

# 데이터 가치평가 방법론

## I. 서론

## II. 데이터 가치평가 방법론 개관

1. 데이터의 분류
2. 데이터의 가치평가 방법 개관

## III. 기업의 관점에서 본 개인정보 가치평가

1. 전통적 자산가치 평가 방법의 적용
2. 부채로서의 데이터 - 데이터 유출 위험
3. 데이터에 부가된 법적 제약의 고려 필요성

## IV. 정보주체의 관점에서 본 개인정보 가치평가

## V. 정책결정자의 관점에서 본 개인정보 가치평가

## VI. 결론



김병필

KAIST 기술경영학부  
초빙교수·변호사

# I. 서론

2016년 마이크로소프트는 링크드인(LinkedIn)을 262억 달러(한화 약 30조 원)에 인수하였다. 링크드인의 가치가 이렇게 높게 평가된 데에는 링크드인이 보유한 방대한 사용자 정보가 중요한 고려사항이었을 것이다. 인수 당시 링크드인의 등록 이용자 수는 4억 3,300만 명이었고, 그 중 1억 명 정도가 월 활동 이용자(Monthly Active User)에 해당하였다. 단순화하여 보면, 마이크로소프트는 활동 이용자 1명당 약 260달러(한화 약 30만 원)을 지급한 셈이다.

마이크로소프트의 링크드인 인수는 성공적인 것으로 평가되고 있다.<sup>1)</sup> 2017년 23억 달러였던 링크드인의 매출은 2018년 53억 달러로 증가하였고, 2022년에는 100억 달러(한화 약 12조 원)에 이를 것으로 전망되고 있다. 링크드인은 비즈니스 네트워킹 겸 구인·구직 플랫폼이므로 이용자들이 다른 소셜미디어에 비해 신뢰할 수 있고 정확한 정보를 게시한다는 특징이 있다. 링크드인은 사용자 데이터를 바탕으로 구인·구직 솔루션을 유상으로 제공한다. 이용자 맞춤형 광고도 중요한 수익원이 된다. 이러한 링크드인의 사업 모델을 고려하면, 링크드인의 급속한 성장에는 양질의 사용자 데이터가 중요한 역할을 하였음을 알 수 있다.

데이터가 중요한 가치를 갖는 것은 사용자 정보를 활용하는 소셜 미디어 등 디지털 플랫폼 기업에만 국한되지 않는다. 데이터를 분석한 결과(analytics)를 얼마나 잘 활용하는지는 기업의 경영 성과를 좌우하는 중요한 요소가 되었다. 2011년 미국 MIT Sloan 경영대학원의 연구자들은 100개 국가 30개 이상 산업 분야에서 3,000명에 가까운 기업 경영자 및 데이터 분석가를 대상으로 데이터 활용의 가치에 관한 설문조사를 실시하였다. 조사 결과, 실적이 우수한 기업들은 낮은 기업들에 비해 데이터 분석 결과를 5배 더 많이 활용하고 있는 것으로 나타났다.<sup>2)</sup> 또한 PwC의 2019년 보고서에 따르면, 데이터 분석 결과에 기반하여 경영하는 기업은 동일 업종 내의 다른 기업에 비해 기업 가치가 더 높고, 데이터 분석 역량이 우수한 기업이 해당 업종에서 상위의 성과를 낼 가능성이 2배 이상 높다고 한다.<sup>3)</sup> 데이터 활용을 통해 기업 생산성을 개선한 사례도 많이 보고되고 있다. 예컨대 세계적 정유사인 BP는 데이터를 활용함으로써 일 생산량을 3만 배럴 향상시킬 수 있었다고 한다.<sup>4)</sup>

데이터의 가치가 크다는 점에는 이견이 없다. 그렇다면 그 가치는 어떻게 평가될 수 있을까? 이 문제는 기업 경영자나 정책결정자뿐만 아니라 법률가에 있어서도 점차 중요한 문제가 되어가고 있다. 기업 도산 절차에서 채무자가 보유한 사용자 정보의 가치는 어떻게 평가되어야 하는가? 개인정보 유출 사고가 발생한 경우 정보주체에 대한 보상금은 어떻게 산정되어야 하는가? M&A 거래에 있어 인수자는 데이터의 가치가 적절히 산정되었다는 점을 보장받기 위해서는 어떠한 점들을 고려해야 하는가? 이러한 질문들은 데이터 가치평가 방법에 대한 이해를 바탕으로 할 때 적절히 답변될 수 있을 것이다. 이 글은 법률가를 대상으로 하여 데이터 가치평가 방법론을 개괄적으로 소개한다.

1)  
머니투데이, 2019. 11. 1, "링크드인이 독보적인 비즈니스 플랫폼이 된 비결"

2)  
LaValle, S., Lesser, E., Shockley, R., Hopkins, M., and Kruschwitz, N., (2011), "Big Data, Analytics and the Path from Insights to Value", MIT Sloan Management Review, Winter 2011, Vol 52 (2).

3)  
PwC, (2019), "Putting a value on data"

4)  
Financial Times, 2019. 10. 6, "BP's Bernard Looney takes oil major into energy transition"

## II. 데이터 가치평가 방법론 개관

### 1. 데이터의 분류

기업은 데이터의 가치를 평가하기에 앞서 구체적으로 어떠한 유형의 데이터를 얼마만큼 보유하고 있는지 분류해야 한다. 그런데 방대하고 복잡한 데이터를 보유한 기업들에게는 데이터 분류 작업부터가 쉽지 않은 일이 된다. 일례로 어떤 금융 분야 대기업은 수백 테라바이트(TB)의 데이터를 보유하고 있는데, 기업 차원의 데이터 활용 전략 수립을 위해 보유 데이터를 “매우 중요”, “중요” “기타”로 분류하는 프로젝트를 진행하였다. 위 프로젝트를 담당했던 팀은 9개월 동안 수백 테라바이트 데이터 중 20테라바이트도 분류하지 못하였다고 한다.<sup>5)</sup> 더욱이 위 조사에 따르면 기업이 보유하고 있는 데이터 용량은 연간 40%씩 증가하고 있다. 이와 같은 데이터의 복잡성, 기하급수적인 용량 증대로 인해 데이터를 분류하는 작업부터 쉽지 않은 과제가 되고 있다.

