

2020년 제4호 (통권4호)



데이터산업 동향 이슈 브리프

ISSUE BRIEF

2020.09

일본의 AI 사회 실현을 위한 데이터 경제 정책

일본의 AI 사회 실현을 위한 데이터 경제 정책

I. 검토 배경	1
II. AI 사회의 데이터 활용	2
1. 일본의 데이터 활용 상황	2
2. AI 사회 실현을 위한 데이터 활용	5
III. AI 시대의 데이터 경제 정책	7
1. ‘새로운 자산’으로서의 데이터 기능과 역할 및 효과와 가치 측정 방법	7
2. 데이터 효과와 가치에 상응하는 정당한 보상 방식	11
3. 데이터 거래의 시장화와 가시화를 위한 과제	16
IV. 결론 및 시사점	20

요약

- AI의 잠재력 실현을 위해서는 무엇보다 데이터가 '새로운 자산'으로서 그 가치를 극대화하기 위해서는 바람직한 가치 측정 및 분배 방식에 대한 검토가 필요
 - 데이터를 생산요소의 하나로 평가하고, 생산성 향상에 미치는 영향을 측정하며, 데이터에 대한 정당한 보상 논의를 심화할 필요성이 제기
 - 그러나 데이터가 생산성 향상에 구체적으로 어떻게 결부되어 있으며, 데이터의 효과 및 가치에 따라 어떻게 분배되어야 하는 것인가는 일본뿐만 아니라 국제적으로도 확립되지 않은 상황
 - 일본 AI 네트워크사회추진회의 산하AI 경제검토회는 '20년 발간한 보고서*'를 통해 AI사회 구현을 위한 데이터 가치 측정과 분배 방식을 검토
 - * AI 네트워크사회推進會議, AI 經濟検討會, 報告書 2020, 2020.7
- 데이터는 일반적으로 자본으로 취급되는 경우가 많지만 AI는 관련 데이터를 생성하는 개인의 적극적인 참여에 의존한다는 특성을 감안하여 데이터를 노동으로 취급하고 데이터를 생성하는 개인에게 금전 등의 방식으로 보상해야 한다는 의견도 존재
 - '자본으로서의 데이터 (Data as Capital)'는 기업가정신과 혁신의 촉진이라는 관점으로부터 데이터가 산출하는 가치의 기여자로 데이터 AI기업·플랫폼 사업자·데이터 과학자 등의 활용자를 중시
 - '노동으로서의 데이터 (Data as Labor)'는 데이터를 생성하는 개인의 공헌을 일반 노동자와 동일하게 인정하는 관점으로부터 데이터가 산출하는 가치의 기여자로 데이터 생성자를 중시
- 한편, 데이터로부터 발생하는 가치에 대한 참여라는 측면에서 데이터 생성자를 중시하는 접근이 최근 다양한 조직과 활동에서 활용
 - EU의 '일반데이터보호규정(GDPR)'에 따라 보호 대상이 되고 있는 IoT 디바이스나 커넥티드 카 등 기계를 통해 수집된 비개인 데이터와 관련해, '17년 EU집행위원회가 제시한 '미래 데이터 접근 프레임워크'의 한 항목으로 데이터 생산자 권한(Data producer's right)이 제시
 - 개인 데이터에 관해서는 '18년 '마이데이터 글로벌(MyData Global)'이 설립되어 개인이 자신의 데이터에 대해 충분히 이해하며 주체성과 주도권을 갖고 자신을 위해 개인 데이터를 활용할 수 있는 방안이 추진
- 현재 데이터 거래에 따른 시장 메커니즘이 충분히 작동하고 있지는 않지만, 불완전한 거래 방식을 보완하면서 미래의 이상적 시장 메커니즘을 중심으로 자원 배분을 조속히 달성하는 것이 바람직
 - 각 이해관계자에 대한 합리적 보상을 바탕으로 데이터 거래 시장 메커니즘을 최대한 활용해 데이터로 인한 경제적 이익을 극대화하기 위해 △데이터 기반 구축 △데이터 통합의 실현 △제3자에 의한 기업 보유 데이터 관리·정보은행 추진 △데이터 활용 촉진을 추진하기 위한 법제도의 검토 등 구체적 방안이 요구

- ▶ 미국의 구글, 아마존, 페이스북, 애플, 중국의 알리바바, 텐센트와 같은 세계적인 기업들은 독자적으로 AI를 개발하거나 AI의 개발·이용의 기반이 되는 툴을 제공하며 AI 사회 실현의 주도권을 장악
- 일본 정부는 '19.6월 '통합혁신전략추진회의'에서 'AI 전략 2019(AI 戦略 2019)'를 발표하는 등 이러한 세계적인 흐름에 뒤처지지 않기 위해 노력
- ▶ AI의 잠재력 실현을 위해서는 무엇보다 데이터가 '새로운 자산'으로서 그 가치를 극대화할 수 있어야 함
 - 데이터 가치의 극대화를 위해서는 바람직한 가치 측정 및 분배 방식에 대한 검토가 필요
 - 그간 일본 정부의 다양한 전략에 따른 시책과 일부 기업들의 빅데이터 분석, 기계학습 등에 대한 대응이 추진되면서 생산 활동에서의 데이터 역할에 대한 이해가 심화
 - 그러나 데이터가 생산성 향상에 구체적으로 어떻게 결부되어 있으며, 데이터의 효과 및 가치에 따라 어떻게 분배되어야 하는 것인가는 일본뿐만 아니라 국제적으로도 확립되지 않은 상황
 - 데이터를 생산요소의 하나로 평가하고, 생산성 향상에 미치는 영향을 측정하며, 데이터에 대한 정당한 보상 논의를 심화할 필요성이 제기
- ▶ 이번 코로나 19 팬데믹을 계기로 AI 사회 실현이 가속화될 것으로 전망되는 가운데 그 실현 방식에 대한 새로운 검토 관점이 요구
 - 코로나 19로 인해 경제 활동과 라이프 스타일이 변화하고 통신 네트워크와 ICT 툴이 주목받게 됨에 따라 AI를 활용한 데이터의 수집과 활용의 중요성이 가중
- ▶ 금번 이슈 브리프에서는 일본 AI 네트워크사회추진회의 산하 AI 경제검토회가 '20년 발간한 보고서*'의 주요 내용을 요약하고 정책적 시사점을 제시
 - * AI ネットワーク社会推進会議, AI 経済検討会, 報告書 2020, 2020.7
 - AI 경제검토회는 AI 사회 실현을 위해 어떤 사회·경제를 지향해야 하는가를 논의하고, 기본 정책과 중장기 전략 방향을 검토할 목적으로 '19.1월에 설치
 - 이번 보고서에서는 △AI 이용을 둘러싼 산업 구조 △지속가능한 경제 성장과 생산성 향상을 위한 AI 투자 방식 △AI 경제를 지탱하는 산업 기반(노동·연구개발)의 형태 △AI 이용을 둘러싼 일본 기업의 국제 경쟁력 △AI 경제에 관한 기본 정책과 전략의 방향 등을 논의

