균형 있는 데이터 경제와 사회를 위한 과제





함유근
건국대 교수, 한국빅데이타학회 회장
ykhahm@konkuk.ac.kr

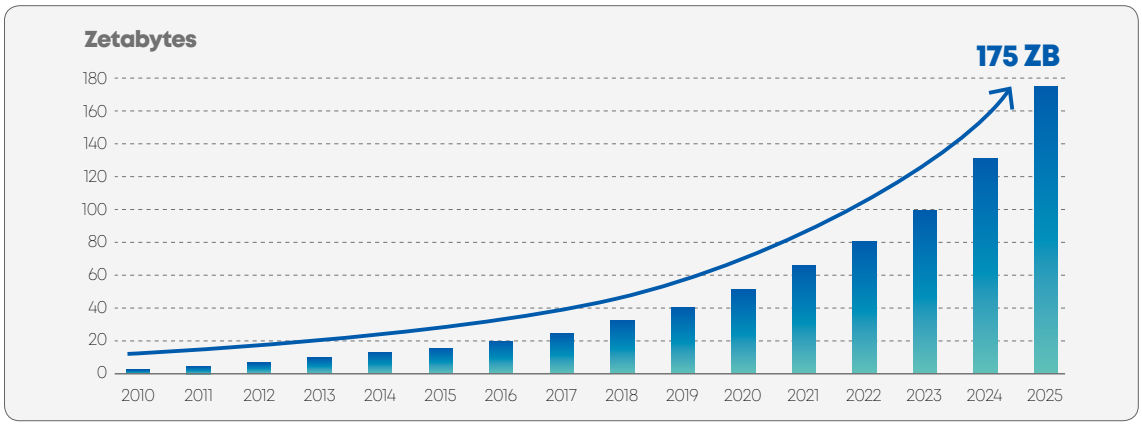
서론

2021년은 혼돈의 시대이다. 예상하지 못했던 혼란으로 기업은 물론 개인과 사회 전반에 과거와는 다른 미래를 맞이하는 불안감이 흐른다. 많은 사람이 뉴노멀을 말하지만 무엇이 그 실체인지는 불투명하다. 그러나 앞으로 우리 삶이 과거와 다름은 분명해 보인다. 그러면서 사람들은 코로나19 전과 후에 대해 얘기할 것이다. 그리고 그런 변화는 좋든 싫든 디지털화와 데이터로 초래되는 경제와 사회의 모습일 가능성이 크다.

IT 시장 조사기관인 IDC에 따르면 2025년경에는 전 세계적으로 1,500억 개의 기기들이 상호 연결된다고 한다. 데이터 규모도 2017년 23 제타바이트(기가바이트의 1조 배)에서 175 제타바이트로 급증할 전망이다([그림 1]). 이런 데이터의 상당 부분이 실시간 데이터이다. 이는 차량 등 자율적으로 작동하는 각종 기계와 사물 인터넷이 제조업 현장은 물론 일상생활 속에서 확산되고 있기 때문이다.