기업들은 산업 특성이나 데이터 분류의 맥락에 따라 각기 다른 기준을 자체적으로 수립하여 데이터를 분류하고 해당 데이터에 대한 가치평가 작업을 수행할 수 있을 것이다. 일용 기업이 활용할 수 있는 기본적인 분류 기준으로는 데이터의 출처, 데이터의 내용 등이 있다.<sup>6)</sup>

우선, 기업이 보유한 데이터의 출처를 기준으로 분류하면 ① 저술 데이터(도면, 사진, 음악 등 창작적 과정을 거쳐 생성된 데이터) ② 이용자 제공 데이터(소셜 미디어, 제품 리뷰 등 이용자가 시스템에 입력한 데이터) ③ 기록 데이터(재무 거래, 웹사이트 방문 이력, CCTV 촬영물 등 온라인 또는 오프라인 데이터 기록) ④ 파생 데이터(신용점수, 통계 등과 같이 여러 데이터를 결합하거나 취합하여 만들어진 데이터) 등으로 구분할 수 있다.

한편, 데이터의 내용을 분류 기준으로 한다면, ① 마스터 데이터(고객 데이터, 협력사 데이터, 제품 데이터, 직원 데이터 등 기업의 운영에 있어 중요한 사람, 장소 및 사물에 관한 정보) ② 거래 데이터(판매 데이터, 지급 데이터, 고객 접촉 데이터, 지리공간 데이터 등 기업 내부 또는 외부 거래 사건 기록) ③ 참조 데이터(지역, 통화, 산업 표준 등 다른 데이터 분류를 위한 목적으로만 활용되는 정보) ④ 메타 데이터(생성일 태그, 저자 태그, 수정 이력, 필드 설명, 감사 이력 등 데이터의 추출, 해석 또는 이용을 보조하기 위한 데이터의 특성 정보), ⑤ 비구조화된 데이터(오디오, 텍스트, 비디오, 사진 등 일관된 양식이나 문법이 결여되어 있는 정보)로 구분될 수 있다.

개인정보에 초점을 맞출 경우, 정보가 생성되는 방식에 따라 구분될 수도 있다.<sup>7)</sup> 즉, ① 정보주체가 자발적으로 생성한 데이터(소셜미디어에 등록 또는 업로드한 데이터 등), ② 정보주체의 행동을 관측한 데이터(웹사이트 방문 이력 등), ③ 추론된 데이터(자발적 데이터와 관측 데이터를 분석하여 얻어진 데이터 등)로 구분될 수 있을 것이다.

5) Short, J., and Todd, S. "What's your data worth?", MIT Sloan Management Review, Spring 2017.

6) Infocomm and Media Development Authority (IMDA) and Personal Data Protection Commission (PDPC), "Guide to data valuation for data sharing", 2019.

7) Julia Wdowin & Dr Stephanie Diepeveen, "The Value of Data Policy Implications Report, Accompanying Literature Review," 2020. 2.

앞서 나열한 여러 데이터 중에서는 별도의 법적 권리의 대상이 되는 경우도 있다. 예컨대 사진, 음악, 영상 데이터는 저작권의 보호 대상이 될 수 있고, 비밀로 관리된 기술상·경영상의 정보는 영업비밀로 보호될 수 있다. 이하에서는 특히 여러 데이터의 유형 중 특히 개인정보 데이터의 가치평가 방법에 대해 초점을 맞추고자 한다.

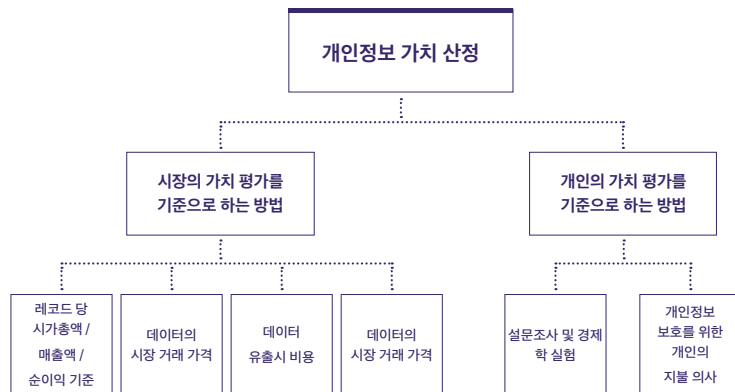
## 2. 데이터의 가치평가 방법 개관

데이터(특히 개인정보)의 가치를 평가하는 방법에 대해 기업, 규제 당국, 조세 당국 간에 합의된 표준적 방법은 아직 존재하지 않는 것으로 보인다.<sup>8)</sup> 따라서 현재로서는 어떠한 방법론이 존재하고, 각기 어떻게 활용될 수 있으며, 어떤 장단점이 있는지 이해할 필요가 있다.

2013년 OECD는 개인정보 가치평가 방법에 대한 문헌 조사 결과를 정리한 보고서를 발간한 바 있다.<sup>9)</sup> 위 보고서는 개인정보의 가치를 평가하는 방법을 ① 시장 가치에 기반한 방법과 ② 정보주체의 가치평가에 기반한 방법으로 구분하였다. 시장 가치에 기반한 방법은 (i) 개인정보 기록을 보유한 기업의 주가, 매출액, 순이익을 기준으로 하는 방법, (ii) 데이터의 정상시장 거래 가격을 기준으로 하는 방법, (iii) 데이터 유출 사고시 보상액을 기준으로 하는 방법, (iv) 암시장에서의 데이터 가격을 기준으로 하는 방법으로 구분된다. 한편, 정보주체의 가치평가에 기반한 방법으로는 (i) 정보주체가 개인정보에 대해 주관적으로 부여한 가치를 조사하는 방법과 (ii) 정보주체가 자신의 프라이버시를 지키기 위해서 지불할 의사를 기준으로 하는 방법으로 구분된다.

8) PwC, 앞의 보고서(주 3)

9) OECD, "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value", OECD Digital Economy Papers, No. 220 (2013. 4. 2.), OECD Publishing, Paris.

