OPINION

연구위원
노성호

시장 데이터 산업의 전망과 발전 과제*

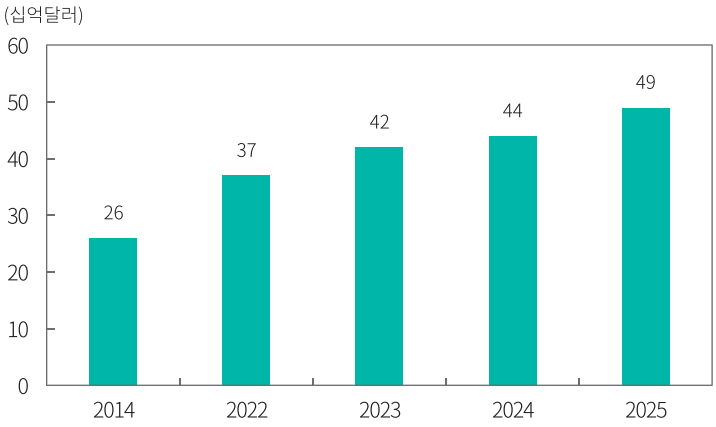
글로벌 시장 데이터 산업은 AI 기반 분석툴의 확산과 투자자 저변 확대를 배경으로 성장세를 이어가 총지출 기준 2025년 492억 달러로 사상 최대치를 기록하였다. 이와 같은 양적인 성장과 더불어 데이터의 축적이 수익 체증으로 이어지는 산업의 특성으로 인하여 성장의 수혜가 소수 대형 사업자에게 집중되고, 대체 가능성이 큰 범용 원자료와 독점적으로 공급되는 IP 간에 가치가 분화될 것으로 예상된다. 이에 대응하여 글로벌 사업자들은 가치사슬(value chain)의 핵심 단계를 통합해 경쟁 우위를 확보하고 있으며, 이는 데이터 원천과 유통 단계의 수직통합 또는 데이터 가공 계층에서의 수평적 확장의 크게 두 방향으로 이루어지고 있다. 국내 시장 데이터 산업은 정합성 높은 고품질 원천 데이터 등 가치사슬 상류의 강점을 보유한 반면, 글로벌 표준 지위를 가진 벤치마크 IP가 없는 가공·유통 계층은 상대적으로 약점이 될 수 있다. 이는 (1) 원천 데이터의 상품화와 유통 경로 개선, (2) 데이터 자산의 결합 및 제휴를 통한 규모의 경제 확보, (3) 데이터 접근성과 생산 유인이 균형을 이루는 제도 설계를 통해 시장 데이터를 AI 시대의 핵심 자산으로, 데이터 사업을 독립적인 성장 산업으로 재정의할 필요가 있음을 시사한다.

증권의 거래 과정에서 발생하는 호가, 체결, 거래량 등의 정보를 의미하는 시장 데이터(market data)는 시장의 가격 발견 기능이 함축된 산출물로서 투자 판단과 집행, 위험관리 등에 활용 가치가 매우 높다.[1] 글로벌 시장 데이터 분석 및 판매 산업의 규모는 2014년 265억 달러에서 2024년 443억 달러로 크게 성장하였으며 2025년에는 492억 달러로 사상 최대치를 기록하였다(〈그림 1〉). 같은 기간 연평균 성장률은 2014~2022년 4.4%에서 2022~2025년 9.7%로 높아졌다는 점에서 성장세는 지속될 것으로 보인다.

* 본고의 견해와 주장은 필자 개인의 것이며, 자본시장연구원의 공식적인 견해가 아님을 밝힙니다.

1) 거래 과정에서 발생하는 정보 외에도 공시, 뉴스, 신용평가 등 거래 정보와 결합되었을 때 가치가 높아지는 정보 또한 넓은 의미에서 시장 데이터로 볼 수 있다. 본고에서는 논의에 필요한 경우에 한하여 제한적으로 이들을 포함한다.

