

데이터 거래산업과 정보보호

- 데이터 거래산업 주요 정책동향 및 데이터 정보보호를 위한 관련 기술-

과학기술정보통신부는 2020. 7. 한국판 디지털 뉴딜의 일환으로 '데이터 댐' 사업을 발표했습니다. 미국 대공황시기 뉴딜정책의 대표 사업인 댐 구축 사업처럼, 공공·민간 데이터를 수집하고 이를 축적·가공하여 활용할 수 있도록 하는 사업입니다. 이에 따라 빅데이터 플랫폼이 구축되고 공공 데이터가 공개되는 등 데이터 개방이 가속되었을 뿐만 아니라, 기업활동의 부산물인 데이터에도 경제적 가치가 매겨지는 이른바 '데이터 모네타이제이션(Data Monetization, 데이터 기반 수익창출)' 시대가 도래하였습니다. 본 뉴스레터를 통해 데이터 거래산업과 정부 정책 및 산업 동향에 대해 알아보고, 데이터 활용도만큼이나 중요해지는 데이터 정보보호에 대해 자세히 살펴보겠습니다.

1. 데이터 거래 산업 개요

산업 전반에서 디지털 전환이 가속화되면서, 본격적으로 데이터를 활용한 다양한 산업 및 서비스가 등장하고 있습니다. 과거에는 데이터가 기업활동의 결과물로서 재고관리 등의 목적으로 사용되었으나, 최근 들어서는 수집·가공한 데이터를 이용하여 새로운 서비스 또는 비즈니스 모델을 창출하기 위해 활용되고 있습니다. 나아가 데이터 자체가 가진 경제적 가치가 인정되면서 데이터가 시장에서 거래되는 등, 데이터가 중요한 경제적 자산으로 인정받고 있습니다. 1996년부터 매년 데이터산업 백서를 펴내는 한국데이터산업진흥원의 2023년 자료에 의하면, 미국의 데이터 시장 규모가 3,685억달러로 전세계에서 가장 크며 2020-2023년까지 4년간 연평균 17.4%의 성장률을 보이는 중입니다. 그 외에도 유럽연합, 일본, 영국 등이 10% 이상의 성장률을 기록했으며, 중국의 경우 21.5% 수준의 한층 빠른 성장률을 보이고 있습니다. 해외 시장 조사 기업인 Research and Markets의 Global Data Marketplace Platform Market 2022 보고서에 따르면 글로벌 데이터 거래 마켓 규모는 2028년 \$35 billion으로 연 24.3% 증가될 것으로 예상되고 있습니다.

우리나라는 2019년부터 디지털 뉴딜의 일환으로 구축한 빅데이터 플랫폼을 시작으로 금융, 환경, 교통 등 각종 빅데이터를 통합데이터지도에서 모아 운영중이나, 주요국에 비해서는 성장률이 상대적으로 낮은 편입니다. 그 이유로는 관련 규제의 엄격성 및 상대적으로 낮은 데이터 처리 기술 수준 등이 언급됩니다. 데이터의 비경합성(non-rivalry), 복제 가능성(replicability), 정량적 가치평가의 어려움 등 데이터 고유의 특성으로 인해 데이터의 정량적인 가치평가가 어려울 뿐만 아니라, 재산권·영업비밀보호·개인정보 해외이전 등의 법률쟁점 및 개인 사생활(privacy)과 직결될 수 있다는 우려 등으로 인해 국내에서는 시장이 활성화되지 못하고 있었습니다.

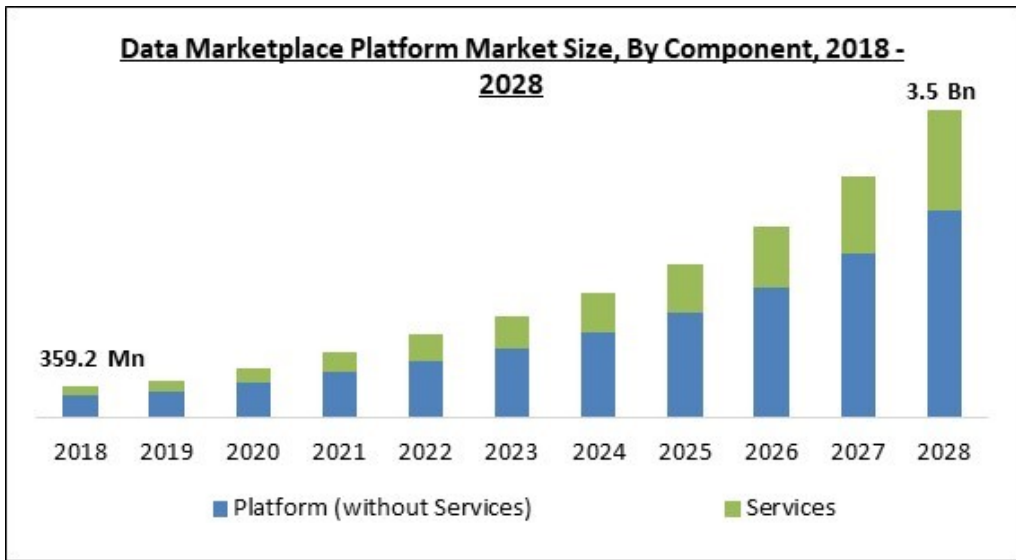


그림 1. 글로벌 데이터마켓플레이스 규모 변화 /2018~2028 (출처: www.researchandmarkets.com)