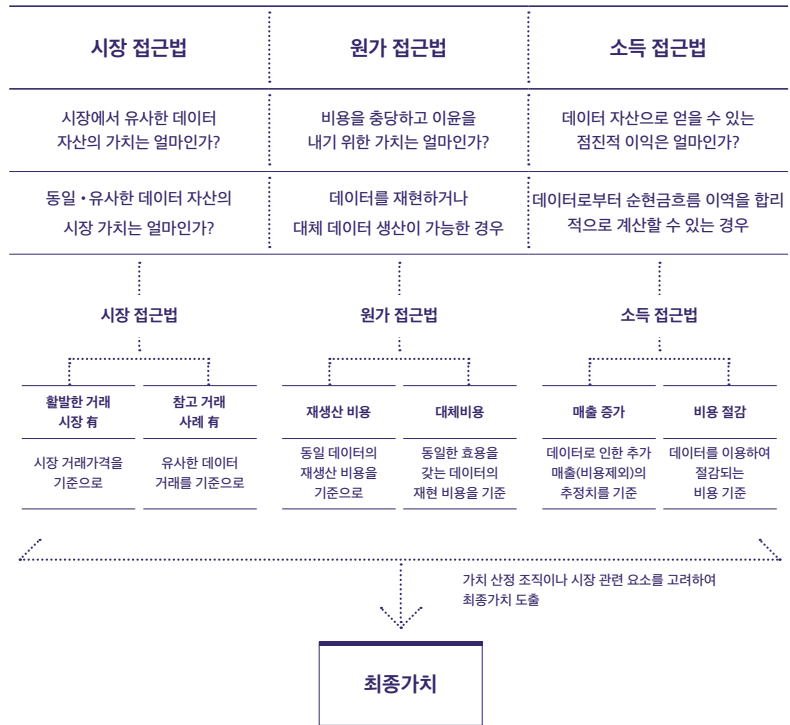


[그림 1. 개인정보 가치 추산 방법. 출처: OECD(2013) 번역]

한편, 재무적 관점에서 본다면 데이터는 수익을 창출해 낼 수 있다는 점에서 지적재산권이나 영업권 등의 다른 무형자산과 다르지 않다고 볼 수 있다. 이러한

관점에서 본다면, 전통적인 자산 평가 방법을 데이터 가치평가에도 그대로 적용할 수 있다.<sup>10)</sup> 일반적으로 자산의 가치평가 방법은 ① 소득 접근법(income approach), ② 시장 접근법(market approach), ③ 원가 접근법(cost approach)<sup>11)</sup>이 있다. 이러한 방법은 유가증권, 부동산, 동산 등의 유형자산뿐만 아니라 소프트웨어나 데이터베이스 등의 무형자산의 가치를 평가하는데도 활용되어 왔다.<sup>12)</sup>

싱가포르 정보통신 개발청(IMDA)와 개인정보보호위원회(PPDC)가 2019년 발간한 “데이터 가치평가 가이드”는 위와 같은 전통적인 자산평가 모델을 적용하여 시장 접근법, 원가 접근법, 소득 접근법을 적용한 결과를 종합하여 데이터의 최종적 가치를 평가할 수 있다고 설명하고 있다(그림 2 참조).



[그림 2. 시장, 원가, 소득 접근법 개관. 출처: IMDA·PPDC(2019)<sup>13)</sup> 번역]

한편, 정책 결정자의 관점에서는 데이터의 사회적 편익을 계산하는 것이 필요하다. 이 경우 데이터 활용을 통해 특정 경제적 또는 사회적 목표에 어떠한 인과적 영향을 낼 수 있을 것인지(가령 사회 후생의 증대 효과) 또는 데이터를 활용하지 못함으로써 비효율이 발생할 경우 그 사회적 비용이 얼마나 되는지 측정하는 방법이 활용될 수 있다. 이러한 목적으로는 이제까지 경제학에서 활용되어 온 비용·편익 분석 방법론을 도입할 수 있다.<sup>14)</sup> 다만 이러한 방법은 데이터 활용의 사회적·경제적 효과를 평가하는 것이므로, 특정한 데이터에 대한 가치평가와의 구분할 필요가 있다.

이상에서 살펴본 여러 방법론을 종합하면, 데이터(특히 개인정보)의 가치평

10) PwC, 앞의 보고서(주 3) 6면.

11) 원가 접근법은 '자산 접근법(asset-based approach)'라고도 한다. 기본적으로 재무제표의 자산은 역사적 원가를 기준으로 하여 측정되기 때문이다.

12) 성태웅·변정은·박현우, “데이터베이스 자산 가치평가 모형과 수명주기 결정”, 한국콘텐츠학회논문지 Vol. 16 No. 3 (2016),

13) IMDA&PPDC, 앞의 보고서(주 6). IMDA&PPDC의 방법론은 PwC, 앞의 보고서(주 3)와 동일하다.

14) 예컨대 Global Partnership for Sustainable Development의 2018년 보고서는 데이터의 가치 평가 방법으로 원가 기반, 시장 기반, 소득 기반 방법론 이외에도 편익 현금화(benefit monetization) 접근법과 영향 기반(impact-based) 접근법이 있다고 설명한다. Jenna Slotin, (2018), “What Do We Know About the Value of Data?”, Global Partnership for Sustainable Development. <http://www.data4sdgs.org/news/what-do-we-know-about-value-data> (2020. 8. 17. 방문)

가는 누구의 관점에서 이를 평가할 것인지에 따라 그 적용 방법론이 달라진다는 점을 확인할 수 있다. 우선 개인정보를 보유한 기업의 관점에서는 전통적인 자산 가치평가 모델(시장 접근법, 원가 접근법, 소득 접근법)을 활용하거나, 데이터 유출 사고 발생시 부담할 수 있는 비용을 기준으로 삼을 수 있다(이하 III장에서 상술). 정보주체의 관점에서는 주관적으로 개인정보에 부여하는 가치 또는 프라이버시를 유지하기 위해 지불할 의사가 있는 금액을 설문조사 또는 실험 등의 방법으로 평가할 수 있다(이하 IV장에서 상술). 정책결정자의 관점에서는 데이터를 활용함으로써 발생시킬 수 있는 사회적 편익 또는 데이터를 활용하지 못함으로써 발생하는 사회적 기회비용을 측정하는 것이 유용할 수 있다(이하 V장에서 상술). 이하에서는 각각의 적용 방법 및 이를 활용한 사례를 살펴본다.

