커뮤니케이션 이론
2020년 16권 4호 5~45쪽
https://doi.org/10.20879/ct.2020.16.4.005

플랫폼 데이터 생태계에서 데이터 격차
: 디지털 불평등을 넘어

김미경 청운대학교, 미디어커뮤니케이션학과, 교수*

본 연구는 플랫폼 데이터 생태계에서 플랫폼 데이터 생산구조와 데이터 거래구조를 분석하면서 플랫폼 데이터 집중으로 야기될 격차를 검토하였다. 데이터의 자원화와 플랫폼과 데이터의 압축적 확장으로 플랫폼 데이터 빈자와 부자의 구조적인 불평등이 발생할 수 있는데, 본 논문은 플랫폼의 데이터 생성에 따른 데이터 접근격차, 데이터 통제격차와 이용자 간의 데이터 표현격차를 제시하였다. 능동적인 이용자의 데이터 표현과 수동적인 데이터 표현이 데이터의 재현으로 인한 편향적 정보생산의 문제를 지적하였다. 플랫폼 데이터 생태계의 네트워크 효과로 인해 데이터 접근과 통제, 표현의 격차는 편익과 위험의 두 상반된 측면에서 성과와 불평등을 야기하기 때문에 데이터 주권 교육의 큰 틀에서 데이터 리터러시 전략을 구상하도록 제안한다.

KEYWORDS 플랫폼 데이터 생태계, 데이터 접근격차, 데이터 통제격차, 데이터 표현격차, 데이터 리터러시

* mkqueen67@gmail.com

1. 서론

우리 시대의 가장 값진 자원은 오일이 아니라 데이터라고 한다(이코노미스트, 2017). 2016년 다보스 포럼 이후 화제가 된 제 4차 혁명이라는 개념은 디지털 세계, 생물학적 영역, 물리적 영역에 걸친 기술융합에 기반한 사회경제적 혁신을 예고하였고, 데이터는 더욱 중요해졌다. 데이터 기반 행정활성화에 관한 법률에 따르면, 데이터는 정보처리능력을 갖춘 장치를 통하여 생성 또는 처리되어 기계에 의한 판독이 가능한 형태로 존재하는 정형 또는 비정형의 정보로 정의한다. 데이터는 현실세계에서 단순히 관찰하거나 측정하여 수집한 사실(fact)에서 얻은 값(value)이다. 이에 비해 정보(information)는 데이터 수집과 분석을 통한 그래프와 해석처럼 상황설명을 위한 데이터이다. 즉 의사 결정에 활용할 수 있도록 처리하여 체계적으로 조직한 결과물이다.

정보는 실생활에서 의사 결정에 활용된다. 이러한 정보를 조직할 데이터를 어떻게 생성하는가와 표현되는가는 정보의 속성에 영향을 미친다. 데이터가 반영한 정보는 편향과 왜곡을 낳을 수 있다. 실례로 저널리즘 측면에서 데이터를 활용할 경우 이용자들은 데이터가 사실(fact)에서 얻은 값이므로 중립적인 것으로 간주한다. 그러나 데이터가 만든 정보가 순수한 질량을 가진다고 볼 순 없다. 데이터 과학자인 캐시 오닐(Cathy O´Neil, 2016)은 데이터가 편견에 사로잡힌 인간보다 공정하며 개인의 권리와 이익을 보호한다고 믿고 있지만 데이터 기반 알고리즘 모형은 인간의 편견과 차별, 오만을 코드화해 불평등을 확대하고 민주주의를 위협하고 있다고 단언하였다. 버너스-리(Berners-Lee, 2010)가 테드(Ted) 강연에서 고립된 개인이 데이터로 얻을 수 있는 성과와 데이터 수집가가 얻을 수 있는 성과의 차이에 대해서 언급한 것처럼, 데이터의 수립역량, 데이터 접근역량, 데이터 통제 역량의 격차에 의해서 발생하는 성과의 차이는 새로운 사회적 불평등을 야기할 것이다. 또한 데이터 활용에 의해 발

생하는 격차는 통합된 데이터베이스에서 발견할 수 있는 예측의 성과뿐만 아니라 데이터 감시(Data surveillance)와 검열(censorship)에도 영향력을 미칠 것이다(Ted, 2010. 3).

이에 데이터 분석과 활용이 보편화된 생태계가 구축되어 가고 있는 현실에서 사회전반의 개방화와 민주에 기여하는 측면도 있겠지만, 데이터 행위와 성과에 따라 예측하기 어려운 불평등도 낳을 수 있기 때문에 데이터 격차에 대한 개념을 설정해 볼 필요가 있다고 하겠다.

현재 데이터를 자원으로 하는 데이터 생태계가 구축되고 있다. 데이터 생태계는 데이터에 접근하고 활용할 수 있도록 협업하는 과정에서 데이터를 생산하고 인프라를 제공하고 연구조사를 행사하는 등 서로 다른 역할을 담당하는 구성원으로 이루어진다(European Commission, 2017). 데이터 생태계는 데이터의 생산, 분배, 소비를 둘러싼 일련의 순환체계이다. 데이터 생성자(Data Producer), 데이터 수집자/관리자(Data Aggregator/Data Custodian), 플랫폼 소유자(Platform Owner), 인사이트 공급자(Insight Provider), 데이터 표현자(Data Representer)의 역할이 선순환하는 체계이다(Opher, Chou, Onda, & Sounderrajan, 2016).

데이터 생태계의 자원인 데이터는 시민을 대상으로 하는 개인데이터와, 정부가 생산하는 공공데이터, 기업이 생산하는 산업데이터로 구분된다. 개인데이터는 의식주, 취미, 지식활동, 의사소통 및 관계형성에서 발생하는 것이며, 공공데이터는 개인, 치안, 외교, 남북교류, 의약품/식품안전, 방역, 국토, 자원, 교통항만 등 국가 인프라 및 환경, 국방, 법률에 관련된 것이다. 산업데이터는 제품 및 서비스의 생산, 판매, 마케팅 관련 데이터이며, 고객과 경영, 인사, 재무와 관련된 데이터이다. 그러나 전체 디지털 데이터 중 개인데이터 비율이 75%를 차지하고 있다(IDC Digital Universe study, 2011 Jan.).

이렇듯 개인 데이터의 비중이 커지고 경제·사회적 활용가치가 높아

짐에 따라, 개인이 정보 권리의 주체로서 스스로 자신의 데이터가 어디에서, 어떻게, 어떤 목적으로 사용될 지 결정하는 문제에 관심이 커지고 있다. 이것은 '데이터 주권(Data Sovereignty)' 논의의 출발점이 되기도 한다. '데이터 주권'이 현실화되기 위해선 데이터 생태계 각 협업부문에서 어떻게 불평등이 발생하는지를 인식해야 한다. 이를 통해서 '데이터 주권'이 실현될 수 있을 것이다. 특히 데이터 생성(Data Production)과 데이터의 표현(Data Representation) 차원은 정보의 불평등을 낳고 이에 따른 경제적이며 비경제적 성과의 불평등을 확대시킬 수 있기 때문에 집중적으로 검토할 필요가 있다.