PART II

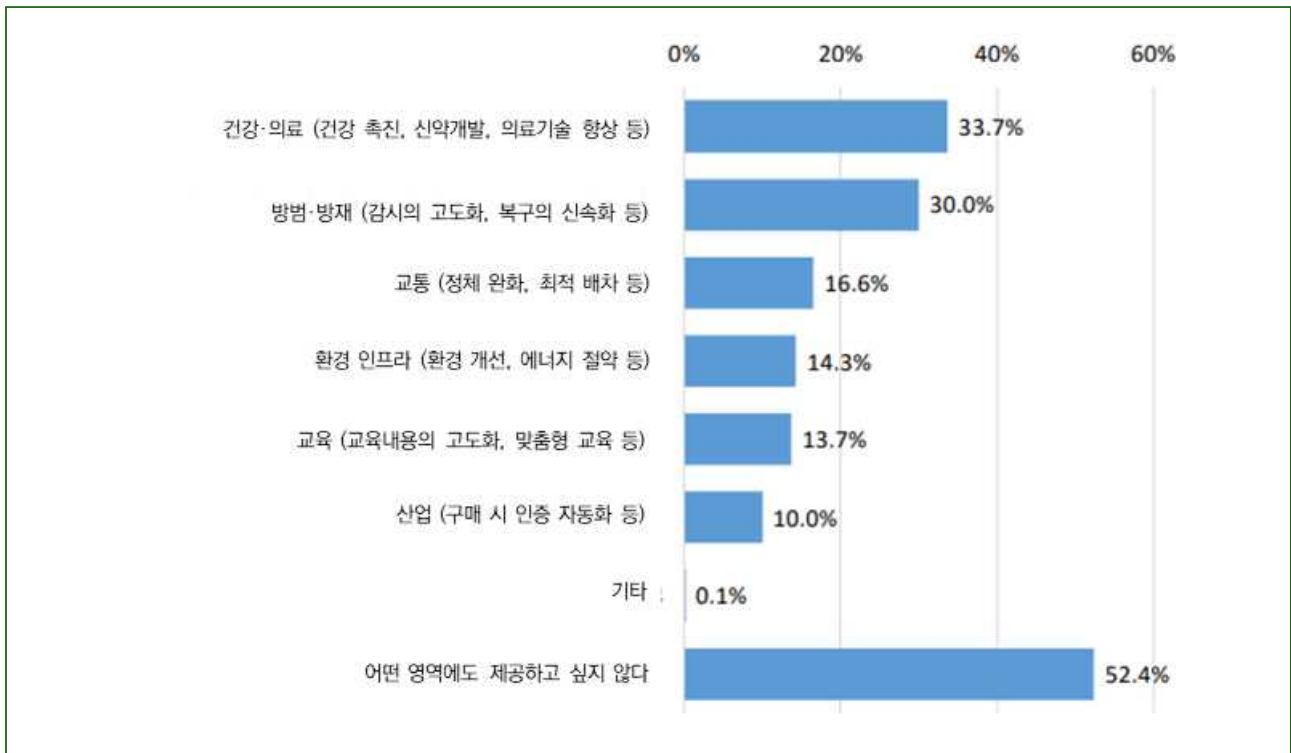
AI 사회의 데이터 활용

1 일본의 데이터 활용 현황

- ▶ 금융 주도의 ICT와 데이터를 활용한 디지털 기업이 글로벌 경제를 주도하고 있으나 일본의 디지털 경쟁력은 높지 않은 편
 - 세계 시가총액 상위 10대 기업을 살펴보면, 1989년에는 일본 기업이 상위를 차지하며 금융기관이나 에너지 관련 기업이 다수 포함되었지만, '18년에는 미국의 구글, 아마존, 페이스북, 애플이나 중국의 알리바바, 텐센트 등 사용자 데이터를 수집·축적해 활용하는 기업이 대부분을 차지
 - IMD의 세계디지털경쟁력 순위(World Digital Competitiveness Ranking)에 따르면, '19년 일본의 디지털 경쟁력 순위는 63개국 중 23위로, 디지털 경쟁력의 결정 요소의 하나인 '비즈니스 민첩성'은 41위
 - 특히 '비즈니스 민첩성'의 구성 항목인 '빅데이터 활용·분석'에서는 63위로 최하위 차지
- ▶ 일본 기업들의 데이터 활용에서도 대기업과 중소기업의 규모 격차와 대도시권 기업과 지방 기업의 격차가 존재
 - 총무성이 실시한 '디지털 데이터의 활용에 관한 설문 조사'에 따르면, 조직 데이터를 수집·저장하지 않는 기업은 대기업(종업원 1,000명 이상)이 7.1%인데 비해 중소기업 (종업원 300명 미만)은 45.5%로 집계
 - 수도권 기업과 지방 기업 사이에도 비슷한 격차가 나타나고 있는데, 본사가 도쿄나 오사카 등 대도시에 위치한 기업에 비해 지방 기업은 데이터에 대한 대응이 늦어지고 있는 것으로 파악
- ▶ 일본 기업들은 데이터의 수집과 분석을 포함한 데이터 관련 대응이 늦어지고 있다고 인식
 - AI는 대규모 데이터를 바탕으로 학습을 진행하므로 AI의 활용을 위해서는 데이터의 수집과 분석이 필수적이거나 일본 기업의 약 60%는 자사의 데이터에 대한 대응이 다른 나라 기업에 비해 늦어지고 있다고 인식
 - 데이터를 활용하고자 하는 의욕이 있어도 데이터 활용과 관련된 인력 부족, 비용 대비 효과의 불확실성, 충분히 수집·축적되지 못한 데이터 등으로 인해 대응이 늦어지는 것으로 조사
 - 반면 데이터를 활용한 경험이 있는 기업 중 60% 이상의 기업이 사내 데이터 활용을 통해 효과를 거둔 것으로 평가
 - 특히 생산·제조 영역의 기업들에서 데이터 활용 효과가 높았던 것으로 분석

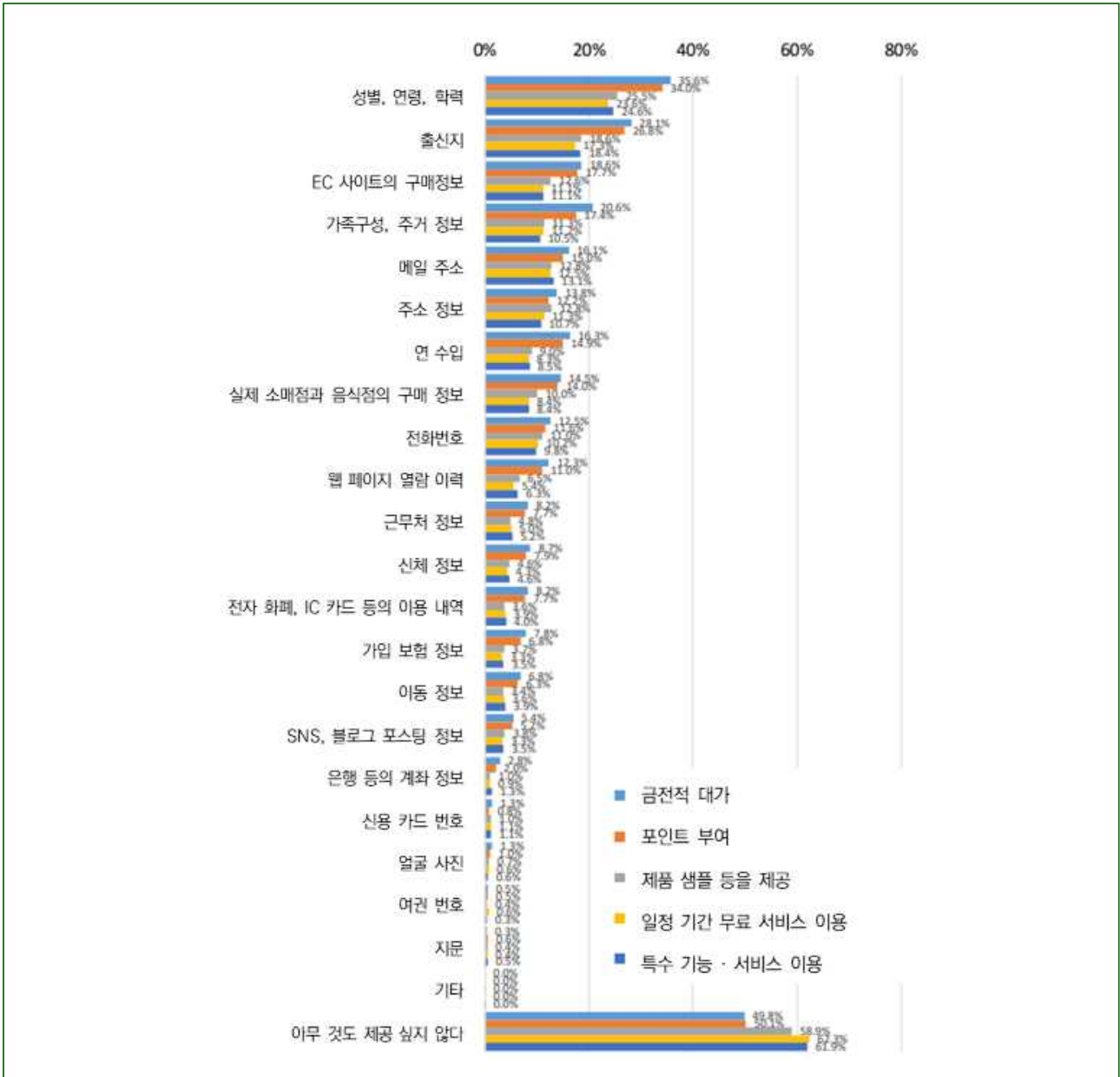
▶ 반면, 기업에 데이터를 제공하는 소비자들은 자신의 개인 정보나 행동 이력 등을 제공하는 것에 신중한 태도를 견지

- 총무성 ‘디지털 데이터의 활용에 관한 설문 조사’에 따르면, 자신의 개인 정보나 행동 이력 등의 데이터가 사회 전체 영역에 도움이 된다면 어떤 대가 없이도 제공하겠다는 소비자보다 어떤 영역에도 제공하고 싶지 않다는 소비자 비율(52.4%)이 높은 것으로 조사
 - 건강·의료 (건강 증진, 신약 개발, 의료 기술의 향상 등) 영역에 도움이 된다면 어떤 대가 없이도 제공할 수 있다고 응답한 소비자의 비율은 33.7%
 - 방재·방법과 교통, 교육 등의 영역에 비해 건강·의료 영역의 데이터의 활용에 대한 소비자의 저항이 적은 것으로 분석



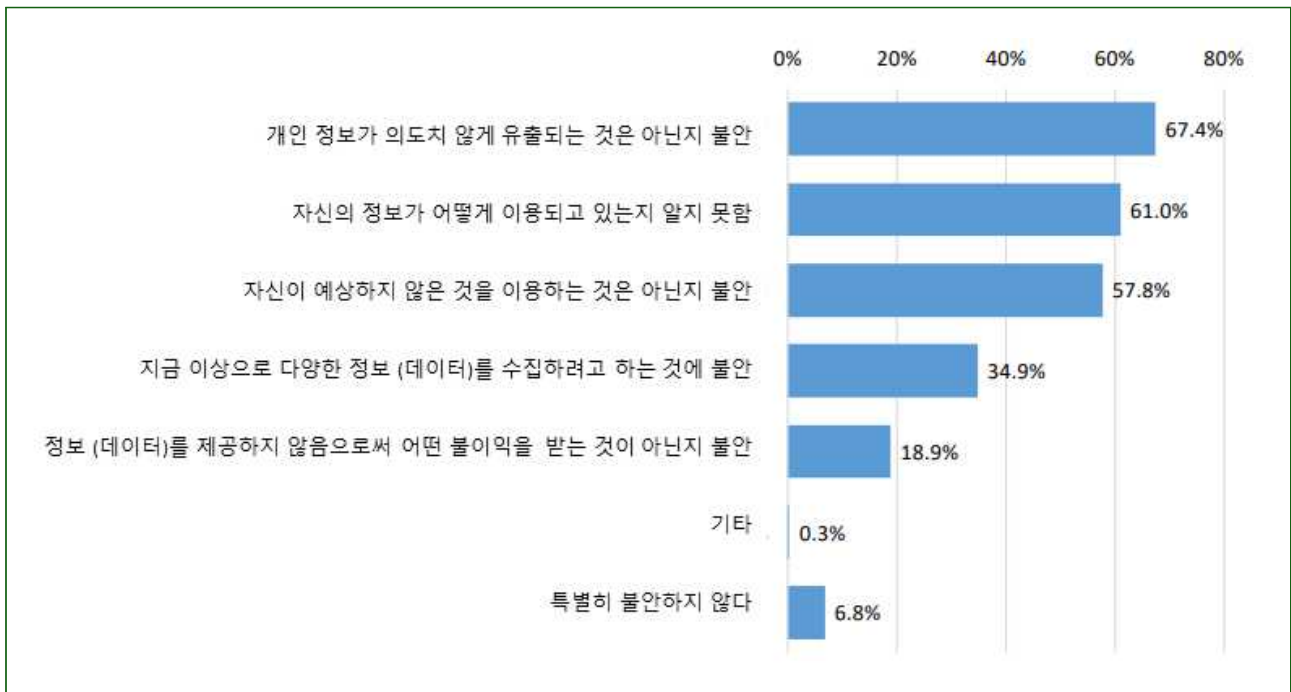
| 그림 1 | 데이터 제공 의사 (기여 영역별)

- 금전적 대가나 포인트 부여와 같은 보상이 있을 경우, 자신의 개인 정보나 행동 이력 등의 데이터를 제공할 수 있는지에 대해 ‘아무 것도 제공하고 싶지 않다’를 선택한 소비자가 약 50% 이상으로 높은 비율을 차지