〈그림 1〉 글로벌 시장 데이터 산업 규모



주 : 연도별로 집계된 금융 데이터 산업의 총 지출 규모
자료: Burton-Taylor

시장 데이터 산업의 성장 배경

시장 데이터 산업의 성장 배경은 크게 두 가지 요인으로 나누어 볼 수 있다. 첫째로, AI에 기반한 데이터 분석툴의 고도화로 인하여 고빈도, 대용량 시장 데이터에 대한 수요가 증가한 점을 들 수 있다. 글로벌 자산운용사 131개사 중 55%가 이미 운용 프로세스에 AI를 통합하였고 91%가 향후 12개월 내 활용 확대를 계획하고 있으며[2], 공급자들도 LSEG-Microsoft 파트너십[3], S&P Global의 Kensho LLM-ready API 도입[4] 등 데이터와 AI의 접점을 선점하기 위하여 경쟁하고 있다. 이와 같은 경쟁이 시사하는 바는 데이터를 이용하는 방식이 변화하면서, 사람이 조회하는 방식이 아닌 알고리즘 · AI 시스템이 직접 소비하는 기계가독형 데이터, 즉 "AI-ready" 데이터의 중요성이 강조된다는 점이다.[5]

시장 참여자의 확대 또한 시장 데이터 산업의 또 다른 성장 요인이 되고 있다. 미국은 주식을 직접 보유한 가구의 비율과 평균 보유 금액이 장기적인 증가 추세를 보이고 있으며(〈그림 2〉), 국내에서도 주식 소유자 수가 2019년 618만 명에서 2024년 1,423만 명으로 크게 증가했다(〈그림 3〉). 참여자 저변의 확대는 시세, 지수 등의 데이터를 소비하는 계좌 · 단말 · 앱 등 데이터 이용료 과금 대상의 수를 직접 늘린다는 점에서 데이터 산업의 양적 성장 기반이 되고 있다. 더불어 거래량의 증가로 인하여 원천 데이터의 생산량 및 실시간 처리 수요를 키워 시장 데이터 생산과 배포를 위한 인프라 투자 수요 또한 확대될 수 있다.

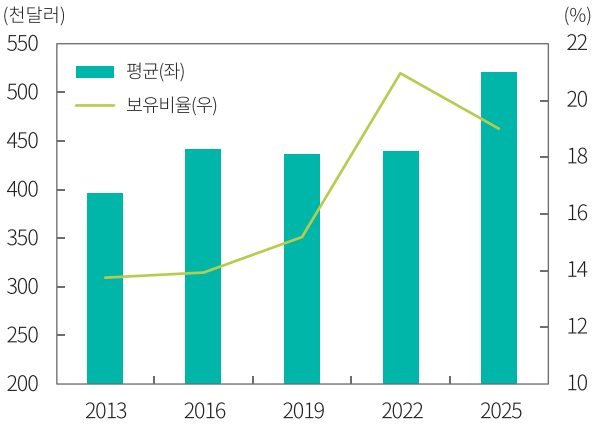
2) Mercer, 2026. 2, Moving Beyond the AI Pitch: Asset Managers' Use of AI.

3) LSEG, 2022. 12. 12, LSEG and Microsoft launch 10-year strategic partnership for next-generation data and analytics and cloud Infrastructure solutions.

4) S&P Global, 2024. 11. 13, S&P Global launches Kensho LLM-ready API (beta).

5) 일례로, 노성호, 2024, AI의 시대에 기업 재무 공시의 구조화가 가지는 의미, 자본시장연구원 『자본시장포커스』 2024-06에서는 AI를 활용한 기업 분석 수요가 증가하면서 구조화된 재무 공시 정보의 중요성이 높아지고 있음을 지적하였다.