데이터 생성(Data Production)의 차원에서, 데이터 접근(data access)과 데이터 통제(data control) 및 데이터 수집(data collection)의 격차가 발생될 수 있다. 이러한 소유권과 운영권의 격차는 데이터 소유권자와 비소유권자간, 운영자와 대상자간의 통제와 지배의 분화를 가속시킬 것이다. 데이터의 표현(Data Representation)의 차원에서 플랫폼 이용자 혹은 시민들의 인구사회학적 요인에 따른 데이터 관여(data engagement)와 경험(data experience)에 의한 표현의 행동력은 능동적인 데이터 표현시민과 수동적인 데이터 표현시민간의 격차를 낳을 것이다. 이러한 격차는 표현된 정보의 품질 뿐만 아니라 헤게모니적 정보 생산의 차이를 낳음으로써 정보의 불평등을 야기할 수 있다. 이렇듯 정보 불평등과 더 나아가 경제적이며 비경제적인 성과의 불평등의 잠재성에도 불구하고 데이터 산업의 육성적 측면에서 접근은 활발하지만 데이터 격차로 인해 발생할 불평등과 그 이후의 성과의 격차에 대해선 아직 이렇다 할 만 한 논의가 부족한 실정이다. 산업적 진흥 논의도 중요하지만 시민 기회에 대한 논의도 점검하는 것이 구조적 전체를 그려 볼 수 있기 때문에 본 연구는 데이터와 관련된 행위와 성과의 관계를 격차의 논의로 검토해보려고 한다.

본 연구는 디지털 격차 논의 지평 위에서 플랫폼 데이터의 역할에 주목한다. 플랫폼에서 이용자들은 서비스 수혜자일 뿐 만 아니라 데이터를

생성하고, 표현하고 거래하는 주체로서 힘을 갖게 된다. 플랫폼 이용자들이 표현하는 데이터가 어떻게 생성되는가에 따라 데이터가 그려내는 세계가 달라질 수 있으며 통제와 지배의 분화가 가속화 될 수 있다. 따라서 플랫폼의 데이터 생성과 표현의 행위를 중심으로 격차에 따른 불평등에 대해 문헌적으로 접근해보고자 한다.

이를 위해 플랫폼과 데이터 공존의 본질을 살펴보고, 디지털 격차에 의한 불평등의 순환구조에서 플랫폼 데이터 격차가 어디에 위치할 수 있는지를 살펴보고, 데이터 격차의 세부격차의 특징이 어떠한 지 검토하고자 한다. 그리고 마지막으로 데이터 격차를 통해서 발생할 수 있는 불평등의 지평을 열어보고자 한다. 또한 데이터 주권 인식 강화를 위한 데이터 리터러시를 시사점으로 제안하고자 한다.

2. 플랫폼 데이터 생태계와 작동의 원리

1) 자원화된 데이터의 플랫폼 압축과 확장

데이터 생태계의 프레임워크는 데이터 생태계에서 활동하는 주체들의 전반적인 역할과 역량 및 추이를 그리면서 만들어졌다. 데이터 주체들은 데이터 표현자, 인사이트 공급자, 플랫폼 소유자, 데이터 수집자와 데이터 관리자, 데이터 생성자등으로 구성된다(〈표 1〉 참고). 데이터 표현자(Data Representer)는 플랫폼 이용자와의 직접적인 접점을 형성한다. 플랫폼 이용자들이 플랫폼을 이용하면서 자발적으로 데이터를 표현하게 된다. 이때 데이터 표현자는 이용자의 자발적 관여에 의해 생성된 데이터를 이용하여 데이터를 표현하는 것이다. 따라서 데이터 표현이 이용자에게 어떤 편익과 위험이 있는지 교육시킬 필요가 있다. 인사이트 제공자(Insight Providers)는 수집된 데이터를 읽는 전문적 영역이다. 플랫폼 소유자(Platform Owner)는 하드웨어(HW), 소프트웨어(SW),

서비스 모듈(Service Module), 아키텍쳐(Architecture)와 같은 컴포넌트(Component)와 룰(Rule)을 소유하고 결정한다. 이를 테면 스마트폰 등의 단말기, 안드로이드 등의 운영체제인 컴포넌트와 네트워크 참여자나 플랫폼에 직간접적으로 관여하고 있는 이해관계자를 조율하거나 조정하는 규칙에 대한 소유와 운영권이 있는 자를 의미한다. 데이터 생산자(Data Producer)는 비즈니스 애플리케이션, 소셜미디어, 웹사이트, 오픈소스, 금융거래, 조사, 검열 및 디지털화된 하드 카피와 같은 범주와 임베디드 칩, 부착된 센서와 웨어러블, 휴대전화등의 사물인터넷 데이터 소스에서 데이터를 양산한다(Opher, Chou, Onda & Sounderrajan, 2016).

현재 데이터 생태계의 주체들은 하나의 부문이 아니라 여러 부문이 중층되고 중첩되어 가치를 만들어 내고 있는데, 대체로 선도적 기업은 다양한 층위에서 중첩되어 있다. 플랫폼 회사는 데이터의 생성, 데이터 분석, 데이터 표현을 수렴하고 있다. 따라서 플랫폼 데이터 생태계라고 해도 과언이 아닐 것이다.

표 1. 데이터 생태계 프레임워크

<table>
<tr><td rowspan="2">데이터 표현자
Data Representer</td><td colspan="2">사용자 인터페이스
User Interface</td><td colspan="2">사용자 경험
User Experience</td></tr>
<tr><td colspan="2">조사와 발견
Investigation and Discovery</td><td colspan="2">사용자 관여
User Engagement</td></tr>
<tr><td rowspan="2">인사이트 제공자
(Insight Providers)</td><td>통계방법론
Statistical & computational methods</td><td>분석개발환경
Development environment for analytics</td><td>알고리즘/로직/규칙
Algorithms/logic/rules</td></tr>
<tr><td>시멘틱 모델
Semantic model</td><td>분석 라이브러리
Analytics library</td><td>머신러닝
Machine Learning</td></tr>
<tr><td rowspan="2">플랫폼 소유자
(Platform Owner)</td><td colspan="2">개발환경
Development</td><td colspan="2">앱용 클라우드/호스트
Cloud/host for apps</td></tr>
<tr><td colspan="2">연계형 API
API for connectivity</td><td colspan="2">장치발견절차
Device discovery</td></tr>
<tr><td>데이터 수집자
Data Aggregator/</td><td colspan="2">데이터 표준화
Data normalization for common</td><td colspan="2">이종 데이터수집
Heterogeneous data collection from</td></tr>
</table>