[그림 2] 데이터 제공 의사 (데이터 내용별)

- 데이터 제공에 대해 소비자가 신중한 자세를 취하는 이유는 자신의 개인 정보나 행동 이력이 수집·이용되는 것에 대한 불안이 있기 때문
 - ‘개인 정보가 의도치 않게 유출된 것은 아닌지 불안’하다고 응답한 소비자의 비율이 67.4%로 가장 높음
 - ‘자신의 정보가 어떻게 이용되고 있는지 알 수 없다’는 응답이 61.0%, ‘자신이 의도하지 않은 이용 방식을 이용한 것은 아닌지 불안’하다는 응답이 57.8%로 자기 정보의 유출 위험과 이용 방법에 대한 불안이 높음
 - 또한 ‘지금 이상으로 다양한 정보 (데이터)를 수집하려고 하는 것에 불안’하다고 응답한 소비자의 비율이 34.9%로 집계



| 그림 3 | 데이터가 이용되는 것에 대한 개인의 불안 형태

2 AI 사회 실현을 위한 데이터 활용

가. 기업

- ▶ 데이터는 경영 자원 중 하나이며, AI는 도구이므로 기업은 데이터와 AI를 경영 전략과 사업 아키텍처 안에 포함시킬 필요성 제기
 - 단순히 데이터나 AI만을 고려하는 것이 아니라 사람·물건·돈·데이터(AI) 등 경영 자원 전체에서 검토하는 것이 중요
 - 사람·물건·돈·데이터(AI)가 회사의 재무 및 성과에 기여할 수 있도록 사업 전략에 반영할 필요
 - AI 사업의 실현을 위해서는 아키텍처 설계도 요구되는데, ICT 시스템뿐만 아니라 사업 측면의 아키텍처를 설계하면서 반영하는 것이 중요
 - * (예시) △가치사슬(기업의 핵심 강점을 발견) △물리적 아키텍처(자원 배분 방식) △디지털 아키텍처(데이터 플랫폼 구축) △레거시 시스템(재구조 방식) 등
- ▶ 기업은 데이터 활용의 전제가 되는 물리적 데이터 수집 체제, 데이터 활용 조직 체제 및 조직 데이터 기반을 구축할 필요
 - (물리적 데이터 수집 체제 구축) 물리적 데이터를 디지털 서버에 반영하고 데이터 수집 체제를 구축해 나가는 것이 중요

- 방대한 데이터 자원을 축적했지만 디지털 데이터의 수집이 이루어지지 않거나 어떻게 물리적 데이터를 수집해야 하는지 알지 못하는 기업들이 다수
- * (예시) 물류 사업자는 드라이버 등의 인력과 영업 거점을 효율적으로 배치하기 위해 수십억 개의 화물을 실시간으로 추적할 수 있는 데이터 수집 방식의 설계가 필요

- (데이터 활용 조직 체제 구축) 데이터 활용에서는 조직의 설계와 인력 배치가 중요
 - 기업 경영진의 이해와 약속을 바탕으로 데이터 활용 담당 조직에 권한과 예산이 배정될 필요
- (조직 데이터 기반 구축) 조직 데이터 유통 및 공유 기반 구축이 요구
 - 구체적으로 △조직 데이터 카탈로그에 의한 데이터의 가시화 △권리, 지식재산권, 계약, 보안 등의 절차 △구조화 및 비구조화 데이터 등 다양한 데이터에 대한 대응 등의 요구 사항을 충족시키는 것이 필요

▶ 기업은 데이터 현행법에 대응하는 동시에 데이터를 제공하는 소비자의 불안감을 해소하거나 경감시킬 필요

- 소비자가 자신의 개인 정보나 행동 이력 등을 기업에 제공하는 것에 신중한 자세를 취하므로 기업은 데이터 유통을 촉진시키기 위해 데이터 제공에 대한 소비자의 불안감을 해소 또는 완화하려는 노력이 요구
- * (예시) 정보의 수집, 분석 및 제3자 제공에서 정보 제공자 본인의 의사 확인, 이용자 동의의 취득이나 정보 이용 목적의 명시 등의 노력

나. 정부

▶ 다양한 주체가 데이터 활용의 이익을 누리기 위해서는 정부가 행정의 디지털화, 즉 디지털 정부(digital government) 추진 노력을 가속화할 필요

- 일본 정부는 ‘디지털 정부 실행 계획’에서 데이터 표준의 보급 등 행정 데이터의 제휴 추진과 보유 행정 데이터 100% 공개, 지방 공공단체의 오픈 데이터 추진에 의한 지역 과제 해결 촉진 등을 제시
- △데이터 기술 형식, 공통 해석 가능 어휘, 사용 문자 통일 등의 표준화와 같은 데이터 상호 운용성 확보 △데이터 취득용 API가 구현된 DATA.GO.JP 및 e-Stat (통계 GIS 기능 구현 완료) 공공 데이터의 민간 활용 기회 촉진 △지방 공공 단체의 오픈 데이터-바이-디자인 개념을 기반으로 정보 시스템의 설계와 정비를 포함하는 오픈 데이터의 추진 등이 필요

다. 코로나19 사태에 따른 과제

▶ 코로나19 사태로 인해 원격근무 및 재택근무가 증가함에 따라 통신 네트워크와 ICT 툴의 중요성이 확인되고 있으며, 데이터 관리 방법에 대한 재검토도 요구

PART III

AI 시대의 데이터 경제 정책

- ▶ AI 사회가 본격적으로 실현되는 AI 시대에는 데이터가 가치를 창출하는 원천으로 생산성 향상에 기여
 - 현재는 데이터를 활용해 산출된 가치의 정도나 그 프로세스 등의 실태를 파악하기 어렵고, 데이터 활용이 사회·경제에 미치는 효과나 제도·기술이 미치는 영향력에 대한 규명이 어려운 실정
 - 측정 대상인 데이터 및 데이터 효과를 명확히 하고, 선행 연구를 중심으로 데이터의 기능과 역할을 확인하며, 데이터 가치 측정 시 고려해야 할 요소를 정리할 필요
 - 또한 데이터 가치를 측정하기 위한 방법을 유형화하고, 각각의 사고방식과 과제에 대해 정리할 필요성도 제기
 - 각 산업 및 각국에 적용할 수 있는 다양성과 실행성 확보에 유의하면서 실현 가능한 방법으로 데이터의 가치 측정 방법을 검토하는 것이 중요

1 '새로운 자산'으로서의 데이터 기능과 역할 및 효과와 가치의 측정 방법

- ▶ (측정 대상) 데이터의 효과와 가치를 측정하기 위해서는 측정 대상이 되는 데이터 및 데이터의 효과를 명확히 하는 것이 필요
 - 데이터는 단위와 가치가 다양하므로 측정 방법, 측정 범위 및 목적 등에 따라 측정 대상을 검토

| 표 1 | 측정 대상

측정 대상	일반적인 디지털 플랫폼 구축 프로세스
데이터	<ul style="list-style-type: none"> • 본 보고서에서 데이터는 '실제 디지털화되어 있거나 쉽게 디지털화 가능한 상태의 것'으로 정의 • 데이터는 다양한 측면에서 분류 가능하다는 점을 고려하면서, 채택하는 가치 측정 기법을 바탕으로 대상을 검토하는 것이 필요 • 데이터의 종류에 따라 단위가 다르기 때문에(바이트, 인력, 건, 분 등) 여러 데이터를 다룰 때는 단위에 주의할 필요
데이터 효과	<ul style="list-style-type: none"> • 데이터에 가치를 부가하기 위해서는 처리와 분석이 필요하며, 아이디어와 혁신의 향상, 예측 정밀도의 향상 등을 기대 <ul style="list-style-type: none"> - 또한 이러한 과정에서 △사회 전체(CO₂ 감축 등) △기업(매출, 이익 등) △개인(효용 등)에 대한 효과가 기대됨 • 다양한 효과를 생각할 수 있으므로 어떤 효과를 측정 대상으로 삼을지를 명확히 할 필요 • 대상 범위와 결과 해석의 용이성은 트레이드오프(trade-off) 관계로, 개별 효과만을 대상으로 삼으면 결과 해석이 용이하지만 전체를 파악할 수 없으며, 전체를 대상으로 삼으면 결과 해석이 곤란

▶ (데이터의 기능과 역할) 데이터 가치 측정에 대한 검토 시에는 [표 2]와 같이 데이터의 기능 및 역할에 대한 다양한 측면의 요소를 고려할 필요

- △데이터가 가치를 창출하기 위해 요구되는 보완 요소 △시간에 따른 데이터 가치의 변화 △부정적인 경제적 가치 △ICT와의 관계 △무형 자산과의 관계와 같은 데이터의 기능과 역할을 고려

[표 2] 데이터의 기능·역할에 대한 검토 사항

데이터의 기능·역할	검토 사항
데이터가 가치를 창출하기 위해 요구되는 보완 요소	<ul style="list-style-type: none"> • 데이터를 활용하는 주체는 다양하며, 각각의 주체가 동일한 데이터를 활용하더라도 동일한 가치가 부가되는 것은 아님 • 데이터에 가치를 부가하기 위해 인력, 투자, 조직, 사업 형태, 시장 환경 등 필요한 보완적 요소를 검토할 필요
시간에 따른 데이터 가치의 변화	<ul style="list-style-type: none"> • 단독 데이터는 오늘 데이터, 한 달 전 데이터, 10년 전 데이터 등 데이터가 나타내는 시점에 따라 사업의 데이터 활용 효과가 상이 <ul style="list-style-type: none"> - 데이터를 조합해 활용할 경우 기존 데이터 덕분에 예측 정확도 등이 향상 가능 • 활용 용도 및 분석 방법에 따라 시간에 따른 데이터 가치의 변화는 상이할 수 있으며, 이를 어떻게 다룰지를 검토할 필요
경제적 가치에 부정적 영향	<ul style="list-style-type: none"> • 데이터 보유로 인해 보안 대책 등의 관리 비용과 정보 유출에 따른 이미지 하락 등의 부정적 위험, 즉 경제적 가치에 부정적 영향이 발생 가능 • 데이터 가치 측정 시에 이러한 경제적 가치에 대한 부정적 영향을 어디까지 고려할 것인가를 검토할 필요
ICT와의 관계	<ul style="list-style-type: none"> • 데이터 분석에는 ICT(하드웨어, 소프트웨어)가 필요하므로 데이터 가치는 광범위하게 파악하면 ICT의 가치로 대체해 생각해볼 수 있다는 견해도 존재 • 데이터의 가치와 ICT의 가치를 얼마나 명확히 구별할 것인지 검토할 필요
무형자산과의 관계	<ul style="list-style-type: none"> • 데이터 또는 데이터의 일부는 무형자산에 포함되는 것으로 파악 가능 <ul style="list-style-type: none"> - 데이터와 무형자산 양자에 공통된 특성도 있으므로, 기능, 역할 등에 관한 유사점 및 차이점을 고려하면서 무형자산에 대한 분석 방법을 참고하는 것이 유용 • 무형자산이 특히 기업 실적에 미치는 영향 등에 대한 선행 연구가 있으며, 데이터 가치 측정 참고가 가능