세상이 디지털화되면서 우리는 인류 역사에 유례가 없을 많은 양의 데이터를 빠르게 쏟아내고 있다. 그만큼 데

[그림 1] 전 세계 데이터 규모 추이

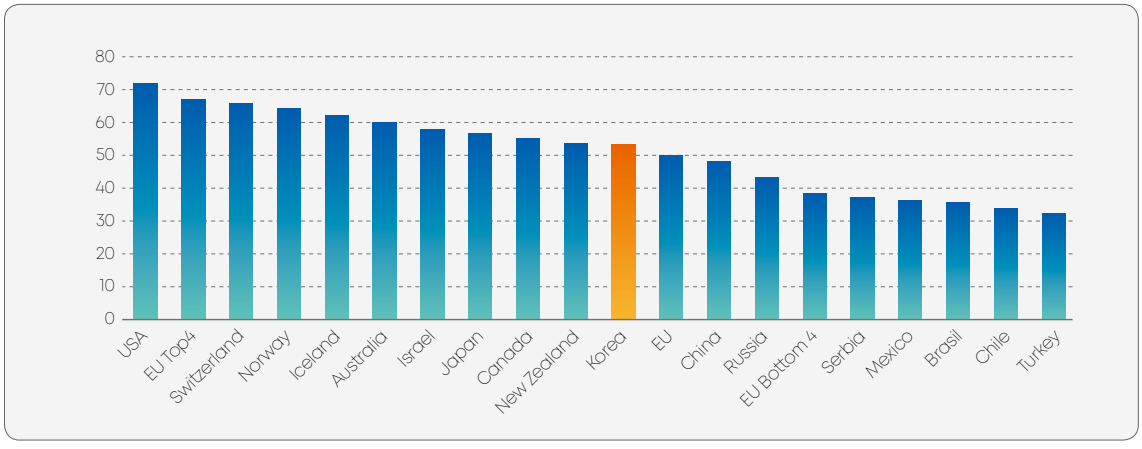


출처: IDC, 2018

이터의 활용 가능성은 커졌지만 이를 사용하는 과정에서의 윤리나 프라이버시 등의 부작용도 커지고 있다. 이에 우리가 해결해야 할 과제는 경제와 사회를 위해 데이터가 더욱 유용하게 사용되도록 하는 것과 동시에 그 과정에서 수반되는 부작용이나 위험을 최소화하는 것이다.

유럽연합이 2020년 발표한 International Digital Economy and Society Index (DESI) 순위에서 우리나라는 이웃 일본보다도 낮은 14위를 차지했다([그림 2]에서 'EU Top 4'는 핀란드, 스웨덴, 덴마크, 네덜란드). 이는 우리의 데이터 경제 및 사회 수준을 간접적으로 보여준다. 네트워크 연결 정도, 인적 자본, 인터넷 사용, 디지털 기술의 통합, 디지털 공공 서비스 등 5가지 기준으로 순위를 평가했는데 우리나라는 생각보다 순위가 높지 않다. 그 배경은 인적 자본과 기업들의 디지털화를 나타내는 디지털 기술의 통합 측면에서 상대적으로 낮은 평가를 받았기 때문이다. 특히 디지털 기술의 통합은 개인정

[그림 2] 글로벌 디지털 경제와 사회 인덱스 순위



출처: European Commission, 2020

보보호 등 법제도와 관련이 높아서 여전히 관련 규제가 심한 우리나라 상황을 반영한 것으로 보인다. 우리나라는 상대적으로 데이터의 공개(Open Data) 수준은 높으나 이를 활용하기 위한 장애가 여전히 많다는 얘기다. 아직 우리나라는 데이터를 하나의 자원으로 활용하고 있지는 못하다.

데이터 활용의 새로운 패러다임

데이터를 자원으로써 활용하기 위해서는 데이터가 가진 차별적인 특성을 고려할 필요가 있다.

첫째, 데이터는 무한정 공유될 수 있다. 달리 표현하면 쉽게 복제되지 않아 한 곳에 사용하면 다른 곳에는 사용하지 못하는 다른 자원들과 달리 데이터는 동시에 여러 곳에 사용이 가능한 비경쟁적(non-rivalrous) 자원이다. 기계와 같은 다른 자산은 여러 명(기관)이 공유할 때 그 가치가 감소하지만, 데이터는 여러 소프트웨어나 알고리즘이 동시에 사용할 수 있다. 전통적 경제학에서 말하는 자원의 한계 효용(marginal utility) 법칙이 적용되지 않는다.

둘째, 데이터는 사용하는 것에 비례해 가치가 증가한다. 대부분 자원(예를 들어 석유)은 사용할수록 가치가 감소하지만, 데이터는 오히려 증가한다. 소위 한 단위를 더 생산하는 데 드는 비용인 한계 비용(marginal cost)이 거의 제로에 가깝다. 그리고 데이터는 사용한 경험을 통해 가치를 부여하는 경험재라는 특성도 있다[1]. 즉 사용해 봐야 그 가치를 정확히 알 수 있다.

셋째, 데이터는 부패할 수 있다(즉, 시간에 따라 가치가 감소한다). 다른 자산들과 마찬가지로 시간 경과에 따라 데이터의 가치도 감소할 수 있는데 그 감소 속도는 데이터의 종류에 따라 다르다. 전자상거래를 위한 현재 제품의 재고 등 운영이나 서비스 제공을 위한 데이터의 가치는 시간 경과에 따라 급속히 감소한다. 하지만 같은 데이터가 의사결정이나 예측에 활용될 경우는 그 가치는 계속 유지되거나 증가할 수 있다. 대표적으로 환자의 과거 진료 및 증상 기록은 신약을 개발하는 제약업체나 진료를 하는 의사에게는 시간이 갈수록 가치가 높아질 수 있다.