### III. 기업의 관점에서 본 개인정보 가치평가

#### 1. 전통적 자산가치 평가 방법의 적용

##### (1) 소득 기반 접근법(income-based approach)

소득 기반 접근법은 데이터를 활용함으로써 얻을 수 있는 미래의 소득이 얼마나 될 것인지 추정된 다음 이를 현재가치로 할인한 금액을 데이터의 가치로 평가하는 방법이다. 소득 기반 접근법에 있어 가장 대표적으로 활용되는 방법으로는 현금흐름할인(Discounted Cash Flow, DCF) 방법이 있다.

데이터를 활용함으로써 기업이 이익을 얻는 경로는 다양하다. 예컨대 맞춤형 광고를 통해 더 많은 고객을 유치할 수 있고, 데이터 분석 결과에 기반하여 제품이나 서비스의 가격을 조정하는 등의 방법으로 매출을 증가시킬 수도 있다. 또한, 데이터를 활용하여 기계의 생산성을 향상시키거나, 고객의 사기를 예방하여 손실을 방지할 수도 있으며, 맞춤형 광고의 결과 전체 광고비용을 줄일 수도 있다.

소득 기반 접근법은 평가 대상 데이터의 유용성이 유지되는 수명 기간 동안의 미래의 매출 증가 혹은 비용 절감 효과를 추정하고, 데이터 활용에 소요되는 비용을 차감함으로써 데이터 활용에 따라 얻어질 수 있는 현금 흐름을 추정한다. 여기에 적절한 할인율을 적용하여 현재가치로 할인한 다음 이를 합산한 것이 해당 데이터의 가치가 된다. 싱가포르 IMDA&PDPC (2019)는 보험회사가 고객 데이터를 활용하여 보험사기를 탐지하여 지급 보험료를 절감할 수 있는 경우, 해당 데이터의 가치를 산정하는 예시 사례를 다음과 같이 제시하고 있다. 예시 사례에서 데이터 활용을 통한 지급 보험료 절감액(A)은 1년차 200만 달러, 2년차 490만 달러, 3년차 680만 달러로 예상되었다. 데이터 활용 비용(B)은 위 기간 동안 매년 20만 달러가 발생한다고 가정하였다. 이 경우 지급 보험료 절감액(A)에서 데이터 활용 비용(B)을 차감한 것이 세전 현금흐름이 된다. 여기에 세금을 차감하고 할인계수를 적용하여 현재가치를 산정한 다음, 이를 합산한 것이 위 데이터의 현재가치가 된다.

\$\$	예측 1년차	예측 2년차	예측 3년차
지급 보험료 절감액(A)	2,000,000	4,900,000	6,800,000
데이터 활용 비용(B)	200,000	200,000	200,000
세전 현금흐름(A-B)	1,800,000	4,700,000	6,600,000
세후(17%) 현금흐름	1,494,000	3,901,000	5,478,000
할인계수	0.91	0.83	0.75
현재가치	1,358,182	3,223,967	4,115,702
현재가치 합계(= 데이터 가치)			<b>8,697,851</b>

[표 1. 소득 기반 접근법에 따른 데이터 가치 산정 예시. 출처: IMDA·PPDC(2019)]

소득 기반 접근법은 이론적으로 정교한 방법이지만, 미래의 현금흐름을 예측하기 위해 많은 추정을 해야 한다는 어려움이 있다. 위 보험회사 사례에서 고객 데이터를 활용함으로써 미래 지급 보험료를 얼마나 절감할 수 있을 것인지 예측하기란 쉽지 않을 것이다. 따라서 평가자의 주관적 요소가 작용할 여지가 크다. 또한, 현재가치는 할인율에 따라 민감하게 변화하므로, 다른 할인율을 적용하면 데이터의 현재가치 액수가 상당히 달라질 수 있다는 문제점도 있다.

## (2) 시장 기반 접근법(market-based approach)

시장 기반 접근법은 일반적으로 시장에서 동일 또는 유사한 자산이 얼마의 가격에 거래되었는지 확인하는 것이다. 만약 해당 자산이 활발하게 거래되는 거래소(또는 판매자)가 있다면 그 거래소의 가격을 기준으로 하고, 그렇지 않다면 유사한 거래 사례를 찾는 방법을 사용한다.

미국의 경우 여러 출처로부터 개인정보를 수집·취합하여 판매하는 이른바 ‘데이터 브로커(data broker)’가 매우 활성화되어 있다. 데이터 브로커들은 소비자가 온라인에 게시한 정보, 온·오프라인 쇼핑 정보, 웹사이트 등록 정보, 제품 등록 정보, 부동산 등기부, 유권자 명부 등 다양한 출처로부터 데이터를 수집·가공하여 판매하는 사업을 수행한다. 미국 연방거래위원회(Federal Trade Commission, FTC)의 2014년 데이터 브로커 산업 실태 보고서에 따르면 미국 내에는 600 내지 700여 개의 데이터 브로커가 활동하고 있고 산업 규모는 약 150억 달러(한화 약 18조 원)에 이른다.<sup>15)</sup> 이들이 판매하는 개인정보의 가격을 통해 개인정보의 가치를 산정해 볼 수 있다.

### 1) 개인정보의 정상 시장 거래 가격

종래 개인정보 거래는 통상 기업 간 사적으로 이루어져 왔으나, 현재는 (적어도 미국의 경우에는) 특정 유형의 개인정보 가격은 인터넷을 통해 공개되어 있거나 데이터 브로커의 영업 사원을 접촉하여 견적을 받을 수 있는 경우가 많다.

15)

USA Federal Trade Commission, "Data Brokers: A Call For Transparency and Accountability: A Report of the Federal Trade Commission". 2014.