- 데이터의 기능과 역할과 관련해 데이터는 가치가 감소하지 않고 반복적으로 사용할 수 있다는 점, 한계비용(추가적인 비용)이 거의 0에 가깝고, 복제가 가능하다는 점 등의 [표 3]과 같은 특성도 고려될 필요
 - 추가 비용이 발생하는 경우로는 ICT 인프라의 증설, 처리 능력 확장 등에 따른 하드웨어 및 소프트웨어 구입 비용 등의 경우가 상정
 - 이외에도 데이터의 직접과 활용을 통한 가치의 발생, 데이터의 내용에 따른 가치나 취급 방법 차이, 데이터 양이 적을 경우 분석효과의 저하, 데이터 유출 위험, 동일 데이터의 경우에도 사용 영역과 방법에 따라 발생하는 효과의 차이 등과 같은 특성을 감안하는 것이 중요

| 표 3 | 데이터의 특성

데이터의 특성	개요
한계비용이 0에 가까움	• 복제가 용이하며 거의 비용을 들이지 않고 복제가 가능해 동일한 데이터를 여러 기업에서 활용하는 것이 기술적으로 용이
집적과 활용에 의해 비로소 가치가 발생	• 데이터를 많이 모으는 것 자체가 반드시 가치 있는 것은 아니며, 데이터에서 추출되는 다양한 의미와 지식을 활용하여 가치가 창출
데이터 내용에 따라 가치와 취급 방법이 상이	• 데이터 내용은 개인정보, 산업정보, 오픈데이터로 분류 가능 - 이 중 개인정보는 이용가치가 크나 엄격한 관리가 요구
데이터 양(건수)이 많지 않으면 분석효과(데이터 가치)가 작음	• 데이터의 양이 적을 경우 분석 효과는 제한적 - 단, 데이터의 양에 비례하여 효과가 무한대로 커지는 것은 아님
데이터 유출 위험 존재	• 개인 정보가 유출될 경우 기업은 부정적인 가치가 발생
동일한 데이터도 사용 영역과 방법에 따라 가치(효과)가 상이	• 오늘 데이터, 한 달 전 데이터, 10년 전 데이터 등 데이터가 나타내는 시점에 따라 실제 사업에서 데이터를 활용하는 효과가 상이

- ▶ (데이터의 효과와 가치 측정 방법) 데이터 가치 측정을 위해 다양한 시도가 이루어지고 있으며, 선행 연구 등을 기초로 비용, 시장 및 임팩트 측면에서 데이터의 가치 측정 방법 분류가 가능
 - 데이터는 다양한 기능·역할과 특성이 있기 때문에 데이터의 가치를 고유하게 정의하고 표현하는 것은 쉽지 않음
 - △데이터의 생성·관리·활용비용에 근거에 데이터의 가치를 산출하는 비용 기반 접근 △유사제품의 시장 가격 또는 사용자의 지불 의사에 기반한 시장 기반 접근 △데이터 활용 효과로 추정하는 임팩트 기반 접근 등을 고려 가능

| 표 4 | 데이터 가치측정 방법의 유형

접근법	사고방식	분류	과제	논점
비용 기반 접근	<ul style="list-style-type: none"> • 데이터의 생성, 관리, 활용 비용(인력, 시간 등)에 근거해 산출하는 방법 • 비용을 들이는 만큼의 (비용과 같은) 가치가 있다는 가정에 근거 	<ul style="list-style-type: none"> • 인건비 산출 • 시간 비용 산출 • 보안 대책 비용 산출 	<ul style="list-style-type: none"> • 품질과 생산성의 차이를 고려하지 못함 • ICT 기기 등의 가격과 성능에 좌우되기 때문에 데이터 가치가 다른 요소로 변화 할 가능성 존재 • 비용과 가치 관계의 직관적 시각화가 어려움 	<ul style="list-style-type: none"> • 어느 측면의 비용을 데이터의 가치로 보는 것이 적절한가?
시장 기반 접근	<ul style="list-style-type: none"> • 유사 제품의 시장 가격 또는 사용자의 지불 의사를 기반으로 가치를 계산하는 	<ul style="list-style-type: none"> • 시장 가격 (데이터 매매 등) • 표명선호법¹⁾ 	<ul style="list-style-type: none"> • 측정에 이용할 데이터가 부족 • 직접 진술의 경우 조사 	<ul style="list-style-type: none"> • 개별 데이터 없이 어떻게 데이터 전반을 측정할 것인가?

접근법	사고방식	분류	과제	논점
	방법	<ul style="list-style-type: none"> 기업 인수, 손해 배상액 등의 사례 분석 실험적 대응 	<ul style="list-style-type: none"> 대상 사용자의 이해도에 좌우 기업 관점, 소비자 관점에 따라 가치가 상이 	<ul style="list-style-type: none"> 기업 관점과 소비자 관점 중 어느 것으로 데이터의 가치를 파악하는 것이 적절한가?
임팩트 기반 접근	<ul style="list-style-type: none"> 데이터 활용에 따른 효과(생산성, 수익 등)를 추정함으로써 가치를 계산하는 방법 	<ul style="list-style-type: none"> 계량 분석(생산 함수 분석, 회귀 분석 등) 실험적 대응 	<ul style="list-style-type: none"> 통계적인 추정이 필요하며, '데이터' 효과와 다른 요소의 효과를 명확하게 분리하는 것이 어려움 	<ul style="list-style-type: none"> 데이터의 가치 측정 대상(데이터 정의, 가치 정의)의 명확화 데이터가 가치를 창출하는데 필요한 보완적 요소 등에 대해 검토한 후, 어떻게 적절한 분석 모델을 구축할 것인가?

- 데이터에서 얻은 정보와 지식은 일반적으로 시장 거래의 대상이 되지 않기 때문에 IMF에서도 다양한 계측 방법이 제안
 - 접근법에 따라 다양한 과제가 존재하고 다른 가치를 나타낼 가능성도 있기 때문에 아직까지는 확정된 방법이 존재하지 않는다고 할 수 있음

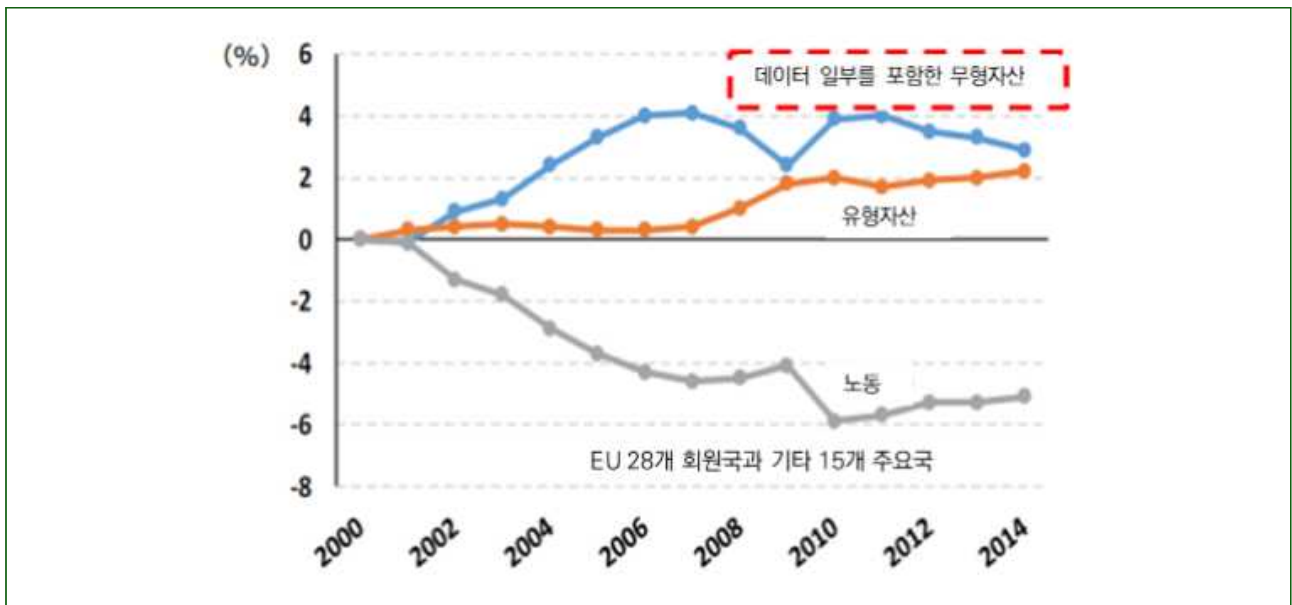
【 표 5 】 거시경제 통계에서 디지털 경제의 계측

접근법	사고방식	과제
시장 기반 접근	<ul style="list-style-type: none"> 시장에서 비교 가능한 제품의 시장 가격을 기반으로 평가 	<ul style="list-style-type: none"> 시판되고 있는 타사의 데이터베이스 외에 시장에서 판매되는 실제 비교 가능한 제품이 존재하지 않음 '시판되는 원시 데이터'와 '정리, 삭제, 사용에 적합한 가공이 이루어진 데이터'는 가치가 다르며, 시장 거래 가격이 본래의 가치와 상이할 수 있음
비용 기반 접근	<ul style="list-style-type: none"> 데이터에서 얻은 정보와 노하우를 산출한 비용으로 평가 	<ul style="list-style-type: none"> 비용 범위의 설정 <ul style="list-style-type: none"> - 데이터 과학자의 비용 고려 등 데이터 자산 전체로 확장할 필요 플랫폼 비용을 데이터 취득 비용으로 처리할 경우 플랫폼이 사용자로부터 얻는 기타 이익을 무시할 수 있음
소득(수입) 기반 접근	<ul style="list-style-type: none"> 데이터에서 얻은 미래의 현금 흐름을 추정하여 평가 	<ul style="list-style-type: none"> 데이터에서 기대되는 수익의 할인 현재가치(discounting present value)¹⁾는 용도에 따라 비용 기반 접근과는 다른 가치를 나타낼 수 있음 <ul style="list-style-type: none"> * 미래에 취득될 가치가 만약 현재 취득된다면 어느 정도의 가치를 가지는지를 나타내는 것 2008 SNA(System of National Accounts, 국민소득통계)는 자산 평가에 수익을 사용할 경우 주의가 필요하다고 밝히고 있음 <ul style="list-style-type: none"> - 자산의 내용(耐用) 년수, 미래 현금 흐름, 할인율에 대한 적절한 가정 등을 결정하기가 쉽지 않음 대부분의 경우 플랫폼 수익은 기술 및 사업 능력, 고객과의 관계, 네트워크 효과, 혁신 등 기타 유·무형의 자산에 속함

