 물류 서비스의 혁신이 어떻게 이루어졌는지 생각해 보자.

가. 전자상거래 시장의 성장

[그림 4-7-1]은 최근 3년간 온라인 쇼핑물의 거래 규모와 순수 온라인 쇼핑물이 차지하는 비중을 보여준다.

[그림 4-7-1] 온라인 쇼핑물 거래 규모



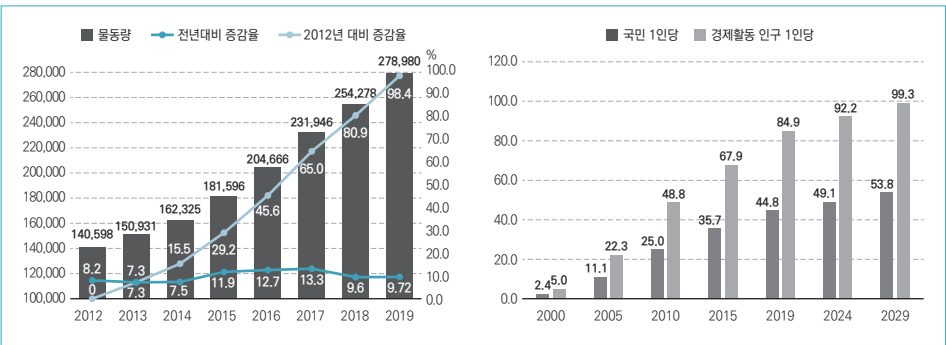
출처: 온라인쇼핑 동향조사-온라인 쇼핑물 운영형태별, KOSIS 국가 통계 포털

분기별 온라인 쇼핑물의 거래 규모는 2020년 2사분기 기준 37조 5,000억 원에 달하며, 2019년 전체는 135조 원 규모다. 이중 온라인 전용 쇼핑물의 비중은 60% 중반대를 유지하다가 코로나19가 대유행하기 시작한 2020년 1사분기부터 급격하게 증가해 70%를 넘어섰다. 2018년 정보통신정책연구원의 분석 보고서에서 예측한 2019년 111조 원, 2020년 130조 원 규모는 이미 초과달성했으며(2018년 113조 원, 2019년 135조 원), 이 속도라면 2020년에는 200조 원을 훌쩍 넘어설 것으로 예상할 수 있다.

온라인 쇼핑물 거래 규모의 성장은 택배 시장의 급속한 성장으로 이어졌으며, 2019년 기준 우리나라 경제 인구 1인당 연 99.3회 택배를 이용한 것으로 나타났다. 특히 우리나라의 경우 스마트폰 등 모바일 기기의 대중화로 모바일 쇼핑이 중요한 비중을 차지하고 있다. 온라인 쇼핑 중 모바일 거래가 차지하는 비중은 2017년 1월 53.6%에서 12월 58.8%, 2019년 2월에는 64.4%에 육박했다⁰¹. 모바일 기기의 보급은 언제 어디서나 인터넷에

01. 「국내 택배산업 해법 찾기: 디지털과 비즈니스 혁신을 중심으로」, 통계청·삼정 KPMG 경제연구원, 2020.03.

[그림 4-7-2] 택배 물동량 및 이용 현황



출처: 국가통합물류정보센터, 생활물류통계

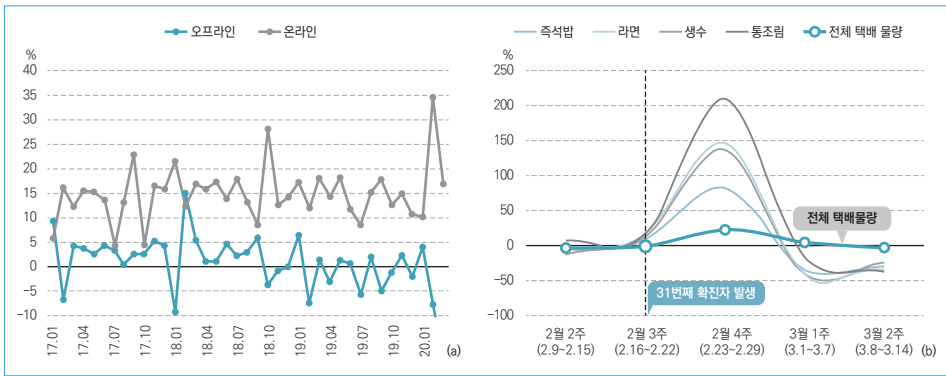
접속해 필요한 물건을 구매할 수 있는 서비스의 선결조건으로, 이제는 온라인 및 오프라인 유통업체들이 모두 모바일 기기 전용 애플리케이션을 제공하고 있다. 이러한 변화는 더 가속화될 것이고, 모바일 중심의 새로운 유통 생태계를 형성하게 될 것이다.

나. 코로나19에 의한 언택트 시장의 등장과 전자상거래 규모의 급격한 성장

전자상거래 시장은 2000~2010년 온라인 쇼핑 채널의 등장 이후 모바일 기기의 확산으로 폭발적으로 성장했다. 그러나 2020년 1월 코로나19의 확산은 오프라인 중심의 유통 시장을 온라인으로 전환하는 가장 중요한 계기가 됐다. 실제로 2020년 상반기 택배산업의 물동량 증가율은 이전에 비해 2~3배 이상 높은 증가세를 보이고 있다. 우리나라의 2020년 1분기 택배물동량은 전년동기대비 약 20% 증가한 것으로 추정되며, 1~2위 사업자인 CJ대한통운과 한진의 택배처리량은 각각 26.1%(yoy), 24.8%(yoy) 증가했다.⁰² 산업통상자원부에 따르면 코로나19가 발병한 2020년 1~3월 오프라인 월별 매출액 증감률은 1월 +4.1%(yoy)를 기록한 반면, 2월에는 -7.5%(yoy), 그리고 코로나19가 전국적으로 심각한 상황에 이른 3월에는 -17.6%(yoy)를 기록했다. 반면 온라인 쇼핑은 2020년 2월 +34.3%(yoy), 3월에는 +16.9%(yoy)를 기록하며 급격히 성장했다. 2월에 사회적 거리두기 캠페인이 확산됨에 따라 비대면거래 선호가 높아지며 모바일로 구매가 가능한 온라인 쇼핑 매출 증가세가 뚜렷하게 나타난 것을 알 수 있다. 2020년 3월의 전년동월대비 상품군별 매출증감률을 살펴보면 패션·잡화가 -30.1%(yoy)를 기록하며 감소했으나, 생필품 수요가 증가하며 식품 +9.7%(yoy), 생활·가정 상품군 매출이 +6.2%(yoy)로 증가했다.

02. 「Post 코로나19-Ⅱ.택배: 비대면 소비의 핵심요소로 물량 켜점점프, 대신증권

[그림 4-7-3] 코로나19 이후 온라인 쇼핑 및 택배물동량 변화



출처: 산업통상자원부, 대신증권 리서치센터, CJ대한통운

코로나19가 전 세계적으로 확산되는 과정에서 가장 큰 사회적 이슈가 생필품의 사재기 현상이었다. [그림 4-7-3-(b)]에서 알 수 있듯이, 31번 확진자가 발표된 이후에는 즉석밥과 라면, 생수 등의 주문량이 급격하게 증가한 것을 확인할 수 있다. 그러나 이러한 현상은 일주일 정도 지속된 후 다시 일상적인 수준으로 회복됐는데, 이는 다른 나라에 비해 안정적으로 운영되는 온라인 쇼핑물과 택배 서비스 덕분이라고 해도 과언이 아니다.

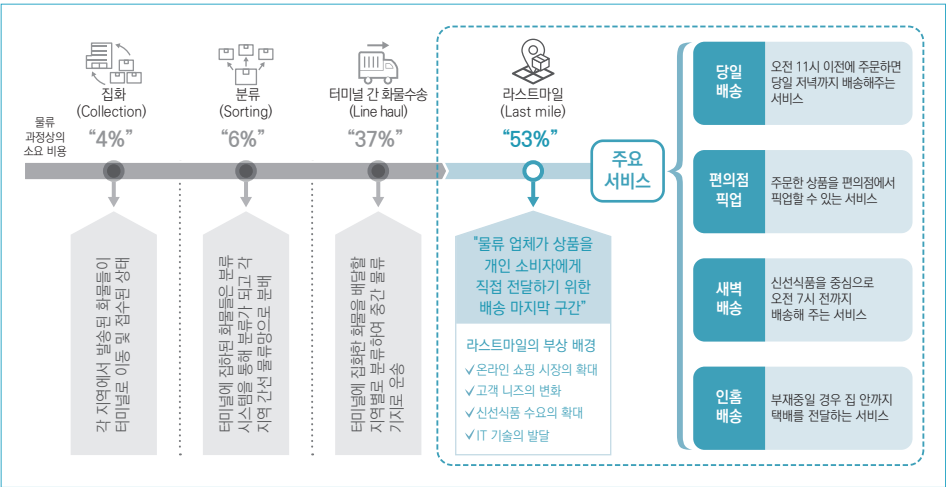
2. 생활 물류서비스의 성장, 새벽배송과 배달 대행 서비스

가. 말단 배송 서비스의 중요성

앞에서 살펴본 바와 같이 온라인 쇼핑의 성장에 따라 택배 시장의 규모도 함께 성장하고 있으나, 택배 기업의 수익성은 여전히 개선되지 못하고 있다. 한국통합물류협회에 따르면 택배 단가는 2015년 2,392원에서 2017년 2,248원까지 하락했으며, 최근 2,000원 초반까지 떨어지고 있다. 택배물동량이 급격히 늘어남에도 불구하고, 규모의 경제를 실현하기 어려운 서비스의 특성상 영업이익률은 여전히 제자리걸음이거나 오히려 적자를 보이는 경우도 존재한다. 이렇게 택배 서비스가 가진 한계를 극복하고, 서비스의 혁신적 개선을 위해 유통 및 물류기업들이 집중하는 곳이 바로 말단 배송(Last Mile Delivery)이다. 택배 서비스의 전체 소요 시간과 비용 측면에서 가장 높은 비중을 차지하기 때문이다. [그림 4-7-4]와 같이 택배 전체 서비스 중 말단 배송에 투입되는 비용이 53%를 차지하고 있는 데다 고객과의 최접점에 위치하기 때문에 전체 서비스의 만족도를 결정하는 가장 중요한

요소이다. 많은 택배사뿐만 아니라 온라인 유통 업체들 모두 말단 배송에 대한 중요성을 인식하고 비용 절감과 서비스 수준 향상을 위해 편의점 픽업, 인출 배송과 같은 다양한 방법을 제시하고 있다.

[그림 4-7-4] 택배 물류서비스와 말단 배송 서비스 다양화



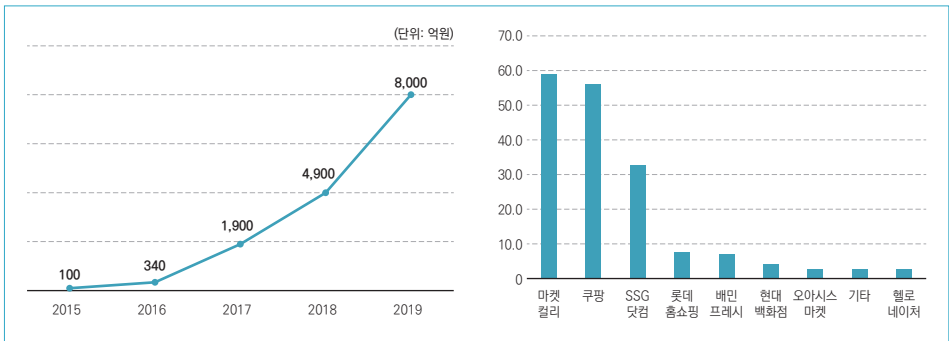
출처: Business Insider Intelligence, 삼정KPMG 경제연구원, 재구성

나. 새벽배송 시장의 성장과 신선식품의 말단 배송 서비스

아마존도 실패한 영역이라고 알려진 신선식품의 배송은 2015년 마켓컬리의 ‘샐러드배송’을 시작으로 쿠팡이 로켓프레쉬 서비스를 도입했다. 이후에 롯데, 신세계 등이 가세하면서 경쟁터를 방불케 하고 있다. 상대적으로 새벽배송 시장에 뒤늦게 참여한 롯데그룹과 현대백화점 그룹은 롯데아이몰에 전문관 ‘새로배송’을, 현대백화점과 현대홈쇼핑 H몰에서 각각 새벽배송 서비스를 시작했다. 업계 자체적인 추산 규모지만, 새벽배송 시장은 2015년 100억 원 규모에서 2018년 4,000억 원을 넘어 2019년 1조 원까지 성장한 것으로 나타났다. 특히 코로나19로 인한 비대면 비즈니스의 성장은 신선식품 새벽배송의 폭발적 성장을 견인할 것으로 기대할 수 있다.

주요 새벽배송 서비스 기업들 중에는 마켓컬리의 사용 경험이 가장 높았고, 이어서 쿠팡 및 SSG닷컴으로 나타났다. 매출 규모에서는 쿠팡이 앞서있지만, 신선식품의 새벽배송에 있어서는 쿠팡과 마켓컬리가 경쟁 구도를 형성하고, 나머지 기업들이 이를 뒤쫓는 형세다.

[그림 4-7-5] 새벽배송 시장의 변화

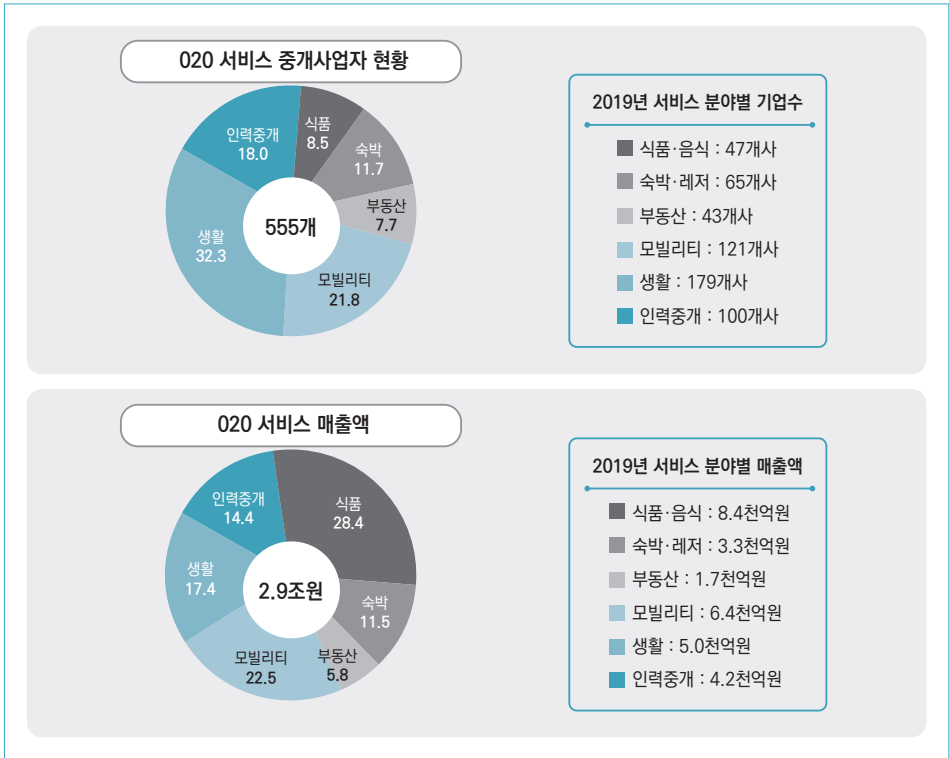


출처: 업계추산

다. 배달대행 서비스 - 플랫폼 전쟁

온라인 쇼핑몰과 함께 4차산업혁명 이후 플랫폼 기반의 대표적 서비스가 바로 음식주문 플랫폼이라고 할 수 있다. 대표적인 기업으로는 지금은 하나의 기업으로 인수된 배달의민족과 요기요를 들 수 있다. 실제로 2020년 과학기술정보통신부에서 발표한 O2O

[그림 4-7-6] 새벽배송 시장의 변화



출처: 과학기술정보통신부

서비스 시장현황을 살펴보면, 국내 O2O시장은 음식주문 서비스가 견인하고 있다고 볼 수 있다. 이러한 식품·음식 관련 O2O 시장의 성장은 생각대로, 바로고(Barogo)·부릉(VROONG)과 같은 배달대행 업체의 성장에 기반하고 있다고 해도 과언이 아니다.

3. 데이터 분석 기반의 유통 및 물류 서비스 혁신과 이슈

최근 말단 배송 중심의 유통 및 물류 시장의 변화는 국가 간 글로벌 물류시장이 아닌 일상생활에서 물류서비스의 집중적인 혁신이 주를 이루고 있다. 특히 신선식품의 경우 타 공산품에 비해 유통기한이 짧아 폐기율이 높을 뿐만 아니라 냉동·냉장 기능을 갖춘 물류 시설과 운송 차량이 필요해 수익성을 높이기가 쉽지 않다. 신선식품 새벽배송의 강자인 마켓컬리와 말단 배송의 혁신 사례인 쿠팡의 로켓배송 서비스가 이러한 문제를 어떻게 해결하는지에 대해 살펴보고자 한다.

가. 신선식품 배송 서비스의 성공 - 마켓컬리

마켓컬리의 새벽배송은 오후 11시까지 주문된 상품을 다음날 오전 7시까지 고객의 문 앞으로 배달하는 서비스이다. 제한된 지역에서만 제공되는 서비스이기는 하지만 다양한 종류의 신선식품을 8시간 내에 고객의 집 앞으로 배송하기 위해서는 가장 먼저 어떤 상품이 얼마만큼 판매될 것인지를 정확하게 예측할 수 있어야 한다. 마켓컬리에서는 상품별 수요 예측은 빅데이터 기반의 예측 시스템인 '데이터를 물어다 주는 멍멍이(데멍이)'를 활용한다. 빅데이터를 기반으로 머신러닝 알고리즘을 활용해 주간 단위 수요를 예측하며, 이를 상품을 준비 및 배송하는 팀에게 제공한다. 단순히 수요 예측 결과를 제공하는 데 그치지 않고 전날 매출과 고객 수 등과 같은 주요 판매 실적 데이터를 전사에 공유한다. 30분 단위로 현재 매출액과 당일 운영 마감 시 예상 매출액도 공유한다. 이렇게 전사에 공유된 데이터를 기반으로 각 기능 담당자들은 최적의 의사결정을 내릴 수 있게 된다.

과거 데이터로부터의 수요 예측에서 그치지 않고 실시간 데이터 공유를 통해 전체 조직의 운영 효율성을 높일 수 있었던 것이 바로 새벽배송을 가능하게 했던 성공 열쇠라고 이해할 수 있다.