넷째, 데이터는 정확성이 높아지면 가치도 증가한다. 오류가 있는 데이터의 가치는 당연히 매우 낮다. 정확성의 중요도는 데이터 종류와 데이터가 사용되는 방법/상황에 따라 다르다. 예를 들어 거래, 신용평가, 의료진단을 위해서는 100% 정확한 데이터가 필요하나, 고객군별 마케팅을 위해서는 개별 고객 선호도가 100% 정확할 필요는 없다.

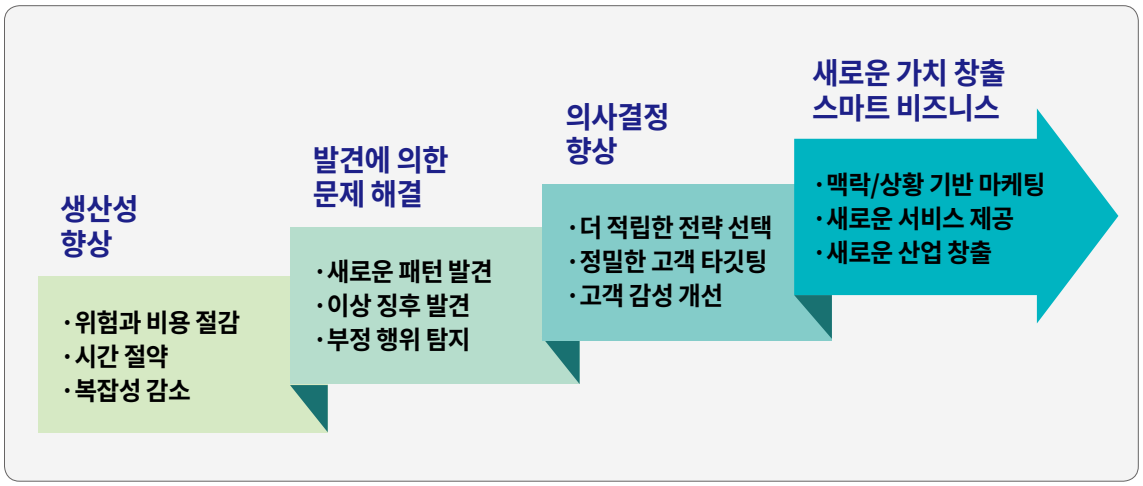
다섯째, 데이터는 다른 데이터와 결합할수록 가치가 커진다. 마케팅에서는 성별, 연령과 같은 고객의 인구 통계학적 데이터와 평소 자주 찾는 장소 등 행태 데이터를 결합하면 고객의 구매 선호도 및 가능성에 대한 보다 정확한 예측이 가능하다. 그러나 데이터 통합은 그 범위와 정도에 따라 예상보다 많은 작업과 인프라 투자가 필요하여 관련 비용이 수반된다. 다른 측면에서 보면 데이터가 얼마나 표준화됐느냐에 따라 가치가 높아질 수 있다.

여섯째, 데이터가 많은 것이 반드시 좋은 것은 아니다. 데이터가 많으면 필요한 문제의 모든 답을 얻을 수 있어 좋지만, 너무 많은 경우 사용되지도 않을뿐더러 (특히 의사결정에 활용되는 경우) 판단에 장애(information overload)가 되어 가치가 감소한다.