<table>
<tr><td colspan="2">데이터 관리자
Data Custodian</td><td colspan="2">transmission</td><td colspan="2">disparate devices</td></tr>
<tr><td colspan="2" rowspan="2">데이터 생산
Data Producer</td><td colspan="2" rowspan="2">데이터 접근
Data Access</td><td colspan="2" rowspan="2">데이터 통제
Data Control</td><td colspan="2" rowspan="2">데이터 수집
Data Collection</td></tr>
<tr><td>IoT</td><td>Big Data</td></tr>
<tr><td colspan="8">조력자</td></tr>
<tr><td colspan="2">네트워크</td><td colspan="2">센서/칩</td><td colspan="2">분석엔진</td><td colspan="2">하이브리드 클라우드</td></tr>
</table>

출처: Opher, Chou, Onda & Sounderrajan(2016).

〈표 1〉에서 나타난 각 데이터 행위자들은 플랫폼 위에서 역할과 역량을 확대하고 있다. 첫 번째 유형은 수평적 확장(Move across the layer)으로 콘텐츠 역량을 강화한다. 예를 들어 피트니스 추적기를 만들어 스텝과 심박수를 모니터링하는 업체가 수면 모니터링을 추가하는 것처럼 콘텐츠 역량을 강화하는 것이다. 두 번째 유형은 수직적 확장(Move up or down the stack)으로 기능적 역량을 강화한다. 예를 들어 운송산업의 RFID/GPS 장치회사가 소비자 상호작용과 상호운용성을 증대시켜 날씨와 텔레마틱스와 같은 데이터를 취합하여 운송경로 최적화를 한다든지 연비를 개선하는 등 데이터 표현자가 데이터 수집가가 되기도 한다. 세 번째 유형은 압축적 확장유형(Compress or expand the stack)으로 플랫폼으로 데이터의 패권이 수렴된다. 플랫폼 소유자는 더 많은 데이터를 획득하여 패권을 획득한다. 예를 들어 구글이나 아마존과 같은 플랫폼은 이용자들이 데이터로 플랫폼 경험을 최적화하도록 한다. 플랫폼은 데이터 추출기구이기 때문에 기업은 아무런 생산수단을 소유하지 않더라도 플랫폼 만은 반드시 소유해야 한다.

플랫폼은 더 많은 이용자를 유인하고 더 많은 데이터를 양산함으로써 더 많은 주체들(공급자와 수요자)이 이용자들과 상호작용하도록 돕는다. 플랫폼에서의 공급자는 제품과 서비스를 개선하는데 이용자 데이터를 이용하고 제품과 서비스 판매를 통해 더 많은 데이터를 수집할 수 있는 것이다. 따라서 플랫폼과 데이터의 공존은 막대한 데이터를 독점하고 추출, 분석, 활용하여 자본주의 이윤을 낳지만 정작 이용자, 소비자, 노

동자, 광고주등 참여 행위자가 플랫폼을 지탱하는 힘이 되고 있음을 보여주고 있다. 플랫폼에 데이터가 많이 쌓일수록 수익이 생기기 때문에 플랫폼 데이터의 확장이 가속화될수록 시장 경쟁과 독점이 필연적이다. 따라서 플랫폼 데이터 확장의 본질을 확인함으로써 데이터 소유와 운영에서의 불평등이 어떠한 격차를 낳고 장기적으로 어떠한 성과의 차이를 양산할 것인지에 대한 문제의식을 낳게 한다.

2) 플랫폼 데이터 가치창출구조와 데이터 집중

플랫폼 데이터 집중 구조에서 플랫폼이 매개하는 서비스는 기업 소유의 알고리즘에 의해 관리되는 중앙 통제적 구조를 가지고 있다. 이는 언제든지 감시구조가 될 수 있다는 의미이다. 플랫폼 데이터의 확장형 역량이 강화될수록 데이터 수집, 관리, 운영의 권한이 독점화되고 개인정보보호 문제와 감시자본주의적 폐해도 발생할 수 있다.

일부에서 플랫폼의 가치창출 특성으로 인해 플랫폼을 중립지대로 상정하기도 하지만 약탈적 성격이 나타나기도 한다. 플랫폼은 양면시장(Two-sided Market) 이다. 양면시장은 제품 생산자(공급자)와 소비가가 다양하게 '연결'되는 '양방향'에 있으며, 한 플랫폼에서 만나 하나의 커뮤니티를 구성하고 그 안에서 서로가 서로를 도와가며 혁신과 가치창출에 기여하는 구조이다. 양면 시장은 플랫폼 내에서 다양한 판매자와 구매자 간에 상시적인 거래가 나타나기 때문에 네트워크 효과가 필수적이다. 네트워크 효과(network effect)란 특정 상품에 대한 어떤 사람의 수요가 클수록 해당상품의 가치가 높아져 보다 높은 이익이나 효용을 얻는 효과를 말한다. 충성도가 높은 사용자가 많다면 해당 플랫폼에 입점하려는 판매자 집단을 비교적 쉽게 확보할 수 있다. 이러한 플랫폼 시장의 매개적 특징으로 인해 코일(Coyle, 2018a, p.51)은 플랫폼이 기업이면서 시장이고, 생산과 교환 모두를 포함하는 것으로 인식되기 때문에 개념 정립이 어렵다고 한다. 또한 비정형의 온라인 공간에서 '연결자(connector)'

로서 모호하면서 중립적인 지위를 누리기도 하기 때문에(Napoli & Caplan, 2017), 독점력을 발휘하기 어렵다고 한다. 플랫폼은 전력회사처럼 경쟁자가 현실적으로 등장하기 어려운 '자연독점'도 아니고, 누구나 오픈된 네트워크 상에 서비스를 만들 수 있고, 이용자도 쉽게 다른 서비스로 옮겨갈 수 있기 때문에 독점의 폐단이 발생하기 어렵다고 한다. 그래서 시장 점유율을 이용하여 자사의 다른 제품을 끼워 팔거나 공급하여 경쟁사의 시장 진입을 막는 경우에 한해서만 반독점 행위로 처벌한다. 또한 플랫폼 안에서 공급자와 이용자는 플랫폼에 종속된 관계라고 볼 수 없고 계약을 통하여 자유롭게 진입을 할 수 있기 때문에 플랫폼 내의 공급자는 종속적 관계가 아닌 독립적 관계로 보기도 한다(이상기 외, 2018).