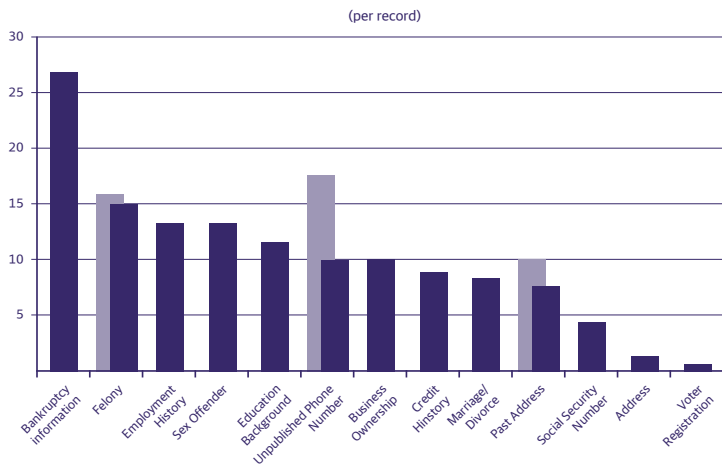
16)  
OECD, 앞의 보고서(주 9) 26면.

17)  
Financial Times, 2013. 6. 13. "Financial worth of data comes in at under a penny a piece"

OECD의 2013년 보고서는 당시 미국의 주요한 데이터 브로커들이 판매하는 개인정보의 가격을 종합한 결과를 제시하고 있다. 위 보고서에 따르면, 미국 개인의 주소 정보는 건당 0.5센트, 생일 정보는 건당 2달러, 사회보장 번호는 건당 8달러, 운전면허 번호는 건당 3달러, 군복무 기록은 35달러 등에 이른다.<sup>16)</sup>

한편, 위 OECD의 조사 결과와는 달리 2013년도의 파이낸셜 타임즈는 데이터 브로커를 통해 구할 수 있는 개인정보의 건당 가격은 1달러도 채 되지 않는다고 보도한 바 있다. 위 기사에 따르면, 이용자의 연령, 성별, 지역 등 기본 정보의 가치는 건당 0.0005달러, 자동차 구매 의향이 있는 고객 정보는 건당 0.0021달러, 어떤 여성의 임신 여부와 임신 기간에 대한 정보는 건당 0.11달러에 불과하였다. 또한, 건당 0.26달러를 지불하면 특정한 처방약을 복용하고 있거나 특정한 건강 상태에 있는 사람의 리스트를 구할 수 있다고 한다.<sup>17)</sup> 데이터 브로커가 개인정보를 복제하여 판매하는 한계 비용은 거의 0이므로, 데이터 브로커는 이미 다량으로 수집한 개인정보를 위와 같이 낮은 가격으로 판매하고 있는 것이다. 한편, 개인정보의 시장 가격은 거래가 이루어지는 맥락에 따라 달라질 수 있다는 점에 유의해야 한다. 일반적으로 개인의 휴대전화번호 1건의 가격은 높지 않다고 하더라도 만약 그 개인의 소득 수준이나 특정한 관심사 등과 함께 제공된다면 그 휴대전화번호의 가치는 높아질 수 있다. 또한, 데이터 브로커에 의해 제공되는 데이터의 품질을 사전에 확인하기 어렵다는 문제가 위 가격에 영향을 끼친다는 점도 간과되어서는 안 된다. 거래 대상이 되는 데이터가 부정확하거나 그 정확도를 신뢰할 수 없는 경우가 있을 수 있고, 앞서 조사된 개인정보의 시장거래 가격에는 이러한 위험성이 반영되어 있는 것으로도 이해될 수 있다.

### Market prices for personal data by type



[그림 3. 미국 개인정보의 건당 시장 가격. 출처: OECD(2013)]



## 2) 개인정보의 암시장 거래 가격

개인정보는 사이버 범죄를 위한 암시장에서도 거래되고 있으므로, 암시장에서의 가격도 참고로 고려될 수 있다. 특히 암시장에서는 신용카드 정보가 특히 많이 거래되고 있는데 2009년 기준으로 건당 1달러에서 30달러 사이의 가격으로 제안되었다고 한다.<sup>18)</sup>

한편, 암시장에서 거래되는 특정인에 관해 완전히 식별된 보건의료 정보의 가치는 건당 대략 1달러에서 15달러 사이인 것으로 추산된다.<sup>19)</sup> 영국에서는 의료전산기록을 인도로 외주를 주었다가 해당 정보가 암시장으로 유출된 사례가 있었는데, 해당 정보는 건당 4파운드에 거래되었다고 한다. 보건의료 데이터의 개별 건당 가치는 이처럼 그리 높지 않을 수 있으나, 방대한 데이터가 축적되어 있는 경우 데이터베이스 전체에 대한 가치는 상당히 높을 수 있다. 미국에서는 2009년 한 해커가 버지니아 주 의료국 처방전 데이터베이스에 침입하여 1,000만 달러의 랜섬을 요구한 사례가 있다. 이 경우 1건의 환자 데이터는 1.2달러에 불과하였지만, 전체 데이터베이스에 대해서는 상당한 금액이 된 것이다.

## 3) 데이터 브로커 기업의 가치

이상에서 살펴본 개인정보의 시장가격은 해당 개인정보를 구매하여 이용하고자 하는 수요 기업의 입장에서 개인정보의 가치에 해당한다. 그런데, 데이터는 비경합적(non-rivalry) 재화이므로 한 고객이 데이터를 구매하였다고 해서, 데이터를 판매하는 브로커 입장에서 데이터의 가치가 감소하는 것이 아니다. 따라서 특정 데이터의 개별 시장 거래 가격은 판매자의 입장에서 해당 데이터의 가치라고 평가하기 어렵다.

데이터 브로커가 보유한 개인정보의 가치는 해당 데이터 브로커의 기업 가치 혹은 해당 데이터로부터 발생하는 매출을 통해 추정해 볼 수 있다. OECD의 2013년 보고서는 다국적 소비자 신용정보 회사인 Experian에 관해 이러한 방법을 적용한 사례를 제시하고 있다.<sup>20)</sup> Experian은 2011년 개인 6억 명 및 사업체 6,000만 개에 대한 신용정보를 보유하고 있었고, 이 데이터를 활용하여 미화 42억 달러(한화 약 5조 원)의 매출액을 얻었다. 당시 위 기업의 기업가치(시가총액)는 미화 약 100억~120억 달러(한화 약 12~14조 원) 사이였다. 따라서 Experian이 보유한 신용정보는 건당 연간 평균 약 미화 6달러(연간 전체 매출액 ÷ 보유개인정보 건 수)의 매출을 발생시켰고, 신용정보의 시장가치는 건당 평균 약 미화 19달러(Experian의 기업가치 ÷ 보유개인정보 건 수)에 이른다고 추산해 볼 수 있다.