나. 로켓배송 - 쿠팡

앞서 택배 서비스의 구조와 말단 배송에서 살펴본 바와 같이 온라인 쇼핑몰을 이용해

상품을 주문하면, 판매업체가 상품을 준비하고 택배회사로 인계한다. 택배사는 집하된 전체 물량에 대한 최적의 배송 계획을 수립하고, 이를 고객에게 전달한다. 이렇게 운영되는 택배 서비스로는 배송 시간을 하루 혹은 8시간 내로 줄이는 데는 물리적 한계가 존재할 수밖에 없다. 쿠팡의 로켓배송은 당일 배송을 위해 주로 판매되는 제품을 중심으로 미리 구매해서 물류센터에 보관하고 있다가 주문이 접수되는 당일 저녁 배송을 시작한다. 쿠팡은 랜덤스토(Random Stow) 방식을 이용하고 있다. 인공지능 기반의 알고리즘을 활용해 비슷한 제품군을 대량으로 보관하는 전통적인 방식에서 벗어나 각기 다른 제품을 소량 보관하는 방식을 적용하고 있다. 랜덤스토 방식이 가능한 이유는 소비자의 구매 패턴을 미리 예측할 수 있기 때문이다.

다. 데이터 기반 풀필먼트 센터 운영 시스템의 중요성

쿠팡과 마켓컬리의 물류센터는 전통적인 물류센터의 개념을 고객의 요구사항을 가장 가까운 곳에서 만족시킨다는 의미의 풀필먼트 센터로 발전시켰다. 풀필먼트 센터는 대량 입고 대량 출고를 위한 자동화 시스템을 통한 효율성을 강조하기보다는 빅데이터를 기반으로 수요를 정확하게 예측하고 고객에게 상품을 전달하는 배송 과정의 효율성을 높이는 데 집중한다. 소량 다품종의 고객 주문을 빠르고 정확하게 처리하기 위해서는 실시간으로 판매와 재고 데이터를 동기화할 필요가 있으며, 동기화한 데이터 기반의 최적 배송 계획을 수립해야 한다. 배송은 전통적인 택배가 아닌 자체적인 배송 네트워크를 운영하거나 쿠팡 플렉스(미국의 아마존 플렉스)와 같이 공유물류 서비스를 이용하거나, 매쉬코리아의 부릉과 같은 이륜차 배송 서비스와의 결합을 통해 이루어질 수 있다. 특히 자체 물류 네트워크를 확보하기 어려운 영세 및 소규모 인터넷 쇼핑몰의 경우 풀필먼트 센터를 운영할 수 있는 노하우나 시스템을 갖추기 어렵기 때문에 마이창고나 두손컴퍼니에서 제공하는 솔루션에 위탁 운영하는 방식을 도입하고 있다.

라. 유통 물류 서비스의 발전과 관련된 사회적 이슈들

인공지능과 빅데이터를 기반으로 온라인 중심의 유통과 물류 서비스가 확대되는 과정에서 소비자의 만족도는 높아지고, 다양하고 높은 수준의 서비스를 제공받을 수 있는 기회가 만들어지고 있는 것은 사실이다. 그렇지만 지속적으로 택배 단가가 하락하고, 물류 기업들이 수익성 측면에서의 문제를 겪고 있다는 점은 아직까지 전체 산업 생태계가 성숙되지 못했다는 사실을 보여주는 하나의 증거라고 볼 수 있다.

유통과 물류 서비스가 플랫폼을 만나서 그리고 플랫폼에 축적되는 데이터 기반으로 더 높은 수준의 서비스를 제공할 수 있게 된 것은 부정할 수 없는 사실이다. 그러나 O2O 시장의

경우 소비자와 서비스 제공자를 연결하는 플랫폼의 운영자에게 모든 데이터가 집중될 수밖에 없다. 데이터가 축적되는 위치가 플랫폼이고 플랫폼의 운영사가 이에 대한 독점적 권한을 가지기 때문에 플랫폼에 참여하는 소비자와 판매자, 플랫폼 노동자에 대한 정보 불균형은 다양한 사회적 문제를 만들어내고 있다. 만약 플랫폼 운영자가 독점적 지위를 활용해서 요금 책정과 소비자에 대한 추천 기준 등의 관점에서 불공정한 서비스 운영 기준을 적용한다면 그 피해는 소비자와 판매자 몫이다.

2019년 발의된 생활물류서비스산업발전법안은 이런 사회 이슈에 기반하고 있다. 내용은 택배서비스에 대한 선진화 방안이 주를 이루고 있으나, 소화물 배송 대행 서비스, 배송 기사의 근무조건과 계약 구조, 소비자 보호를 위한 제도 등이 함께 포함돼 있다. 이와 함께 이륜차 배송 기사들과 같이 플랫폼에 참여해 직접 서비스를 제공하는 노동자의 근무조건과 서비스 제공에 대한 대가 지불 기준에 대한 이슈도 지속적으로 제기되고 있다. 법안 세부 내용의 옳고 그름이나 실효성 문제는 제외하더라도, 전통적인 화물운송 서비스를 기준으로 만들어진 법안이 생활물류산업에는 적용하기 어렵기 때문에 새로운 적용 기준이 필요하다는 점은 모두 동의하는 바이다.

4. 데이터 기반의 유통 및 물류 산업의 혁신 가속화

4차산업혁명에는 세상의 모든 정보가 디지털로 전환되는 데서부터 출발한다. 유통과 물류산업 역시 이러한 디지털 전환을 통해 다양한 서비스가 등장했고, 물리적 한계를 극복하기 위한 여러 방법이 구체화됐다. 코로나19의 대유행은 디지털 전환을 가속화시키는 촉매제 역할을 하고 있으며, 더 많은 사람들이 빠른 속도로 오프라인에서 온라인 서비스로 이동하게 됐다. 결국 더 많은 데이터가 확보될 것이고, 개별 고객에 대한 이해 수준도 높아지고, 가까운 미래를 예측하기 위한 기술과 시스템이 완성될 것이다. 데이터에 기반한 유통과 물류서비스는 어떻게 변화할 것인지 정리해 본다.

가. 수평 및 수직 통합을 통한 서비스 확대

글로벌 공급사슬의 확대 과정에서 확인한 바와 같이 유통 물류 시장에서도 수직적 통합과 수평적 통합이 확산될 것이다. 현재 제공하는 서비스의 수익성을 높이기 위해서는 규모의 경제 실현이 필수적이기 때문에, 주요 경쟁자들은 수평적 통합을 통해 서비스의 양적 증대를 위해 노력할 것이다. 11번가가 신선식품의 새벽배송을 위해 헬로 네이처를 인수한 것과 같이

온라인 중심의 유통업체들은 서비스의 다각화를 위해 시장에 존재하는 중소 규모 서비스를 인수하는 방식을 활용할 가능성이 높다. 즉, 새벽배송과 같이 수많은 기업들이 참여하고 있는 전쟁터는 몇몇 주요 기업들에 의해 경쟁 구조가 재편될 것이다.

나. 예측 배송의 현실화

지금까지 그래왔듯이 많은 온라인 유통 기업들은 더 빠른 배송을 위해 다양한 방법을 시장에 내놓기 위해 노력할 것이다. 그 와중에 소비자들은 지금 이용하고 있는 새벽배송이나 당일배송, 로켓배송에 만족하지 못하고 더 빠르고 정확한 배송 서비스를 요구할 것이 당연하다. 결국 더 빠른 배송을 위해서는 더 일찍 주문과 배송이 이루어져야 하므로, 2014년 아마존이 이야기한 예측 배송(Anticipatory Shipping)이 실제 구현될 것으로 예상된다.

지금과 같은 속도로 정확하고 많은 데이터가 실시간으로 확보할 수 있는 기술과 환경이 구축된다면, 어떤 고객이 언제 어떤 제품을 주문할 것인지를 상당히 높은 확률로 예측할 수 있을 것이다. 결국 내가 주문도 하기 전에 풀필먼트 센터에서는 이미 픽킹과 포장과 같은 배송을 위한 작업 과정이 먼저 수행될 것이고, 주문과 동시에 배송 완료 메시지를 받을 수도 있을 것이다.

다. 개별 고객을 위한 서비스의 상품화 실현

고객에 대한 데이터가 실시간으로 수집되면, 이제 전체 시장을 몇 개의 그룹으로 나누는 데 시간과 에너지를 소비하기보다는 단 한 명의 고객을 위한 개인화된 서비스가 유통산업에 등장하게 될 것이다. 현재 대부분의 온라인 쇼핑몰이 제공하고 있는 상품의 단순 추천 서비스를 넘어 패키지 형태의 상품 구매와 연관된 서비스들이 결합돼 하나의 상품으로 제공되는 큐레이션 서비스가 빠른 시일 내에 등장할 것이다. 이를 큐레이션 커머스(Curation Commerce)라고 부를 수도 있고, 이미 유사한 개념이 2010년 초중반 소개된 바 있다. SSG닷컴이나 롯데ON, 현대백화점그룹과 같이 온오프라인의 다양한 유통 채널을 보유한 기업일수록 큐레이션 서비스를 선제적으로 도입해서 시장 경쟁력을 높일 수 있을 것이다. 다양한 상품을 직접 매입하는 쿠팡과 마켓컬리, 전자상거래 플랫폼을 제공하는 11번가나 지마켓 등의 기업 역시 입점 기업의 상품을 대상으로 유사한 서비스를 제공할 수 있을 것이다.

제 5 부

데이터산업 기술 동향

제1장.
데이터 생산 및 수집 기술

제2장.
데이터 저장 및 관리 기술

제3장.
인공지능 기술

제4장.
데이터 보안과 비식별 기술

제5장.
데이터 분석 및 시각화 기술

제1장. 데이터 생산 및 수집 기술

필자: 안필용
(인젠트 상무)

현대 산업에서 데이터는 중요한 경쟁요소다. 빅데이터나 인공지능 기술은 대량의 데이터 축적이 필수적으로 요구되고 있다. 대량의 데이터를 축적하기 위해서는 자동화한 데이터 생산 기술과 데이터 수집 기술이 필요하다. 최근에는 외부에 흩어져 있는 무수히 많은 비정형 데이터를 어떻게 수집해서 저장·관리하는지가 생산이나 수집에서 핵심기술로 주목받고 있다.

1. 데이터 생산 및 수집 기술 개요

데이터를 업무에 활용하기 위해서는 먼저 필요한 데이터를 확보해야 한다. 데이터를 확보하는 방법으로 데이터를 직접 생산하거나 외부 데이터를 수집하는 방법이 있다. 전통적인 데이터베이스(DB) 환경에서는 프론트엔드 애플리케이션에서 필요한 데이터를 생산하는 방법을 주로 사용했다면, 최근에는 내부에서 생산되는 데이터보다 외부에서 수집되는 데이터 비중이 훨씬 올라가고 있다. 과거에는 정형 데이터의 자체 생산에 초점이 맞춰져 있었다면, 앞으로는 외부의 비정형 데이터를 수집하는 방향으로 전환할 것이다.

외부로부터 수집되는 데이터는 장비에서 기계적으로 생성되는 데이터가 상당 부분 포함돼 있어 그 양에서 차이가 크게 난다. 따라서 데이터 수집 단계에서는 성능이 시스템의 중요한 요소가 된다. 또한 외부의 데이터는 소스가 다양해, 이를 수집하기 위해서는 여러 수집 방식을 동원해야 한다. 수집 대상이 되는 외부 데이터는 DB나 파일의 정형 데이터뿐만 아니라 텍스트, 로그, 음성, 영상 등 비정형 데이터가 포함돼 있어 비정형 데이터를 수집해 분석 목적에 적합하도록 저장하는 기술이 필요하다.

2. 데이터 생산

조직의 업무 처리나 의사결정은 내부와 외부에 존재하는 여러 유형의 데이터를 생산해 축적하는 것부터 시작된다. 애플리케이션을 활용해 데이터를 입력하거나 이를 가공해 재생산하는 전통적인 방식과 기기나 센서를 통해 자동으로 데이터를 생산하는 방식이 있다.

가. 데이터 입력 방식

데이터를 입력해 생산하는 방식에는 조직 내에서 자체적으로 개발한 업무용 애플리케이션을 통해 생산되는 정형 데이터를 이용하는 방식과 SNS나 블로그 등과 같이 개인이 생산하는 비정형 데이터를 이용하는 방식이 있다.

조직 내에서 데이터가 생산되는 환경에서는 데이터 생산 시점부터 완료 시점까지 정합성 보장을 위한 트랜잭션 관리가 매우 중요하다. 만약 특정 사용자의 정보를 수정할 경우, 수정이 완료되기 전까지는 과거 정보로 일관되게 표현돼야 한다. 이후 이 작업이 완료된 것이 확정된 때부터 새로운 정보로 나타나야 하는 것이 원칙이다.

데이터는 업무 목적에 맞게 사용되도록 일관되고 표준화돼야 하므로 데이터는 정형화한 형태를 띠게 된다. 이런 성격 때문에 정형화한 구조의 데이터를 관리할 수 있고 락(Lock 또는 Enqueue)과 같은 오브젝트로 트랜잭션을 보장해주는 관계형 DB가 데이터 생산에 중요한 기술로 쓰였다.

SNS나 블로그 같은 웹을 통해 개인이 생산하는 데이터는 조직 내 데이터와 달리 제약 없이 자유로운 형태를 띠고 있다. 이러한 불특정 개인이 웹을 통해 생산하는 데이터는 여기저기 분산돼 있으므로 생산보다는 수집이라는 관점에서 다시 다루도록 하겠다.

나. 데이터 자동 생산 방식

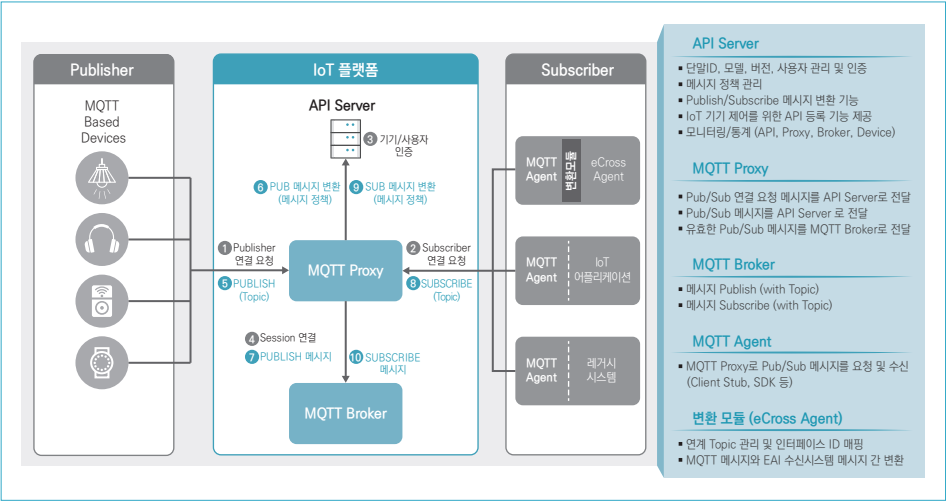
사물인터넷(Internet of things, IoT) 기술이 확산되면서 데이터 생산 방식이 다양화되고 있으며, 센서·로그 데이터·SNS가 제공하는 스트리밍 데이터 등 생산되는 데이터의 저장 및 관리 기술이 점점 더 중요해지고 있다.

최근 IoT 플랫폼은 서비스를 구성하는 데 필요한 공통 요구 기능을 포함하고 있으며, 개별 사물과 서비스에서 독립적으로 동작할 수 있다. 이 플랫폼은 서버나 클라우드 형태로 제공될 수 있으며, Pub/Sub, Topic 방식의 메시지 송수신 기능을 제공하는 IoT 어댑터 등을 제공해 디바이스에 직접 위치할 수도 있다.

[표 5-1-1] 사물인터넷 데이터 생산 방식

구분		데이터 생산 방식
스마트 홈	구글 네스트	CCTV, 기상정보 연계, LED/온도 센서 네트워킹
	애플 홈킷	잠금장치, 조명, 카메라, 온도 조절, 플러그, 스위치 등과 아이폰 연결
스마트시티	스마트 주차장	바닥 센서 네트워킹
	스마트 표지판	디지털 방향조절 표지판과 연결
스마트 물류	아마존 물류 창고 로봇 '키바'	제어센터와 운반 로봇 통신
	스마트 사물함	디지털사물함과 스마트폰 연결
스마트 카	볼보 커넥티드 카	충돌센서 정보 수집, 차량 제어 기능 등 스마트폰 연결
	폭스바겐 커넥티드 카 + 스마트 홈 서비스	차량 제어 기능에 집안의 조명, 난방, 가전기기 제어 기능 연결

[그림 5-1-1] IoT 플랫폼 구성



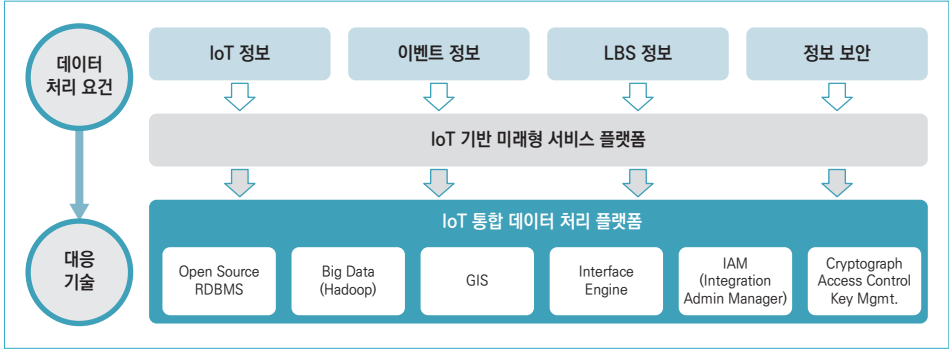
IoT 플랫폼 기술은 초기에는 센서·액추에이터(Actuator)·데이터 중심의 특징을 기반으로 수백만 개의 인터넷 기반 서비스 도메인 수준의 규모를 지원했으나, 점차 데이터·프로세스·지능 중심으로 바뀌어 수백억 개 이상의 글로벌 규모로 발전하고 있다.

이러한 IoT 플랫폼에 대한 표준은 IoT 글로벌 협의체를 바탕으로 개방형으로 운영 중이다.

다. 데이터 저장

IoT 기반의 미래형 서비스에 적합한 기술을 구현하기 위해서는 데이터를 저장하는 요건들이 필요하다. IoT 데이터 저장을 위해 고성능 대용량 데이터 처리가 가능한 오픈소스 RDBMS를 제공해야 하며, LBS(Location Based Service) 데이터 저장을 위해서는 국제표준(ISO)을 따르는 지리(위치) 정보 처리 및 관리 기능을 위한 오픈소스 GIS-DBMS를 제공해야 한다. 또한 방대한 데이터를 처리하기 위한 스트리밍 프로세스(Streaming process)와 분산 처리⁰¹를 위한 오픈소스 하둡(Hadoop)⁰² 플랫폼을 제공해야 한다. 새로운 데이터 보안 위협에 대응하기 위한 국가 표준 알고리즘(ARIA-128/256)의 적용 및 엄격한 키 관리를 통한 데이터 암호화 기능과 사용자, 애플리케이션, IP 등의 조건에 따른 데이터 접근 제어/관리 기능을 제공해야 한다. 마지막으로 실시간 데이터 흐름 및 시스템 상태 모니터링을 위한 통합 관제 기술을 제공함으로써 안정적인 데이터 저장을 유도할 수 있다.