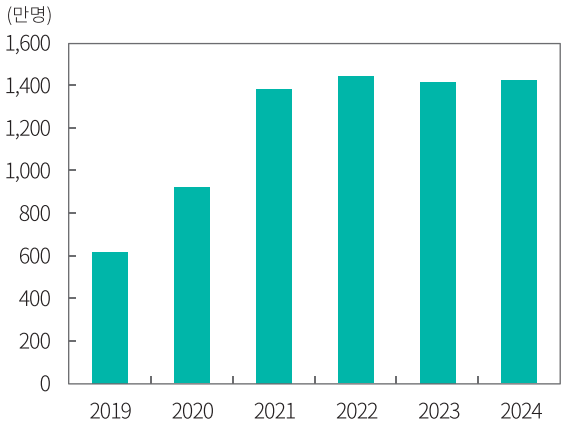
〈그림 2〉 미국 주식 보유 가구



주 : 설문 대상 가구 중 주식 계좌를 직접 소유한 가구의 비율과 평균 보유 금액

자료: Survey of Consumer Finances

〈그림 3〉 국내 주식 소유자 동향



주 : 12월 결산 상장법인 주식 기준, 2024년은 발표된 개인 소유자 수에서 추정

자료: 한국예탁결제원

데이터 산업의 경제적 특성

데이터 산업의 경제적인 특성을 앞서 설명한 성장 동인과 연결지어 볼 때 향후 경쟁 구도의 변화 양상을 전망해볼 수 있다. 우선, 주 생산물인 데이터는 비경합적이어서 복제비용이 거의 없고, 축적이 추가 수집 · 처리의 유인과 능력을 키우는 수익 체증의 성질을 지닌다.[6] 동시에 정보의 가치는 희소성에서 나오기 때문에 분석 도구가 범용화되면 기존 신호에서 발생하는 수익이 감소하게 된다.[7] 이는 최근의 시장 데이터 산업의 변화 양상을 설명할 수 있는데, 수익 체증은 AI 기반 수요의 확대와 결합해 소수 대형 사업자 중심의 성장을 유도하고, 희소성은 범용 원자료의 가격을 누르는 대신 독점 원천과 가공 IP의 가치를 높일 수 있다. 복수 시장구조, 대체 데이터 등 연관된 데이터 자원이 늘어날수록 범용 원자료의 협상력은 약해지면서 업무 흐름의 표준을 선점하는 방식으로 경쟁의 양상이 변화하는 것을 예상할 수 있다.

데이터의 다른 특징으로는 외부성과 공공재적 성격을 들 수 있다. 고객 정보의 거래 과정에서 제3자의 정보가 노출되는 경우 데이터의 가격은 사회적 최적 수준보다 낮아져 과잉 공급이 발생하고 효율적인 배분이 보장되지 않는다.[8] 또한 가격 발견의 산물인 시장 데이터는 준공공재적 성격으로 인하여 배타적 공급 유인과 접근 형평성이 충돌하게 되어 데이터 중립성 규제는 경쟁을 촉진하는 대신 데이터

6) Farboodi, M., Veldkamp, L., 2020, Long-run growth of financial data technology, *American Economic Review* 110(8), 2485-2523.

7) Grossman, S., Stiglitz, J., 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70(3), 393-408.

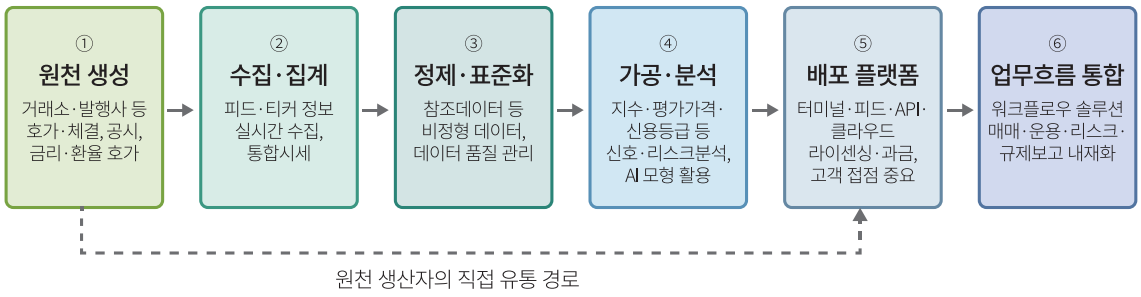
8) Acemoglu, D., Makhdoumi, A., Malekian, A., Ozdaglar, A., 2022, Too Much Data: Prices and Inefficiencies in Data Markets, *American Economic Journal: Microeconomics* 14(4), 218-256.

생산 유인을 줄일 수 있다.[9] 이와 같은 특성은 유럽 통합 시세(Consolidated Tape: CT), 미국 Market Data Infrastructure Rule 등 데이터 접근 및 가격 규제의 배경이 되고 있다.