그러나, 플랫폼 데이터 가치창출은 순다라라잔이 규정한 바처럼(Sundararahjan, 2016). '군중 기반의 자본주의(Crowd-based Capitalism)[1)]'처럼 생산수단이 탈중앙화되는 것이 아니고 탈계급적인 것도 아니다. 플랫폼은 데이터를 방대하게 수집한다. 플랫폼은 서비스 제공자를 통제하며 더 경쟁력 있는 비즈니스를 제공할 수 있도록 알고리즘을 최적화한다. 수집된 데이터는 플랫폼에서 경쟁의 주요 수단이다. 방대한 양의 데이터를 수집하는 것이야말로 비즈니스 모델의 중핵이다. 플랫폼 데이터 비즈니스 모델이 성공하기 위해선 사업에 참여하는 이해관계자를 모으고 차별화된 가치를 제공하여 네트워크 효과를 창출해야 한다. 우선 검색 서비스, 소셜 네트워킹 서비스, 메신저 서비스, 인공지능서비스, 가상현실 서비스 등 편리하면서도 혁신적인 킬러 서비스를 제공해 소비자를 끌어 모은다. 소비자들에 서비스를 제공하는 대신 데이터를 입수한다. 이 데이터를 분석해 소비자들의 패턴을 예측한다. 데이터 분석 시

---

1) 군중기반 자본주의는 '사회적 관계가 중시되는 네트워크 기반 자본주의'를 정의하는 것으로 시장거래를 기반으로 자본의 영향력이 강화되고, 군중기반 네트워크, 사적·전문적 서비스 간의 경계 및 일과 여유시간 간의 경계가 모두 모호해지는 것을 특징으로 한다.

스템은 초기 구축비용이 높지만, 분석을 위한 한계비용은 낮고, 데이터 수집량이 많을수록 정확한 분석이 가능해진다. 그 결과를 토대로 더 많은 소비자들을 끌어당길 새로운 서비스를 개발하고, 그럼으로써 관련 시장을 장악한다. 데이터 경제로부터 창출되는 가치는 플랫폼에 귀속되고 감시와 통제를 통해 불평등을 가속할 수 있다.

표 2. 플랫폼 유형과 수익모델

플랫폼 유형	플랫폼 사업자	수익모델
온라인 마켓플레이스	이베이, 아마존	유료회원,리스팅수수료,판매수수료,페이팔수수료
앱스토어	구글플레이, 애플앱스토어	애플: 앱스토어 입점승인후 유료판매시 수수료 30%
공유플랫폼	우버, 에어비엔비	우버:드라이버 매출 0~30% 중개수수료
영상플랫폼	유튜브, 데일리모션, 넷플릿스	구독료, 광고, 솔루션
지불시스템	애플페이, 페이팔, 알리페이	시장독점, 수수료
검색엔진	구글	광고
특화된 검색엔진	엘프, 트립어드바이저	호텔엔조이닷컴:판매액의 7~20%수수료
위치기반	Tmap, 구글맵	구글맵 연동 검색광고 및 로컬광고
뉴스어그리게이션	뉴스나우(NewsNow), 허핑턴 포스트	광고
클라우드서비스	아마존 웹서비스, MS, 구글, IBM	라이센싱
소셜플랫폼	페이스북, 인스타그램	광고
메신저서비스	카카오, 라인	아이템, 캐릭터, 기프트콘 판매

출처: Langley & Leyshon (2017, p.6) , Table 1 재구성

〈표 2〉의 플랫폼 유형과 수익모델을 보면, 플랫폼은 데이터 수집의 중심에 서서 다양한 이용자의 참여를 통해 네트워크 효과를 증대하고 가치를 획득한다. 클릭 건수 당 발생하는 광고수입에 의존하는 검색엔진이나 소셜미디어와 같은 관심 플랫폼(attention platform)은 소비자들에게 개인정보나 관심을 서비스의 댓가로 지불하게 한다. 또한 소비자들의 클릭기록 또는 직접정보제공을 통해 수집된 개인 정보를 맞춤형 광고에 이용하여 새로운 소비자를 유인하고 더 높은 광고수익을 창출하기도 한다. 또한 판매자와 구매자, 고용인과 피고용인의 상호작용 플랫폼인 매칭 플랫폼(matching platform)은 거래과정에서 개인정보를 제공한다. 플

랫폼은 상호운용이 가능한 기술과 표준을 개발하고 플랫폼 매커니즘을 제어할 능력은 네트워크 효과와 전지구적 확산 잠재력과 결합하여 힘의 조건을 갖게 되었다. 메타데이터는 얼마 전까지만 해도 플랫폼 매개 서비스의 가치 없는 부산물로 여겨져 왔으나, 지금은 소중한 제품으로 채굴되고, 농축되고, 용도 변경할 수 있는 귀중한 자원으로 점차 변모해 왔다. 대부분 이용자는 생성한 데이터를 무료 서비스와 거래함으로써 데이터 흐름의 알고리즘 거버넌스를 가능케 했고, 수익모델로 변환했다(Graeff, 2017; van Dijk, Poell & De Waal, 2018).

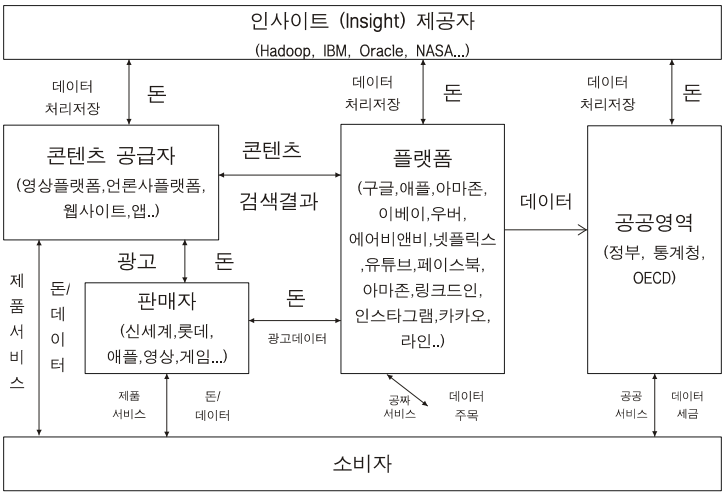

그림 1. 데이터 생태계의 행위자와 데이터 거래구조

출처: 공정거래위원회(2017). 해외경제정책동향, 제 124호

〈그림 1〉 데이터 행위자와 거래구조를 살펴보면, 반다익(van Dijk, 2014)이 지적한 바대로 플랫폼 생태계에서 플랫폼 소유자는 무료 서비스의 대가로 얻은 사용자의 메타데이터를 맞춤형 마케팅을 목적으로 하는 제 3자에게 데이터를 공유한다. 정량화된 사회적 상호작용을 동료 사용자, 회사, 정부 기관 또는 기타 플랫폼에 상관없이 제3자가 접근할 수 있