[그림 5-1-2] IoT 기반 서비스 데이터 저장 및 처리를 위한 SW 기술요소



3. 데이터 수집

데이터 수집 기술은 데이터의 형태나 종류에 관계 없이 수집할 수 있어야 한다. 단 데이터의 형태나 종류에 따라 수집방법이 달라지므로 다양한 수집방법이 필요하다. 수집한 데이터를 가공·재생산할 수 있어야 하며, 수집방법에는 데이터를 가공·재생산하는 과정을 포함해야 한다. 가시적 형태로 서비스를 제공하고 데이터를 효율적으로 활용할 수 있도록 저장해야

01. HDFS, Hadoop Distributed File System
02. Hadoop, High-Availability Distributed Object Oriented Platform

하며, 정형/비정형 형태로 데이터를 저장하는 기능을 제공해야 한다.

- 수집 기술: EAI/ESB, APIM, 크롤링, RSS Reader, Aggregator(Log, RDB), Streaming

가. EAI/ESB: 파일, DB, 웹 서비스에서 데이터 수집

EAI(Enterprise Application Integration)는 기업 내외부의 서로 다른 시스템을 통합하기 위해 사용하는 기법으로, 개별 애플리케이션을 에이전트를 이용해 중앙 허브와 연결하고 중앙 허브를 통해 상호 데이터를 수집한다.

기업 내외부의 데이터 수집은 시스템 통합 차원에서 중요한 의미를 지니고 있다. 즉, 엔터프라이즈 기업의 여러 시스템에 흩어져 있는 정보들은 기업 내 의사결정 시 신속하게 공유돼 합리적인 방안을 수립하는데 중요한 요소로 작용한다. EAI 기술은 이를 효율적으로 구현하기 위한 많은 노하우가 축적돼 있다.

EAI 기술은 TCP/IP 기반 소켓프로그래밍에서 출발해 다양한 인프라 및 리소스에서 데이터를 수집하는 기능, 데이터 매핑 및 변환, 콘텐츠 기반 서비스 라우팅, 무중단 서비스, 보안 등 비즈니스 파트너와의 원활한 협업 시스템을 구축하기 위해 필요한 기능을 필수적으로 제공한다. EAI 영역에서 취급하는 리소스 형태는 데이터베이스, 파일, 애플리케이션, TCP 서비스, 웹서비스 등 그 종류가 다양하다. 인터페이스 아키텍처는 과거의 피어 투 피어(Peer to Peer) 방식을 지양하고, 허브 앤 스포크(Hub & Spoke) 방식, 버스 방식 등 구성 방안의 다양성을 제공해 고객 맞춤형 연계 시스템을 구축할 수 있는 장점이 부각되고

있다. 특히, 애플리케이션 시스템에 제공되는 에이전트는 인터페이스 데이터에 대한 수집부터 수신 시스템 내 적재까지 모든 업무를 대행함으로써 개발이 편리하고 연계 시스템 구축이 수월하다는 장점을 지니고 있다.

이처럼 EAI는 데이터웨어하우스(DW)나 전사적자원관리(ERP) 등 기업에서 운영하는 서로 다른 애플리케이션을 네트워크 프로토콜이나 DB, 운용체제(OS)에 관계없이 비즈니스 프로세스 차원에서 통합하는 것을 말한다. 이를 통해 기업 내 상호 연관된 모든 애플리케이션을 유기적으로 연동해 필요한 정보를 중앙 집중적으로 통합관리할 수 있는 환경을 구현하는 기술이다.

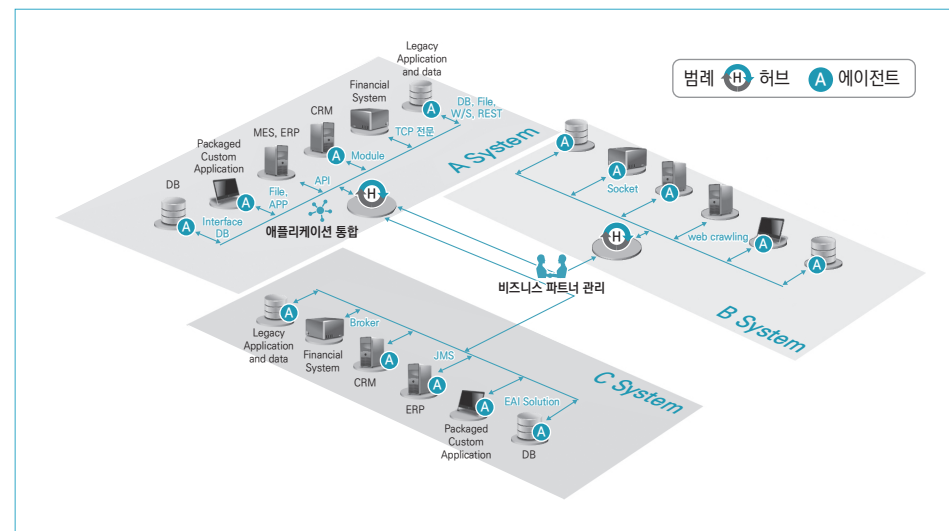
ESB(Enterprise Service Bus)는 애플리케이션 서비스들을 컴포넌트화된 논리적 집합으로 묶는 핵심 미들웨어이며, 비즈니스 프로세스 환경에 맞게 설계 및 전개할 수 있는 아키텍처를 제공한다.

ESB는 기본적인 메시징, EAI, 중개(라우팅) 기술을 사용하거나 J2EE 시스템의 서비스 통합 버스와 같은 플랫폼 내의 구성요소를 사용하는 등 다양한 방법으로 구현될 수 있다. ESB는 EAI와 애플리케이션 서버 기술을 함께 조합해 구현할 수 있지만 이러한 구현 방식이 전반적인 아키텍처에 영향을 미치지 않아야 한다. ESB는 전사 컴퓨팅 인프라스트럭처에 공통으로 분산돼 있는 여러 서비스, 애플리케이션, 다양한 데이터 등을 연결하는 지능형, 분산형, 트랜잭션형, 메시징 레이어로 작용한다. 또한 지능형 전송 기능과 라우팅 기능으로 동기 및 비동기 메시징 백본을 형성해 메시지를 신뢰성 있게 전달한다. ESB는 개발자가 WSDL(Web Service Description Language) 기반 표준 인터페이스에 의해 정의한 서비스, 즉 컴포넌트화된 비즈니스 기능을 API 또는 프로토콜에 관계없이 사용 및 호출하는 것을 가능하게 한다. WSDL은 추상화된 서비스 인터페이스 정의 부분, 재사용 가능한 프로토콜 바인딩 부분 그리고 서비스를 제공하는 종단점(endpoint)으로 나뉘어 있다. 이것은 본질적으로 확장이 가능하며 필요에 따라 다양한 프로토콜을 동일한 서비스에 연결할 수 있도록 포트 및 바인딩을 위한 확장 요소를 제공한다.

ESB는 웹서비스 관련 표준기술을 지원하는 웹서비스 어댑터를 지원해 외부 시스템의 데이터 수집 요청에 대해 표준 웹서비스 방식의 수신 기능을 제공하는 특징을 지니고 있다. 필요 시 타 기관 또는 응용시스템의 웹서비스를 호출하기 위해 위저드 기반 클라이언트 스텝(stub)을 생성하는 기능을 제공한다. 또한 오픈 API 서비스를 위한 Restful 방식의 API 서비스 처리 기능을 지원해 외부의 API 서비스 호출에 대해 표준 REST(REpresentational State Transfer) 방식의 서비스를 제공한다. 하지만 표준 REST 방식에서 제공되지 않는 인증·보안, 유량 제어 등 API 서비스를 위한 기능이 추가로 제공돼야 한다.

이처럼 ESB는 기업 및 기관의 서비스지향 분산 컴퓨팅 기술의 일종으로 기존의 분산

[그림 5-1-3] 엔터프라이즈 기업의 EAI 구성



컴퓨팅 기술인 CORBA(Common Object Request Broker Architecture), DCOM(Distributed Component Object Model), RMI(Remote Method Invocation) 등을 기초로 한 ‘분산 오브젝트’ 환경의 웹 서비스와의 연계를 통한 데이터 수집에서 구현이 용이하고 체계적으로 관리할 수 있다.

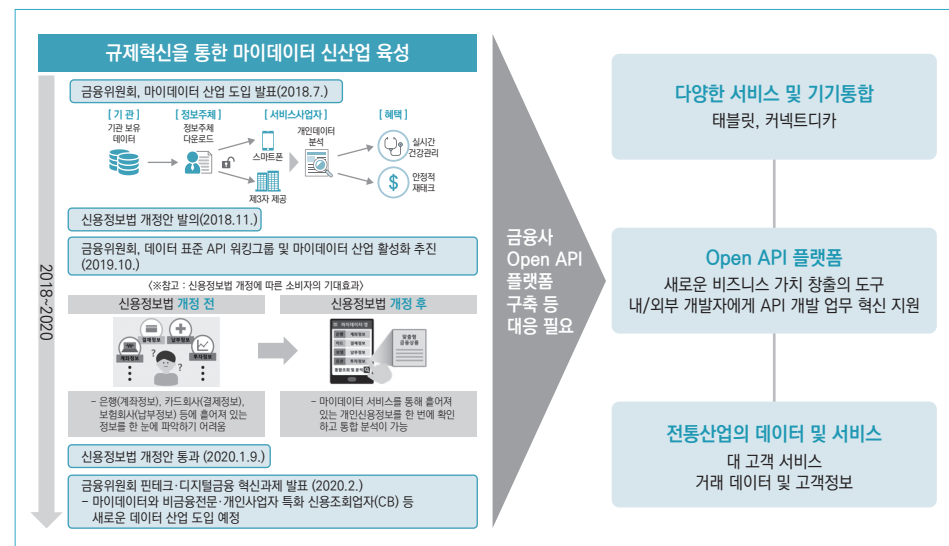
웹 서비스는 SOAP(Simple Object Access Protocol), WSDL(Web Service Description Language), UDDI(Universal Description, Discovery and Integration)와 같은 공개 표준을 정해 이를 근간으로 하여 상호작용이 이루어진다. 그러므로 웹 서비스를 지원하는 응용 프로그램은 상호 작동에 문제가 없고 플랫폼과 언어에 독립적이며, HTTP와 같은 웹 프로토콜을 사용하므로 방화벽과 같은 장애에도 문제가 없어 응용 프로그램에 대한 접근이 용이하다.

나. API 플랫폼

Open API(Open Application Programming Interface, 오픈 API)는 데이터 소유 주체가 웹 개발자나 사용자를 위해 정보/데이터를 정해진 방식으로 공개하는 기술을 뜻한다. HTTP 프로토콜 기반의 웹 서비스로서 제공 기술은 SOAP 기반의 웹 서비스와 REST 기반의 웹 서비스로 구분된다. 다만 개방형 기술의 보안 이슈를 해결하기 위한 인증, 보안, 접근제어, 정책관리 등의 개발 소요가 발생할 수 있으므로 도입 전 해당 기능의 필요성과 구축비용을 면밀히 검토할 필요가 있다.

최근 코로나19로 사회 전반에서 많은 변화가 일어나고 있다. 지난 3월 맥킨지는 ‘Next Normal’이라는 표현을 쓰며 코로나19로 사회 전반에 큰 변화가 생길 것으로 예상했다.

[그림 5-1-4] 데이터3법 개정에 따른 오픈 API 플랫폼 필요성



실제로 세계 금융권에서는 비대면 수요가 급증하고 있으며, 핀테크 기업에 대한 수요가 증가하고 있다. 국내 금융권도 오픈 API를 활용한 다양한 비대면 서비스를 선보이고 있다.

‘Next Normal’ 트렌드에 따라 기업 및 기관들이 플러그 앤 플레이 방식의 단순한 아키텍처 기반의 오픈 API 플랫폼을 구축해 비즈니스 민첩성 강화, 통합비용 감소, 나아가 공급기업에 대한 종속성 방지 등의 요구를 해소하고 있다.

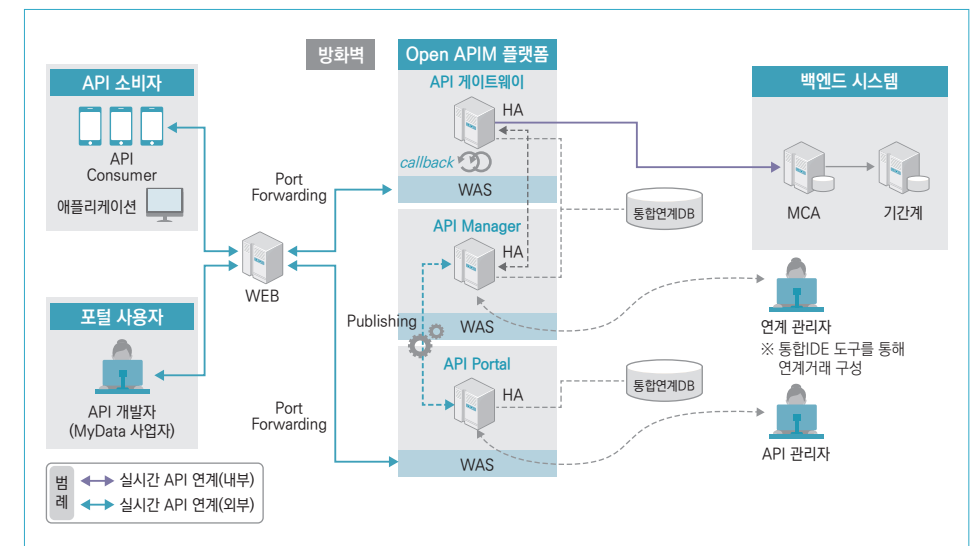
데이터3법 중 신용정보법(신용정보의 이용 및 보호에 관한 법률) 개정안의 국회통과 (2020.1.9.)로 마이데이터 정보 제공 등 API를 통한 개방형 플랫폼 사업들이 진행되고 있으며, 관련 산업 성장에 대한 기대감이 높아지고 있다. 오픈 API는 이미 보편화되고 가속화되고 있으며, 비즈니스적·기술적으로 핵심 영역이 됐다. 이에 따라 효율적이고 안정적인 API 관리를 위한 오픈 API 플랫폼의 역할이 점점 더 중요해지고 있다.

API 관리 플랫폼은 각 서비스의 API를 안전하게 공유하고 쉽게 사용할 수 있도록 개발자 포털을 제공하고, API 서비스들이 어떻게 사용되고 운영되는지를 분석하며, 그 외 API의 트렌드 확인 등 API에 대한 전반적인 관리 기능을 제공해야 한다.

공공기관을 시작으로 일반기업까지 오픈 API 서비스를 늘려가고 있다. 이러한 오픈 API 수요의 결과로 API 관리 및 운영에 대한 효율성을 높이기 위해 APIM(API Management) 기술이 대두되고 있다. APIM 기술에서는 개발 생산성 향상과 기간 단축을 위해 API 게이트웨이, API 라이프사이클 관리를 위한 API 포털 등 다양한 기능이 담긴 소프트웨어를 제공하고 있다.

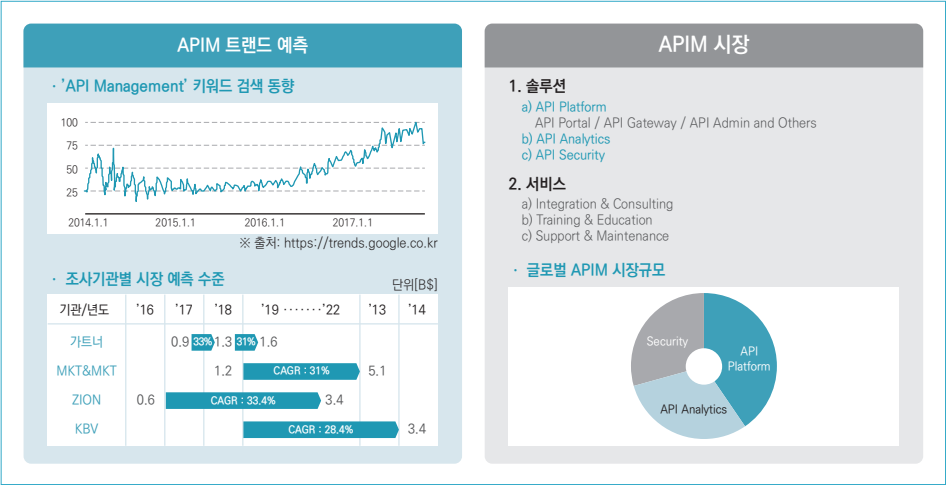
API 라이프사이클 기반의 관리 기능은 API 생성, API 게시, 인증 및 보안, API 정책관리,

[그림 5-1-5] 오픈 API 플랫폼 구성도



API 검색, 서비스 개발, API 사용, 모니터링 및 분석 순으로 구성돼 있다. API 생성에서 요구되는 서비스 오케스트레이션 기능은 소단위 REST API를 조립해 단일 응답으로 처리해야 하는 API 서비스 제공에 적합하며, 백엔드 시스템의 DB, TCP 등 프로토콜로 변환이 요구되기도 한다.

[그림 5-1-6] 글로벌 APIM 시장 현황



출처: Global API Management By Component, KBV Research

오픈 API는 데이터 공유를 위한 표준 방식으로 자리매김하고 있고 SNS나 웹 콘텐츠와 달리 정형화되고 검증된 정보를 수집할 수 있는 방법이므로 데이터 수집에서 중요한 역할을 담당한다. 오픈 API는 공공 분야에서 시작해 자산화한 데이터들을 수집해 새로운 서비스를 제공하려는 핀테크 사업자, 마이데이터 사업자 등으로 확산되고 있다. 이렇게 증가하는 오픈 API의 수요에 대응하기 위해서는 APIM 같은 API 관리 플랫폼의 역할과 비중이 갈수록 커질 것이다.

다. 크롤링

크롤링은 API를 제공하지 않는 기존 애플리케이션과의 연결성을 해결하는 기법으로, 웹 상의 비정형 콘텐츠를 수집한다. 다시 말해 주로 웹 로봇을 이용해 조직 외부에 존재하는 소셜 네트워크 데이터 및 인터넷에 공개된 자료를 수집하는 기술이다. 이를 프로그래밍으로 자동화하기 위해 다음의 세 가지 기법이 활용되고 있다.

1) HTML 페이지를 가져와 HTML/CSS 등을 파싱하고, 필요한 데이터만 추출하는 기법

- 2) 오픈 API(REST API)를 제공하는 서비스에 오픈 API를 호출해 받은 데이터 중 필요한 데이터만 추출하는 기법
- 3) Selenium, PhantomJS 등 브라우저를 프로그래밍으로 조작해 필요한 데이터만 추출하는 기법

이외에도 로그인이 필요한 웹페이지 크롤링 기법, XPATH를 이용한 기법 등이 있어 데이터 수집을 위해 표준화하기 전에 빈번한 커스터마이징 개발이 필요하다는 단점이 있다. 따라서 크롤링 기법을 적용하기 전에 철저한 검토가 필요하다.