일곱째, 데이터는 많이 사용해도 소모되지 않는다. 즉 데이터 가치가 영원할 수도 있다. 한편으로는 시간 경과에 따른 데이터 가치의 감소와는 반대되는 특성이나, 여기서는 존재의 무한함을 강조한 특성이다. 대부분 자원은 소모되지만, 데이터는 그렇지 않다. 심지어 최근 인공지능 알고리즘은 불충분한 데이터로부터 완전한 데이터를 만들어 내기도 한다. 반면에 개인정보보호나 사라지는 이

1 https://www.technologyreview.com/s/601081/the-rise-of-data-capital/

[그림 3] 데이터를 통한 기업의 혁신 단계



출처: 함유근과 채승병, 2012

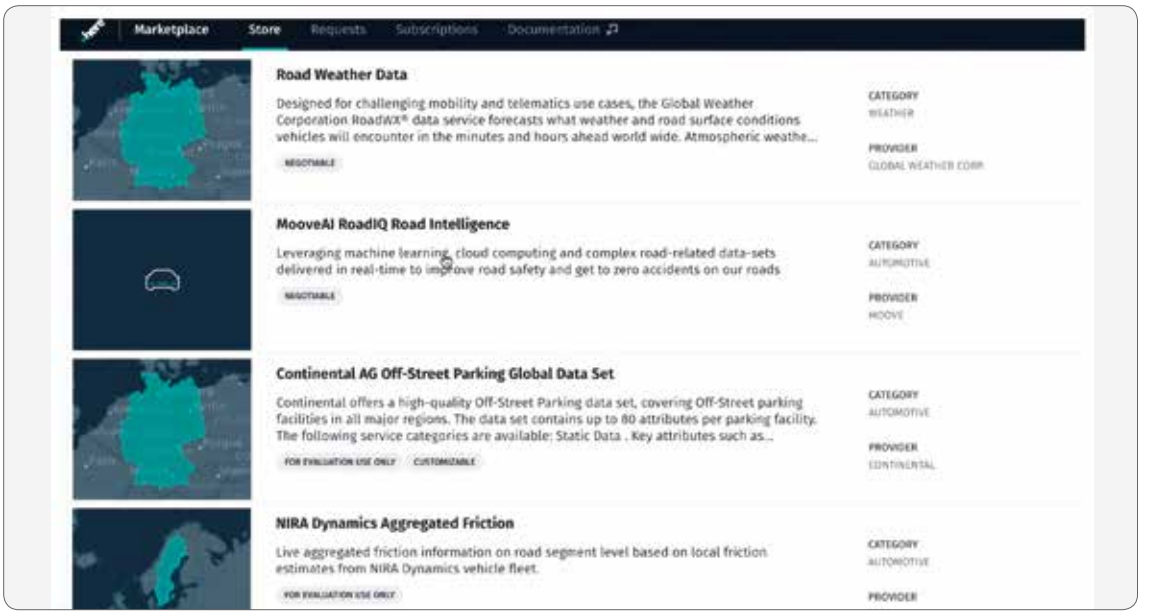
미지로 소통하는 소셜 미디어인 스냅챗(SnapChat)의 경우처럼 데이터가(의도적으로) 소멸되는 경우(국내 금융기관이나 통신사는 수집 후 1개월 후)도 있다.

데이터 경제에서 새로운 자원으로서 데이터는 조직의 다양한 혁신을 초래한다. 이제 기업은 물론이고 어떤 조직이든 크건 작건, 국내건 해외건, 미래의 운명이 데이터의 효과적인 활용에 달려 있다. 특히 조직의 생산성을 높이고, 의사결정을 하며, 새로운 비즈니스 모델을 만들어 내는 것이 가능하다([그림 3]). 데이터 활용은 국내 데이터 3법 개정에서도 나타난 개인 정보의 활용과 데이터 활용 범위의 확대에 큰 영향을 받는다. 그래서 이를 위한 데이터 공유 및 유통 촉진 등 2가지 이슈가 성공적인 데이터 경제를 위한 관건이다. 기업들은 데이터를 공유한다면 가치사슬상의 완전한 투명성 확보나 제품생산 과정의 높은 효율성 및 유연성을 달성할 수 있다. 그리고 아마존이나 중국 알리바바의 사례에서 보듯이 이제는 혁신적인 상품이나 새로운 비즈니스 모델도 데이터 공유 없이는 불가능하다.

기존에는 데이터를 단지 수익원이나 위험요소라고 분리해서 생각했다. 이제는 이 둘을 동시에 고려하는 접근법이 필요하다. 결국 기업의 상업적 활용과 개인이나 사회의 이익을 상호 균형 있게 추진해야 한다. 또한 데이터를 기업 내에 숨겨놓고 마치 “우리만” 이용하는 행태에서 벗어나야 한다. 데이터를 보유하고 있는 주체들끼리 협력적 관계 속에서 데이터를 공유하거나 거래하는 것이 자연스러워져야 한다. 마지막으로 데이터를 이용한 일회성 수익 추구보다는 모두를 위한 데이터 활용이 있어야 장기적으로 기업에게도 유익하다.