게 한다. 사회적 상호작용을 디지털로 변환함으로써 누가 누구와, 어느 위치에서, 그리고 얼마나 오랫동안 커뮤니케이션했는지 보여주는 데이터와 메타데이터의 가치에 기반을 둔 산업을 낳았다. 예컨대 페이스북(Facebook)은 "프렌딩(friending)"과 "좋아요(liking)"과 같은 사회적 활동을 알고리즘적인 관계(Bucher 2012; Helmond and Gerlitz 2013)로 바꾸었다. 트위터(Twitter)는 "팔로잉(following)"와 "리트윗(re-tweet)" 기능을 만들어 사람들의 온라인 페르소나를 대중화하고 아이디어를 홍보했다(Kwak et al., 2010), 링크드인(LinkedIn)은 직원과 구직자의 전문 네트워크를 디지털 인터페이스로 변환하였고(van Dijk, 2013), 유튜브(YouTube)는 시청각 콘텐츠의 일상적인 교환을 데이터화 했다(Ding et al., 2011).

네트워크 효과로 플랫폼 사업은 더 많은 데이터를 확보하고 알고리즘 개발에 의해 지배권을 두고 있는 승자독식의 독점시장을 형성한다. 반다익, 니보그와 포엘(van Dijk, Nieborg & Poell, 2019)은 플랫폼의 데이터 집중이 시장과 사회에 미칠 수 있는 잠재적 영향력을 설명하면서 다섯 가지 문제점을 제시하고 있다. 첫째, 데이터 집중에 의한 플랫폼 지배 사업자는 알고리즘 지식의 입력으로 데이터 흐름을 끝없이 재결합하고 재사용할 수 있는 잠재력을 갖게 된다. 둘째, 온라인 트래픽을 제어하기 위한 다양한 관문(gateway)으로 기능하고 관리할 수 있는 능력을 확보함으로써 데이터 통제력을 갖게 된다. 셋째, 광고주, 앱 개발자, 신문, 임시 근로자, 홈 공유 등 플랫폼 의존적 이해관계자인, "보완자(complementer)"와의 관계를 통제할 수 있는 잠재력을 갖게 된다. 넷째, 이용자가 모든 온라인 활동을 위해 점점 더 의존하고 있는 연결 인프라를 통제할 수 있는 역량을 갖게 된다. 다섯째, 사회적이고 민주주의적 기능을 방해할 수 있는 잠재력을 갖게 된다.

플랫폼의 데이터 생성과 축적 및 통제의 집중화 문제는 정보의 집중화 및 불평등의 심화의 문제가 발생할 가능성을 보이고 있다. 분명 데이

터 플랫폼의 역할이 집중화될 때 자유를 포함한 시민들의 기본권에 영향을 미칠 것이라 우려하지만 실제로 플랫폼이 무엇을 어떻게 하고 있는지 알기란 쉽지 않은 실정이다 (DeVito, 2017). 이에 플랫폼의 데이터 집중에 따른 데이터 빈자와 부자와의 격차로 발생할 불평등의 범주를 검토하고자 한다.

3. 디지털 불평등을 넘어 데이터 불평등 문제 설정

1) 데이터 순환에 따른 일상에서의 데이터 격차의 가시화

디지털 격차 연구는 디지털 미디어에 대한 접근과 이용에 초점을 두고 있다. 사회학은 자원, 각종 자본, 사회 참여 측면에서 사회적 불평등 문제에 관심을 강조한다. 심리학은 디지털 미디어를 이용하려는 태도와 동기를 다루고 컴퓨터 불안과 테크노포비아 현상에 대해 접근한다. 경제학은 기술 혁신의 확산과 성과로서 디지털 격차에 접근하고 교육과학은 정보나 디지털 이용 능력을 강조한다(van Dijk, 2017). 이러한 디지털 격차에 대한 연구는 디지털 리터러시를 구성하는 필수적인 개념이 되고 있다. 기술적 접근과 이용과 활용 역량의 차이로 인해 발생하는 사회적 불평등을 연구하고 경제적 차원과 비경제적 차원의 성과 차이를 진단한다. 이러한 불평등을 완화하기 위해 기술 친밀성 교육을 실시하고 디지털 서비스의 확산과 생산성 향상 및 다양한 사회적 편익 연구를 통해서 격차를 해소한다.

반다익(van Dijk, 2005)은 디지털 불평등이 물리적 접근이 완성된다고 해서 끝나는 것이 아니고 실제로 디지털 미디어의 이용이 일상생활에 통합될 때 디지털 불평등이 시작된다고 한다. 실제 일상에서 미디어의 이용과 활용이 어떻게 이루어지는가가 격차를 판단하는데 중요한 요소라고 한다. 마찬가지로 데이터 불평등도 데이터가 일상에서 편리함과 위험

함으로 깊이 개입될 때 격차가 드러난다고 할 수 있다.

디지털 격차가 데이터 격차로 확장되기 위해서 논의의 초점은 일상의 이용자들이 데이터의 편리함으로부터 배제되거나 위험함을 인지하지 못함으로써 스스로를 보호하지 못하는 '데이터의 이용과 활용'에서 접근해야 할 것이다. 데이터의 이용과 활용을 이해하기 위해 우선 데이터가 일상생활에서 어떻게 순환되어 나타나는지를 살펴볼 필요가 있다.

먼저, 일상생활에서 데이터의 순환은 이용자 피드백(user feedback loop)과 플랫폼 사업자 피드백(monetization feedback loop) 두 차원에서 이루어진다. 이용자 피드백(user feedback loop)은 이용자(user)의 데이터가 서비스의 품질을 개선하여 다시 이용자의 편익으로 돌아온다는 의미이다. 즉, 이용자가 적극적으로 데이터를 표현함으로써 일상의 편리함을 촉진할 정보와 서비스를 개발하게 된다. 이용자의 자발적 데이터로 만들어진 정보가 이용자의 일상의 편익을 증진하는 것이다. 만약 이용자들 간에 데이터 표현 역량이 있는 자와 데이터 표현 역량이 없는 자 간의 격차가 발생한다면 다량의 데이터로 표현된 정보만이 세계로 유통될 것이며, 보편적인 지식으로 확산되면서 다양성이 반영되지 못하게 될 것이다. 이용자 간 생성된 데이터의 양과 품질의 차이는 각 계층·지역·사회에 대한 정보의 차이를 낳고 나아가 삶의 질의 차이를 야기할 수 있다.