라. 기타 데이터 수집 기술

- RSS Reader: 웹 기반 최신 정보를 공유하기 위한 XML 기반의 콘텐츠 배급 프로토콜인 RSS 서비스를 통해 데이터를 수집하는 기술로, RSS 사이트에서 제공하는 주소를 RSS Reader에 등록하면 자동으로 데이터를 전송받아 수집하는 기술
- Log Aggregator: 조직 내부에 존재하는 웹서버 로그, 웹로그, 트랜잭션 로그, 클릭 로그, 데이터베이스 로그 등 각종 로그 데이터를 수집하는 오픈소스 기술(Chukwa, Flume, Scribe 등)
- RDB Aggregator: 관계형 데이터베이스(RDBMS)에서 정형 데이터를 수집해 HDFS(하둡 분산파일시스템)나 HBase와 같은 NoSQL에 저장하는 오픈소스 기술(Sqoop, Direct, JDBC 등)
- Streaming: 인터넷에서 음성, 오디오, 비디오 데이터를 실시간으로 수집할 수 있는 기술

제2장. 데이터 저장 및 관리 기술

필자: 허승재
(티맥스데이터 상무)

데이터가 폭증하는 시대에서 클라우드 환경에 필요한 DBMS 기술은 무엇이 있을까? 클라이언트 접속량과 데이터 처리 부하에 따라 DBMS 컴퓨팅 자원을 탄력적으로 스케일링하고 다양한 형태의 데이터 통합과 AI 기반 DB 관리가 가능한 새로운 DBMS 기술이 필요한 시점이다.

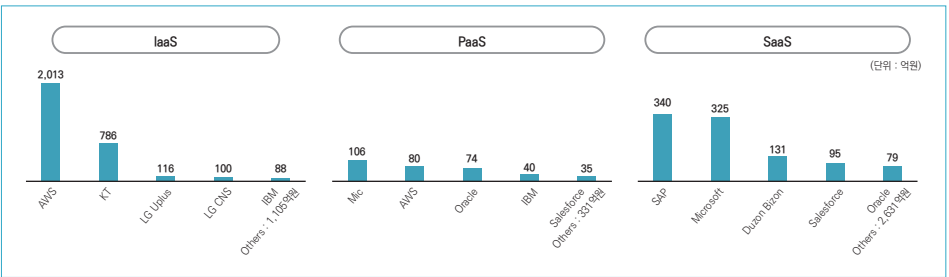
1. 데이터 기술 변화의 동인

5G 시대, 인공지능 기반의 4차산업혁명 시대를 맞아 자동화와 첨단 서비스 구현의 근간이 되는 데이터의 중요성이 더욱 커지고 있다. 디지털 데이터는 과거와 달리 석유와 같은 원자재나 필수 자원으로 대접받고 있다. 최근에는 데이터3법 개정과 같이 국가 차원의 보호 및 활성화 정책을 유도해내는 가장 영향력 있는 미래의 자산으로 다뤄지고 있다.

초고속 네트워크, 사물인터넷(IoT) 기술의 발달에 따라 데이터 생산 주체가 다양해지면서 저장되는 데이터 형태 역시 다양해지고 있다. 이뿐만 아니라 데이터 양은 100배 이상 폭증하고 초 저지연 처리 기술이 구현되며 지능적인 데이터 저장과 활용 기술이 요구되고 있다.

데이터 저장, 관리, 활용을 담당하는 핵심 소프트웨어인 DBMS 기술은 이러한 변화에 적응하고 있는가? 폭증하는 다양한 형태의 데이터 프로세싱을 위한 여러 가지 시도가 진행되고 있다. 대표적인 적용 기술로 인프라 측면에서 컴퓨팅 파워를 탄력적으로 활용해 기존 단일 박스 형태의 하드웨어가 가진 구성의 한계를 병렬화, 가상화 방식으로 극복하는 클라우드 기술이 있다. 또 다른 기술로는 데이터 최적 관리, 보안 관리, 모니터링을 자동화하고 DBMS를 자율 운영할 수 있는 인공지능 기술의 적용이 시도되고 있다.

[그림 5-2-1] 국내 클라우드 시장 현황



출처: 한국IDC, 2018.

클라우드가 도입되며 가장 먼저 인프라 가상화(IaaS) 기술이 적용돼 국내에서만 1조 1,000억 원 규모의 시장을 형성했다. 앱서비스 제공(SaaS) 기술은 특정 서비스에 한정돼 빠르게 성장하는 반면, 클라우드 기반 플랫폼 서비스(PaaS) 기술은 성장이 정체돼 있고 시장에서도 도입이 주저되고 있는 것이 현실이다. 그 이유로, 플랫폼 서비스의 핵심은 데이터 관리 기술이고 현존하는 DBMS 기술로는 클라우드 환경에서 풀지 못하는 기술적 난관이 존재하기 때문이다.

2. 클라우드 DBMS 방향성과 한계

현재 DBMS 벤더들은 각자의 방식으로 클라우드 환경에서 최적 기술을 적용하기 위한 방안을 강구하고 있다. 기존 온프레미스 환경에서는 DBMS를 위한 고정된 서버 하드웨어와 SAN 기반 DB 전용 스토리지, 고속의 네트워크가 제공된다. 이와 달리 저가 범용 서버 하드웨어와 네트워크 기반 스토리지를 기본으로 하는 클라우드 환경에서는 기존 온프레미스 수준의 성능을 보장하면서 확장성, 유연성, 관리 효율성을 달성하는 데 어려움을 겪고 있다.

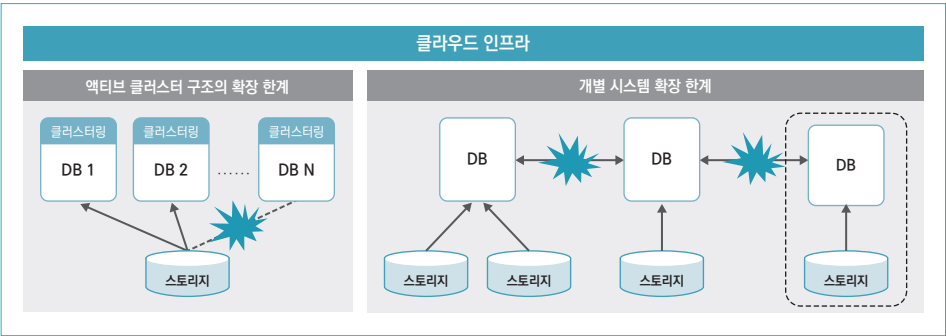
가. DB서버 확장의 한계

정형 데이터를 주로 다루는 관계형 데이터베이스(RDBMS)에서 높이 평가받는 기술은 RAC(Real Application Cluster, 오라클), TAC(Tibero Active Cluster, 티베로), PureScale(IBM DB2)과 같은 DB 서버 액티브 클러스터링 기술이다. 벤더별 용어와 상세 스펙은 조금씩 다르지만, DBMS 단일 스토리지를 공유함으로써 데이터 정합성을 유지하면서, 데이터 처리(저장, 변경, 삭제, 관리)를 수행하는 DB 서버를 다중화하고 클러스터링 함으로써 확장성, 성능, 가용성을 보장하고자

하는 목적은 동일하다.

물리적 스토리지를 여러 DB 서버가 공유한 상태에서 DB 서버 노드 확장을 지속하는 것은 동일 데이터 경합 발생, 노드 간 통신부하 증가 등의 문제를 발생시켜 더 이상의 성능 향상을 보장할 수 없게 만든다. 범용 저사양 컴퓨팅 자원을 병렬화해 사용해야 하는 클라우드 환경에서는 DB 클러스터 노드 확장성 제약이 더욱 심해지므로 DB 성능과 관리 측면에서 한계가 발생하게 된다.

[그림 5-2-2] DB 서버 확장의 한계

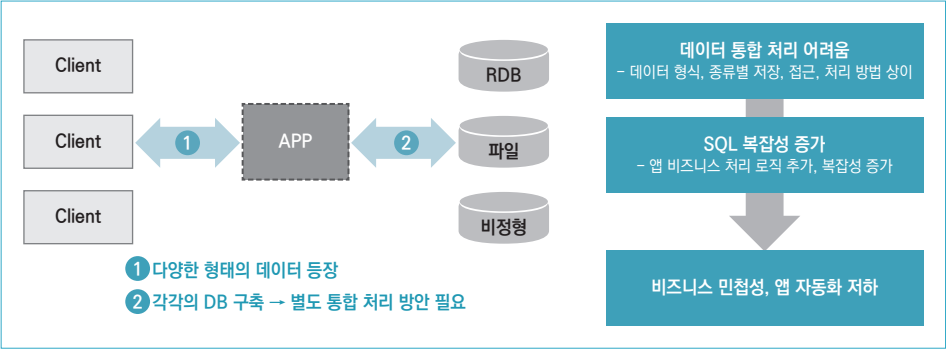


MSA(Micro Service Architecture) 같은 방식으로 DB 업무를 분리해 개별 시스템간 성능을 유지하고자 해도 EAI, ETL, Link와 같은 DB 간 추가 연계 작업, 기업 내 데이터 품질, 정합성 유지 측면에서 여전히 문제가 존재한다

나. 다양한 데이터 통합, 활용의 한계

데이터 생산 주체가 다양해지며 여러 형태의 데이터가 등장했으며, 각각의 DB가 별도 구축되고 별도 처리 기술이 적용되고 있다. 즉 RDB 정형, 반정형 파일, 비정형 콘텐츠 등

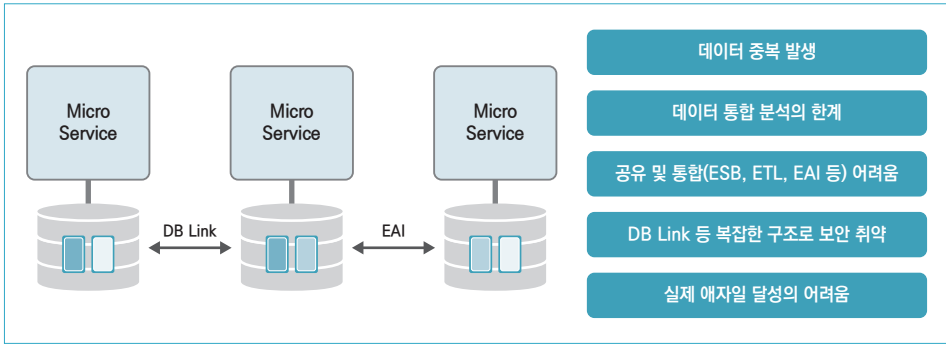
[그림 5-2-3] 데이터 통합의 한계



데이터 형태별로 별도의 저장 방법이 필요하고 조회, 분석을 위한 별도 기술이 사용되고 있다. 결국 데이터 통합과 활용이 어려워지고 그로 인해 업무시스템의 애플리케이션 복잡성이 증가하고 비즈니스 민첩성(Agility)이 저하되는 결과를 가져온다.

비즈니스 민첩성을 달성하기 위해 채택되는 MSA 방법도 데이터 형태별 통합 활용 문제와 다양한 데이터 간 연계 문제로 본래 달성하고자 하는 의미가 퇴색되고 있는 상황이다.

[그림 5-2-4] 현재의 MSA 구조



다. 기존 기술 개선, 신기술 적용 한계

기존 DBMS 벤더와 CSP(Cloud Service Provider)는 클라우드 DBMS 기술을 적용하기 위해 노력하지만, 아직은 중급 이하 업무를 대상으로 하는 클라우드 마이그레이션 위주다. 기업의 핵심업무 이관이 가능해야 하고 이를 위해 클라우드 DB 성능과 가용성이 보장돼야 한다. 또 클라우드 특성과 장점을 활용한 업무 DB로서 부하 상황에 따른 탄력적 자원 활용이 가능한 스케일링 동작, AI 기반의 관리자동화 등 핵심 기술 적용이 필수이지만, 아직까지는 기존 방식의 근본적 제약에서 벗어나지 못하는 상태다.

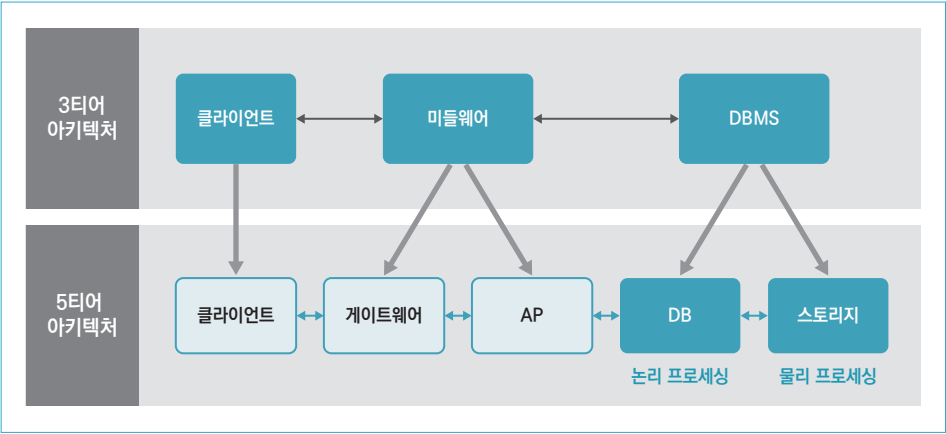
3. 클라우드 DBMS 기술의 필요 요소

클라우드와 AI 시대에 차세대 DBMS가 갖추어야 할 기술 요소는 네 가지가 있다. 성능 저하 없이 무한 확장이 가능한 Scalable DB, 비즈니스 민첩성을 갖춘 DB, 다양한 업무가 수용되는 DB, AI기술 전면 적용이 그것이다.

가. 성능 저하 없는 스케일러블 DB

기존 클러스터 구조의 DBMS 확장 제약 문제를 해결하기 위해 데이터의 논리적 처리(Logical Processing)를 담당하는 DB서버와 물리적 처리(Physical Processing)을 담당하는 스토리지서버를 명확히 분리하고 역할을 나눠 확장이 자유로운 구조로 바꾸는 것이 필요하다. 즉, 1티어 DB구조를 2티어 구조로 바꾸어 티어별 프로세싱 역량을 높이고 논리적, 물리적 확장성을 별개로 갖추어 기존 스토리지 서버 간 답답한 구조를 뜯개어 놓는 것이 필요하다.

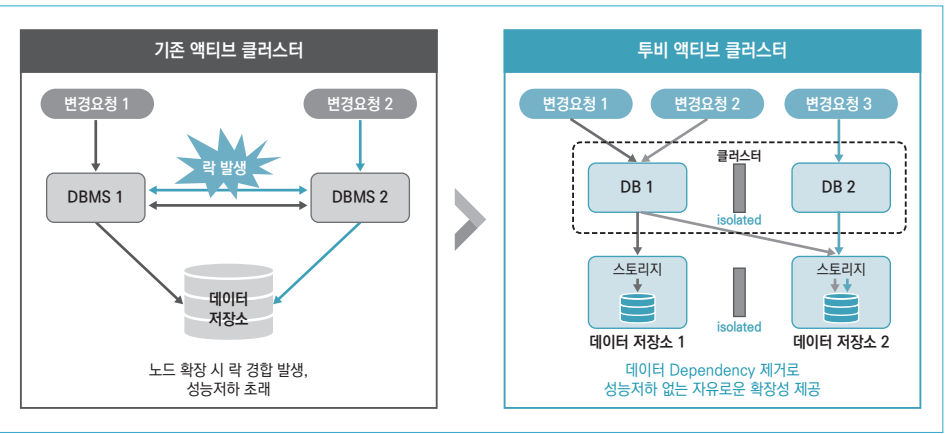
[그림 5-2-5] 5티어 아키텍처와 2티어 DB 구조



클러스터 노드 확장 시 성능 저하를 가져온 락(Lock) 경합 발생과 DB 노드 간 통신 오버헤드를 없애는 기술이 가장 핵심이다. 이를 위해 데이터를 처리할 때 DB 서버 간에는 논리적 처리의 독립성이 보장돼야 하고 타 노드 의존성은 제거돼야 한다.

실질적인 데이터 처리는 스토리지 서버가 담당해야 한다. 기존에 단순 저장용도로 쓰인 스토리지가 아닌 자체 락 관리와 트랜잭션 순서 보장, 디스크 IO 등 기존 DB 서버가 담당한 데이터 처리 기능을 스토리지 레이어에서 구현해야 한다. 이러한 구조를 통해 부하 상황에 맞게 DB 서버(프론트엔드), 스토리지 서버(백엔드) 간 성능 저하 없는 자유로운 확장이 가능해진다.

[그림 5-2-6] DB 노드 간 경합 구조

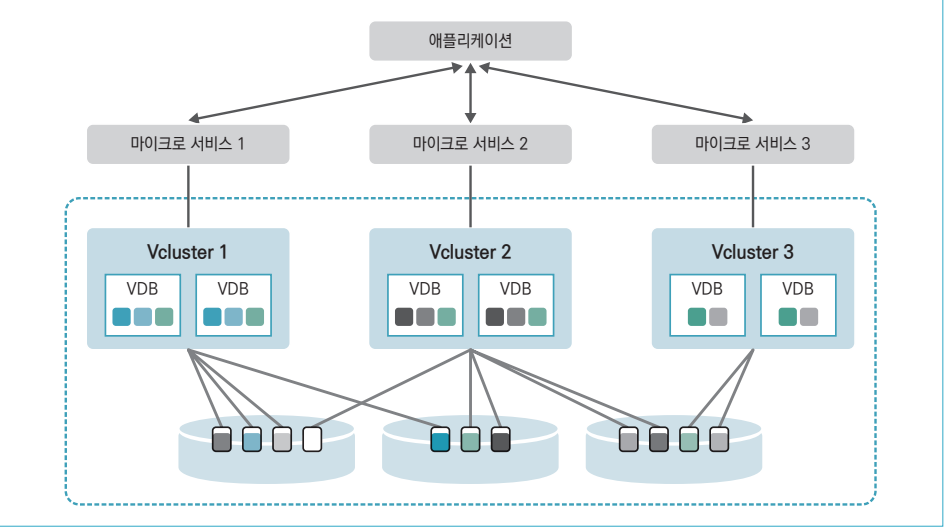


나. DB 가상화를 통한 애자일 구현

비즈니스 민첩성을 달성하기 위해 업무를 잘게 나눠 서비스화해 분산시키고 애플리케이션 개발도 그에 맞는 방법론을 채택하고 있다. 그러나 DBMS 차원의 민첩성 달성을 위해선 DB 가상화 기술이 반드시 필요하다.

DB 가상화(DB Virtualization)는 물리적으로 저장되는 데이터, 테이블과 별도로 사용자나 업무 애플리케이션에서 필요로 하는 테이블을 논리적으로 집합시켜 놓은 개념이다. 기업내 데이터 무결성, 품질 보장 측면에서도 필요하고 복잡한 데이터 연계작업을 제거하는 유일한 방법이기도 하다.

[그림 5-2-7] 가상 DB 구조



DB 서버와 스토리지 서버로 분리함으로써 DB 가상화가 가능하고, 스토리지 서버에 저장된 테이블 메타정보를 이용해 DB 서버 레이어에서는 필요한 테이블만 뽑아 가상화된 데이터베이스를 구성하게 되는 것이다.