데이터 경제는 데이터 사회와 공존해야 한다. 데이터 공유가 기업들만의 이익이 아닌 공공의 이익도 고려하여 이루어져야 데이터 중심의 경제는 물론 사회가 완성될 수 있다. 2010년대 초기부터 일본 이동통신사들의 빅데이터 활용 사례를 보더라도 그렇다. 일본에서는 먼저 통신사의 위치 데이터를 도시재생 등 공익 목적에 사용하면서부터 통신사의 고객이기도 한 일반인들도 기업의 데이터 활용에 대한 반감이 줄어들었다. 현재 국내에서도 이동통신사나 신용카드사들이 코로나19 극복을 위해 자신의 보유 데이터를 제공하고 있다. 이런 시도도 향후 기

[그림 4] 히어 마켓플레이스



출처: HERE 홈페이지

업들의 데이터 공유와 활용도를 높이는데 긍정적으로 작용할 것이다. 단지 법규 몇 개가 바뀌었다고 상업적 데이터 사용에 대한 사회적 반감은 줄어들지는 않는다. 코로나19와 같은 공공보건 문제 이외에도 기업이 참여한 데이터 공유는 필요하다. 금융 소외자 지원이나, 재난/사고 대응 및 극복, 그리고 기후 변화, 자원 보호 등의 문제에 대해서도 데이터 공유 가능성을 높여야 한다.

물론 이러한 데이터 공유나 거래가 잘 이루어지지 않는 다양한 이유가 존재한다. 당연히 개인정보보호 및 신뢰 문제가 먼저 거론된다. 그 외에도 거래 프로세스나 기술적인 한계에 따른 거래 비용의 증가와 불확실성도 저해 요인이다. 많은 데이터를 보유하고 있는 기업들로서는 데이터를 공유하는 것이 자신의 경쟁력을 저해시키거나 자체적인 수익 창출 기회를 상실하는 것이 아닌지 우려하기도 한다. 하지만 이러한 어려움에도 불구하고 전 세계적으로 성공적인 데이터 공유 사례들이 나타나고 있다.

독일 자동차 제조사들 및 이들의 1차 협력업체(인텔, 미츠비씨, NTT 등) 컨소시엄이 소유한 정밀 디지털 지도 서비스 회사인 히어(HERE Technologies)를 보자. 최근 국내 외산 차량의 내비게이션 3D 지도 제공업자로 알려진 히어는 전 세계 유수의 위치 데이터 및 기술 플랫폼 서비스 업체로 대규모로 데이터 공유를 성공시키고 있다. 이 회사의 데이터 소스는 자동차 제조사들을 포함해 독립적인 수천 곳의 협력업체와 파트너들이다. 이 회사는 컨소시엄 내 데이터 공유에 필요한 표준을 수립하고, 데이터의 수집 및 관리는 물론 참여기관 간의 이해충돌 시 조정자 역할도 한다. 이를 통해 정밀한 3차원의 디지털 지도를 만든다. 이런 지도는 T맵 같은 일반 내비게이션용 지도보다 훨씬 고도화된 지도다. 일반 지도가 입체 현실을 평면에 그린 것이라면, 정밀 지도는 도로의 굴곡, 표지판이나 차선 두께 등 도로상 모든 정보를 촬영해 3D 화면으로 만든 것이다.

히어 마켓플레이스라는 플랫폼은 데이터 소스(공급자)들을 데이터 사용자(수요자)와 연결하며 공급자들이 자신

데이터의 가치를 평가하고 보다 많은 가치를 창출하도록 지원한다([그림 4]). 물론 유럽의 개인정보보호 법규인 GDPR을 준수하고 완전한 개인정보보호를 바탕으로 개인 동의를 받아 수용가능한 수준에서 데이터를 사용한다.