다음, 플랫폼 사업자의 피드백(monetization feedback loop)에서 보면, 데이터는 플랫폼 사업자의 수익창출 수단이다. 데이터 보유량은 플랫폼 사업의 결정적 요소가 될 수밖에 없다. 이에 플랫폼 사업자간 데이터 빈자와 데이터 부자의 격차는 승자독식의 불균형 구조를 심화시킨다. 데이터의 보유량이 증가할수록 학습량이 기하급수적으로 증가하는 현상은 장기간 지속되고, 학습 곡선이 평평해지는 지점에 도달하면 데이터의 규모가 너무 커져 승자독식의 상황이 발생한다(〈그림 2〉 참고). 데이터 통제와 지배권을 독식할 경우 데이터는 생산요소가 아니라 다른 생산요소를 압박하면서 시장경쟁을 좌우하게 될 우려가 있다. 플랫폼 사업자의

데이터 생성 역량의 격차에 의해서 데이터 빈자와 부자간의 격차는 심화될 수 밖에 없으며 데이터 지배권으로 인한 집중과 독점을 우려하게 한다.

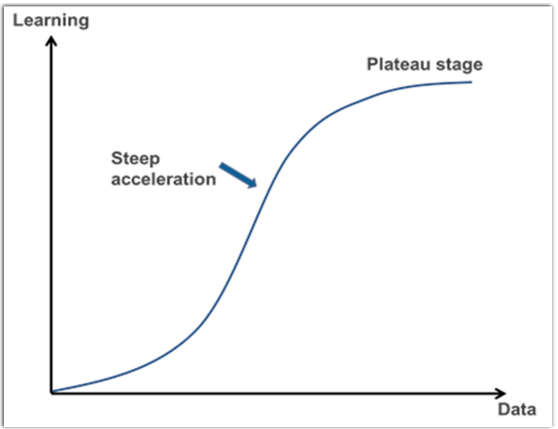

그림 2. 데이터와 학습량 관계

출처: 공정거래위원회(2017). 해외경제정책동향, 제 124호

2) 데이터 불평등, '참여'에 따른 성과의 차이

초기 디지털 격차연구는 디지털 미디어 하드웨어와 소프트웨어와 인터넷 연결과 같은 물리적 접근에 대한 것이었다. 이후 디지털 격차는 '접근을 넘어(beyond access)' 인터넷 이용의 질적 특성의 관점에서 다루어진다. 2단계 디지털 격차(second-level digital divide)는 자원동원이나 시민참여, 사회적 소속과 같은 사회적 이용(use)에 대해 다루며 이용역량의 강화를 제기한다(Hargittai, 2002). 물리적 접근을 넘어 디지털의 이용과 활용에 적극적으로 참여함으로써 불평등을 해소할 수 있다는 것이다.

이미 앞서 언급한 바 있듯이(p.8) 이용자 피드백에서 발생하는 데이터 불평등의 사회적 쟁점을 보면 개인이 데이터를 표현하고 데이터를 생성하는 참여적 행위는 매우 중요하다. 데이터 표현과 생성의 참여자와 비

참여자의 이분법적 차이는 불평등을 심화시킬 수 있다. 정적이고 영구적인 불평등으로 남기지 않기 위해 데이터 격차에 '참여'가 중요한 단계로 상정될 필요가 있다.

다음의 〈그림 3〉 자원전용이론 모델에서 데이터 불평등 극복을 위한 '참여'의 의미를 확인해보고자 한다. 반다익(van Dijk, 2005)은 먼저, 디지털 이용시 필요한 기술의 소유권인 '물리적 접근'과 호의적인 태도인 '동기적 접근'을 구분한다. 물리적 접근의 격차는 자원의 불평등한 배분과 관련되어 있으며, 나이·성별·지능·성격·능력·사회적 지위 등으로 설명된다. 동기적 접근의 격차는 사람들이 기술이 자신 삶과 무관하다고 생각하거나(Dutton & Helsper, 2007; Selwyn, 2006), 부정적으로 인식할 경우 동기유발이 되지 않음으로써 접근이 안되는 격차이다. 두 번째 격차는 디지털 이용능력에 대한 것이다. 물리적이고 동기적인 접근이 일반화되었다고 하더라도 효과적이고 효율적으로 사용하기 위한 필수적인 노하우(knowhow)가 없으면 불평등이 종결되지 않는다고 한다. 여기에서 노하우(knowhow)는 디지털 편익뿐 만 아니라 디지털 위험으로부터 개인을 보호하는 것 까지 포함한다. 두 번째 단계에 더하여 반다익(van Dijk)의 디지털 기술이 끊임없이 변화하고 발전하고 있기 때문에 '운영'과 '공식' 기능 뿐 만 아니라 '정보화'와 '전략적' 지향성 기술에 대해서 주목해야 한다고 강조한다. 구체적으로 정보를 찾고 평가하고 선택하고 특정 목표에 도달하기 위해 효과적으로 이용할 수 있음을 의미한다. 바로 세 번째 격차는 활용에 대한 것이다. 활용은 주도적인 '참여'를 통해서 얻게 되는 욕구 충족과 편익을 의미한다. 반 듀어슨, 반다익와 헬스퍼(van Deursen, van Dijk & Helsper, 2014)는 활용격차를 해소하는 것이 불평등을 극복의 최종 단계라고 한다.

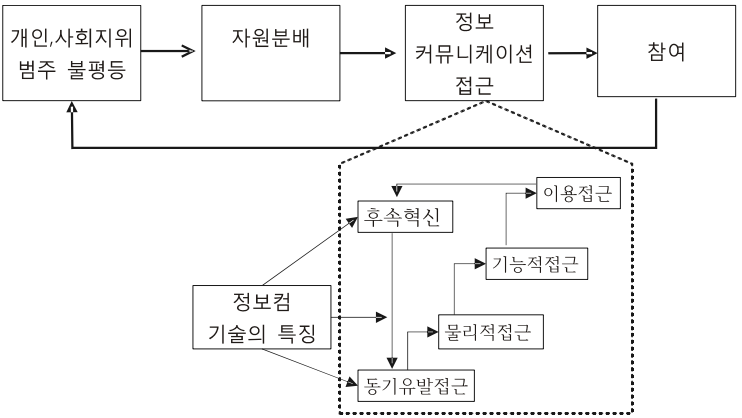

그림 3. 자원 전용이론 모델(A Causal Model of Resources and Appropriation Theory)

출처: van Dijk(2005, p.24) 재구성

데이터 표현과 생성에서 이용자와 플랫폼 소유자, 공급자, 소비자의 생태계 순환적 역할을 고려할 때 '참여'는 더욱 중요해지고 있다. 〈그림 3〉에서 반다익은 정보커뮤니케이션 기술의 특징에 접근하지 못하고 다룰 만 한 기능이 없고, 이용할 만한 역량이 없는 것에 대해 정보커뮤니케이션 접근의 불평등으로 설명하고 있다. 또한 이러한 정보커뮤니케이션 접근이 '참여'에 영향을 미치고 있는 것으로 나타나고 있는데, 결국 참여가 개인과 사회의 불평등을 야기하는 것으로 설명한다. 데이터 격차는 반다익의 자원전용모델의 정보커뮤니케이션 접근의 격차에서 더 나아가 '참여'로 지평을 넓혀 접근할 필요가 있다. 왜냐하면 플랫폼 데이터 생태계에서 시민의 자발적이며 유의미한 데이터 생성과 데이터 축적은 필수적이기 때문에 '참여'는 더욱 중요해진다.