2. 데이터거래 산업 활성화 추진 동향

정부에서는 2018년 6월 발표한 '데이터산업 활성화 전략'을 시작으로, 2023년 11월 관계부처 합동으로 '데이터경제 활성화 추진 과제'를 발표하였습니다. 생성형 AI의 급속한 성장, 자율주행 기술의 고도화 등 신산업 발전이 가속화되면서 데이터 경제 활성화를 위한 정책적·제도적 기반이 필요하다는 산업계의 의견이 적극적으로 반영된 결과로 풀이됩니다. 마이데이터, 보건의료, 자율주행 등 생활에서 체감할 수 있는 분야로 세분화하여 규제와 관행을 정비하고, 기업이 겪는 불확실성을 해소하기 위한 현장 맞춤형 제도가 추진될 것으로 기대됩니다.

추진 분야	항목	내용
자율주행 /로봇	규제 샌드박스를 통한 영상정보 원본활용 허용	개인식별목적활용금지, 데이터접근 최소화, 저장 및 전송시 암호화 등의 안전조치를 전제로 하여 영상정보 원본을 자율주행 AI개발에 활용할 수 있도록 규제 샌드박스 운영
	민간 중심 원본정밀지도 데이터거래시장 조성	정부 중심 정밀지도 데이터에서 민간 주도의 원본정밀지도 데이터거래 중심으로 단계적 전환 및 플랫폼 마련
인공지능	공적데이터 개방 및 AI학습데이터시장 활성화	금융감독원이 보유한 보이스피싱 음성데이터나 개인식별위험이 제거된 의료 AI 학습용 합성데이터를 민간 기업에 제공하고, AI 학습데이터 중개 플랫폼 및 제도를 정비하여 AI 학습데이터 거래시장 활성화 기여
	사전적정성 검토제 도입 및 관련 가이드라인 마련	AI 관련 개인정보 보호법 준수방안을 민관이 함께 마련하고 행정처분을 면제받는 사전적정성 검토제 제도화 및 AI 개인정보 가이드라인 마련
바이오/헬스	의료데이터의 활용 범위 및 절차 개선	공공기관 보유 저위험 가명 의료데이터의 반출 범위 확대, 유전체비정형 의료데이터 가명처리 기준 구체화, 의료데이터 AI 학습 가명처리 기준 등의 가이드라인을 제정하고 데이터 활용에 대한 법적 책임 명확화
	의료 빅데이터 구축개방 확대 및 의료데이터 표준화 추진	건강보험 가명데이터 활용을 위한 지침 개정 데이터 결합을 통한 빅데이터 구축을 위한 데이터 표준화 추진
마이데이터	선도 프로젝트 추진 및 의료 등 분야별 특성 고려한 정책 추진	정부주체가 본인의 의사에 따라 자신의 정보를 활용하는 마이데이터를 확산하기 위해 관련 법령을 마련하고 국민 생활과 밀접한 분야로 선도 서비스를 선정하여 예산 지원
데이터경제 인프라	민간 주도 데이터 생태계 조성 위한 제도적 인프라 강화	공공-민간 데이터 플랫폼을 연계하여 통합하는 국가 데이터 인프라를 구축, 데이터 가치평가 제도 안착, 데이터 거래자문-중개 전문 인력 양성

표1. 데이터경제 활성화 추진 과제 주요내용 (출처: 관계부처 합동 데이터 경제 활성화 추진과제)

해당 추진과제가 실생활과 밀접한 분야를 중심으로 세분화되어 있으므로 국민이 체감할 수 있는 구체적 성과가 기대되는 한편, 과제 적용 대상인 의료데이터 및 건강보험 가명데이터, 가명처리된 영상정보 등은 개인정보 안전조치가 반드시 적절히 수행되어야 하는 데이터에 해당되기도 합니다. 이에 따라 데이터 정보보호 이슈 및 각종 관련 기술이 중요한 쟁점으로 떠오르고 있습니다.

3. 데이터 정보보호 이슈 및 관련 기술

가. 데이터 거래 주체에 따른 정보보호 이슈

데이터 이용자 및 운영자는 데이터 생명주기(생성, 수집, 사용 및 공유, 보관, 삭제)에 따라 정보보호 측면에서 안전하게 관리될 수 있도록 해야 합니다.