히어는 상업적인 목적 이외에 공공 서비스 프로젝트를 위해서도 데이터를 적극적으로 공유한다. 예를 들어 EU Data Task Force 프로젝트는 유럽 모든 도로의 안전을 제고하는 공익사업 컨소시엄이다. 데이터 보유 민간 기업과 테크기업을 포함해 도로교통안전과 관련된 모든 주체가 이를 위한 데이터 공유 생태계의 구축을 목표로 한다. 동 프로젝트는 특정 조건에 따라 저작물 배포를 허용하는 크리에티브 코먼 라이센스싱(creative commons licensing) 하에 데이터를 공유하여 데이터 재산권에 대한 문제 소지를 제거한다.

이런 사례를 보면 우리나라와 같이 개인정보보호 법규가 강한 유럽에서는 상업적 목적의 데이터 공유라고 하더라도 사회적 이익과 조화시키는 체제를 갖추면서 데이터 활용이 추진되고 있음을 알 수 있다. 데이터 공유 과정에서 컨소시엄이 여러 데이터 소스들로부터 데이터를 통합하고 다른 소스들을 결합하여 정제해 더욱 유용하게 만드는 작업을 한다. 이런 조직은 누가 어떤 목적으로 어디에 사용하는지에 대해 통제하는 소위 데이터 거버넌스 역할도 수행한다. 그리고 사회적 합의로 데이터 공유 과정에서 법제도상의 마찰을 피하는 방법을 충분히 고려하면서 사업을 추진한다.

새로운 데이터 거버넌스

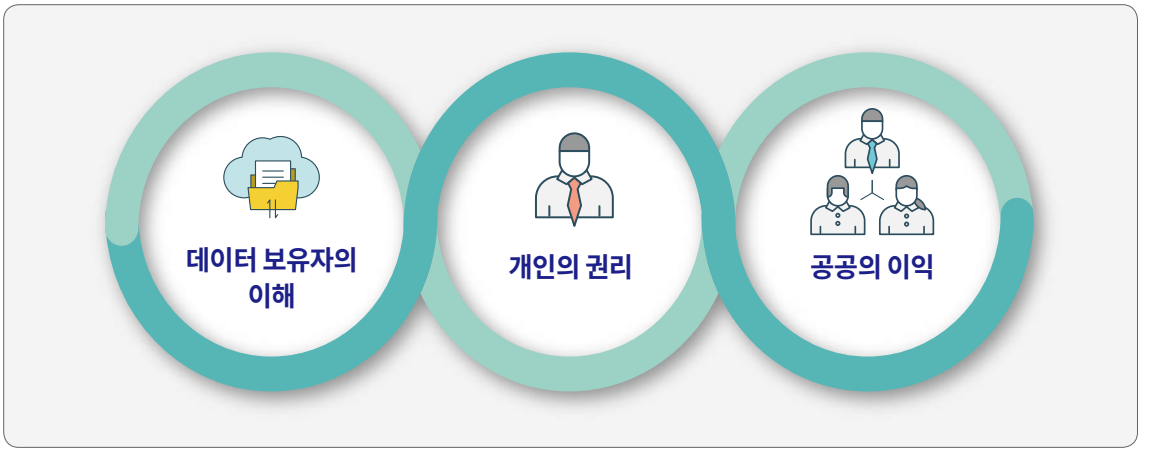
많은 기업이 아직도 데이터를 효과적으로 활용하는데 어려움을 호소하고 있으며 과연 이에 투자해야 하는지도 확신을 못하고 있다, 데이터는 기업 가치창출의 원천이지만 반대로 심각한 위험 요인이기도 하다. 개인정보를 잘못 관리하거나 사용하다가 법적인 책임과 배상은 물론이고 기업의 평판에 심각한 손상을 입을 수 있다. 데이터의 활용이 단기간에 수익으로 연결되지도 않는다. 데이터의 가치와 이에 대한 책임은 동전이 양면이다. 이런 데이터의 양면성으로 인해 상용 및 공공 데이터를 기반으로 하는 혁신들이 진전되지 못할 수도 있다. 어떻게 데이터가 사회 모든 구성원의 이해를 균형 있게 조화시키면서 유용하게 활용될 수 있는지의 문제는 최근 새로운 개념의 데이터 거버넌스 확립 측면에서 논의되고 있다.