1) 표명선호법(stated preference)은 가상의 상황을 제시하고 이에 대한 개인의 의사결정을 유도함으로써 심리적으로 내재되어 있는 개인의 선호, 의식, 의향 등을 조사하는 기법

2 데이터의 효과와 가치에 상응하는 정당한 보상 방식

- ▶ 데이터의 효과와 가치를 사회 전체가 향유하기 위해서는 적절한 주체에 정당한 보상을 배분할 필요
 - [그림 4]에서와 같이 최근 자산과 노동이 소득에서 차지하는 분배율을 보면 데이터의 일부를 포함하고 있는 무형 자산의 분배율이 높아지는 것을 확인할 수 있음
 - 데이터가 산출한 부의 상당 부분이 데이터를 수집·활용하는 기업에 집중적으로 배분될 가능성이 있음
 - 향후 경제 활동에서 중요한 지위를 차지할 것으로 전망되는 데이터의 효과와 가치에 상응하는 정당한 보상 방식을 검증할 필요성 제기



| 그림 4 | 무형자산 · 유형자산 · 노동의 분배율 (누적변화폭)

* 출처: Chen, Los & Timmer(2018) “Factor Incomes in Global Value Chains: The Role of Intangibles,” NBER Working Paper 25242

- ▶ 데이터는 일반적으로 자본으로 취급되는 경우가 많지만 AI는 관련 데이터를 생성하는 개인의 적극적인 참여에 의존한다는 특성을 감안하여 데이터를 노동으로 취급하고 데이터를 생성하는 개인에게 금전 등의 방식으로 보상해야 한다는 의견도 존재
 - ‘자본으로서의 데이터 (Data as Capital)’는 기업가정신과 혁신의 촉진이라는 관점으로부터 데이터가 산출하는 가치의 기여자로 데이터 AI기업·플랫폼 사업자·데이터 과학자 등의 활용자를 중시
 - ‘노동으로서의 데이터 (Data as Labor)’는 데이터를 생성하는 개인의 공헌을 일반 노동자와 동일하게 인정하는 관점으로부터 데이터가 산출하는 가치의 기여자로 데이터 생성자를 중시

표 6 | '자본으로서의 데이터' 및 '노동으로서의 데이터'의 주요 특성

	자본으로서의 데이터	노동으로서의 데이터
데이터 산출 가치의 기여자	• 데이터 활용자를 중시	• 데이터 생성자를 중시
인센티브	• 기업가정신	• 일반적 (노동과 같은) 공헌
노동의 미래	• 보편적 기본 소득	• 데이터 노동
예상되는 우려 사항	<ul style="list-style-type: none"> • AI는 관련 데이터를 생성하는 개인의 적극적인 참여에 의존하는데, 이러한 요인이 AI의 생산성 향상에 장애요인이 될 가능성 존재 • 개인이 생성한 데이터가 특정 기업에 집중되면서 시장의 활력이 정체될 가능성 존재 	<ul style="list-style-type: none"> • 반(反)공유재의 비극 발생 가능성 존재 - 공유되어야 할 재산이 분할되고 사유화되어 사회에 도움이 되는 자원의 활용을 막음 • 금전적 대가를 제공함으로써 활동의 본래적 동기가 저하될 가능성 존재

▶ 데이터로부터 발생하는 가치에 대한 참여라는 측면에서 데이터 생성자를 중시하는 접근이 최근 다양한 조직과 활동에서 활용

- EU의 '일반데이터보호규정(General Data Protection Regulation, GDPR)'에 따라 보호 대상이 되고 있는 IoT 디바이스나 커넥티드 카 등 기계를 통해 수집된 비개인 데이터와 관련해, '17년 EU집행위원회가 제시한 '미래 데이터 접근 프레임워크 (A future EU framework for data access)'의 한 항목으로 데이터 생산자 권한(Data producer's right)이 제시

* (GDPR) '18.5월에 전면 시행된 개인 데이터 보호에 관한 EU 역내에서의 통일적인 규정이며, 데이터 이동의 권리가 포함. 이는 자신의 개인 데이터를 기계가 가독할 수 있는 형식으로 되찾을 수 있는 권리이며, 기술적으로 가능한 경우 자신의 개인데이터를 한 관리자에서 다른 관리자에게 직접 이동시킬 수 있는 권리도 포함

- 데이터 생산자 권한과 관련한 공공의견수렴 과정에서 새로운 데이터 소유권이 생겨나는 것에 대한 우려감이 나타났으며, 오히려 기업 간 공유 데이터 접근성 확보 필요성이 주장
- '미래 데이터 접근 프레임워크 (A future EU framework for data access)'에서 언급된 기업 간 데이터 공유라는 관점은 '유럽데이터전략(A European strategy for data, '20.2)'에서 명시된 '데이터법 (Data Act, '21년 제정 예정)'에 관한 구상으로 계승

* 산업용 IoT 등 공동 생성 데이터의 사용 권한에 대한 지원 및 데이터 공유에 대한 법적 책임의 명확화 등이 포함

표 7 | 미래의 데이터 접근을 위한 EU 프레임워크의 개요

미래의 데이터 접근을 위한 프레임워크(A future EU framework for data access)

- 기업의 데이터 공유 인센티브에 관한 지침 (Guidance on incentivising businesses to share data)
- 신뢰 가능한 데이터의 식별과 교환을 위한 기술 솔루션 개발 촉진 (API 등)
(Fostering the development of technical solutions for reliable identification and exchange of data)
- 기본 계약 규칙 (불공정 계약 조항 규제 포함) (Default contract rules)
- 공익과 과학적 목적을 위한 액세스 (B2G 포함) (Access for public interest and scientific purposes)
- 데이터 생산자 권한 (Data producer's right)
- 액세스에 대한 대가 (FRAND (Fair, Reasonable And Non-Discriminatory) 등) (Access against remuneration)

- 개인 데이터에 관해서는 '18년 '마이데이터 글로벌(MyData Global)'이 설립되어 개인이 자신의 데이터에 대해 충분히 이해하며 주체성과 주도권을 갖고 자신을 위해 개인 데이터를 활용할 수 있는 방안이 추진되고 있음

* (MyData Global) 유럽을 중심으로 세계 각지에 거점이 설치되어 있으며, 40여 개 국가에서 약 100명의 조직 회원과 약 400명의 개인 회원이 가입

- 개인 데이터와 관련해 힘의 균형이 개인 데이터의 수집, 유통 및 이들에 근거해 의사결정권을 갖는 조직에 극단적으로 치우쳐 있고 개인은 자신의 데이터에 일어나는 일을 제어할 수 없다는 현실 인식 아래 힘의 불균형 상태를 회복하고 개인 데이터에 대한 개인 중심의 비전을 추진

| 표 8 | '마이데이터 글로벌'의 혁신과 원칙

MyData에 의한 혁신	
1. 형식적인 권리에서 행사 가능한 권리로	• 액세스와 수정의 권리, 이동의 권리, 잊혀질 권리를 '원클릭 권리' (간단하고 효과적인 권리)로 함
2. 데이터 보호에서 데이터 활용으로	• 개인 데이터 보호뿐만 아니라, 조직이 보유하고 있는 자신에 대한 데이터 이용에 대해서도 개인에게 보호와 권한이 부여되어 상식과 관습의 변화를 기대
3. 폐쇄적 생태계에서 개방적 생태계로	• 오늘날 데이터 경제는 대량의 개인 데이터를 수집, 처리할 수 있는 제한된 플랫폼에 유리한 네트워크 효과를 보유 • 개인의 자유 의지에 따라 국제적인 데이터 구획화(enclosure)에서 개방되고, 진정 자유로운 데이터 유통을 실현하는 디지털 경제에서 균형, 공정성, 다양성, 경쟁 환경을 창조할 것을 기대
MyData의 원칙	
1. 개인 중심의 개인 데이터 제어	• 누가 데이터에 대한 액세스 권한을 갖고 어떻게 사용되며 공유되고 있는지를 이해하고, 효과적으로 관리하기 위한 실용적인 수단을 제공할 필요 • 개인 데이터를 이용하기 위한 약관과 조건을 통해 개인과 조직 간 공정한 방법으로 협상이 가능
2. 결합점으로서의 개인	• 개인 데이터의 가치는 다양성에 따라 기하급수적으로 증가하는 반면, 프라이버시에 대한 위협도 증가 • 이러한 모순은 개인이 개인 데이터의 상호 참조 '허브'가 됨으로써 해결 가능
3. 개인의 통제권	• 개인이 자신에게 맞는 방법으로 개인 데이터를 안전하게 관리할 수 있도록 하고, 개인 데이터를 유용한 정보, 지식, 자주적인 의사 결정으로 변환할 수 있는 툴, 스킬, 지원을 제공
4. 이동성 : 액세스 및 재사용	• 개인이 자신의 개인 데이터를 입수하여 자신의 목적이거나 다른 서비스에서 재사용할 수 있도록 하는 데이터 이동성은 단순히 법적 권리 차원이 아니라 실용적인 수단과 결합될 필요
5. 투명성과 설명 책임	• 개인 데이터를 사용하는 조직은 그 데이터로 무엇을 하고 있는지, 왜 그렇게 하는지를 공표하고 그 내용에 따라 다루어져야 할 필요 • 개인 데이터의 보유 및 이용의 결과에 대해 책임을 져야 하며, 개인이 조직에 책임을 묻는 것을 허용할 필요
6. 상호 운용성	• 오픈 생태계의 효과를 극대화하기 위해 데이터, 오픈 API, 프로토콜, 애플리케이션, 인프라의 상호 운용성을 항상 추구하며, 개인이 자신의 데이터에 관한 제어를 잃지 않고 모든 개인 데이터를 운반, 재사용 할 수 있도록 할 필요