구분	이슈사항	내용
이용자 (제공자)	개인정보노출문제 (개인정보재식별)	데이터내에개인정보가포함된 경우또는데이터의특정레코드에특정인식별이가능할 유일한특징을 가진 데이터가포함된 경우개인정보 재식별 위험이 있으며, 비식별개인정보라 하여도 다양한 데이터를 결합할 때 정보주체를 식별할 가능성이 있음
	저작권 및 부정경쟁방지법 등 영업비밀문제	데이터는 민법상 물건이 아니므로 소유권의 대상이 될 수 없으나, 데이터 자체가 기업의 노하우 또는 영업비밀이거나 혹은 그러한 방법으로 생성된 데이터일 경우 안전 및 적법하게 거래되어야 함
	데이터 소유권 및 부당사용금지	개인·기업·정부 측면에서의 데이터 소유권(Data Ownership) 및 부당사용금지청구권, 개인정보자기결정권 등 데이터에 관한 권리제문제
운영자	데이터 보안 및 기밀유지	기밀 또는 민감한 데이터가 무단으로 공개되지 않도록 안전한 보관조치 필요
	데이터 무결성	데이터는 저장된 위치나 작업하는 사용자수, 또는 복사본의 수에 관계없이 정확하여야 하며 신뢰할 수 있어야 함
	데이터 가용성	승인된 사용자는 원하는 시간, 장소에서 데이터에 접근할 수 있어야 함

표2. 데이터 거래 정보보호 이슈 (TechTarget.com 참고)

앞서 언급한 내용 이외에도 데이터의 해외이전관점에서 EU-미국의 DPF(Data Privacy Framework)를 참고할 수 있습니다. 미국의 개인정보보호 관련법이 EU의 GDPR(General Data Protection Regulation)에서 요구하는 수준에 적합하지 않다고 판단하여 EU가 개별적인 인증절차를 요구하는 등의 번거로움이 지속되자, 미국 바이든 대통령과 EU 집행위원회의 협이에 따라 2023년 7월 DPF가 효력을 발생하게 되었습니다. DPF는 EU의 우려사항을 해소하고 더욱 강력한 개인정보보호 수준을 보장하기 위해 아래의 내용을 포함하고 있습니다.

정보 접근권한의 제한	미국 정보기관의 접근을 국가안보 보호에 필요하고 비례적인 범위 내로 제한할 것
2단계 법적 구제 시스템	EU 정보주체의 권리보장에 대하여 조사 및 민원이 있는 경우 미국 정보기관 내 담당 부서에서 처리하며, 불복절차가 발생한 경우 전담 법원이 독립적으로 검토할 것
정보주체의 권리 강화	미국 기업이 개인정보를 잘못된 방식 또는 불법적으로 처리하는 경우, GDPR의 기본 원칙과 유사하게 EU 정보주체가 개인정보 접근, 정정, 삭제할 권한을 부여할 것
개인정보보호 원칙	목적제한, 개인정보 최소화, 보안, 개인정보 정확성, 투명성 등 GDPR의 기본 원칙과 유사한 의무 원칙을 명시

표3. DPF 주요 내용

본 내용은 일반적인 정보의 제공만을 목적으로 발행되므로 이에 수록된 내용은 저희 법인의 공식적인 견해나 구체적 사안에 관한 법률의견이 아님을 밝혀드립니다.

나. 데이터의 프라이버시 보호 기술

간접식별정보 등 추가정보가 결합될 경우 데이터내에서 개인을 재식별 할 수 있는 확률이 높아지므로, 빅데이터 환경에서 데이터 보호 및 활용을 위해서는 '프라이버시 보호기술(PETs, Privacy Enhancing Technologies)'이 필요합니다. 프라이버시 보호기술을 활용한다면 데이터를 암호화하거나 통계적인 방법으로 추론이 어렵도록 처리하여 개인정보 및 프라이버시를 보호할 수 있습니다.