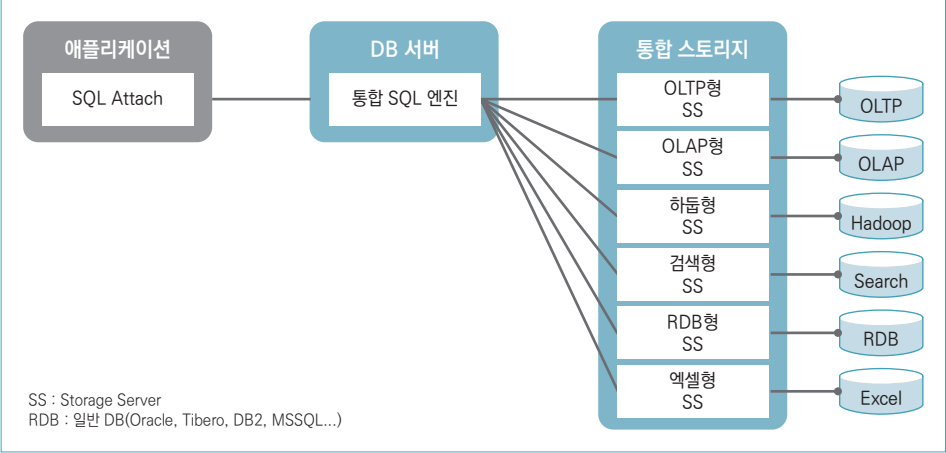
DB 가상화 기술을 이용하면 신규 업무 생성 시 빠른 DBMS 구성이 가능하고, 저장 위치에 상관없이 데이터 접근과 활용이 용이하며, 업무간 간섭 없는 유연한 데이터 관리가 가능해진다.

다. 다양한 데이터 형태의 통합

기업에서 생산되고 저장되는 다양한 형태의 데이터는 사용자나 애플리케이션 관점에서 동일한 도구를 통해 접근, 조회, 활용할 수 있도록 해야 한다.

OLTP형, OLAP형, 하둡형, 서치형 등 데이터 형태별 물리적 저장 스토리지 공간과 스토리지 프로세싱 기술은 각기 존재하고, 애플리케이션에서는 여러 종류의 데이터를 마치 동일한 DB 내에 존재하는 데이터처럼 보이게 하고 통합 처리를 가능하게 하는 하나의 강력한 인터페이스, 즉 통합 SQL 엔진 기술이 필요하다.

[그림 5-2-8] 다양한 데이터 형태 통합

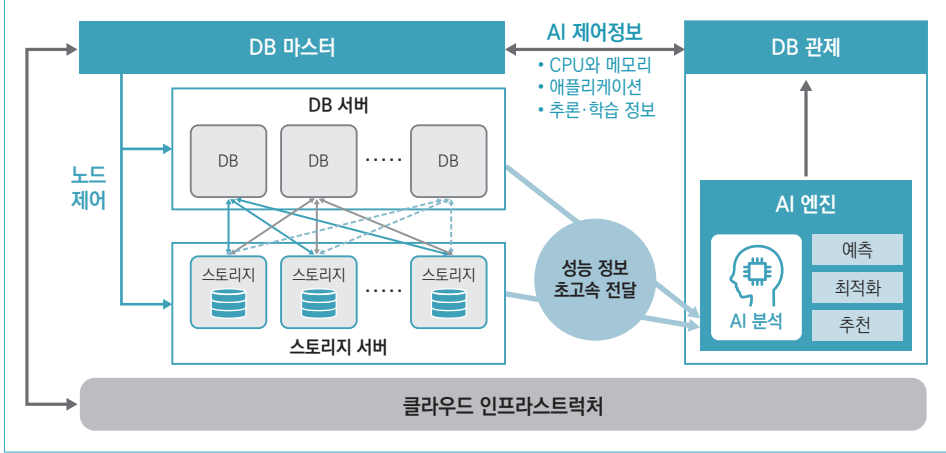


통합 SQL 엔진 기술은 클라우드 내 존재하는 정형, 비정형, 상용, 오픈 등 다양한 DB를 통합해준다. 또 다양한 데이터 형태를 수용할 수 있어 PaaS 기반의 애플리케이션 개발을 쉽게 하도록 해 진정한 님프로그래밍(Non-Programming)을 앞당기고 애자일한 비즈니스를 가능하게 해준다.

라. AI 기반의 지능화 DB

클라우드 내에는 다수의 DB 인스턴스가 생성과 소멸을 반복하고 그에 따라 DB 관리 복잡성이 높아진다. 기존의 DBMS 전용 모니터링 도구나 DB 벤더에서 제공하는 툴은 자체 뷰나 로깅, 자체 리포트를 통해 현 상태를 실시간 감시하고 과거 이력과 추이를 분석하는 툴로서 역할을 하고 있다. 자동 진단과 튜닝 어드바이저 기능도 일부 제공하고 있으나 섬세하고 선제적인 관리 기능으로 사용하기엔 부족한 측면이 있다. 지능화된 DBMS 감시와 제어를 위해 다양한 성능과 운영 정보를 실시간으로 수집하고 AI 기반 학습을 통한 예측과 자동 최적화 기술이 접목돼야 한다.

[그림 5-2-9] 지능화한 DB 관제



DB 이외의 관련된 애플리케이션 서비스, 클라우드 인프라 관리 솔루션과의 초고속 인터페이스를 통해 데드락(Dead Lock), 악성 쿼리, 성능 저하 등 DBMS에서 발생 가능한 심각한 장애를 사전에 예측하고, 사전 방지 가이드 제공, DB 노드 직접 제어까지 가능한 수준까지 도달해야 한다.

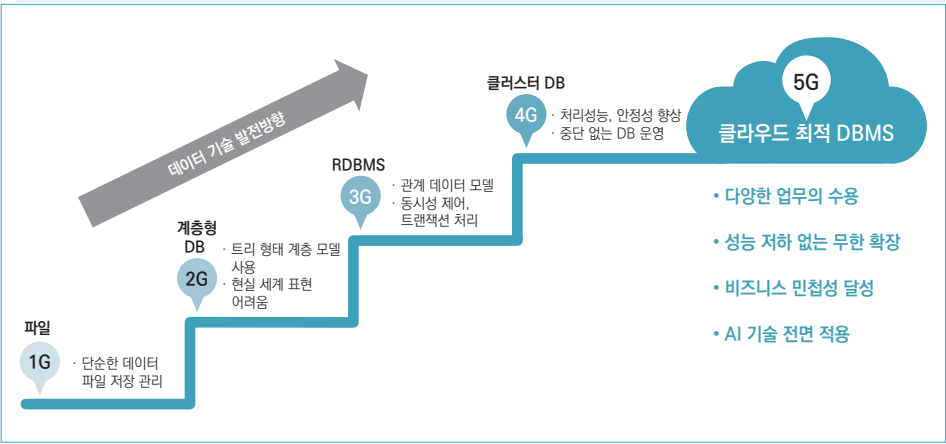
4. DBMS 진화의 선택압, 클라우드, AI

단순 데이터 파일 DB(1G), 계층형 모델 DB(2G), 관계형 DB(3G), 클러스터 DB(4G)를 거쳐 5G 시대는 한단계 진화한 DBMS를 요구하고 있다. 진화를 강요하는 환경은 클라우드로의 IT 이동과 AI 기술의 보편화이다.

변화에 대응하기 위해 DBMS는 스케일링 가능한 구조와 노드 제어 기술, 가상화 DB 기술, 다양한 데이터를 통합하는 SQL 엔진, 클라우드 인프라와 연계한 AI 기반 관제와 모니터링 기술을 구현해야 한다.

이러한 기술을 통해 등장할 5G DBMS는 클라우드 환경에 최적화된 DB로서 다양한 업무의 수용, 성능 저하 없는 무한확장과 탄력적 자원 활용, 비즈니스 민첩성 달성, 선제적이고 자동화된 관리가 가능한 신개념 DB가 될 것이다.

[그림 5-2-10] 5G DBMS로의 발전



제3장. 인공지능 기술

필자: 나동원
(비투엔 데이터 지능본부 이사)

인공지능이란 인지, 추론 등을 통한 학습과 문제해결 등 인간의 사고 능력을 기계적으로 구현하고 자동화한 시스템을 의미한다. 인공지능이라는 용어에 대한 기대와 기술 한계로 그동안 많은 변화를 겪었으나 빅데이터 발전, 컴퓨팅 능력 향상으로 딥러닝 기술이 발전함에 따라 인공지능 분야가 도약하고 있다.

인공지능과 관련해 최근 주목할 만한 흐름으로 딥러닝 분야의 진화, 자연어처리 기술의 도약, 엣지컴퓨팅의 확산 등이 있다.

1. 인공지능 개요

인공지능은 학자들 사이에서 다양하게 정의되고 있다. 다양한 정의를 종합하면 인지, 추론 등을 통한 학습과 문제해결 등 인간의 사고 능력을 기계적으로 구현해 자동화한 시스템을 의미한다.

인공지능이라는 용어는 1950년대 다트머스 회의에서 처음 사용됐다. 다트머스 회의 이후 인공지능은 활발히 연구됐고 논리적인 분야에서 많은 성과를 냈다. 그러나 사람처럼 보고, 듣고, 움직이는 등의 기본적인 행동과 연관된 현실 세계의 문제는 엄청나게 어려운 일이라는 것을 깨닫고 1974년에 혹독한 첫 번째 겨울을 맞는다.

1980년대에 전문가의 지식을 논리적인 규칙으로 생성해 특정 영역에 대한 사람의 질문에 답할 수 있는 인공지능으로서 전문가 시스템이 부상한다. 전염성 혈액 질환을 진단하는 마이신 등 몇 가지 성공사례가 있지만, 사람이 규칙을 일일이 생성하고 이를 조합해 논리적인 결과를 얻어내는 과정이 복잡하고 유지보수가 어려워 기대만큼의 성과를 내지 못하고 1987년에 두 번째 겨울을 맞는다.

2012년에 제프리 힌튼(Geoffrey Hinton) 캐나다 토론토대학 교수 연구실이 IMAGENET(이미지인식 경진대회)에서 딥러닝 알고리즘을 활용해 우승하면서 인공지능의 부활의 신호탄을 쏘았다.

인공지능이 다시 도약할 수 있었던 것은 검색엔진을 통해 인터넷 혁명이 일어나고 이에 따른 빅데이터 발전과 GPU(graphics processing unit)로 인한 컴퓨팅 능력 향상, 딥러닝 알고리즘 발전에 기인한다. 제프리 힌튼 교수가 발표한 딥러닝 연구는 앤드류 응(Andrew Ng) 미국 스탠퍼드대 교수와 얀 르쿤(Yan Lecun) 미국 뉴욕대 교수 같은 AI 구루들에 의해 발전됐다. 이후 세 명은 구글, 페이스북, 바이두 같은 글로벌 IT 기업에 영입돼 기술 발전을 촉진시켰다.

인공지능과 관련해 최근 주목할 만한 흐름인 딥러닝 분야의 진화, 자연어처리 기술의 도약, 엣지컴퓨팅의 확산 등에 대해 살펴보자.

2. 딥러닝

딥러닝(Deep Learning)은 머신러닝의 일종으로, 은닉층을 깊게 쌓은 인공신경망을 사용해 기계가 학습한다. 머신러닝에서는 사람이 특징 추출을 해야 하나, 딥러닝에서는 기계가 특징 추출을 할 수 있다. 딥러닝의 연속된 은닉층에서 점진적으로 의미 있는 특징을 학습할 수 있기 때문이다. 신경망의 은닉층이 깊어질수록 기계가 데이터에서 특징을 잘 추출할 수 있으나 매개변수가 늘어남에 따라 필요로 하는 학습데이터의 크기가 머신러닝보다 훨씬 커야 한다.

최근 딥러닝 기술의 진화는 세분화돼 다양한 분야에서 이뤄지고 있다. 딥러닝 연구를 동인으로 구분하면 데이터, 시스템, 프로세스 및 알고리즘으로 정리할 수 있다.

가. 데이터 측면

딥러닝 구축 시 품질에 영향을 주는 요소는 학습데이터의 양과 질이다. 양질의 학습데이터를 확보하는데 많은 비용과 시간이 소요되므로 인공지능을 활용하는 데 큰 장벽이 되고 있다. 딥러닝에서 학습데이터에 대한 비용을 절감할 수 있는 방향으로 알고리즘이 진화하고 있다. 기계가 스스로 레이블을 만들어 학습하는 방법인 자기지도학습과 다른 분야에서 학습한 모델을 이용해 적은 학습데이터로도 좋은 성능을 낼 수 있도록 하는 방법인 전이학습에 대한 연구가 활발히 진행되고 있다.

1) 자기지도학습

자기지도학습은 레이블이 없는 데이터를 활용하기 위해 제안됐다. 레이블이 있는 데이터 세트를 생성하는 것은 비용이 많이 드는 반면 레이블이 없는 데이터는 항상 생성되기 때문이다. 자기지도학습을 글자대로 유추하면, 스스로 지도하는 학습이다. 스스로 레이블을

만들어서 학습한다는 의미다. 자기지도학습의 주된 아이디어는 레이블이 없는 데이터들을 이용해 사용자가 새로운 문제를 정의하며 레이블도 자동으로 생성하는 것이다. 이를 자기지도학습 방법이라고 하며, 인공신경망이 문제를 학습하게 해 데이터 자체에 대한 특징들을 이해할 수 있도록 한다. 데이터의 잠재 특징을 학습할 수 있으므로 표현학습에 널리 사용된다.

기존 AI 훈련 기술은 주어지지 않은 정보까지 알아내지 못하지만, 자기지도학습으로 만든 AI는 상식선에서 생각할 수 있는 정보까지 추측해낼 수 있다. 딥러닝 분야 최고 권위자로 꼽히는 얀 르쿤 교수는 “세상이 어떻게 돌아가는지에 대한 배경지식을 쌓고, 인간과 소통할 수 있게 하는 것이 미래 AI 연구 방향이어야 할 것”이라며 인공지능의 미래를 이끌 기술로 ‘자기지도학습’을 꼽았다.⁰¹

2) 전이학습

전이학습은 신규 데이터 세트의 크기 및 데이터 세트 간의 유사성에 따라 사전학습된 모델에서 가져온 특징 추출 네트워크의 매개변수 전체를 고정하고 모델을 학습할 것인지, 일부만 고정하고 모델을 학습할 것인지, 특징 추출 네트워크의 매개변수를 고정하지 않고 초기값으로만 사용해 모델을 학습할 것인지를 결정한다.

전이학습은 학습에 사용할 만한 양질의 데이터가 충분히 확보되지 못하거나, 학습시간을 단축시키고자 할 때, 다른 분야의 풍부한 데이터를 바탕으로 한 좋은 성능의 모델에서 일부 계층을 재활용해 모델을 구축하는 학습방식이다.

앤드류 응 교수에 의하면, 전이학습은 산업에서 머신러닝의 상업적인 성공을 이끄는 핵심동인이 될 것이다.

나. 시스템 측면

엣지컴퓨팅 기술이 급격히 발전하고 있고 이에 따라 딥러닝 모델이 엣지에서 적용될 수 있도록 하는 움직임이 활발하다.

딥러닝은 대용량의 데이터가 모일 수 있는 환경이 클라우드 기반에서 조성돼 많은 컴퓨팅 리소스와 시간을 투입해 학습을 해왔다. 최근에는 실시간성 및 개인정보보호 요구 증가로 스마트폰 및 가정이나 현장의 종단에서 직접 수행할 필요성이 증가함에 따라 엣지컴퓨팅 환경에 맞춰 실시간으로 소량의 학습데이터를 점진적으로 학습에 적합하도록 알고리즘 구조를 변경하는 작업들이 진행되고 있다. 텐서플로우 라이트(Tensorflow lite),

01. 삼성AI포럼, 2018.

카페2(caffe2), 파이토치(Pytorch), 엠엑스넷(MxNet) 등의 딥러닝 프레임워크가 엣지컴퓨팅을 지원하고 있다.

다. 프로세스 및 신경망 알고리즘 측면

모델 개발에 시간이 많이 걸린다는 문제점과 환경 변화 시 모델 성능이 떨어져 다시 설계해야 한다는 점이 현장에서 딥러닝을 활용하는 데 걸림돌이 되고 있다. 이를 해결하는 방향으로 연구가 활발히 진행되고 있다.

1) AutoML

딥러닝 모델 개발은 시간이 많이 걸리는 반복적인 작업이 필요하다. 딥러닝 모델의 설계요소를 변경하고 모델의 학습결과를 비교해가며 최적의 성능을 발휘하는 딥러닝 모델 구조를 찾아야 하기 때문이다. AutoML(Automated Machine Learning)은 머신러닝으로 자동 설계하는 머신러닝을 의미하며, 아키텍처 탐색(Network Architecture Search)과 초매개변수 최적화(Hyperparameter Optimization) 방향으로 연구가 진행되고 있다.

아키텍처 탐색이란 인공지능망의 아키텍처를 사람이 직접 하나하나 설계하는 대신에 아키텍처 탐색공간에서 학습을 통해 최적의 아키텍처를 찾는 방법이다. 주로 강화학습이나 유전 알고리즘 등을 이용한 연구 및 경사하강법에 기반한 차분구조탐색법(Differential Architecture Search)이 활발히 연구되고 있다.

초매개변수 최적화란 학습을 시키기 위해 필요한 초매개변수들을 학습을 통해 추정하는 방식이다. 예를 들어 학습률(learning rate), 배치 크기(batch size) 등 학습에 큰 영향을 주는 초매개변수들을 기존에는 사람이 하나하나 값을 바꿔 모델을 학습시켜서 가장 성능이 좋았던 초매개변수를 사용하는 방식이었다면, AutoML에서는 학습을 통해 최적의 초매개변수를 추정하는 방법을 제안한다.

2) 적응형 기계학습

전통적인 인공지능 시스템에서 프로덕션에 배포된 학습 방법은 시스템 운영 환경의 변경, 시스템의 입력 변경, 조직이 원하는 결과물이 변경될 경우 이를 해결하는 데 큰 어려움을 겪었다. 환경이 변하지 않는다고 가정하고 학습하기 때문이다. 환경에 대한 변화를 감지하고, 이에 따라 작은 데이터로 학습해 환경 변화에 빠르게 적응하게 하는 학습 방법이 적응형 기계학습(Adaptive Machine Learning)이다. 적응형 기계학습은 환경과 상호작용하며 학습하는 강화학습, 실시간 데이터를 연속적인 시퀀스로 처리하는 온라인 학습, 작은 샘플 크기로 새로운 학습을 하게 하는 메타학습 등을 결합하는 방식으로 연구되고 있다.

3) 트랜스포머

트랜스포머(Transformer)는 2017년 구글이 발표한 논문인 ‘Attention is all you need’에서 나온 모델로 기존의 시퀀스-투-시퀀스(seq2seq)의 구조인 인코더-디코더를 따르면서도, 논문의 이름처럼 어텐션(Attention)만으로 구현한 모델이다. 어텐션 메커니즘은 입력 시퀀스의 매 스텝마다 컨텍스트를 저장하고, 출력과 연관성이 높은 컨텍스트에 집중해 긴 시퀀스일지라도 앞부분의 중요 정보를 보존할 수 있도록 한 방식이다. 트랜스포머는 순환신경망 계열을 사용하지 않고, 인코더-디코더 구조를 설계했음에도 성능이 순환신경망 계열보다 우수하다. 가장 큰 장점은 순환신경망 계열과 달리 GPU 기반에서 병렬 처리를 효율적으로 할 수 있는 구조로 학습시간을 단축할 수 있다는 것이다. 최근 자연어처리 분야의 딥러닝 모델에서 널리 사용되고 있다.