- ▶ 데이터 생성자에 대한 보상 지급과 관련하여 ‘노동으로서의 데이터’ 입장은 긍정적인 반면 ‘자본으로서의 데이터’ 입장은 찬반양론이 존재
 - ‘노동으로서의 데이터’ 접근에서는 데이터 생산자에게 보상함으로써 그들의 노동 소득을 보완할 수 있다고 보고 있으나, 기여도 측정 및 측정 기술 실현에 한계가 있다는 점이 지적
 - ‘자본으로서의 데이터’ 접근에서는 데이터는 분석을 통해 가치가 발생하기 때문에 데이터 생성자가 아닌 분석가가 보상을 받아야 한다는 부정적 의견과 데이터 제공에 대한 인센티브가 있어야 한다는 긍정적인 의견이 공존

표 9 | ‘데이터를 생성한 사용자에게 보상을 지급한다’는 접근법의 논점

데이터를 생성한 사용자에게 대한 보상 지급	자본으로서의 데이터		노동으로서의 데이터
	긍정적	부정적	긍정적
보상의 지급/비지급 이유	<ul style="list-style-type: none"> AI 발전으로 데이터의 중요성이 높아짐에 따라 데이터를 제공할 인센티브가 필요 데이터가 만들어내는 가치의 대부분을 사적 이익으로 확보하고 있는 기업의 존재 	<ul style="list-style-type: none"> 무료 데이터 세트 및 기계 학습 알고리즘을 무료로 얻을 수 있음 데이터는 분석을 통해 비로소 가치를 창출하며, 데이터 생성자가 아닌 데이터 분석가가 보상을 받아야 함 금전적 대가를 제공함으로써 활동의 본래적 동기를 저하시킬 수 있음 	<ul style="list-style-type: none"> 노동으로서의 데이터는 소득을 보완하는 중요한 기회가 되며, 격차 확대로 어려움을 겪는 시민에게 사회에 기여하고 있다는 의식을 갖게 함
보상 지급 프로세스	<ul style="list-style-type: none"> 시장 메커니즘 (소액 결제 · 정보신탁 기능(정보은행)) 규제 (경쟁 정책/세제/사회보장 등) 	-	<ul style="list-style-type: none"> (한계) 기여를 측정하여 개인 사용자가 만들어내는 가치를 소급 추적하기 위한 적절한 기술 시스템을 구축
예상되는 과제	<ul style="list-style-type: none"> 시장 메커니즘 : 사업 모델의 확립 규제 : 도입의 객관적 정당성 (정량적 지표 등) 	<ul style="list-style-type: none"> 기업, 데이터 분석가 및 기타 개인의 부의 편재 (노동 분배율 저하) 가능성 	<ul style="list-style-type: none"> (한계) 기여의 측정과 기술의 실현 가능성 기여의 측정이 감시 사회와 격차·차별 조장을 가져올 가능성

- ▶ 데이터의 적극적인 활용과 그에 따른 효과와 가치를 사회 전체가 향유하기 위해서는 데이터를 생성하는 개인에게 서비스 무료 이용 등 이외에 대가를 주는 방법을 검토할 필요
 - 측정된 데이터의 가치를 가장 정확한 형태로 개인에게 분배하기 위해서는 개별 데이터에 대한 개인의 기여 정도를 측정할 필요가 있지만, 데이터에 대한 개인의 다양한 기여를 측정하기가 쉽지 않은 것이 현실

- 개인의 기여를 정확하고 안전하게 파악하는 데에는 현재 기술적으로 한계가 있으며, 기여도를 측정 시, 감시사회 심화와 격차·차별 조장 위험 등이 지적
- 데이터 활용자인 플랫폼 사업자·데이터 과학자 등과, 데이터 생성자인 개인 간에 데이터 전체 가치에 대한 기여 정도를 측정할 수 있다면, 개인에게 분배되는 데이터의 총 가치도 측정할 수 있게 되므로 그에 따라 개인에게 ‘평균적인’ 보수를 지급하는 것이 가능하나 현재는 기술적 어려움이 존재
- 데이터를 생성하는 개인의 관점에서 대가를 고려하는 방법으로는 자신의 정보를 기업에 제공할 때 최소한의 금전적 대가를 요구하는 방법이 고려
 - 개인 설문조사를 통해 개인이 요구하는 금전적 대가를 검토한 결과, 많은 데이터에서 중앙값은 500엔~1,000엔으로 분석
 - 상대적으로 민감한 지문, 신용카드 번호, 얼굴 사진에 대해서는 평균 10,000엔을 상회하여 기업 제공에 높은 장벽이 있음이 확인
 - 어떤 데이터도 평균과 중앙값 사이에 괴리가 있고 고액의 대가를 요구하는 사람이 일정 정도 있다는 점을 고려하면서 대가를 주는 시스템을 검토할 필요

【 표 10 | 기업에 정보를 제공할 때 요구되는 최소한의 ‘금전적 대가’

(단위 : 엔)

구분	평균값	중앙값
지문	10,879	5,000
신용카드 번호	10,641	1,000
얼굴 사진	10,114	3,000
여권 번호	7,905	1,000
은행 등의 계좌 정보	7,710	1,000
이동 정보	5,705	1,000
전화 번호	4,654	1,000
SNS, 블로그 포스팅 정보	4,648	500
신체 정보	4,565	1,000
주소 정보	4,378	1,000
실제 소매점이나 음식점에서의 구매 정보	3,364	500
회사 정보	3,163	500
전자 화폐, IC 카드 등의 이용 이력	3,028	1,000
이메일 주소	3,004	500
웹페이지 열람 이력	2,915	500
EC 사이트의 구매 정보	2,866	500
가족 구성, 주거 정보	2,445	500
성별, 연령, 학력	2,409	500
가입 보험 정보	2,177	500
연수입	1,627	100
출신지	1,575	100

3 데이터 거래의 시장화와 가시화를 위한 과제

가. 일본의 데이터 거래 현황

- ▶ 총무성 ‘디지털 데이터의 활용에 관한 설문 조사’에 따르면, 일본의 기업들은 데이터 거래가 활발하지 않은 것으로 분석

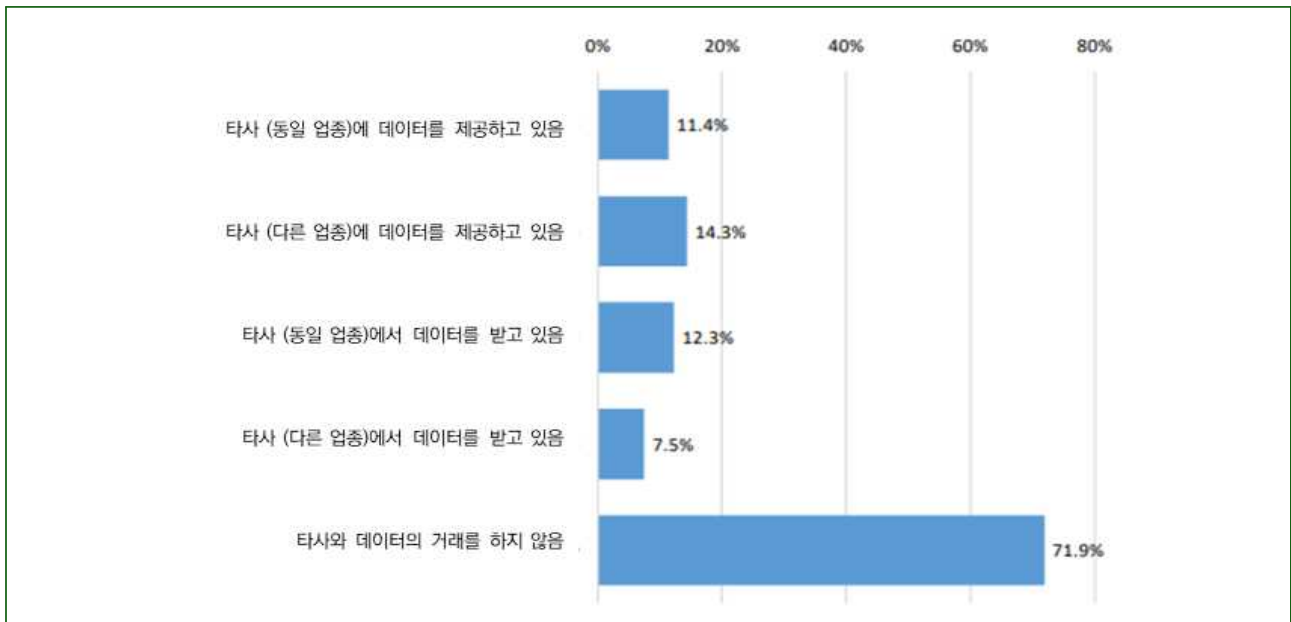
* 기업의 데이터 활용 실태를 파악하기 위한 조사

①조사대상: : 총 4,286사(전체 상장사(3,819개 社)+ 비상장사(467개 社))

②조사시기: 2020년 2월 7일(금) ~ 3월 27일(금)

③조사내용: 기업이 활용하는 데이터양(양, 다양성 등)과 분석 상황(체제, 수법, 빈도 등) 등

- 타사에 조직이나 개인에 관한 데이터를 제공·수령하지 않는 기업이 70% 이상으로, 데이터 송수신이 거의 이루어지지 않는 실정
- 이에 따라 데이터 거래의 시장화·시각화를 위한 제도 설계 시 데이터의 제공·수령을 저해하는 요인을 파악할 필요



