

통계서비스의 혁신적 전환을 위한 「통계정보플랫폼 및 원포털 구축」

Column



통계서비스정책관 정구현 / 통계학 박사

디지털시대인 오늘날 데이터는 석유나 댐에 비유되곤 한다. 이는 데이터를 석유나 댐처럼 산업 시대의 산물로서 산업혁명을 이끄는 주요 자원으로 보기 때문이다. 하지만 데이터는 사실상 인류 역사와 맥을 같이해 왔다고 해도 과언이 아니다. 구약성경 '민수기'에 나오는 인구조사를 비롯해 '신라 장적'에서 찾아볼 수 있는 호구, 전답, 가축 등에 관한 상세 자료, 일본에 국권을 빼앗긴 이후 1925년에 최초로 실시된 인구센서스 '간이국세조사' 등 오래전부터 데이터가 인간의 의사결정에 기초 자료로 활용되어 왔음을 보여 주는 기록이 많다. 20세기 이후 컴퓨터의 발달과 함께 데이터를 활용하려는 현상이 더욱 가속화되었다. 매일 금융·유통·마케팅 등에 관한 데이터가 대량으로 컴퓨터를 통해 축적되고, 산업 전반에서 이를 활용하며 마케팅이 활발하게 이루어지고 있다. 기존에는 생산량, 취업자, 소득 등 정형화된 데이터 위주로 취급했다면, 21세기에는 기술의 발달로 음성, 이미지, 영상, 텍스트 등 비정형화된 데이터까지 폭넓게 수집하고 있다.

그럼 왜 데이터인가? 우리는 경제위기, 기후변화, 신종질병 등 예상치 못한 거대한 사건뿐 아니라, 일상생활에서조차도 불확실성이 지배하는 세상에 살고 있다. 따라서 우리는 이러한 불확실성을 이해하고 적절하게 대비해야 하는데, 데이터는 불확실한 사건에 대해 가장 합리적인 결정을 내리는 데 도움을 준다. 예를 들어 오늘 비가 올 확률이 얼마인지 알고 싶을 때 각종 데이터를 고려한다면 예측의 정확성을 높일 수 있다. 오늘 비가 올 확률은 얼마일까? 평소에 비가 올 확률이 0.3이라면 오늘도 같을까? 월별 강수량을 참작하면 비가 올 확률이 어떻게 될까? 경험적 자료의 정보에 따라 실제 발생할 확률에 가까운 수치를 얻을 수 있다. 베이지안(Bayesian) 이론에 따르면 데이터로부터 얻은 사전정보는 결과에 대한 정확한 확률을 제공하는 데 중요한 역할을 한다. 특히 불확실한 세상에서 데이터는 정부가 정책을 수립하고, 기업과 개인이 의사를 결정하거나 문제를 해결하는 데 핵심 자원이 된다.

하지만 데이터의 변화와 양적 증가에도 불구하고 공공·민간기관의 경우 데이터가 여기저기 흩어져 있을 뿐만 아니라 공개 방식과 입수 여부도 제각각이고 제한적이어서, 기업이나 일반 국민이 이용하는 데 아직도 어려움이 많다. 또한 방대한 데이터를 결합했을 때 발생할 수 있는 민감한 정보 노출 위험과 비표준화된 구조, 메타데이터의 부재는 데이터 활성화에 큰 장벽이 되고 있다.

공공데이터의 공유는 2017년에 2만 5,000건에서 2021년에 5만 9,000건으로 크게 늘어났다. 따라서 공공데이터를 대중이 손쉽게 활용할 수 있게 하려면 데이터에 대한 복합적 연계지원, 개인 정보 보호, 자료 간 표준화 등을 갖춘 '허브 플랫폼'이 필요하다.

통계청은 통계 데이터의 대중적 활용과 통계정보 서비스 환경의 현대화, 지능화, 자동화를 위해 「통계정보플랫폼 및 원포털 구축」 사업을 준비하고 있다.

통계청은 현재 통계작성기관 400여 곳이 작성하는 통계 1,300여 종의 생산·관리·서비스를 담당하며, 전 국민과 모든 사업체에 대한 모집단인 통계등록부를 표준화하여 관리하고 있다. 통계청은 통계법에 따라 민감한 식별정보를 취급할 수 있을 뿐만 아니라 개인과 사업체에 대한 고유식별번호를 보유하여 다양한 분야의 연계·융합이 가능하다. 연계·융합으로 생산한 통계의 대표적인 사례로 신혼부부통계와 육아휴직통계가 있으며, 이 통계는 통계등록부와 타 부처 행정자료의 상호연계로 작성되었다. 또한 통계청은 공공·민간데이터를 연계한 작성 기법을 보유하고 있으며, 특히 데이터 연계와 집적을 통계적 목적으로만 활용하는 중립성을 확보하고 있어 빅 브러더 문제에 대한 우려가 가장 적은 기관이라고 할 수 있다.

사업의 대상 중 하나인 「통계데이터 허브플랫폼」 서비스는 각 기관으로 흩어진 데이터를 연계하고, 각 기관에서 보유한 데이터를 조회할 수 있는 '데이터 카탈로그' 기능과 암호 기술을 활용하여 민감 데이터를 보호하고 분석할 수 있는 '민감데이터 안전구역' 등을 갖춘 통계정보플랫폼 서비스이다.

「원포털」 서비스는 국가통계포털(KOSIS), 통계지리정보서비스(SGIS) 등 별도로 제공되던 통계 서비스포털을 하나로 묶어 원사이트를 구축하는 서비스이다. 이를 통해 통계의 생산·분석·서비스를 아우르는 원스톱 통계정보 서비스를 제공할 것이다. 특히 이 서비스는 데이터 활용이 생소한 일반 국민도 손쉽게 목적에 맞는 통계를 탐색·분석·활용할 수 있도록 지원하는 지능화된 통계활용 환경을 제공한다. 예를 들어 이용자는 자연어 통계문답, 연구에 활용 가능한 통계 데이터 추천, 챗GPT와 같은 초거대 AI를 활용해 2차 가공 및 분석한 통계정보를 실시간으로 받을 수 있게 된다.

동 사업은 대국민 통계정보서비스를 한 곳에서 찾을 수 있다는 이점과 더불어 선제적 통계 AI 구현을 통해 어려운 통계 데이터에 대한 상세한 설명을 제공하고 항목별 비교가능성을 높이는 등 통계서비스의 패러다임을 바꿀 것이다. 또한 각 부처의 데이터를 허브플랫폼 내에서 분석하여 연금, 저출산·고령화 문제와 같은 정부의 정책에 대해 과학적인 결정을 내릴 수 있도록 지원할 것이다. 결론적으로 동 사업은 통계서비스의 민간 개방과 활용을 확대함으로써 통계산업의 활성화를 촉진하며, 데이터 가치를 제고하여 디지털 경제 전환에 기여할 것으로 기대된다.