### (3) 원가 기반 접근법(cost-based approach)

원가 기반 접근법은 데이터를 생성하기 위해 소요된 비용을 기준으로 데이터의 가치를 산정하는 방법이다. 비용 접근법은 크게 두 가지로 나뉘는데, (1) 평가 대상 자산을 자체적으로 다시 생산 또는 재현해 내는 현재 비용을 기준을 하는

18)

OECD, 앞의 보고서(주 9) 27면.

19)

Symantec, Symantec Global Internet Threat Report — Trends for July-December 07 (Symantec Enterprise Security, 2008) Khaled El Emam & Luk Arbuckle, 2014, "Anonymizing Health Data", O'Reilly Media, Kindle Edition에서 재인용

20)

OECD, 앞의 보고서(주 9) 21면.

21)  
IMDA&PDPC, 앞의 보고서(주 6), 21면.

22)  
IMDA&PDPC, 앞의 보고서(주 6), 21면.

방법(재생산비용, reproduction cost)과 (2) 평가 대상 자산을 그대로 재생산 또는 재현할 수 없는 경우, 동등한 효용을 갖는 유사한 자산을 생산하는 비용을 기준으로 하는 방법(대체비용, replacement cost)으로 구분된다.<sup>21)</sup> 이를 데이터 가치평가에 적용하면, 해당 데이터베이스를 직접 구축하기 위해 현재 소요되는 비용을 기준으로 그 데이터의 가치를 평가하는 방식이 된다. 다만, 원가 기반 접근법은 해당 데이터를 통해 장래에 얼마나 되는 경제적 가치를 얻을 수 있을 것인지를 고려하지 못한다는 한계가 있다.

IMDA&PDPC는 원가 기반 접근법을 적용하여 데이터의 가치를 평가할 경우 다음과 같은 비용을 고려할 수 있다고 설명한다.<sup>22)</sup>

- 입력 비용 - 데이터 수집 장비의 구입, 설치 및 유지 관리 비용
- 소프트웨어 및 스토리지 비용 - 데이터 전송, 처리, 관리 및 저장에 필요한 도구 비용
- 인적 자본 비용 - 데이터 시스템과 아키텍처를 구축, 개발 및 관리하고 데이터 자산을 수집 및 처리하기 위한 인력 고용 비용
- 판매 및 마케팅 비용 - 잠재적 소비자에게 데이터를 마케팅하고 홍보하는 비용
- 기타 간접비 - 임대료, 법률 및 관리 비용, 재무 비용 등

## 2. 부채로서의 데이터 - 데이터 유출 위험

데이터는 잠재적으로 부채가 될 수 있다. 방대한 용량의 데이터를 저장·보호·접근·분석하는 데는 상당한 비용이 든다. 특히 저장된 데이터는 언제든지 사이버 공격의 대상이 될 수 있다. 데이터의 용량이 증가하면서 데이터 보안을 위한 비용도 증가한다. 데이터 보안을 위한 조치를 하더라도 개인정보의 유출이 발생할 위험을 완벽히 방지하기란 어렵다. 개인정보를 수집, 보관하는 기업의 입장에서는 개인정보 유출 사고(data breach) 발생 시 부담하게 될 비용을 기준으로 기업이 보유한 개인정보 데이터의 (잠재적 부채로서의) 가치를 추정해 볼 수도 있다.<sup>23)</sup>

예컨대, 2017년 미국 Equifax 사가 보유한 개인정보(이름, 생일, 주소, 사회보장번호 등) 1억 4700만 건이 유출된 사고의 경우, Equifax 사는 1인당 최소 125달러를 지급하도록 FTC와 합의한 바 있다. 국내에서는 흠플러스가 부정한 방법으로 개인정보를 수집하여 보험사에 판매한 행위 등에 대한 손해배상 책임을 인정하면서 정보주체에 대해 5만원~20만원의 지급을 명한 사례가 잘 알려져 있다.<sup>24)</sup>

하지만 정보주체에게 지급하는 손해배상액은 데이터 유출 사고로 인해 기업이 지출하는 비용의 일부에 불과하다. 기업은 많은 경우 상당한 벌금을 내야 하고(국내에서는 형사처벌의 위험이 따른다), 추가적인 보안 조치를 취하는 등의 비용을 지출해야 할 뿐만 아니라, 개인정보 유출 결과 특정 브랜드가 훼손되고, 해당 제품이나 서비스 전체에 악영향을 입을 수 있다. 예컨대 2011년 소니의 플레이

23)  
OECD, 앞의 보고서(주 9) 5면.

24)  
서울고등법원 2018. 8. 31 선고 2018나 2014586 판결

스테이션 네트워크의 개인정보 유출 사고로 7,700만 건의 이용자 정보가 유출되었고, 24일 동안 시스템이 중지되었다. 위 정보 유출 사고로 인해 발생한 복구 비용이나 및 벌금 등의 직접 비용은 미화 1억 7,100만 달러 상당으로 추산되었다(이용자 당 2.2 달러). 그러나 가입자 감소, 브랜드 이미지 손상 및 주가 하락 등으로 인한 간접적 비용을 포함한 전체 손실은 12억 5천만 달러 상당으로 추정되었다(이용자 당 16 달러).<sup>25)</sup> 이처럼 개인정보 유출 사고로 인한 비용의 범위를 어디까지로 정하느냐에 따라 개인정보의 가치 추산 결과가 크게 달라질 수 있다.

### 3. 데이터에 부가된 법적 제약의 고려 필요성

데이터 가치 산정 시에는 해당 데이터에 부가된 법적 제약도 고려될 필요가 있다. 예컨대 개인정보를 수집하면서 정보주체로부터 동의를 얻은 이용목적의 범위가 매우 제한적이라면 그 개인정보의 가치가 낮게 평가될 수 있다.