4) 적대적 생성 신경망

적대적 생성 신경망(Generative Adversarial Networks, GAN)은 이안 굿펠로우(Ian Goodfellow)가 2014년 NIPS(Neural Information Processing Systems) 학회에서 발표한 비지도 학습 기반 신경망 알고리즘으로, 진짜 같은 가짜를 만들어내는 기술이다.

적대적 생성 신경망은 기본적으로 두 신경망 모델이 경쟁해 학습하면서 결과를 만들어낸다. 두 모델은 생성자와 감별자로 불리는데 각각은 서로 상반된 목적을 갖고 있다. 생성자는 실제 데이터를 학습해 거짓데이터를 만들고, 감별자는 생성자가 만든 데이터가 실제인지 거짓인지 판별하는 과정을 거치며 실제에 가까운 거짓 데이터를 생성한다.

최근 적대적 생성 신경망을 이용한 이미지 변환 기술이 많이 연구되고 있다. 엣지나 세그멘테이션 맵을 이용해 이미지를 생성하는 연구, 레이아웃을 이용해 이미지를 생성하는 연구, 저해상도의 이미지를 고해상도로 복구하는 연구, 이미지의 사라진 부분을 복구하는 연구, 사람의 머리색과 나이, 성별, 피부색 등을 변경하는 연구 등 사용자의 의도를 반영한 이미지 생성 및 고해상도 이미지를 생성하는 방향으로 적대적 생성 신경망의 기술들이 빠르게 발전하고 있다.

3. 자연어처리

자연어처리(Natural Language Processing, NLP)란 넓은 의미로 말이나 글 같은 자연어를 소프트웨어를 이용해 자동으로 기계가 다룰 수 있게 처리하는 기술로서 정의되며,

1950년대부터 인공지능의 한 분야로서 연구되기 시작했다. 규칙기반으로 연구되기 시작해 머신러닝 도입을 통해 한차례 기술 도약이 있었고, 2010년대에 딥러닝을 기반으로 자연어이해(Natural Language Understanding, NLU) 및 자연어생성(Natural Language Generation, NLG)에서 비약적인 발전을 하고 있다.

가. 자연어이해와 자연어생성

자연어이해는 음성 또는 텍스트의 의미 이해에 초점을 맞춘 기계학습의 유형분류 기법으로 볼 수 있다. 동일한 질문을 하는 방법에는 여러 가지가 있을 수 있고, 각각 다른 단어, 구문 및 문장 구조를 포함할 수 있다. 자연어이해는 다양하게 표현될 수 있는 질문을 표준화된 형태로 변환하고, 질문의 바탕에 깔린 의도를 이해하는 역할을 한다. 자연어이해 기능의 예로는 문장의 의도 분류, 서로 다른 언어 간 번역 문장 생성, 자연어 질문에 대한 답변 추출 등이 있다.

자연어생성은 쓰여진 내용이나 말한 내용을 이해하는 데 사용되는 대신에 자연어 응답을 생성하는 데 사용된다는 점에서 자연어이해와 반대개념을 이룬다. 자연어생성은 자연어이해 및 지능형 반응 시스템과 결합되면 대화형 AI 또는 자연어 상호작용을 뒷받침하는 중요한 요소가 된다. 대화형 AI는 사용자의 질문을 이해한 후에는 답변하기 위해 무언가를 분석하고, 사용자에게 그 결과를 다시 보내기 위해서는 답장을 보낼 텍스트를 만들어야 하는데, 텍스트를 만드는 것이 바로 자연어생성의 역할이다.

자연어생성은 3단계로 나뉜다. 구조화된 데이터에서 기본 콘텐츠의 순서가 이뤄지는 텍스트 계획, 문장이 구조화된 데이터에서 결합돼 정보의 흐름을 나타내는 문장 계획, 마지막으로 텍스트를 표현하기 위해 문법적으로 정확한 문장이 생성되게 구현하는 단계로 구분된다.

자연어생성은 통계 분석 및 비즈니스 인텔리전스 결과에 대한 설명 요약뿐만 아니라 블랙박스로 간주되는 기계학습 모델에 대한 해석에도 활용된다. 데이터 과학자나 전문가가 아닌 사람에게 분석결과에 대한 설명을 쉽게 제공하는 것은 머신러닝 활용 측면에서 매우 의미가 있다.

나. 자연어처리 기술

심층신경망이 등장하기 전에 자연어처리는 텍스트 데이터 내의 단어 수에 의존하는 언어모델을 사용해 주제를 분류했다. 이 모델은 특정 단어가 특정 주제 아래에 나타나는 횟수를 계산하고, 특정 단어가 새 데이터의 특정 주제 아래에 나타날 베이저안 확률을 계산한다. 그러나 이러한 기술은 단어 순서를 무시하기 때문에 컨텍스트를 활용하지 못한다.

자연어처리에서 심층신경망의 한 종류인 순환신경망(Recurrent Neural Networks)을 적용하면서 단어 순서를 반영한 컨텍스트를 활용할 수 있게 됐다. 그러나 순환신경망 계열의 단점 중 하나는 시퀀스가 길어지면 컨텍스트를 표현하는 벡터의 크기가 고정됐기 때문에 시퀀스의 시작 부분에 있는 중요한 정보를 잊어버릴 수 있다는 것이다. 이 문제를 해결하기 위해 어텐션 메커니즘을 적용했다. 어텐션 메커니즘은 컨텍스트를 매 스텝마다 저장하고, 출력과 연관성이 높은 컨텍스트에 집중해 긴 시퀀스일지라도 앞부분의 중요 정보를 보존할 수 있도록 한 방식이다.

순환신경망 계열은 구조적으로 병렬연산이 어려워 처리속도에 제약이 있다. 트랜스포머는 순환신경망 계열 없이 다층구조의 어텐션 메커니즘으로 심층신경망을 구성한 모델로서, 순환신경망 계열보다 성능이 대체로 우수하고 처리속도에서도 월등한 우위를 보인다.

최근에 발표된 모델에서는 우수한 성능으로 트랜스포머와 전이학습을 이용하는 것이 하나의 트렌드다. 전이학습은 언어모델이 어떠한 방식으로 사용될지 확실하지 않더라도 미리 양질의 방대한 양의 학습 데이터로 미리 학습을 시켜두고 이렇게 학습된 모델(사전학습모델)을 간단하게 조정해 다른 목적으로 활용한다. 이러한 사전학습모델은 사람이 레이블링 작업을 할 필요가 없는 자기지도학습 방법을 이용해 학습된다. 자연어처리 분야에서 OpenAI GPT, BERT, OpenAI GPT-2, XLNet, RoBERTa, ALBERT, GPT-3 등과 같은 다양한 모델들의 강력한 성능 덕분에 전이학습은 큰 인기를 얻고 있다.

최근에는 자연어처리에 대한 품질을 더욱 향상시키기 위해 신경망 아키텍처 최적화, 비주얼 컨텍스트 활용, 비지도 및 준지도 학습방식의 기계번역에 대한 연구, 언어모델과 이를 이용한 전이학습에 대한 연구 및 딥러닝에 언어학을 접목해 데이터 기반 접근방식에 대한 해석 가능성을 개선하는 방향으로 연구가 활발히 이뤄지고 있다.

4. 엣지 컴퓨팅

클라우드 이후 시대에 사물인터넷(IoT)의 확산과 4G, 5G의 대중화는 대중의 데이터 액세스 및 처리 습관을 점진적으로 변화시키고 있다. 엣지 컴퓨팅(Edge Computing)은 응답 시간을 개선하고 대역폭을 절약하기 위해 네트워크 엣지에서 데이터를 처리하는 새로운 컴퓨팅 패러다임이다. 이러한 분야에 대해 빠르게 증가하는 수요와 관심에 힘입어 엣지 컴퓨팅 시스템과 도구가 확산되고 있다.

딥러닝을 네트워크 엣지에서 지원하는 시스템을 구축하기 위해서는 컴퓨팅 성능 제한, 데이터 공유 및 공동 작업, 엣지 플랫폼과 딥러닝 알고리즘 간의 불일치 등을 고려해야 한다. 이러한 노력의 일환으로 네트워크 엣지에서 딥러닝 모델로 데이터 처리를 지원하기 위해 구글, 페이스북, 마이크로소프트 등이 엣지 기반 딥러닝 프레임워크 및 도구를 출시했다.

하드웨어로는 에너지 효율성이 높아진 엣지용 필드 프로그래머블 게이트 어레이(FPGA), 그래픽처리장치(GPU) 기반 하드웨어들이 출시되고 있다. 특히 주문형반도체(ASIC)는 더 작은 크기, 더 낮은 전력 소비, 더 높은 성능, 더 높은 보안을 제공하므로, 엣지 시나리오에 적합해 딥러닝 컴퓨팅 속도와 에너지 효율을 더 높일 수 있다. 엣지 TPU는 엣지 컴퓨팅을 위해 특별히 제작된 구글의 주문형 반도체로서, 개방형 소프트웨어 및 최신 딥러닝 알고리즘을 결합해 낮은 전력 소비로 고성능을 발휘한다.

아마존, 마이크로소프트, 구글 등 업계의 리더들은 네트워크 엣지에서 딥러닝을 지원하는 서비스를 진행하고 있다. 아마존은 5G 초지연 전송 애플리케이션을 지원할 수 있는 엣지컴퓨팅 서비스로 웨이브랜스(Wavelength)를 발표했는데, 이동통신사와 파트너십을 맺어 이동통신 사업자가 운영하는 데이터센터를 이용해 초저지연을 실현하는 방식이다. 마이크로소프트도 통신사와 협력해 5G와 애저(Azure)를 통합해 기업고객에게 엣지 컴퓨팅을 제공하는 NEC(Network Edge Compute) 서비스를 발표했다. 구글은 아마존이나 마이크로소프트와는 다른 방식으로 엣지 컴퓨팅에 접근하고 있다. 엣지 TPU로 엣지 컴퓨팅 능력을 향상시키고, 쿠버네티스(Kubernetes⁰²)로 엣지와 클라우드 간 워크로드 분배를 용이하게 하는 방식을 핵심으로 추진하고 있다.

실시간 처리 능력 향상과 개인정보보호에 대한 요구는 딥러닝과 엣지 컴퓨팅을 결합하는 움직임을 만들었다. 딥러닝과 결합된 엣지 컴퓨팅은 전형적인 중앙화된 사물인터넷 모델보다 실시간에 가깝게 반응할 수 있으며, 대부분 인사이트가 즉각적으로 전달되고 처리될 수 있고, 데이터를 보내고 받는 과정에서 정보 노출 위험을 최소화할 수 있다.

엣지 컴퓨팅에 적합한 딥러닝 알고리즘 연구, 저전력 및 고성능의 병렬처리 하드웨어 개발, 이를 패키지화한 통신사 및 플랫폼 사의 서비스 개발이 점점 고도화될 것으로 기대된다.

02. 컨테이너화된 워크로드와 서비스를 관리하기 위한 이식성이 있고, 확장가능한 오픈소스 플랫폼

제4장. 데이터 보안과 비식별 기술

필자: 윤경구
(파수 개발1본부 전무)

컴퓨팅 환경이 클라우드와 온라인 비대면 시대로 접어들면서 데이터가 기업 내부를 벗어나서 존재하는 시간이 길어졌다. 한편으로 국내에서는 개인정보 보호와 관련한 ‘데이터3법’이 통과돼 개인정보를 활용한 가치 창출이 법적으로 허용됐다. 기업 밖으로 나간 데이터를 보호하기 위해 클라우드 서비스의 여러 영역에서 데이터 보안이 강화되고 다양한 데이터를 지원하는 형태로 발전하고 있다.

국내에서 개인정보를 활용하기 위해 기관과 기업들이 데이터 전문 기관, 결합 전문 기관이 되려고 나서고 있으며 가명화된 개인정보 데이터를 결합·활용한 부가가치 서비스와 개인 맞춤형 금융 서비스인 마이데이터 서비스 등 본격적인 데이터 유통 산업이 꿈틀대고 있다.

1. 데이터 보안 기술의 새로운 흐름

코로나19 팬데믹으로 전 세계가 격리되는 가운데 각 조직은 재택 근무(Work For Home)에 대한 전략, 클라우드 서비스 도입 가속화라는 큰 흐름을 채택하고 있다. 이에 따라 데이터가 물리적으로 조직 외부에 존재하는 시간이 길어지면서 외부의 데이터에 대해서도 보안을 유지할 수 있는 솔루션에 대한 요구가 커졌다.

유럽과 미국에서 개인정보 및 사생활 보호 흐름이 확산되는 가운데 국내에서도 데이터3법이 국회를 통과하면서 개인정보에 대한 보안과 활용이 화두로 떠올랐다.

가. 재택 근무와 클라우드 도입 양상의 변화

기존의 데이터 보안 전략은 데이터가 기업 내부에 있다는 가정 하에 외부 유출을 막는 데

치중해 왔다. 이와 달리 재택 근무의 급속한 확산과 클라우드 도입 확산, 그리고 국가별로 데이터의 소재에 따라 다른 법적 규제 적용 등은 데이터의 물리적 위치에 따른 보안 전략 수정을 필요로 하고 있다.

재택 근무와 클라우드 서비스의 적극적인 채택은 기존 보안 역량을 기업 외부까지 확장할 것을 요구하고 있다. 또 하나가 아닌 여러 클라우드 제공자 서비스를 활용하는 추세와 각 국의 개인정보 및 사생활 관련 보호 법률들은 데이터의 지리적 소재, 저장소 위치, 사용자 접속 위치 등에 대해서도 좀 더 엄밀한 관리를 필요로 한다. 머신러닝 기술과 블록체인, 익명화, 가명화 같은 기법들은 개인정보와 사생활을 보호하기 위한 실질적인 조치를 요구하고 있다.

나. 떠오르는 주요 보안 기술⁰¹

1) 데이터 위협 평가

데이터 위협 평가(Data Risk Assessment)는 데이터 보안과 개인정보에 대한 통제가 효과적으로 적용되는지에 대한 평가를 뜻한다. 보안 제품과 응용 프로그램에서 위협 수준이 기준을 만족하는지를 평가하며, 개인정보나 중요 데이터 유출 및 법령 미준수로 인한 위협을 완화하는 것이 목적이다.

초기에는 특정한 데이터베이스 플랫폼과 같은 스토리지 저장소나 특정 데이터에 대한 위협 평가부터 우선 진행하고 주요한 데이터 세트의 범위를 확대하는 것이 적절한 전략일 것이다.

2) 기밀 컴퓨팅

기밀 컴퓨팅(Confidential Computing)은 하드웨어를 활용해 신뢰할 수 있는 실행 환경에서 코드를 실행하는 보안 메커니즘이다. 이 실행 환경은 보통 인클레이브(enclave)라고 부르는 보호된 메모리 영역이거나 CPU 차원에서 메모리 접근 시 암호화를 처리해주는 기법(AMD CPU의 Secure Memory Encryption 등⁰²)을 사용한다.

기밀 컴퓨팅은 처리 중인 데이터가 암호화돼 비공개로 유지되기 때문에 사용 중이더라도 메모리 접근에 의한 데이터 유출 우려가 없다는 장점이 있다. 클라우드 인프라 서비스 벤더들은 이런 하드웨어를 적용한 가상머신을 제공하기 시작했으며, 이를 통해 클라우드 컴퓨팅의 보안 우려가 해소될 것으로 기대한다.

01. Hype Cycle for Data Security, Gartner, 2020.

미국, 캘리포니아주 소비자 프라이버시법(CCPA) 집행 시작 및 안면인식사용기술 관련 법안 발의

02. Helping Secure the Cloud with AMD EPYC Secure Encrypted Virtualization, AMD, 2019.4.

3) 금융 데이터 위협 평가

FinDRA(Financial-based Data Risk Assessment, 금융 데이터 위협 평가)는 데이터 보안에 대한 위협이 비즈니스에 미치는 영향을 평가해 재무적인 기회 판단과 전략 수립에 보안 위협을 반영하는 것을 목적으로 한다.

데이터 보안과 개인정보 및 사생활 보호에 필요한 데이터 관리 비용 등을 사업 전략과 예산에 반영하기 위해 데이터 책임자, 정보보안 책임자, 개인정보 보호 책임자 등이 참여하는 기법이다.

4) 데이터 보안 거버넌스

보안 기술, 기준데이터의 위치 소재, 개인정보 보호를 포함한 데이터에 영향을 미치는 다양한 사건들이 내재한 비즈니스 위협을 평가하고 우선순위를 정하며 대응하는 전반적인 데이터 보안 프레임워크를 데이터 보안 거버넌스(Data Security Governance)라고 한다.

여기에는 앞에서 언급한 데이터 위협 평가의 지속적인 수행을 기반으로 데이터별 목적과 위험도 관리, 실제 사용자와 각종 장치, AI 등이 사용하는 사용자 계정의 관리, 접근 제어를 사용하는 정책이 필요하다.

5) 동형 암호화

동형 암호화(Homomorphic Encryption) 기법은 데이터가 암호화된 상태에서 연산을 수행하고 그 결과도 암호화된 상태로 유지되는 방식의 암호화 기법을 뜻한다. 이 기법은 복호화를 하지 않고 연산을 수행할 수 있어 데이터가 외부에 있는 클라우드 환경에서 주목받고 있다. 그러나 연산 속도 문제와 지원하는 연산 부족 등으로 아직 전면적으로 사용하기에는 한계가 크다.

최근 제한된 연산 기능에 대해서만 동형 암호화를 지원하는 부분 동형 암호화 상용 솔루션이 암호 상태 검색이나 다자 간 컴퓨팅 등에 실험적으로 사용되고 있다.

6) 차분 프라이버시

차분 프라이버시(Differential Privacy) 기술은 개인정보 등을 포함하는 데이터 세트를 공유하기 위한 목적으로 사용하기 위해 특정 정보에 소음을 추가하거나 왜곡해 데이터 분석이 개인별 정보에 의존하지 않도록 하는 기법이다. 데이터 세트를 통계 목적으로 활용할 때 개인의 민감 정보 유출을 방지하기 위한 익명화 기법의 하나로 볼 수 있다.

차분 프라이버시라는 이름은 데이터 세트에서 특정 개인의 데이터를 빼고 진행한 분석과 포함한 분석이 큰 차이가 없도록 하기 위한 알고리즘을 뜻한다.

7) 데이터옵스

데이터옵스(DataOps)는 여러 조직에 걸친 데이터 관리자와 데이터 사용자 사이의 데이터 흐름을 관리, 통합, 자동화하기 위한 협업적 데이터 관리 실무를 뜻한다.

데이터옵스는 다양한 환경에서 데이터 사용성과 가치를 향상시키기 위해 보안, 품질, 메타데이터를 적절한 수준으로 사용해 데이터 전달을 오케스트레이션하고 자동화하는 기술들을 필요로 한다.

8) 데브옵스 테스트 데이터 관리

응용 프로그램 테스트에 사용되는 데이터에 대해서도 개인정보 보호와 가상화 등의 기법(DevOps Test Data Management, 데브옵스 테스트 데이터 관리)이 반드시 필요해졌다.

데이터의 선택적 복사, 익명화, 마스킹, 데이터 가상화, 인위적 데이터 합성 등이 기법적으로 법령 준수와 개인정보 보호를 위해 사용된다.