시장 데이터 산업의 가치사슬(Value Chain) 전략 사례

시장 데이터 산업의 가치사슬은 크게 6단계로 구분하여 볼 수 있다(〈그림 4〉). 거래소, 발행사 등의 기관이 호가 · 체결 · 공시 자료를 생산하면(①) 피드(feed)와 티커(ticker) 등을 포함해서 자료를 실시간으로 수집한다(②). 다음 단계로 이와 같이 수집된 데이터는 오류를 정정하고 참조 데이터를 관리하는 정제 · 표준화 단계(③)를 거쳐 지수, 평가가격, 신용 등급 등으로 가공되거나 신호 · 리스크 분석(④)의 원재료가 된다. 이렇게 사용 가능한 형태로 가공된 데이터는 플랫폼을 통해 사용자에게 배포된다(⑤). 나아가 사용자가 매매, 운용, 규제 보고 등 실무에 데이터를 통합하는(⑥) 과정에서 추가적인 부가가치가 발생할 수 있다. 최근 클라우드 · API의 확산으로 원천 데이터 생산자는 중간 단계를 거치지 않고 직접 유통하는 경로(①→⑤)도 형성되고 있다. 가치사슬의 각 단계에서 발생하는 부가가치는 해당 기능이 확보한 데이터의 희소성 · 품질과 더불어 서비스 제공자의 표준 지위, 고객 전환비용에 의해 결정된다.

〈그림 4〉 시장 데이터 가치사슬 구조



자료: 저자 작성

단계별 매출을 보여주는 공개 통계는 제한적이지만, 글로벌 데이터 서비스 매출(2022년 약 450억 달러)의 구성은 원천 생산자(거래소 등) 약 17%(74억 달러) 대 벤더 83%(373억 달러)로 추정되며, 하류의 유통 · 플랫폼 시장은 Bloomberg와 LSEG이 약 50%를 점유하는 과점 구조를 보인다.[10]

9) Fang, H., Kim, S.J., 2026, Data neutrality, data supply, and market competition, *NBER Working Paper* No. 35159.

10) DataBP & Quinlan, 2023, *Reclaiming the Market Data Value Chain: Leveraging the Cloud to Optimise Distribution and Product Innovation.*

주요 시장 데이터 벤더는 가치사슬의 핵심 요인을 통합하는 성장 전략을 취하고 있는데 이는 크게 두 방향으로 나타나고 있다. 수직통합은 상 · 하류의 인접 기능을 결합해 거래 생애주기를 장악하는 전략이고, 수평통합은 동일 계층에서 상품 · 도메인의 폭을 넓혀 규모 · 범위의 경제를 노리는 전략이다. 각 방향의 대표 사례와 쟁점은 〈표 1〉과 같다.

〈표 1〉 데이터 벤더의 가치사슬 통합 사례 비교

구분	수직통합	수평통합
검토 초점	거래소와 데이터 · 유통 플랫폼의 종적 결합	가공(벤치마크 · 데이터) 계층의 횡적 확장
대표 거래	LSEG의 Refinitiv 인수 (2021년, 약 270억 달러)	S&P Global과 IHS Markit 합병(2022년, 약 440억 달러)
기대 효과	거래 생애주기 장악, 구독 수익 전환, 유통망 내부화	범위의 경제, 교차판매, 벤치마크 IP 포트폴리오 강화
제약 사항	통합 비용, 플랫폼 의존도 및 전환비용	데이터 정합성 확보 비용, 결합 데이터 수요 예측
정책적 고려 사항	상류 독점력의 하류 전이, 데이터 중립성	벤치마크 · 평가데이터 시장의 집중도

자료: LSEG, S&P Global의 공시 자료를 참고하여 저자 작성

LSEG은 거래소가 데이터의 생산과 유통 기능을 결합한 사례이다. 2021년 1월 Refinitiv를 약 270억 달러 전액 주식교환으로 인수[11]하면서 원천 데이터 생산자가 글로벌 유통망과 지수 및 분석 기능을 함께 확보하여 가치사슬의 전반을 통합하였다. 인수의 명시적 목적은 거래량 연동 수익에서 구독 수익으로의 전환이었으며, 이후 데이터 관련 부문이 그룹 매출의 약 61%를 차지하고 있다.[12]

S&P Global은 데이터 가공 단계에서 포함하는 상품의 범위를 횡적으로 확장한 사례이다. 신용등급, 지수, 원자재 가격평가 등 보유한 벤치마크 자산에 더하여 2022년 IHS Markit 합병으로 에너지 · 공급망 · 사모시장 데이터를 더했고, 이후 선별적 인수와 매각, 분사로 포트폴리오를 최적화하였다. 반복형 수익 비중이 79.5%를 차지하는 구조[13]와 지수 부문의 AUM 연동 수익화는 벤치마크 IP 결합 전략의 성과로 평가할 수 있다.