미국 시저스 엔터테인먼트의 도산 절차에서 문제 된 고객 빅데이터 가치평가 사례는 이러한 점을 잘 보여준다. 세계 최대 카지노·호텔 업체인 시저스 엔터테인먼트는 2014년 경영난과 과도한 부채 부담으로 인해 사업 관리·운영 자회사에 대해 미국 연방파산법 제11장 도산 절차를 신청하였다. 시저스는 1998년부터 Total Rewards라는 회원 프로그램을 시행하여 마케팅 및 고객 유지를 위해 활용해 왔고, 이에 포함된 회원 4,500만 명에 이르렀다. 월스트리트 저널은 시저스의 가장 가치 있는 자산은 고객 빅데이터라 보도하기도 하였다.<sup>26)</sup> 위 도산절차에서 채권자들은 위 회원 프로그램 데이터가 10억 달러 이상의 가치를 갖는다고 주장하였다. 그러나 도산 법원의 검사자는 다른 회사가 위 데이터를 취득하더라도 인수자의 기존 고객 데이터에 통합시켜 활용할 수 없으므로, 이를 분할하여 가치평가할 수 없다고 판단하였다.<sup>27)</sup> 그 회원 데이터는 시저스의 사업에 있어서는 가장 중요한 자산일 수 있으나, 제3자의 이용상 제약이 있는 경우 제3자에 대해 동등한 가치를 갖는다고 평가할 수 없었던 것이다. 이처럼 데이터에 대한 가치평가는 해당 데이터의 활용에 부가된 조건에 대한 법적 평가가 수반되는 것이라는 점을 유의할 필요가 있다.

## IV. 정보주체의 관점에서 본 개인정보 가치평가

개인정보의 가치를 평가하는 또 다른 방법은 정보주체가 자신의 개인정보를 얼마나 가치있다고 평가하는지 설문조사 또는 경제학적 실험을 통해 확인하는 것이다.<sup>28)</sup> 2000년대 이후 개인정보의 주관적 가치를 측정하는 다양한 경험적 연구가 진행되어 왔다. 개인정보의 주관적 가치를 측정하기 위해 자주 활용되는 방법은 개인정보에 대한 지불의사금액(willing to pay, WTP)와 수용의사금액(willing to accept, WTA)를 확인하는 것이다.<sup>29)</sup>

25)

van Lieshout, M., "The value of personal data," IFIP Advances in Information and Communication Technology, 2015, p. 8.

26)

The Wall Street Journal, 2015. 3. 19. "Real Prize in Caesars Fight: Data on Players"

27)

Short, J., and Todd, S., "What's your data worth?" MIT Sloan Management Review, Spring 2017

28)

OECD, 앞의 보고서(주 9) 29면.

29)

이러한 방법은 금전으로 평가하기 쉽지 않은 것의 가치를 측정하는데 활용할 수 있는데, 예컨대 오정일, "조건부가치 추정을 통한 징역형의 금전적 가치에 관한 연구, 형사정책연구 제29권 제3호(2018)은 위와 같은 방법을 적용하여 평가한 결과 우리나라 국민의 징역 1개월에 대한 지불 의사 금액을 평가하면 300~430만원 사이라고 한다.

30)

이하는 오정일, 앞의 논문(주 29)를 참조하여 정리한 것이다.

31)

Winegar, A. G., & Sunstein, C. R., (2019), "How much is data privacy worth? a preliminary investigation", Journal of Consumer Policy, 42(3), pp. 425-440.

32)

참고로 63명의 응답자들은 지불의사금액으로 월 25,000달러 이상의 금액을 응답하였고, 그 중 25명은 100만 달러 이상으로 응답하였는데, 이는 비현실적으로 높은 금액이므로 최대 25,000달러로 상한을 정하였다고 한다.

지불의사금액과 수용의사금액은 서로 차이가 있다.<sup>30)</sup> 가령 “손가락 하나를 잃을 위험을 제거하기 위해 돈을 얼마나 지불할 것인가?” 라는 질문은 손가락 하나에 대한 지불의사금액을 측정하는 것이고, “손가락 하나를 잃은 대가로 돈을 얼마나 받을 것인가?” 라는 질문은 수용의사금액을 측정하는 것이다. 일반적으로 수용의사금액은 지불의사금액보다 적게는 2배, 많게는 15배까지 높다. 사람은 자신의 소유물에 대해 일반적으로 더 높은 가치를 매기는 경향이 있기 때문인데, 행동경제학에서는 이를 소유효과(endowment effect)라고 한다.

개인정보의 지불의사금액과 수용의사금액을 측정한 최근 연구로는 하버드 대학의 Angela Winegar와 Cass Sunstein이 2019년 미국인 2,416명을 대상으로 설문조사를 진행한 사례가 있다.<sup>31)</sup> 위 연구에서 지불의사금액을 묻는 설문 문항은 “대다수 온라인 플랫폼(예: 페이스북, 구글, 기타 디지털 사업자)은 이용자 개인정보를 수집하는 것으로 알려져 있습니다. 당신은 당신의 개인정보를 보유한 모든 사업자로부터 이를 삭제하기 위해 매월 얼마(US 달러 기준)를 지불하겠습니까?” 였다. 이에 비해 수용의사금액을 묻는 설문 문항은 “대다수 온라인 플랫폼(예: 페이스북, 구글, 기타 디지털 사업자)은 이용자 개인정보를 수집하는 것으로 알려져 있습니다. 당신은 매월 얼마(US 달러 기준)를 받으면 이러한 사업자들이 당신의 개인정보를 얻는 것을 허용하겠습니까?” 였다.