9) 영지식 증명

영지식 증명(Zero-Knowledge Proof) 방식은 양자 간에 정보를 전송하거나 공유하는 일 없이 서로가 가진 정보가 정확함을 증명해 개인정보를 보존하는 메시징 규약이다. 예를 들어 어떤 답을 알고 있는지를 답을 전송하지 않고 타인에게 증명하는 방법이라고 할 수 있다.

블록체인에 영지식 증명을 적용하면 어떤 코인을 소유하고 있는지를 밝히지 않고도 코인을 소유하고 있다는 사실을 알릴 수 있다. 비트코인의 확장인 ZCash나 이더리움의 확장인 ZoE 등에서 영지식 증명을 사용해 개인정보 노출을 최소화하고 있다.

10) 클라우드 제공자 고유 데이터 손실 방지

SaaS 혹은 IaaS 클라우드 서비스 제공자가 직접 데이터 손실 방지 기능을 함께 제공하는 경우이다(Cloud Provider-Native DLP, 클라우드 제공자 고유 데이터 손실 방지). 이를 통해 클라우드 서비스에서도 민감 정보에 대한 제어를 강화할 수 있다.

이 경우 온프레미스 환경이나 멀티 클라우드 서비스를 포함한 복합 환경에 대해서는 단일 정책으로 제어하지 못할 수 있다.

11) 데이터 보안용 블록체인

블록체인을 데이터 보안에 활용하면 중앙집중식 방식에서 핵심 관리자에게 의존하는 문제를 완화할 수 있다. 데이터 무결성을 보장하는 보호 체인을 구성하거나 탈중앙 방식의 PKI를 위한 키 저장소로 블록체인을 활용하는 등 다양한 사용 시나리오가 나오고 있다.

2. 개인정보 보호 및 활용 시대의 절차와 기술

개인정보 보호에 대한 여러 가지 입법들이 국제적으로 진행되는 가운데 국내에서도 2020년 1월 데이터3법이 진통 끝에 통과돼 개인정보 보호를 위한 방안과 가명화를 통한 데이터를 활용하는 방안이 구체적으로 명시됐다.

여기에서는 개인정보를 포함한 데이터를 보관·관리하는 법령 준수 관점의 절차와 이를 활용하기 위한 절차, 그리고 기술적 요구 사항을 짚어본다.

가. 데이터3법과 주요 내용

데이터3법은 데이터 이용을 활성화하는 ‘개인정보보호법’, ‘정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 정보통신망법)’, ‘신용정보의 이용 및 보호에 관한 법률(이하 신용정보법)’ 등 3가지 법률을 통칭한다.⁰³ 데이터3법 개정안은 2020년 2월 4일 공포됐다.

개인정보보호법은 2011년에 제정된 이래 2018년 EU의 일반 데이터 보호 규정(General Data Protection Regulation, GDPR) 시행과 국내에서의 개인정보 활용에 대한 요구사항 등을 반영해 전면 개정됐으며, 기존 정보통신망법에 포함된 개인정보 관련 규정을 삭제하고 ‘개인정보보호법’ 내에 특례 규정으로 흡수했다. 금융권을 중심으로 적용되는 특별법인 ‘신용정보법’은 일반법인 ‘개인정보보호법’과 정합성을 맞추도록 함께 개정됐다.

데이터3법의 실행을 촉진할 시행령과 가이드라인 등 후속 조치들이 차근차근 진행되고 있다. 금융위원회는 신용정보법에 따라 데이터 전문 기관, 마이데이터 사업 관련 지정 및 심사 공고를 내고 가명, 익명 처리 안내서를 발간했다.

- 금융위원회, ‘데이터 전문 기관 지정(신용정보원, 금융보안원)’, 2020.8.6.
- 금융위원회, ‘금융 분야 가명, 익명 처리 안내서’ 발간, 2020.8.6.
- 금융위원회, ‘본인신용정보관리업(마이데이터) 허가 심사 일정 안내’ 공고, 2020.8.19.

개인정보보호위원회는 ‘개인정보보호법’에 따라 결합 전문 기관 지정 공고를 내고 가명 처리 가이드라인을 발간했다.

03. 대한민국 정책 브리핑 ‘데이터3법’, 2020.3.

- 개인정보보호위원회, ‘개인정보보호위원회 결합 전문 기관 지정계획’ 공고, 2020.9.1.
- 개인정보보호위원회, ‘가명정보 처리 가이드라인(최종)’ 발간, 2020.9.24.

데이터3법의 주요 내용은 다음과 같다.

- 데이터 이용 활성화를 위한 가명정보 개념 도입
- 관련 법률의 유사·중복 규정을 정비하고 추진체계를 일원화하는 등 개인정보 보호 협치(거버넌스) 체계의 효율화
- 데이터 활용에 따른 개인정보 처리자의 책임 강화
- 모호한 개인정보 판단 기준의 명확화

신용정보법과 개인정보보호법의 각 시행 가이드라인을 살펴보면 가명화, 결합 등과 관련해 준수해야 할 프로세스는 내용면에서 매우 유사하다. 하지만 금융기관을 대상으로 하는 신용정보법의 시행령은 개인을 알아볼 수 없게 하는 익명화 처리를 포함하는 반면, 개인정보보호법의 시행령은 가명화에 국한해 언급하고 있다.

가명정보의 결합에 필요한 전문 기관은 신용정보법에서는 ‘데이터 전문 기관’을 통해야 하고, 개인정보 보호법에서는 ‘결합 전문 기관’을 통해야 한다는 점이 다르다. 즉 ‘데이터 전문 기관’은 금융 기관을 위한 ‘결합 전문 기관’이라고 생각할 수 있다.

[표 5-4-1] 신용정보법에 따른 개인정보의 개념과 활용 가능 범위

구분	개념	활용 가능 범위
개인정보	특정 개인에 관한 정보, 개인을 알아볼 수 있게 하는 정보	사전적이고 구체적인 동의를 받은 범위 내 활용 가능
가명 정보	추가 정보의 사용 없이는 특정 개인을 알아볼 수 없게 조치한 정보	다음 목적에 동의 없이 활용 가능 (EU GDPR 반영) ① 통계 작성(산업적 목적 포함) ② 연구(산업적 연구 포함) ③ 공익적 기록 보존 목적 등
익명 정보	더 이상 개인을 알아볼 수 없게 (복원 불가능할 정도로) 조치한 정보	개인정보가 아니기 때문에 제한 없이 자유롭게 활용

출처: 「금융분야 가명 익명 처리 안내서」, 금융위원회, 2020.8.6.

나. 가명 처리의 통제와 활용

개인정보 보호법에서 익명화에 대한 구체적인 가이드라인이 없다. 그러므로 개인정보 활용에는 금융 기관을 제외하고는 주로 가명화 기법을 사용한 활용 및 결합, 반출 등이 활성화될 것으로 예상할 수 있다.

개인정보를 가명, 익명화할 때 주요 개념으로 ‘식별 정보’와 ‘식별 가능 정보’가 있다. 신용정보법에서는 식별 정보를 ‘식별자’, 식별 가능 정보를 ‘개인 식별 가능 정보’로 언급하고 있지만 기본적인 개념은 동일하다.

[표 5-4-2] 식별 정보와 식별 가능 정보(예시)

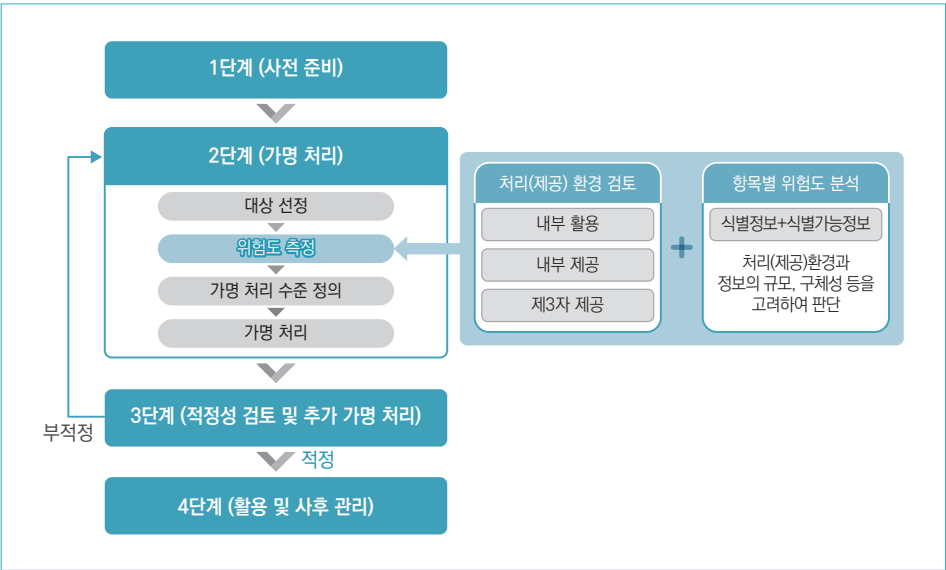
식별 정보	고유식별정보(여권번호, 외국인등록번호, 운전면허번호), 성명, 전화번호, 전자메일 주소, 의료기록번호, 건강보험번호, 자동차 등록번호 등 외부 연계(식별)를 목적으로 생성된 정보 등
식별 가능 정보	성별, 연령(나이), 국적, 혈액형, 신장, 몸무게, 직업, 위치정보 등 가명정보 처리자 입장에서 개인을 알아볼 수 있는* 정보 * 개인을 '알아볼 수 있는지'는 해당 정보를 처리하는 자(정보의 제공 관계에 있어서는 제공받는 자를 포함)를 기준으로 판단해야 함
	특이정보 • 국내 최고령, 최장신, 고액체납금액, 고액급여수급자 등 전체적인 패턴에서 벗어나 극단값이 발생할 수 있는 정보 • 희귀 성씨, 희귀 혈액형, 희귀 노동자 색깔, 희귀 병명, 희귀 직업 등 정보 자체로 특이한 값을 가지고 있는 정보

출처: 「가명정보 처리 가이드라인」, 개인정보보호위원회, 2020.9.

가명 처리는 ‘개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가정보 없이는 특정 개인을 알아볼 수 없도록 처리하는 것’을 의미한다. 여기에서 추가 정보(Additional Information)는 GDPR에서 정의하고 있는 개념과 동일하며 개인정보의 전부 또는 일부를 대체하는 데 이용된 수단이나 방식(알고리즘 등), 가명정보와의 비교·대조 등을 통해 삭제 또는 대체된 개인정보를 복원할 수 있는 정보(매핑 테이블 정보, 가명 처리에 사용된 개인정보 등) 등을 나타낸다(이상 ‘가명정보 처리 가이드라인’).

- 1단계 사전 준비: 가명 정보 처리를 위해 처리 목적의 적합성을 확인하고 필요 서류를 작성하는 단계
- 2단계 가명 처리: 데이터를 실제로 가명 처리하는 단계로, 데이터 속성별로 식별 정보와 식별 가능 정보 여부를 확인한 후 위험도를 측정하고 위험도 평가 결과를 작성한다. 위험도 평가 결과에 따라 가명 처리 수준을 정의해 어떤 항목에 대해 어떤 알고리즘을 사용해 가명화할 것인지를 결정한다(가명 처리 수준 정의표 작성). 가명화 방법이 결정됐으면 이에 따라 가명 처리를 실행한다.

[그림 5-4-1] 가명 처리 단계별 세부 절차도



출처: 「가명정보 처리 가이드라인」, 개인정보보호위원회, 2020.9.

- 3단계 적정성 검토: 2단계에서 작성한 '가명 처리 수준 정의표'의 기준에 따라 적절히 가명 처리가 됐는지를 확인하고, 가명 정보의 활용목적에 달성할 수 있는지, 또 재식별 가능성은 없는지를 검토한다. 이때 개인정보 처리자의 판단에 따라 외부전문가로 구성된 적정성 평가단을 구성해 검토할 수 있다. 적정성 평가 결과 부적정 판단이 내려질 경우 2단계로 돌아가 가명 처리 수준을 재판단하고 다시 가명 처리를 할 수 있다.
- 4단계 활용 및 사후 관리: 가명 처리를 하고 나면 기록 작성, 보관을 해야 하며 주기적으로 개인 식별 가능성이 높아지는지 여부를 모니터링해야 한다. 가명 처리에 사용된 추가 정보는 재식별에 필요한 정보이므로 분리 보관하고 접근 권한도 분리해야 한다.

가명 처리 시 많이 사용되는 방법은 식별 정보를 가명화하는 데 사용되는 해싱, 마스킹, 암호화, 삭제 등의 기법과 식별 가능 정보에 대해 식별 가능성 수준을 검사하는 k-익명성과 같은 알고리즘의 검사 기준을 충족하기 위해 사용하는 삭제, 범주화(예: 나이를 연령대로 변경) 등의 기법을 들 수 있다. 원칙적으로 가명 처리된 결과물은 분리 보관된 추가 정보의 도움 없이는 개인 식별이 돼서는 안 된다.

다. 결합 서비스와 기술

개인정보보호법은 가명 처리된 데이터를 결합한 데이터를 활용할 수 있도록 결합에 대한 절차를 정의하고 있다. 서로 다른 개인정보 처리자 간의 가명정보 결합은 결합 전문 기관을 통해야 한다.

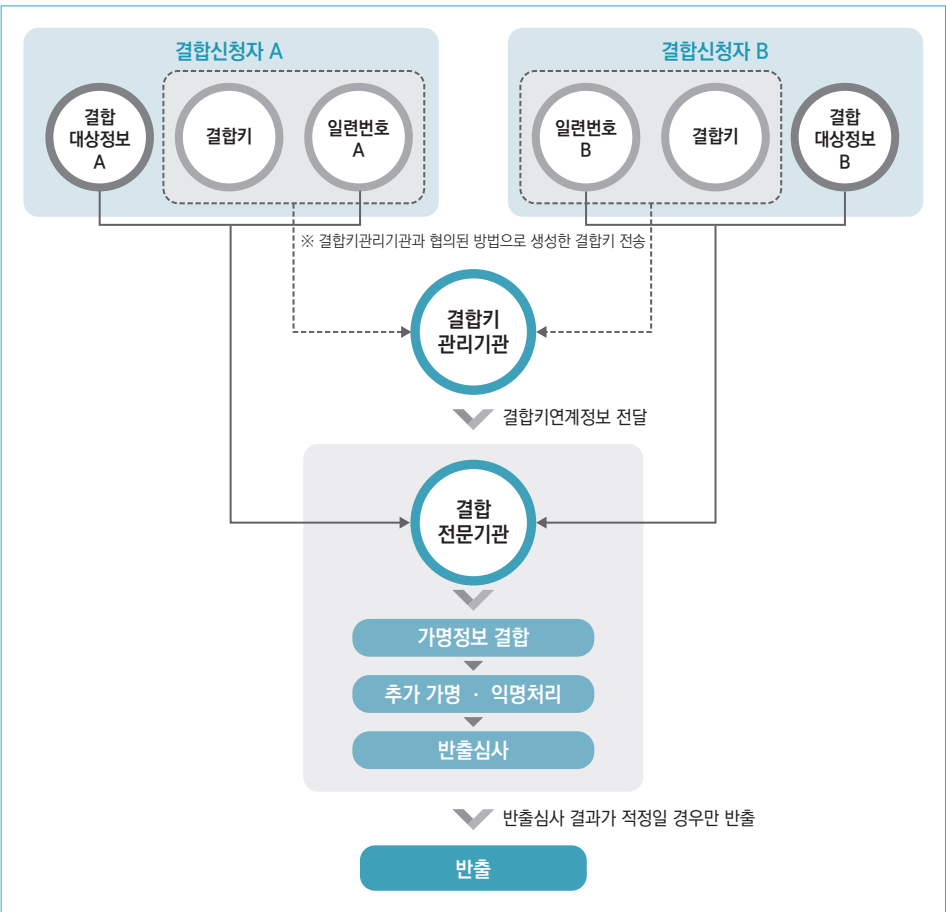
결합 전문 기관은 결합키의 생성에 관여하지 않고 '결합키 관리 기관(한국인터넷진흥원이 결합키 관리 기관으로 지정됨)'을 통해 결합키 생성 방법을 정의하고 이를 결합 신청자들에게 전달한다.

결합 전문 기관을 통한 가명 정보의 결합 절차는 다음과 같다.

- 1) 결합 신청자는 각자 보유한 개인정보를 활용해 ① 일련번호, ② 결합키, ③ 결합 대상정보를 생성한다.
- 2) 결합 신청자는 결합 신청이 접수된 후 1)단계에서 생성한 결과물(①, ②, ③) 중 결합키 관리 기관에는 ①, ②를 송신하고 결합 전문 기관에는 ①, ③을 송신한다.
- 3) 결합키 관리 기관은 결합 신청자로부터 수신한 정보를 이용해 '결합키 연계 정보'를 생성해 결합 전문 기관에 전달한다.
- 4) 결합 전문 기관은 결합키 연계 정보를 이용해 결합 신청자로부터 수신한 '결합 대상정보'를 결합한다.
- 5) 결합 신청자는 결합 전문 기관 내에서 결합된 정보에 대한 추가 가명·익명 처리를 하고 결합 전문 기관에 반출을 요청한다.
- 6) 결합 전문 기관은 반출심사위원회를 구성해 반출 여부를 심사하고 적정인 경우 반출을 승인한다.

결합 절차 상에서 특이한 점은 실제 결합된 가명 데이터에 대해서는 반출 전에는 '결합 전문기관' 안에서만 볼 수 있다는 점이다. 또 결합키에 대해서는 결합 신청자와 '결합키 관리 기관'만, 실제 결합에 사용되는 결합키 연계 정보에 대해서는 '결합키 관리 기관'과 '결합 전문 기관'만 볼 수 있다. 최종적으로 반출 승인되지 않은 결합 정보 혹은 결합키가 외부로 유출되는 것을 방지하기 위한 절차라고 판단된다.

[그림 5-4-2] 가명정보 결합, 반출 절차



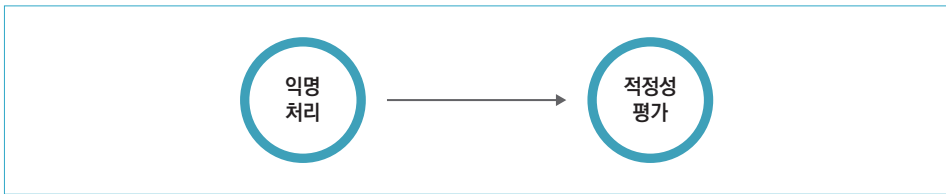
출처: 「가명정보 처리 가이드라인」, 개인정보보호위원회, 2020.9.

금융 분야의 경우 결합키 관리 기관 없이 ‘데이터 전문 기관’이 결합 전문 기관 역할을 하게 돼 있다. 결합 의뢰 기관은 결합 대상 정보집합물을 가명 처리한 후 결합 의뢰 기관 간에 직접 협의한 방식으로 결합키를 생성하고 데이터 전문 기관에 정보집합물 결합을 신청한다. 즉 금융 분야는 결합을 신청하는 금융 기관끼리 어떤 정보를 기반으로 어떤 알고리즘으로 결합키를 생성할지 결정한 후 결합키와 가명 처리된 결합 대상 정보를 데이터 전문 기관에 신청한 후 보내는 방식으로 진행한다.