네트워크 효과로 강화되는 플랫폼 데이터 생태계에서는 '참여' 격차의 개념이 중요해지고 있다. 플랫폼에서 습관적으로 행해지는 활동의 종류, 이용활동의 범위, 콘텐츠 생성 및 공유 등 '참여'에 따라 성과의 차이가 발생한다. 보다 특권적인 사회 경제적, 문화적 배경을 지닌 사용자들

은 더 많은 관련 데이터를 표현하고 창조적으로 활용함으로써 경제적이며 비경제적 성과를 낳는다. 정교한 형태의 플랫폼 참여를 하는 사람과 그렇지 않은 사람간의 생산하는 데이터와 정보의 차이를 낳게 되며 더 나아가 장기적인 정보의 편향을 만들어 낼 수도 있다. 예컨대, 남성들이 부지런히 유흥가에 있는 자신의 위치데이터를 표현한다면 그 장소는 유흥가로 검색된다. 지역적 정보의 왜곡이 확정되는 사례이다. 얼마나 많은 이용자들이 데이터 표현에 참여하는가에 따라 정보의 다양성이 약화되고 불균등한 정보가 강화되는지 관심을 가져야 할 것이다. 예를 들어, 플랫폼 소비자 참여의 수준에 따라 소셜리스트, 시장기피자, 조용한 안티 소비자, 수동적 회원등으로 구분한다면, 소셜리스트는 플랫폼에 적극적이고 완전한 참여를 수행하겠지만 시장기피자나 조용한 반소비자는 다른 이용자와의 연결과 공유에 관심이 적고 회피하게 된다. 또한 수동적인 행위자는 유의미한 수준의 참여 혹은 관여를 하지 않음으로써 플랫폼 데이터 생태계에서의 혜택을 얻지 못할 것이다(Ozanne & Ballantine, 2010). 여기에서 '참여'는 데이터 생성에 주체적이고 자발적인 일련의 행위뿐 만 아니라 데이터 생성 격차에 의해 발생하는 불평등을 인식하고 해결하기 위한 모든 활동도 포함한다.

이렇게 데이터 참여의 격차는 경제적 측면과 비경제적 측면에서 성과의 차이를 발생시킨다. 경제적 측면에서 데이터 접근성이 있는 사람들은 제품을 사고파는 데 있어서 낮은 가격뿐만 아니라 인터넷을 통해 일자리를 쉽게 구한다. 정치적 입장을 개진하기 위해서 데이터 관여도를 높이고, 문화적 데이터 참여는 취향과 맞춤형의 혜택을 얻게 한다. 제도적 측면에서 공공데이터의 참여는 공공복지뿐만 아니라 다양한 공공서비스의 혜택을 얻게 한다. 비경제적 측면에서 데이터 참여격차는 사회적 자본과 신뢰와 평판의 성과의 차이를 낳게 될 것이다.

4. 데이터 격차의 설정

1) 데이터 생성(Data Production) 격차 : 데이터 접근격차와 데이터 통제격차

'정보커뮤니케이션기술'의 변화가 새로운 불평등을 야기하고 기능성, 콘텐츠 가용성, 개방성 등의 새로운 문제를 낳는 것처럼 데이터에 접근할 수 있는 자와 분석·의미 추출을 할 수 있는 힘을 갖는자가 소수에 집중됨으로써 '데이터 불평등'이 현실화되고 있다(Andrejevic, 2014; Boyd & Crawford, 2012; McCarthy, 2016). 데이터 기반으로 정보를 생산하는데 있어서 데이터에 접근하는 자와 수집된 데이터를 분석 처리하고 배치할 수 있는 알고리즘 능력의 차이는 새로운 차원의 불평등을 야기할 것이다.

플랫폼 데이터 집중으로 인해서 데이터 부자와 데이터 빈자간의 격차가 발생한다. 데이터 부자는 대규모 데이터 셋을 생성하거나 구매 및 저장할 수 있는 기업 및 대학과 같은 조직이고 '데이터 빈자는 데이터, 전문성, 처리 능력에 대한 접근권에서 제외된 자들이다'(Boyd & Crawford, 2011). 미디어 이론가 레프 마노비치는 최근 데이터의 폭발과 연산 데이터 분석의 출현으로 인해 완전히 '새로운 종류의 격차가 만들어지고 있다고 주장'했다(Manovich, 2012, p.470). 이러한 격차로 인해 '데이터 계급(Data Class)'이 양산될 것이라고 한다. 그는 빅데이터 수집·분석이 가능한가와 빅데이터 (수동적인) 생산자인지에 따라 데이터 계급을 구분한다. 또한 보이드와 크로포드(Boyd & Crawford, 2011)는 데이터베이스 '가진자(have)'와 '가지지 못한자(have-not)'간의 데이터 격차가 발생하고 비대칭적인 분류 과정과 데이터의 정보화 및 지식화 적용과정에서도 데이터 격차는 발생할 수 있다고 한다.

이미 앞에서 설명한 바와 같이(p.4), 플랫폼 데이터 압축적 확장유형(Compress or expand the stack)은 플랫폼에 데이터를 최적화한다. 플랫폼 힘은 데이터의 생성과 맥을 같이하기 때문에 플랫폼의 데이터

생성 역량은 더욱 소수에 집중되며 네트워크 효과에 대해 승자독식의 경향은 강화되고 있다. 예를 들어 애플, 알파넷-구글, 마이크로소프트, 아마존, 페이스북은 많은 사람들이 쓰면 쓸수록 엄청난 데이터가 쌓이게 되고 플랫폼의 힘은 강화된다. 데이터 집중으로 인해 데이터 집중으로 인해 빅 5는 새로운 비즈니스 기회를 포착하기 쉽다. 또한 구글과 애플은 데이터 분석을 통해 자신들의 서비스를 적용하거나 개별 행동을 분석해 서비스를 개선할 수 있다(Farhad Manjoo, 2016.1.20). 플랫폼의 데이터 집중은 데이터 수집과 대표성 뿐 만 아니라 데이터를 통제할 수 있는 알고리즘의 활용에 영향을 미치며 결국 '데이터 생성'에 영향을 미친다(Boyd & Crawford, 2011). 데이터 생성은 데이터 접근, 데이터 통제, 데이터 수집의 세부적인 작업으로 이루어지는데, 플랫폼의 데이터 생성과 관련하여 데이터 접근과 데이터 통제에 의한 불평등이 심화되고 있다.