11) LSEG, 2021. 1. 29, LSEG all-share acquisition of Refinitiv transaction - completion and admission to trading, Press Release.

12) LSEG 2025년 잠정 실적(Preliminary Results) 기준(2026. 2. 26. 발표)

13) S&P Global 2025년 연차보고서(Annual Report) 기준

국내 시장 데이터 사업의 발전 방향

앞선 사례들은 성장하는 시장 데이터 산업에서 대체 불가능한 데이터 자산 → 플랫폼 · 워크플로우 결합 → 전환비용과 구독 수익 → 재투자로 이어지는 선순환 구조의 구축이 중요한 성공 요인임을 시사하고 있다. 이를 국내 시장 환경에 적용하면 발전 과제를 크게 세 가지로 제시할 수 있다.

첫째, 원천 데이터의 상품화와 유통 경로의 개선이다. 모든 데이터를 단순히 'AI-ready'로 전환하기보다 권리와 수요가 분명한 상품부터 우선순위를 확립할 필요가 있다. 신뢰할 수 있는 호가 · 체결 데이터, 복수 시장 통합 시세 등 정제된 데이터부터 단순 조회 또는 AI 학습 등 목적별로 이용권을 명확히 규정한 라이선스 체계를 도입할 수 있다. 유통 단계에서는 클라우드 · API 직접 판매가 확산되고 있으나 반드시 비용 효율적인 방식은 아닐 수 있으므로 판매 · 지원 · 재배포 관리 비용과 기존 벤더와의 제휴에서 발생하는 성과를 비교하여 고객군별로 유통 경로를 다변화하는 것이 현실적이다.

둘째, 데이터 자산의 결합 및 제휴를 통한 규모의 경제 확보이다. 주요 시장 인프라 기관에 분산된 데이터의 결합은 개별 기관이 제공하기 어려운 분석과 업무 지원을 가능하게 한다. 다만, 분산된 구조가 반드시 비효율성을 의미하지는 않으며 결합을 통해 얻을 수 있는 (사용자의) 편익이 표준화, 운영, 권리관계 조정 등의 비용을 초과하는지를 살펴볼 필요가 있다. 해외 사례는 조직의 대형화보다 선순환을 가능하게 하는 데이터 생산-가공-배포 기능의 연결이 안정적인 수익을 창출하는 핵심 요인임을 시사한다.

셋째, 데이터 접근성과 생산 유인의 균형을 갖춘 제도 설계이다. 시장 데이터 산업에 대한 규제의 강도를 정할 때는 접근 형평성과 데이터 생산 및 정제 인프라에 대한 투자 유인의 상충을 고려할 필요가 있다. 나아가 개인정보가 포함된 데이터의 경우 정보 노출에 의한 외부성을 고려하여 데이터에 대한 권리, 프라이버시 관리 체계를 별도로 구축해야 한다. 이와 같은 규제 정책의 성과는 데이터 산업의 양적 성장과는 별개로 지속적인 평가와 관리가 필요한 영역일 것이다.

국내 시장 데이터 산업은 증권의 유형에 따라 한국거래소, 넥스트레이드(이상 주식), 금융투자협회(펀드 및 채권)의 원천 생산, 한국예탁결제원 · 코스콤 등 인프라 기관의 처리와 분배, 민간 벤더 · 평가사의 가공 · 유통의 계층 구조로 이루어져 있다. 정합성 높은 시장 데이터와 고빈도 데이터, 표준화된 전자공시(예: DART) 등 희소성 있는 고품질 원천 자산과 더불어 데이터 매출 증가 추세, 높은 해외 고객 비중은 긍정적이지만, 가공 계층에는 글로벌 표준 지위의 지수 · 벤치마크 IP가 없고 유통 · 플랫폼 계층에서는 글로벌 벤더가 고객 접점을 넓히고 있는 위협 요인 또한 공존하고 있다. 시장 데이터를 거래의 부산물이 아닌 AI의 시대에 핵심 자산으로, 데이터 사업을 부수 기능이 아닌 독립적 성장 산업으로 재정의하여 국내 시장 데이터 가치사슬 상류의 강점을 중 · 하류의 부가가치로 연결할 필요가 있다.