라. 익명 처리 기술

신용정보법에서는 익명화 처리에 대해 [그림 5-4-3]과 같이 2단계 과정과 익명화 처리 기록 보존을 통해 회사 내부나 제3자에게 제공할 수 있도록 하고 있다.

[그림 5-4-3] 익명화 처리 2단계



[표 5-4-1]에서 언급한 대로 익명 정보는 개인정보가 아니기 때문에 자유롭게 유통하거나 활용할 수 있다. 따라서 익명화 수준을 판단할 때 신중해야 한다.

금융위원회가 발간한 ‘금융분야 가명익명 처리 안내서’에 따르면 대상 정보에 대해 식별자와 개인 식별 가능 정보로 분류한다. 이 후 식별자는 목적에 필요한 경우를 제외하고 원칙적으로 삭제하고 식별 가능 속성들을 가능성 정도에 따라 k-익명성 모델을 포함한 익명화 기법들의 수준을 높여 처리하도록 요구하고 있다.

마. 가명·익명 및 결합 처리 시스템 구축 방향

데이터3법에 따른 개인정보의 가명화, 익명화는 여러 가지 준수해야 할 의무와 절차를 정의하고 있다. 기술적인 핵심 사항은 데이터 자체 혹은 결합된 데이터 등의 가명화, 익명화 처리와 적정 수준을 판단하기 위한 정보 생성이다.

데이터3법은 기관에 구축해왔던 빅데이터의 효율적 활용에 대한 요구를 반영하고 있으며 가명·익명 처리 시스템은 개인정보의 유출 방지를 위해 분리된 시스템에서 엄밀한 절차를 걸쳐 처리될 필요가 있다. 빅데이터의 보안 송수신부터 가명화, 익명화 및 결합 처리, 적정성 지표 계산은 데이터 크기의 수준과 트랜잭션 양에 따라 충분한 동시 연산 능력을 가진 시스템을 구축하도록 요구한다. 기업 및 기관은 처리할 데이터 용량과 동시 처리 필요량에 따라 확장 가능한 분산 빅데이터 플랫폼을 구축할 필요가 있다.

원 데이터는 개인정보를 담고 있어 가명화, 익명화 처리 전에는 반출이 허용되지 않는다. 통계, 연구, 기록 보존 등의 목적으로 개인정보를 다루고자 하는 기관들은 보안성과 고성능 분산 처리 능력을 함께 갖춘 전문 가명, 익명 처리 시스템을 앞다투어 구축하고 이러한 기능을 구현한 전문 솔루션 시스템들이 경쟁할 것으로 예상된다.

결합의 경우 내부 데이터의 결합이 있을 수 있고, 결합 전문 기관에서 사용할 결합이 있을 수 있다. 결합은 원본 데이터를 여러 개 합해 더 큰 데이터를 만드는 것이므로 데이터 처리 용량을 고려해야 한다. 그러나 결합 후 평가에서 기본적인 가명, 익명 처리에 사용하는 평가 알고리즘과 추가로 가명, 익명 처리 알고리즘이 모두 필요할 수 있기 때문에 가명, 익명 처리 시스템과 비교해 다른 추가 요구사항은 없다.

제5장. 데이터 분석 및 시각화 기술

필자: 안동욱
(미소정보기술 대표)

‘데이터를 보다’라는 표현은 다소 추상적일 수 있다. 글은 보는 것이 아니고 읽는 것이고, 읽는 행위는 이해하고 공유하기 위한 기본적인 절차다. 그렇다면 데이터를 본다는 것은 상대를 이해시키고 공유할 수 있는 형태로 전달돼야 하는데, 이것이 바로 ‘시각화’다. 이런 시각화의 시도는 주로 정적인 차트 또는 인포그래픽으로 발전해왔으며, 다양한 대시보드 솔루션들이 활용돼 왔다. 하지만 빅데이터와 인공지능에 이르는 광범위한 데이터의 활용은 좀 더 다른 의미의 시각화를 요구하기 시작했다.

1. 데이터 바로 보기

데이터는 모든 행위를 숫자나 문자의 기록으로 남긴 것을 말한다. 이를 정형 데이터와 비정형 데이터로 구분하곤 하는데 최근에는 분석 대상이 되는 숫자, 문자, 기호, 이미지, 소리, 영상 등을 모두 데이터로 칭한다. 의미의 확대라기보다는 쓰임새가 넓어지고 다양해지면서 인간의 행위를 이해하고 기계를 인간화하는 데 필요한 재료들을 통칭하기에 이르렀다. 내가 보고 듣고 입고 먹고 자고 읽는 모든 것의 기록에 라벨링을 하면 인공지능에 필수적인 학습데이터가 된다. 마치 이제 세상이 영화 ‘매트릭스’의 한 장면처럼 모든 것이 ‘0’과 ‘1’로 보이는 세상이 돼버린 것 같다. 이러한 데이터 경쟁시대에서 중요한 것은 ‘무엇을 보느냐’와 함께 ‘어떻게 보느냐’다. 이것이 바로 데이터 시각화의 중요성이다.

2. 데이터와 대화하기

데이터 시각화는 데이터와 대화하기 위한 기술 또는 태도다. 데이터 시각화의 결과는 사실 결과가 아니고 새로운 질문으로 볼 수 있다. 이 질문에 다시 답변하기 위한 탐구를 하고 질문하는 방식의 대화가 이어지는 것이 시각화 분석의 기본적인 접근 방식이다. 따라서 시각화한 차트는 정적인 것이 아니라 재질의가 가능하도록 반응해야 한다. 그 질의의 형태는 시각적인 형상화와 대화형의 쉬운 인터페이스여야 하고, 시각화 설계 요소들은 하나의 분석에 다양한 관점을 가지게 할 수 있는 유연함을 제공해야 한다.

3. 데이터 시각화 시장 동향

시각화를 통해 데이터와 대화하고, 새로운 질문에 답변하고, 꼭 스토리를 담아야 한다. 데이터는 온갖 이야깃거리다. 데이터는 담겨져 있을 뿐 보는 이마다 다른 해석을 하고 저마다의 이야기를 만들어낸다. 데이터는 사실에 근거하기 때문에 다른 이야기를 하더라도 공감하게 되고 충분한 설득력을 가진다. 이 기본적인 열개를 잘 풀고 해석하고 섞고 공유하는 과정이 시각화다.

글로벌 데이터 시각화 시장 규모는 다양한 기업의 시각적 분석, 과학적 시각화 및 정보 시각화에 대한 수요가 증가함에 따라 상당한 성장을 보일 것으로 예상된다. 데이터 시각화는 정보 또는 데이터를 그래픽 형식으로 표현하는 것이며 여기에 사용된 도구는 데이터를 액세스 가능한 방식으로 레이아웃하고 통계, 그래프 및 아웃 라이어를 잘 이해하는 데 도움을 준다. 포천비즈니스인사이트(Fortune Business Insights)에서 최근 발표한 보고서인 ‘데이터 시각화 시장규모 2020~2027’을 따르면, 2019년에 88억 5,000만 미국 달러를 기록했으며 연평균 10.2% 성장해 2027년 말까지 192억 미국 달러에 이를 것으로 예상된다.⁰¹

시각화에 대한 기업 및 비즈니스의 요구가 증가하고 있다. 이로 인해 기업은 시각화 도구에 투자하고 시장의 유명한 기업들과 합작 투자를 한다. 예를 들어 2019년 8월 세일즈포스닷컴은 분석 기능 향상과 인공지능을 제공하기 위해 타블로소프트웨어(Tableau Software)를 인수했다.

01. 글로벌 뉴스 와이어, 2020.7.9.

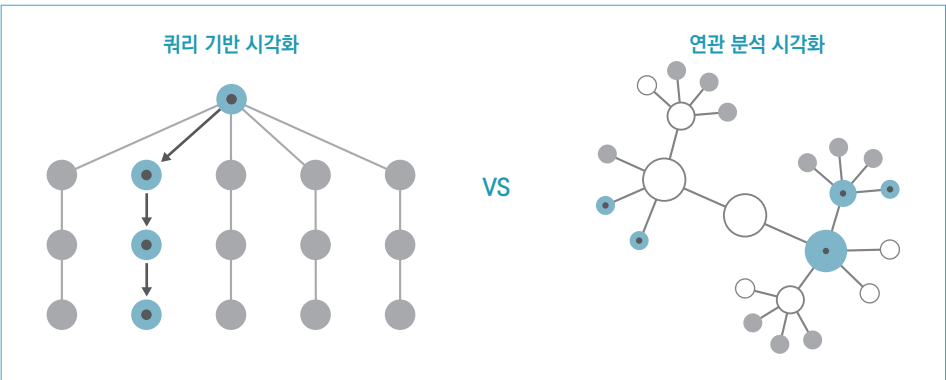
4. 데이터 시각화 기술 동향

가. 데이터 분석을 위한 연관 분석 기법

대화형, 비정형 탐색 및 분석을 위해 최근에는 연관 분석 기법이 활용되고 있다. 다수의 데이터 소스를 완전히 결합하고 인덱싱해 가능한 연관을 조사하며, 어떤 데이터도 제외되지 않는다. 연관 분석 엔진은 강력하고 신속한 계산 및 집계를 통해 분석을 즉시 업데이트하고, 각각을 클릭 후 관련된 값과 관련되지 않은 값을 모두 노출시키며, 데이터 내의 연관성을 모두 강조해서 보여준다. 이제 전문가에게 문의하고 기다릴 필요 없이, 보는 것을 바탕으로 자유롭게 제한 없이 검색, 탐색, 피벗을 할 수 있다. 이것이 쿼리 기반 도구에서는 볼 수 없었던 예상치 못한 통찰을 연관 분석 사용자가 지속적으로 발견해 엄청난 가치를 창출하는 이유다.

연관 데이터 모델링 기술을 강력하게 만들어주는 주요 특성은 다음의 세 가지와 같다. 첫 번째는 다양한 소스에서 데이터를 가져오는 능력이다. 사용되지 않는 데이터가 없고, 잘못 처리되는 데이터도 없다. 둘째, 사용자는 단순한 검색 및 선택을 통해 경계 없이 탐색하고, 모든 질문과 후속 질문을 할 수 있다. 마지막으로 사용자는 전문가에게 문의하고 새로운 질문을 기다릴 필요 없이 생각의 속도로 작업할 수 있다. 이러한 주요 속성을 결합해 모든 기술 수준의 사용자에게 비교할 수 없는 경험을 제공한다.

[그림 5-5-1] 쿼리 기반 시각화와 연관 분석 시각화 비교



나. 쿼리 기반 도구 - 한계와 사각지대

쿼리에 의존해 정보를 분석하는 시각화 도구에서는 사각지대가 생겨 결국 정보를 완전히 이해할 수 없다. 이러한 도구는 구조적인 선형 접근 방식을 사용해 데이터의 부분 집합만을 시각화하지만, 연관 엔진은 모든 데이터에 대해 자유형 탐색 및 검색 기능을 제공한다. 쿼리 기반 도구가 연관 분석의 탐색 경험을 모방하려 해도 유연성과 성능 면에서 금세 한계에 봉착한다. 쿼리 기반 도구에서 예상치 못한 질문에 대답하려면 복잡한 SQL 쿼리를 다시 작성해야 한다. 그런데 일반 비즈니스 사용자는 쿼리를 작성하지 못하므로 대개 숙련된 데이터 전문가가 필요하다.

쿼리 기반 도구는 기본 소스에 직접 연결하거나, 데이터 정적 영역의 일부 영역을 이용한다. SQL 조인을 사용해 소스를 결합하는 경우, 특히 여러 개의 소스를 결합하는 경우에는 데이터 손실이 일어나거나 부정확성이 발생한다. 데이터 웨어하우스와 정적 영역은 상당한 시간을 투자해 사전 모델링을 해야 한다. 또한 사용자의 상호작용을 지원하려면 클릭·질문마다 쿼리를 작성해 유연성을 제한하고 사용자를 데이터 부분 집합으로 제한한다.

시각화 및 쿼리는 일반적으로 고급 사용자에게 의해 구축되며, 다른 사람들은 개발자가 생각한 사전 형성된 질문으로 제한된다. 간단한 필터링 외에 비즈니스 사용자가 객체 내에서 새로운 질문을 검색 및 탐색할 수 있는 방법은 없다. 유일한 선택은 데이터 전문가가 새로운 쿼리 또는 시각화를 구축하기를 기다리는 것이다.

다. 경계 없는 탐색

연관 엔진을 이용하면 제한있는 경계에 구애받지 않고 어떠한 방향에서든 전체 데이터 소스를 자유롭게 탐색할 수 있는 연관 사용자 경험을 제공한다. 사용자는 시각화, 차트, 그래프, 필터 창, 전역 선택 인터페이스 등 어디서나 상호 작용할 수 있다. 연관 엔진은 클릭이 발생할 때마다 즉시 분석을 재계산하고 데이터 내 연관을 강조한다. 또한 엔진이 레코드 수준 데이터 세트에 기반해 동적으로 계산하므로, 사용자는 사전 정의된 질문이나 계층에 제한을 받지 않고 어떤 세부 수준에서든 원하는 항목을 선택해 새로운 아이디어나 데이터로 피벗할 수 있다. 이 과정을 계속 진행하면서 사용자가 더 많은 질문을 던지고, 컨텍스트를 추가하고, 각 단계에서 정보를 얻을 수 있다.

라. 생각의 속도

시각화하는 데에는 쿼리 기반 도구로 충분할지 모른다. 그러나 이러한 작업을 할 수 있는 경험을 갖추고, 작업을 담당할 의사가 있는 사용자는 소수에 지나지 않는다. 더구나 이러한 시각화를 지원하는 쿼리를 생성할 수 있는 사람의 수는 훨씬 더 적다. 그렇다면 대부분의

비즈니스 사용자는 어떨까? 결국 데이터의 부분 집합에 대한 제한적인 선형 분석만을 제공하는, 느린 정적 시각화와 보고서를 이용하게 돼 있다. 사전에 질문을 정확히 예측해 적절한 쿼리를 구축할 수 없으면 데이터 전문가에게 도움을 요청할 수밖에 없다. 이를 두고 ‘요청, 대기, 응답 사이클’이라고 한다.

연관 엔진은 분석을 동적으로 계산하고, 사용자가 질문을 떠올리는 것과 같은 속도로 연관성을 강조해 보여준다. 이렇게 즉시 응답을 받으면 시스템이나 전문가의 도움을 기다릴 필요 없이 필요한 응답을 바로 받을 수 있다. 이는 가치를 신속하게 창출하고 매일 수백 개의 의사 결정을 정보에 기반해 처리할 수 있음을 의미한다. 연관 엔진은 쿼리 기반 도구에서는 불가능한 타의 추종을 불허하는 속도와 유연성의 결합이 가능하다.

5. 데이터 시각화의 미래

팬데믹은 대부분의 오프라인 환경을 온라인화하는 언택트 시대를 앞당기고 있다. 이런 현상은 가파르게 늘어나고 있는 데이터의 양을 더욱 가속화시킬 것이다. 이렇게 풍부해진 데이터는 세상의 현상을 해석하고 예측하기에 충분한 자원이고, 이는 우리의 생활과 직업에 절대적인 영향을 끼칠 것이다. 데이터를 다룬다는 것은 일상이 될 것이며, 대중적인 데이터 접근을 위해 데이터 시각화는 더욱 큰 시장을 형성할 것이다. 다만 앞서 기술동향에서 언급했듯이 전문가가 아닌 대중이 데이터 소비를 쉽고 간결하게 할 수 있는 시각화 도구가 지속적으로 요구될 것이다.

2020년 데이터산업 백서 집필진

제1부. 역동적인 데이터산업 생태계 지각변동과 기회

1장. 데이터 권리, 연결 그리고 새로운 가치	김인현 대표(투이컨설팅)
2장. 데이터 주도 경제 성장과 발전	박영일 원장(강원연구원)
3장. 새로운 시장생태계의 지각 변동	조현도 수석연구원(넥스텔리전스)

제2부. 데이터산업 주요 정책 현황

1장. 국내 데이터산업 정책 현황	김동일 실장(한국데이터산업진흥원 경영기획실)
2장. 해외 데이터산업 정책 현황	김희성 이사(넥스텔리전스)

제3부. 데이터산업 시장 현황

1장. 국내 데이터산업 시장 현황	유효라 선임(한국데이터산업진흥원 기업지원팀)
2장. 해외 데이터산업 시장 현황	나영민 이사(날리지리서치그룹)

제4부. 산업별 데이터 활용 비즈니스 현황

1장. 금융 분야 데이터 활용 비즈니스	장혜리 이사(에이젠글로벌)
2장. 헬스케어 분야 데이터 활용 비즈니스	김준연 대표(인바이츠헬스케어)
3장. 제조 분야 데이터 활용 비즈니스	이재혁 공동대표(마키나락스)
4장. 통신 분야 데이터 활용 비즈니스	김정선 박사(SK텔레콤 T3K)
5장. 광고 분야 데이터 활용 비즈니스	안익진 대표(몰로코)
6장. 교육 분야 데이터 활용 비즈니스	김진숙 수석연구위원(한국교육학술정보원)
7장. 유통·물류 분야 데이터 활용 비즈니스	신광섭 교수(인천대학교 동북아물류대학원)

제5부. 데이터산업 기술 동향

1장. 데이터 생산 및 수집 기술	안필용 상무(인젠트)
2장. 데이터 저장 및 관리 기술	허승재 상무(티맥스티베로)
3장. 인공지능 기술	나동원 이사(비투엔 데이터지능본부)
4장. 데이터 보안과 비식별 기술	윤경구 전무(파수 개발1본부)
5장. 데이터 분석 및 시각화 기술	안동욱 대표(미소정보기술)

2020년 데이터산업 백서 [통권 33호]

2020 DATA INDUSTRY WHITE PAPER

편찬위원	김인현 투이컨설팅 대표
	박영일 강원연구원 원장
	장홍성 지능정보산업협회 협회장
	전종훈 명지대학교 교수
	차상균 서울대학교 교수
편집위원	이종서 한국데이터산업진흥원 산업지원실 실장
	하진희 한국데이터산업진흥원 산업지원실 기업지원팀 팀장
	유효라 한국데이터산업진흥원 산업지원실 기업지원팀 선임
발행처	한국데이터산업진흥원
	서울특별시 중구 세종대로 9길 42, 부영빌딩 8층(04513)
	Tel. 02. 3708. 5300
	www.kdata.or.kr
발행인	민기영
발행일	2020년 12월 30일
ISSN	2465-7662

- 본 백서의 내용은 한국데이터산업진흥원의 공식 견해와 다를 수 있습니다.
- 본 백서 내용의 무단 전재를 금하며, 가공 인용 시에는 반드시 「한국데이터산업진흥원, 2020 데이터산업 백서」라고 밝혀 주시기 바랍니다.
- 본 백서는 과학기술정보통신부의 DB산업 육성 사업의 결과물입니다.
- 「2020 데이터산업 백서」와 관련한 문의는 한국데이터산업진흥원으로 연락해 주시기 바랍니다.

2020 2020 DATA INDUSTRY WHITE PAPER

데이터산업 백서

 한국데이터산업진흥원
Korea Data Agency

서울특별시 중구 세종대로9길 42 부영빌딩 7,8층
Tel. 02. 3708. 5300 Fax. 02. 318. 5040
www.kdata.or.kr



ISSN 2465-7662

[비매품]