데이터 관리의 체계 및 절차 그리고 역할과 책임을 말하는 데이터 거버넌스는 자금까지 개인정보보호에 치중해왔으나 이는 결국 기업은 물론 공익을 위한 데이터 활용을 저해해 왔다. 따라서 새로운 데이터 거버넌스는 앞서 히어의 사례와 같이 개인의 권리, 사회적 이익, 그리고 기업들 데이터 보유자의 이해 등 세 가지 요소 간의 적절한 균형을 찾는 데 보다 중점을 둔다([그림 5]). 최근 WEF(World Economic Forum)은 개인의 동의에 대한 과도한 요구, 보유한 데이터를 상업적으로 사용하려는 기업에 대한 (개인정보보호 이외의) 높아진 규제, 상업적 혹은 공익적 데이터 사용 기회 상실 등의 문제 해결에도 데이터 거버넌스가 적용되어야 한다고 주장한다.

앞으로 개인정보보호와 데이터 소유권 이외에도 데이터 주권 등 모호한 개념보다는 좀 더 구체적인 이슈들에 대한 거버넌스와 법제도 정비가 필요하다. 특히 데이터 재산권, 데이터 정확성, 데이터 접근권에 대한 구체적인 거버넌스는 긍정적인 데이터 경제와 사회를 이루기 위한 근간이다. 이들은 상호 연관되지만 각각에 대한 사회적 이해와 합의도 필요하다. 개인정보보호 이외에도 우리가 데이터 거버넌스에 고려해야 할 이슈들은 다음과 같다.

첫째, 데이터의 공유 또는 거래의 필요성과 데이터 자체가 가지는 금전적 가치가 높아지면서 데이터의 재산권을 어떻게 평가하고 인정하느냐가 주목을 받고 있다. 이는 단순히 소유권 문제도 있지만, 데이터의 가치를 어떻게 평가하느냐도 중요하다. 개인에게 소유권이 있는 데이터에 대해서도 마찬가지이다. 민간 기업들 상호 간 혹

[그림 5] 새로운 데이터 거버넌스의 세 가지 핵심 요소



출처: WEF, 2020

은 자체의 문제는 정부가 개입할 필요가 없지만 공공 데이터의 상업적 공유나 개인 데이터 거래에서 이의 가치를 평가하는 것은 점점 더 중요하고, 필요하게 될 것이다. 현재도 국내 일부 공공기관들은 자신들 데이터의 판매를 시도하고 있으나 이에 대한 정부나 지자체 차원의 원칙이나 기준은 물론 구체적인 법제도는 존재하지 않는다. 앞서 언급한 공개 데이터의 활용 문제도 이와 관련된다.

두 번째로 데이터의 오류나 품질 저하로 인한 이해충돌이나 사고 등의 문제 발생 가능성이 커지고 있어 단순한 데이터의 표준화를 넘어 데이터 정확성에 대한 체계도 필요하다. 개인에 관한 데이터를 잘못 수집해 개인에게 피해가 가는 문제도 있지만, 주소 업데이트가 안 되어 개인이 불편을 겪을 수도 있다. 주기적인 데이터 갱신의 의무화 등 데이터의 정확성에 대한 원칙 수립이 분야별로 필요하다. 이런 문제는 데이터의 일관성 유지를 위해 데이터의 단위까지 관련이 된다. 해외 사례를 보면 의료 분야에서 환자 중심의 통합된 의료 정보 제공에서 데이터 단위 표준화의 중요성이 매우 높다.

마지막으로 데이터 접근권에 대한 거버넌스도 구체화할 필요가 있다. 이미 코로나19를 거치면서 개인 동의 없이 이동통신이나 신용카드 데이터 등 개인의 사생활에 관련해 제한 없는 정부기관의 접근이 이루어지고 있다. 하지만 어떤 위급한 상황에서 "사전에 승인된" 접근이 가능한지에 대한 기준 및 프로세스 확립이 필요하다. 국내에서는 개인이 '위치정보수집'에 동의하지 않더라도 '감염병의 예방 및 관리에 관한 법률 제76조의 2항'에 근거하여 확진자들의 신용 카드 기록, 이동통신사 데이터와 같은 위치 데이터를 확보할 수 있다. 이와 관련해서 정부가 모든 데이터를 중앙에 집중해 관리할 필요가 있는지도 사실 더 많은 논의와 검토가 필요하다. 개인정보 보호에 민감한 독일, 스위스 등 일부 유럽 국가들은 코로나19와 같은 심각한 공중보건 상황에서도 개인 데이터를 정부에 집중하기보다는 단지 잠재적인 감염 위험을 개인들에게 통보할 때 사용한다. 이들의 위치 데이터는 추적하지도 않는다. 싱가포르와 같은 통제된 국가에서도 개인의 동의 하에서만 정부가 위치 데이터를 사용한다.