구분	개념	특징	장점	단점
프라이버시 보호 모델 (K-Anonymity)	가능한 추론형태와 프라이버시 노출에 대한 정량적인 위험성을 규정하는 방법	연결공격 (Linking Attack) 대응	직관적이고 단순	재식별 가능성 존재 (동질성, 배경지식, 쓸림, 유사성 등)
연합학습 (Federated Learning)	다수의 클라이언트와 하나의 중앙서버가 탈중앙 환경에서 협력하여 데이터 모델 학습	데이터 분석결과만 외부로 전송하므로 데이터가 직접 유출되지 않음	학습결과 취합해 더 높은 정확도의 모델 도출	모델 수립시 평가 필요
재현/합성데이터 (Synthetic Data)	통계적, 확률적 특징이 원본 데이터와 유사한 임의의 데이터	GAN 활용	무한대로 샘플 수 증가 가능	불일치로 인한 예측 정확도 감소
영지식 증명 (Zero-Knowledge Proof, ZKP)	어떤 내용(Statement)이 참이라는 것을 증명할 때, 참/거짓 여부를 제외한 어느 것도 노출되지 않는 것을 보장	임의성을 사용, 공개하는 것 외의 정보는 제공받는 사람이 알 수 없음	익명화하면서 특정 내용의 검증 가능	증명과 정의의 크기에 따른 오버헤드 증가
동형암호 (Homomorphic Encryption)	암호화된 상태에서 데이터 연산이 가능한 암호 기술	양자내성암호, 튜링완전성	데이터를 암호화한 상태로 전송 및 처리	월등히 느린 처리 속도
차분 프라이버시 (Differential Privacy)	특정인이 데이터베이스에 있든 없든 질의 응답값을 통해 공격자가 이미 알고 있는 정보보다 더 많은 정보를 얻을 수 없게 함	정보주체가 수집을 거부한 경우(opt-out)에 준하는 척도를 수치화(ε)하여, 수집 거부와 비슷한 정도로 프라이버시 보호	비식별화 기법의 한계를 극복할 수 있음	ε값이 작을수록 정보주체 보호 가능하나, 그만큼 정보의 효용 감소

표4. 프라이버시 보호기술 (이글루코퍼레이션 '데이터 활성화를 위한 보안 강화방안', 마이데이터 가이드북 등 참조)

4. 시사점

2022년 데이터산업현황 조사에 따르면 국내 데이터산업 전체 시장 규모는 22조 8,986억원이며, 이 중 데이터 판매·중개 서비스업은 그 시장규모가 2조에 육박하는 산업입니다. 그러나 과학기술정보통신부가 2023년 공개한 '2022 데이터산업현황조사'에 의하면 국내에서 데이터 거래 경험이 있는 기업은 28.7%에 불과하였습니다. 데이터 거래 경험이 없다고 응답한 기업들은 대부분 '필요한 데이터를 자체 구축'(37.2%)하거나, '내부 인력 부족'(13.2%)을 이유로 들었으며, 데이터 거래 경험이 있는 기업이라 하여도 '데이터 소재파악 및 검색의 어려움'(30.3%)이나 '데이터 유통·활용시 법률지원(16.8%)' 등이 데이터 거래의 애로사항으로 응답되었습니다.

그러나 지난 11월 발표된 '데이터 경제 활성화 추진과제'가 규제 샌드박스, 의료데이터 제공 및 데이터 수요-공급 매칭 등의 각종 정책을 발표한 만큼, 데이터 거래의 애로사항이 크게 개선될 뿐만 아니라 이를 통해 기업의 데이터 거래도 더욱 활성화될 것으로 전망됩니다. 뿐만 아니라 데이터 정보보호 기술 연구가 적극적으로 추진되고 있으며, 다양한 프라이버시 보호 모델이 공공기관 및 각종 기업의 실무에서 적용되고 있는 만큼 앞으로 국내에서도 데이터 산업이 활성화될 것으로 기대됩니다.

데이터 개방 흐름에 따라 다양한 환경에서 수집된 서로 다른 데이터를 결합하여 더 큰 부가가치를 만들어낼 것이 기대되는 만큼, 데이터 분석과 활용을 체계적으로 관리하기 위한 민간 및 공공 영역에서의 데이터 거버넌스 체계 구축이 앞으로도 중요해질 전망입니다. 이에 따라 관련 산업계에서는 데이터 거래 및 정보보호에 대한 정부 정책의 변화에 대해 지속적으로 모니터링이 필요할 것으로 보입니다.

화우 정보보호센터는 오랜 경험과 축적된 노하우를 기반으로 기업 고객을 위한 최적의 솔루션을 안내해 드리고 있습니다. 정보보호 관련 법령의 해석 및 그 대응과 정보보호 기술적 자문(해킹 진단, 보안취약점) 등 포괄적인 올인원(All-in-One) 서비스를 제공하고 있습니다. 관련하여 문의사항이 있으신 경우 언제든지 화우에 연락하여 주시기 바랍니다.

Contacts

이근우 T. (+82) 2 6003 7558
센터장/파트너변호사 E. klee@yoonyang.com

이광욱 T. (+82) 2 6003 7535
파트너변호사 E. kwlee@yoonyang.com

이수경 T. (+82) 2 6182 8132
파트너변호사 E. sgyi@yoonyang.com

이유진 T. (+82) 2 6182 8768
변호사 E. lyouj@yoonyang.com

백재환 T. (+82) 2 6182 8366
전문위원 E. jhb@yoonyang.com

지재원 T. (+82) 2 6003 7568
연구위원 E. jwji@yoonyang.com