위 연구는 설문의 대상이 되는 개인정보의 범위를 구체적으로 (1) 신원정보(이름, 나이, 성별, 직업, 소득, 주소, 사진) (2) 건강정보 이외의 민감정보(나이, 성별, 정치 성향, 종교, 성적 지향), (3) 건강정보(나이, 성별, 개인 특징, 신체적·정신적 건강 상태)으로 바꿨을 경우 응답이 어떻게 달라지는지 확인하였다. 위 연구에 따른 설문 응답 결과는 다음과 같다.<sup>32)</sup>

	지불의사금액	수용의사금액
일반적 개인정보	\$5/월	\$80/월
신원정보(이름·나이·성별·직업·소득·주소·사진)	\$5/월	\$50/월
건강정보 외 민감정보(나이·성별·정치 성향·종교·성적 지향)	\$5/월	\$77.5/월
건강정보(나이·성별·개인 특징·건강 상태)	\$5/월	\$100/월

[표 2. 개인정보에 대한 지불의사 및 수용의사 금액. 출처: Winegar&Sunstein (2019)]

즉, 미국 개인들이 온라인 플랫폼 사업자들이 보유한 개인정보를 삭제하기 위해 지불할 의사가 있는 금액(지불의사금액)은 월 5달러인데 비해, 위 사업자들이 자신의 개인정보를 취득한 것에 대해 요구하는 보상금액(수용의사금액)은 50~100달러 수준이었다. 위 연구는 이처럼 지불의사금액과 수용의사금액 사이에 10~20배에 이르는 큰 괴리가 있다는 점에 주목한다. 이는 개인정보에 대한 소유 효과가 매우 크다는 점을 시사하는데, 저자들은 이를 “super-endowment effect”라 표현하고 있다. 위 연구는 많은 이들이 평소에는 개인정보를 보호하기 위

한 노력을 크게 기울이지 않지만, 만약 개인정보가 유출되었을 경우 이를 심각하게 받아들이는 이유를 설명한다.

한편, 위 연구 결과 응답자들의 개인정보에 대한 지불의사금액이 월 5달러에 불과한 것을 두고, 사람들이 개인정보에 대해 크게 가치를 두지 않는 것은 아닌가 하는 의문이 제기될 수 있다. 이러한 응답은 행동경제학의 낙관 편향(optimistic bias)이나 현재중시 편향(present bias)로 설명될 수 있다. 즉 사람들은 자신이 제공한 개인정보가 오용되지 않을 것이라는 기대(낙관 편향)를 갖고 있기도 하고, 개인정보가 장기간 보유됨으로써 발생하는 미래의 위험성은 과소평가하거나 현재의 현금 지출에 대해서는 과대평가(현재중시 편향)하기도 한다. 이러한 행동경제학 이론은 설문 응답자들이 지불의사금액을 낮게 응답하는 현상을 이해하는데 도움이 된다.

한편, 개인정보의 주관적 가치에 대한 평가는 조사의 맥락에 따라 크게 달라질 수 있다는 점에 유의할 필요가 있다.<sup>33)</sup> 예컨대 정보를 수집하는 사업자에 대한 신뢰도에 따라 정보주체의 지불의사금액이나 수용의사금액은 크게 달라질 수 있기 때문이다.

## V. 정책결정자의 관점에서 본 개인정보 가치평가

데이터의 공공재로서의 성격을 갖는다. 정부, 공공기관, 사업자 등 여러 주체가 데이터를 활용하여 다양한 사회적 가치를 창출할 수 있다. 특히 정부가 공공 데이터를 구축하여 제공하는 경우 그 데이터를 활용하여 얻어지는 사회적 편익을 금전적으로 평가함으로써, 해당 공공데이터의 가치로 산정해 볼 수 있다.<sup>34)</sup>

예컨대 정부가 수행하는 인구조사(census) 통계 데이터는 자원 분배, 자본 투자 계획, 정책 수립 및 모니터링, 서비스 계획 수립, 학술 및 시장 조사, 통계 벤치마크, 선거 구획 결정 등 다양한 영역에 널리 활용될 수 있다. 뉴질랜드 통계청이 2013년 발표한 보고서에 따르면 뉴질랜드 정부의 인구조사 데이터의 현재가치는 총 10억 달러에 이른다고 추산한 바 있다.<sup>35)</sup> 한편 영국 국가통계청은 2011년 영국 인구조사 통계의 사회적 편익을 조사한 결과 연간 4억 9,000만 파운드 상당으로 평가한 바 있다.<sup>36)</sup>

데이터 활용의 사회적 가치를 평가하는 또 다른 방법은 데이터가 활용되지 못함으로써 발생하는 사회적 기회비용을 추산하는 것이다. 영국 교통 시스템에 관한 연구소인 Transport Systems Catapult의 2017년 교통 분야 데이터 활용에 관한 공동 보고서는, 영국에서 교통 데이터가 공개·공유되지 못함으로써 2025년까지 150억 파운드의 가치가 실현되지 못할 것으로 추정했다. 위 보고서는 이러한 가치가 실현되고 있지 못한 이유로 사일로에 갇힌(siloed) 사고 방식, 프라이버시, 보안 및 안전 위험, 데이터 공유에 따른 비용이 편익보다 크다는 생각 등이 라고 지적하였다.<sup>37)</sup>

33)

OECD, 앞의 보고서(주 9) 32면.

34)

Slotin, J., 앞의 보고서(주 14)

35)

Bakker, C (2014). Valuing the census.

36)

UK Office for National Statistics. The Value of Census Statistics.

37)

Transport Systems Catapult, (2017), "The case for government involvement to incentivise data sharing in the UK intelligent mobility sector", Briefing paper.

## VI. 결론

이상에서 검토한 바와 같이 데이터의 가치를 평가하기 위해 활용될 수 있는 다양한 방법이 존재한다. 기업의 입장에서는 전통적인 자산 가치평가 모델(시장 접근법, 원가 접근법, 소득 접근법)을 활용하거나, 데이터 보유 기업에 대한 가치평가 기준을 참고로 하거나, 데이터 유출 사고 발생시 발생할 수 있는 비용을 기준으로 삼을 수 있다. 정보주체의 관점에서 프라이버시에 대한 지불의사금액이나 수용의사금액을 측정하는 방법도 있다. 정책결정자의 관점에서는 데이터를 활용함으로써 발생시킬 수 있는 사회적 편익이나 데이터를 활용하지 못함으로써 발생하는 사회적 기회비용을 측정할 수 있다.

국내에서는 아직 ‘데이터 시장’이 활성화되어 있지 않고, 데이터의 ‘시장 가격’이라 부를 만한 것도 형성되어 있지 않다. 현재 일부 데이터 거래가 이루어지더라도 주로 개별 데이터셋마다 개별적 협상을 통해 가격이 결정되고 있는 상황으로 이해된다. 향후 데이터 시장이 성장할 경우 자연스럽게 데이터의 가격이 시장을 통해 형성될 수 있을 것이다. 이 글에서 소개한 여러 방법론이 데이터 거래시 그 가치를 산정하는 데 참고가 되기를 바란다.