코로나19와 같은 사회적 문제 해결에 개인 데이터 사용을 총괄하는 정부 체계는 필요하다. 그러나 개인 데이터의 집중은 이에 상응하는 데이터 거버넌스가 있어야 한다. 코로나19를 위한 감염접촉자 파악 앱은 바이러스

를 추적하는 데 효과적이지만 적절한 데이터 거버넌스는 필수적이다. 때때로 개인 동의 없이도 구체적이며, 광범위한 사회적인 영향을 미치는 문제에만 개인 데이터의 접근과 사용이 가능할 수 있다. 누가 특정 데이터를 소유하고, 누가 어떤 조건에서 데이터에 접근하는 것을 허용하는지는 사전 원칙과 절차가 있어야 한다.

맺음말

현재 정부가 추진 중인 「데이터 기본법」에는 데이터의 가치 평가와 지원, 데이터 이동 촉진을 위한 원칙, 공정한 유통 환경 조성, 데이터 거래소 지원, 데이터 품질 관리, 인력 양성, 세제 지원 등이 포함돼 있다. 그동안 데이터 거버넌스의 3개의 축 중 개인의 권리(개인정보보호)에 치중했던 무게의 중심을 다양한 데이터 관련 이슈들을 고려하면서 데이터 보유자의 이해를 반영하려는 정책으로 보인다. 하지만 「데이터 기본법」은 이미 데이터의 상업적 활용에만 초점이 맞춰져 있다는 시민단체의 반발을 사고 있다. 사전에 공공의 이익에 대한 충분한 고려와 구체적인 대안 제시가 아쉬운 대목이다. 반대로 코로라19 상황에서의 지나친 정부의 개인 데이터 접근과 개입은 개인의 권리를 침해할 수 있다.

결국 데이터의 활용은 공익적 목적이든 상업적 목적이든 먼저 데이터 거버넌스의 세 가지 요소들을 동시에 균형 있게 충족시키면서 추진해야 경제와 사회에 긍정적인 결과를 가져올 수 있다. 그런 의미에서 현재의 개인정보보호법은 개인정보(특히, 위치 데이터)의 공익적 활용 조건 및 용도, 그리고 범위 등을 명시하도록 시행령에서 개선할 필요가 있다. 이와 관련해서 개인정보보호위원회의 역할과 기능도 위에서 소개한 새로운 데이터 거버넌스 요소들을 반영할 여지가 있다. 물론 이런 원칙은 반드시 법을 통해서만 구현되는 것은 아니며 데이터를 사용하는 곳은 어디든지 내부 정책 수립과 의지로 실현할 수 있다.

참고 자료

이민우·김예지·이재진·문규환·황선배·전용주·함유근 (2020). “이동통신 데이터를 활용한 빅데이터 기반 역학조사지원 시스템” 한국빅데이터학회지, 제5권 제2호, pp. 187-199.

함유근, 채승병 (2012). 빅데이터, 경영을 바꾸다 서울: 삼성경제연구소.

European Commission, (2020). International Digital Economy and Society Index (DESI) 2020.

IDC (2018), The Digitization of the World: From Edge to Core.

Kushal, A., Moorthy, S., & Kumar, V., (2012). Pricing for data markets. Technical Report.

Laney, Douglas B., (2017). Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage, NY: Taylor and Francis. Kindle Edition.

Moody, D. L., & Walsh, P. (1999, June). Measuring the Value of Information-An Asset Valuation Approach. In ECIS (pp. 496-512).

Ratner, A., Bach, S. H., Ehrenberg, H., Fries, J., Wu, S., & Ré, C. (2020). Snorkel: Rapid training data creation with weak supervision. The VLDB Journal, 29(2), pp. 709-730.

World Economic Forum, (2020). Technology and Data Governance in Cities Indian Smart Cities at the Forefront of the Fight Against COVID-19.