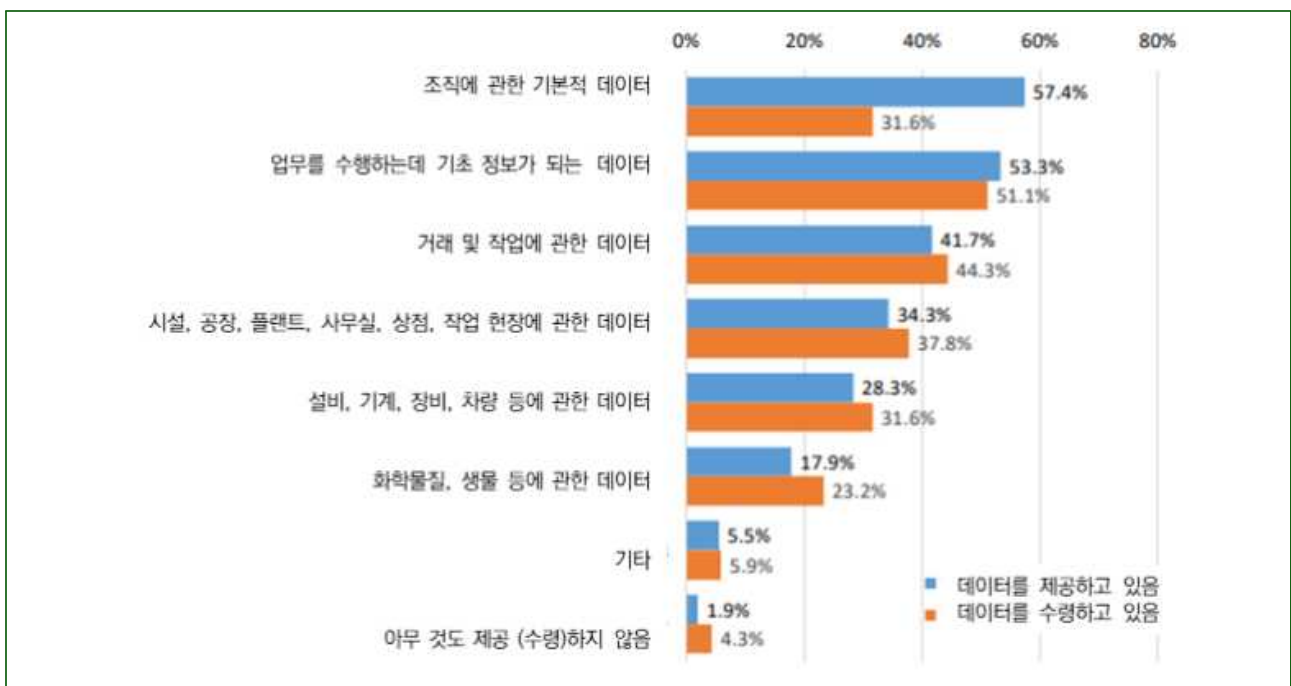
| 그림 5 | 타사 (동일 업종 또는 다른 업종)에 조직이나 개인에 관한 데이터 제공·수령 현황

- ▶ 데이터가 수집·이용되는 것에 대한 개인의 불안감도 높은 것으로 확인

- 앞의 [그림 3]에서와 같이 ‘개인 정보가 의도치 않게 유출된 것이 아닌지 불안’하다는 응답과 ‘자신의 정보가 어떻게 사용되고 있는지를 알지 못함’, ‘자신이 의도하지 않은 이용 방식을 이용한 것은 아닌지 불안’하다는 응답 비율이 모두 절반 이상으로 높은 편
- 데이터 거래의 활성화를 위해서는 개인의 불안을 불식시킬 방법을 고려할 필요

▶ 기업 조직 데이터는 데이터의 종류에 따라 제공·수령 상황이 상이

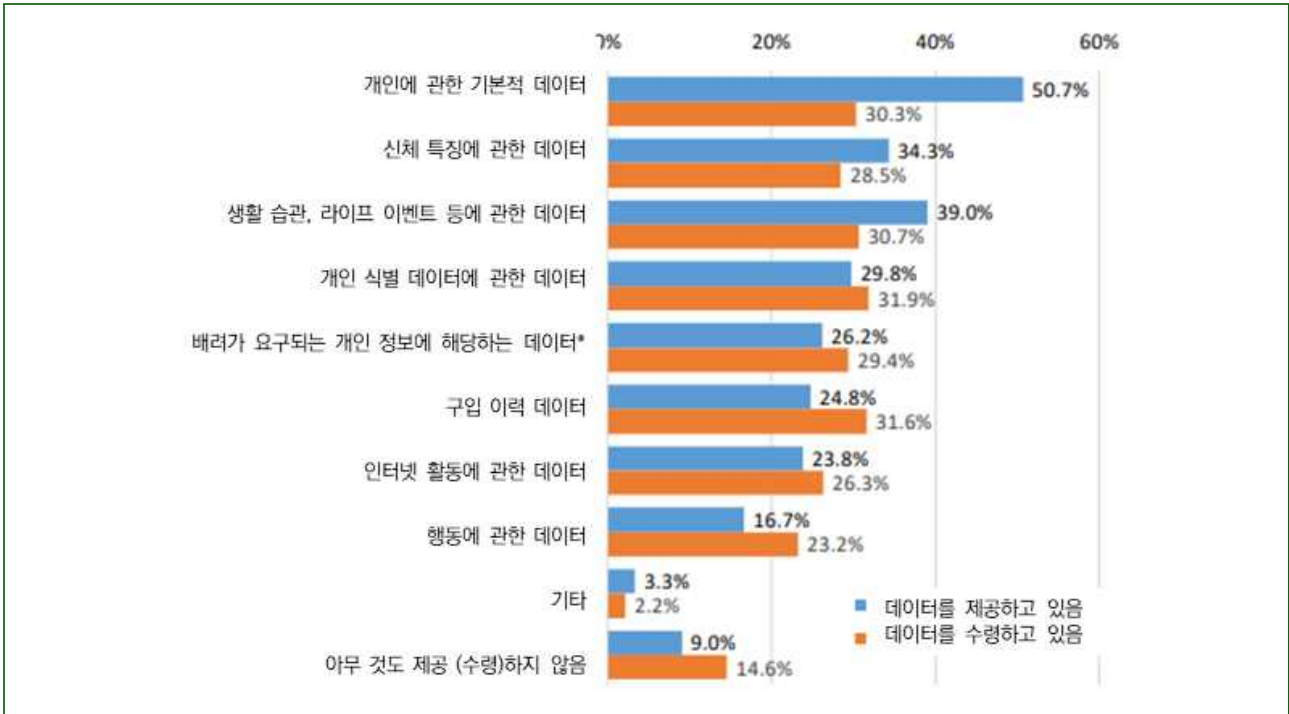
- 제공 비율이 높은 기업 조직에 관한 데이터는 ‘조직에 대한 기본적 데이터’나 ‘업무를 수행하는 데 기초 정보가 되는 데이터’이며, 수령 비율이 높은 기업 조직에 관한 데이터는 ‘업무를 수행하는데 기초 정보가 되는 데이터’와 ‘거래 및 작업에 관한 데이터’로 확인
- 반면 ‘설비, 기계, 장비, 차량 등에 관한 데이터’나 ‘화학물질, 생물 등에 관한 데이터’는 제공·수령 비율이 낮은 것으로 분석
- 데이터 거래의 활성화를 위해서는 데이터 종류에 따라 제공·수령 상황이 다른 점을 고려할 필요



| 그림 6 | 타사에 조직에 관한 데이터 제공·수령 현황

▶ 개인에 관한 데이터 이용은 제공 측과 수령 측 간 비대칭성이 존재

- 개인에 관한 데이터는 ‘개인에 관한 기본적인 데이터’와 ‘생활 습관, 라이프 이벤트 등에 관한 데이터’의 제공 비율 및 ‘개인 식별번호 데이터’나 ‘구입 이력 데이터’의 수령 비율이 높지만, ‘인터넷 활동에 대한 데이터’와 ‘행동에 관한 데이터’는 제공·수령 비율이 모두 낮음
- 데이터 거래의 활성화를 위해서는 데이터의 제공·수령 경향이 비대칭적임을 고려할 필요



| 그림 7 | 타사에 개인에 관한 데이터 제공·수령 현황

* ‘배려가 요구되는 개인정보(要配慮個人情報)’는 일본 ‘개인정보보호법’ 개정에서 ‘개인정보’와는 별도로 “인종, 신념, 사회적 신분, 병력 등 개인에 대한 부당한 차별, 편견 등의 불이익이 발생하지 않도록 그 취급에 주의를 요하는 것으로 정령(政令)이 정하는 기술(記述) 등이 포함된 개인정보로 정의(개인정보보호법 2조 3항)

나. AI에 활용하는 데이터 거래 계약 형식의 표준화의 필요성

▶ 계약내용, 가격협상 및 제품 공급자에 관한 책임

- 기존의 데이터는 경향과 특징의 가시화나 분석으로 자기 역할을 다하지만 AI에 활용하는 학습 데이터는 애초 상정한 범위를 넘어서 가치를 산출하고 형태를 바꾸어 영구적으로 존속 가치를 계속 산출할 가능성이 있으므로 AI에 활용하는 데이터의 특성을 고려한 계약 형식의 표준화가 필요
- (계약내용) AI에 활용하는 데이터의 이런 특성은 기술 특허와 매우 가까우며, 기술 특허와 같은 방식(제3자 판매 등)으로 계약 내용을 작성할 필요
- (가격협상) 라이선스가 최종 사용자인지 또는 부품업체가 다른 업체에 그것을 전매할 것인지 등 다양한 요소가 존재하며, AI의 특성상 데이터를 사용해 시행착오를 겪지 않으면 사업에 활용하기 어려우므로 이를 가격협상 시 고려할 필요
- (제품 공급자에 관한 책임) AI를 활용한 소프트웨어의 경우 책임의 분배(채무 불이행 여부, 귀책성·인과관계의 유무)와 학습된 모델이 갖는 곤란성(개발 좌절의 리스크, 학습된 모델의 품질이나 성능 문제, 외부 시스템과의 통합 문제)에서 그 책임 소재를 파악하는 것이 어렵기 때문에 제품 공급자에 관한 책임은 상류(商流)²⁾에 따라 누가 얼마나 책임이 있는지 등에 대해 계약 형식에 담을 필요

2) 상품이 생산자에서 소비자로 유통되는 과정에서의 소유권, 금전, 정보 등의 흐름

▶ 데이터의 분포, 품질, 추적성 및 작성 시점 파악

- (데이터의 분포) 데이터는 분포가 매우 중요하기 때문에 데이터 거래에서는 언어, 지역, 연령 등의 분포를 파악할 필요
 - 예를 들어, 영어와 일본어의 음성 품질은 영어 음성이 더 높는데 이는 학습 데이터로서의 영어 음성 데이터가 풍부하기 때문
 - 일본어 학습 데이터의 경우 노인이나 특정 방언의 음성 데이터가 부족하므로 음성 인식의 품질을 높이기 위해 이러한 데이터에 대한 구매 니즈가 높은 편
- (데이터의 품질) 데이터 거래에서는 해당 데이터의 품질 파악이 우선될 필요
 - 예를 들어, '자동차를 사람'으로, '칼을 펜'으로, '선로를 횡단보도'로 잘못 라벨링 한 저급한 품질의 데이터가 혼입된 사실을 모른 채 학습 데이터로 활용한다면 AI의 처리에 큰 영향을 주기 때문
- (데이터의 추적성) 데이터의 분포와 품질을 파악하기 위해서는 데이터가 어디에서 생성되고, 누가 라벨링 정보를 부여했는지 등을 확인하는 작업이 필요하기 때문에 데이터 거래에는 추적성(traceability)이 포함될 필요
- (데이터의 작성 시점) AI 데이터 활용 시 시간이 지나면서 데이터 가치의 변화가 발생할 수 있기 때문에 데이터 작성 시점의 파악이 필수적
 - 예를 들어, 5년 전에 수집한 청소년들의 대화 데이터를 활용해 대화 모델을 만들어 현재 청소년들에게 그 모델을 적용하면 좋은 결과를 산출할 수 없는데, 이는 5년 동안 청소년들의 말투와 사용하는 언어가 크게 변화했기 때문

▶ 데이터 라벨링/어노테이션 작업자의 신원 확인 및 보증

- 일본에서 데이터의 수집, 라벨 작업, 어노테이션 작업을 모두 사내에서 실시하는 AI 벤더가 매우 적어 구입 데이터에서 '자동차를 사람'으로, '칼을 펜'으로 표기하는 것처럼 악의적인 정보처리를 하는 문제가 발생할 가능성 존재
- 이에 따라 데이터 거래 계약 시 데이터의 라벨링 및 어노테이션을 담당할 작업자의 신원 확인과 보증을 규정할 필요

다. 제3자에 의한 데이터 관리의 필요성

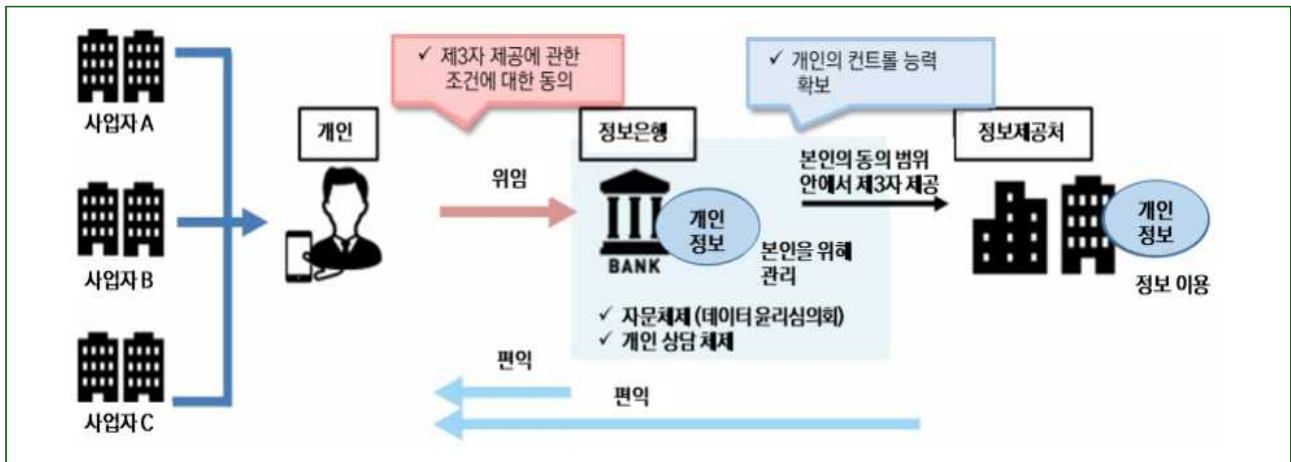
▶ 제3자에 의한 기업 보유 데이터 관리의 필요성

- 기업 보유 데이터는 해당 기업이나 제품·서비스를 통해 사회를 대상으로 악용될 위험이 항상 존재하므로 향후 데이터 경제에서 원활한 기업 활동을 위해서는 기업 보유 데이터의 관리 구조 정비가 요구
 - 대기업은 사내 법률 부서에서 권리 관계 정리, 계약 절차 등의 대응이 가능하지만 스타트업을 포함한 중소기업의 경우 법무부서 등 데이터 관리를 담당하는 체제가 충분하지 않기 때문에, 데이터를 취급하는 제반 절차 및 데이터 거래 자체에서 오류가 생겨 데이터 거래를 포기하는 경우가 빈번

- 따라서 중소기업의 데이터 제공과 이용을 원활히 하기 위해 제3자에 의한 데이터 관리가 필요

▶ 제3자에 의한 개인정보 관리의 필요성(정보은행)

- 개인정보와 관련해 실효적인 본인 관여, 즉 컨트롤 능력을 높이고 개인 데이터의 유통·활용을 촉진하기 위해서는 본인 동의 범위 내에서 본인이 신뢰하는 주체가 개인정보의 제3자 제공을 위임하는 정보은행이 필요
- 정보은행이 취급하는 데이터의 종류, 제공처 제3자가 되는 사업자의 조건, 제공처의 이용 조건 등 정보은행이 개인에게 제공하는 서비스 내용을 정보은행이 개인에게 적절히 제시해 개인이 동의하면 계약을 통해 정보은행이 해당 서비스 내용에 대한 책임을 담보
- 정보은행 사업은 이제 막 시작된 초기 단계로, 향후 데이터 거래의 시장화·가시화에 중요한 역할을 할 것으로 기대되며, 데이터를 생성하는 개인에 대해 대가와 편익을 지급하는 시장 기반 구조로서도 정보은행에 대한 높은 평가가 가능



| 그림 8 | 정보은행의 구조

PART IV

결론 및 시사점

- ▶ 현재 데이터 거래에 따른 시장 메커니즘이 충분히 작동하고 있지는 않지만, 불완전한 거래 방식을 보완하면서 미래의 이상적 시장 메커니즘을 중심으로 자원 배분을 조속히 달성하는 것이 바람직
- 각 이해 관계자에게 보상을 지급하는 합리적인 근거를 검토하고, 데이터 경제 성장을 추진하는 정책을 수립할 필요

【 표 11 】 데이터 거래와 관련된 이해관계자별 보상 개념

	자원 데이터 생성 주체 (인터넷 서비스 이용자)	데이터 분석 주체 (데이터 과학자)	데이터 보유 주체 (기업·플랫폼 등)	데이터 거래 중개인 (가공, 신용평가) (정보은행 등)
보상을 지급하는/ 지급하지 않는 이유	자원 데이터 생성에 따른 기여를 인정할 것인가	분석에 기여	데이터 수익	데이터 거래 수수료
보상 지급 방법 (프로세스)	무료 서비스, 정보 제공, 포인트, 현금 등	분석 결과에 대한 평가	사업 수익	사업 수익
보상 수준	데이터 보유 주체·데이터 거래 중개자의 의사	노동 시장	광고시장 등	당사자 거래

- ▶ 각 이해관계자에 대한 합리적 보상을 바탕으로 데이터 거래 시장 메커니즘을 최대한 활용해 데이터로 인한 경제적 이익을 극대화하기 위해 다음과 같은 구체적 방안이 요구
- (데이터 기반 구축) AI를 중심으로 한 ICT 자산이 경제 성장에 기여하기 위해서는 데이터를 고속 처리할 수 있는 데이터 기반 필요
 - 5G, Beyond 5G, FTTH (Fiber to the home, 광가입자망) 등을 비롯한 액세스 회선의 고속화뿐만 아니라 네트워크 전체의 고속성과 항구성의 확보, 정보 처리의 측면에서 CPU와 GPU 등의 고속화가 요구되며 분산 처리 및 집중 처리의 균형 등에도 유의할 필요
- (데이터 통합의 실현) 데이터 연계를 지속적으로 강화함으로써 활용 폭 확대
 - 일본 정부의 ‘디지털 사회 구축 TF’는 농업, 방재·감재, 인프라, 물류, 자동 운전 등 다양한 분야에서 진행되는 민관의 데이터 연계 방안을 조사하고, 분야를 초월한 협력이 가능하도록 필요한 데이터 형식과 기능, 데이터 취급에 관한 규칙의 공통화를 검토
- (제3자에 의한 기업 보유 데이터 관리·정보은행 추진) 기업 보유 데이터 거래는 이해 당사자의 관계가 복잡하다는 점에서 그 책임 소재를 개별 사례마다 명확히 해둘 필요
 - 제3자에 의한 데이터 거래에 관한 계약 형식의 표준화를 통해 책임 소재를 개별 사례마다 명확히 하고, 조직적인 대응이 어려운 중소기업의 데이터 거래를 촉진하기 위한 지원이 필요
 - 개인에 관한 데이터의 유통을 위해서는 데이터 거래 중개 역으로서 정보은행의 역할이 앞으로 더욱 중요하며, 정보은행이 개인 정보의 신뢰성을 담보하고 자원 데이터를 생성하는 개인에게 대가를 지급하는 방식으로 추진하는 것이 중요
- (데이터 활용 촉진을 추진하기 위한 법제도의 검토) 위와 같은 구체적 방안들을 실시하는 과정에서 현재의 법제도와 충돌하는 상황이 생길 수 있으므로 해당 법제도에 대한 끊임없는 개선이 필요

참 고 문 헌

- AI ネットワーク社会推進会議, AI 経済検討会, 報告書 2020, 2020.7



데이터산업 동향 이슈 브리프

| 발행일 2020년 9월 29일

| 발행처 **K data** 한국데이터산업진흥원

서울시 중구 세종대로 9길 42, 부영빌딩 8층

| 기획 및 편집 산업지원실 기업지원팀

| 문의처 Tel: 02-3708-5371

ISSUE BRIEF

* 본 지에 실린 내용은 한국데이터산업진흥원의 공식 의견과 다를 수 있습니다.
본 내용은 무단전재를 금하여, 가공/인용할 경우 반드시 출처를 밝혀주시기 바랍니다.