(1) 데이터 접근 격차(Data access divide)

데이터 접근격차(Data access divide)는 데이터에 접근하지 못함으로써 발생하는 격차이다. 데이터 기록을 하지 못하고 접근하지 못함으로써 데이터를 가용력 차이를 만드는 격차이다. 데이터 접근(Data access)은 데이터를 확인하는 능력이다. 법률적으로 액세스권(Right of Access)은 열람의 의미로 사용된다. 열람은 데이터를 조회한다는 의미이다. 그 이상의 의미로, 데이터를 처리하는 목적, 데이터를 제공하는 곳, 데이터의 보관주기 등 데이터 주체의 데이터 관련 사항들을 확인하는 것을 포함한다. 데이터 접근은 단순히 데이터를 조회하는 것이 아니라 데이터의 효용성 맥락을 이해하는 능력이 있는가에 대한 것이다. 즉 데이터 접근격차는 데이터를 이해하는 능력이 있는 자와 없는 자의 격차이다.

데이터 접근격차는 플랫폼의 방화벽과 같은 기술적 시스템에 의해 데이터에 접근하지 못함으로써 발생한다. 데이터 접근 가능한 자와 그렇지 못한 사람간의 성과의 차이는 지속적으로 증대할 수 있다. 또는 개발

도상국가처럼 데이터 접근이 원천적으로 불가능한 경우에서도 발생한다. 데이터 접근역량은 경제적이며 사회적 성과의 차이를 낳는다.

데이터에 접근 가능한 자와 그렇지 못한 자의 격차의 이해를 위해 페이스북 데이터 접근한 기업들이 유리한 위치를 차지한 사례를 제시할 수 있다. 페이스북은 개인의 이름, 이름, 성별, 생년월일, 거주지, 친구, 사진, 페이지링크등의 데이터에 접근할 수 있는 플랫폼이다. 페이스북은 수집한 개인데이터를 61개의 앱(app) 개발 기업이 접근할 수 있도록 허용하였는데 데이트 앱(app)부터 택배업체까지 여러 업체가 이 데이터에 접근함으로써 수익모델을 공고히 하였다. 또한 2016년 미국 대선 당시 도널드 트럼프 캠프는 이용자 데이터에 접근하여 유리한 선거운동을 실행하기도 하였다. 데이터에 대한 접근 격차는 자원을 유리한 지위에서 이용함으로써 성과와 이득을 발생시키게 한다.

개발도상국가처럼 원천적으로 데이터 접근이 불가능한 경우는 데이터가 사회경제적 불평등을 증폭시키는 중요한 요소가 된다. 데이터로 기록될 수 있는지, 없는지에 따라 불평등의 수준이 결정되기도 한다. 가난한 국가는 아직도 개인데이터에 접근하지 못한다. 가장 취약한 집단은 여전히 개인 데이터로 기록되지 못하고 있다(ID4D, 2016). 개인 데이터로 기록되지 못하는 계층은 여성, 아이들, 시골과 외딴 지역에 사는 사람들, 난민과 이주민들, 노숙자와 유목민들, 인신매매를 당하는 사람들이다. 데이터 빈곤층에 속하는 사람들은 출생, 사망, 사망원인에 대한 등록은 고사하고 공식적인 기록도 없다. 이렇게 시민권이 등록이 부족한 사회에서도 데이터로 기록이 되는 사람들은 사회경제적 지위가 높거나 시민권의 특권을 부여받은 집단에 속한다. 데이터 접근격차는 데이터 인프라와 데이터 가용성의 차이에 따라 발전의 궤적 차이를 낳는다(The Economist, 2014). 데이터 접근 격차의 영향을 받고 있는 특정 집단을 지명해서 데이터 불평등의 증거를 제시하는 것은 어렵지만, 분명한 사실은 세계적 지리적 불평등이 존재한다는 것이다. 예를 들어, 유럽인의 92%는 완전한

시민 데이터로 등록되지만 아프리카인 중 단지 5%만이 시민 등록되었다(Mahapatra et al., 2007). 데이터 인프라의 견고함과 데이터의 가용성은 국가의 발전과 시민사회의 강화 및 민주주의를 위해 매우 중요하다. 저개발 국가처럼 데이터가 증거로서 제시되지 않는다면 어떤 투자자도 저개발 국가와의 협력을 선호하지 않을 것이다.

한편, 플랫폼의 데이터 집중과 독점으로 데이터 접근이 불가능하면서 데이터 격차가 발생할 수 있다. 이와 같은 데이터 접근 격차를 해결하기 위해서 개인데이터의 무분별한 이용에 의한 폐해를 예방하고 시정하려고 노력하고 있다(수정헌법안 22조 내용). 이를테면 '개인데이터 자기결정권' 또는 '자기정보관리통제권'등은 데이터 접근권의 불평등을 시정하기 위한 것이다. 자신에 관한 정보가 누구에게 어느 범위까지 알려지고 어떻게 이용하게 할 것인지를 정보주체 스스로가 결정할 수 있는 권리를 통해 불평등을 시정하려고 한다. 또한 데이터 접근격차를 줄이기 위해서 '개방형 데이터'를 제안하기도 한다. 연구자와 공공, 민간 기업, 시민사회 및 개발단체들이 과학 및 정부 데이터셋(data set)에 접근할 수 있도록 한다.[2] 개발도상국과 같이 원천적인 '데이터 빈곤'에 있는 국가가 디지털 시스템(웹, 휴대폰네트워크)과 센서(원격감지위성, GPS)에 의해 만들어진 데이터 부산물을 사회경제적, 인구통계학적, 위치와 이동 데이터로 이용하도록 함으로써 격차를 해소하기도 한다(Gurstein, 2011; Kitchin, 2014).

---

2) 베데스다(Bethesda, 2002)와 베를린(Berlin, 2003) 성명, 판톤 원칙(Panton Principles, 2009년)과 같은 개방 자원(Open Science)에 대한 주요 선언문은 공통의 열망을 표현하고 있다: (공적으로 자금을 지원한) 과학적 결과들은 출판물에서 데이터까지 다양한 형태로 제공되어야 하며, 모든 사람들이 자유롭게 이용할 수 있어야 한다는 것이다. Open Data와 관련하여, 팬톤 원칙은 과학이 효과적으로 기능하고, 사회가 과학적 노력으로부터 완전한 이익을 거두기 위해서는 과학 데이터를 개방하는 것이 중요하다고 명시하고 있다. 미국 국가 CODATA 위원회(1997: 10)가 다음과 같이 선언한 보고서 Bits of Power에서 이와 유사한 감정을 표현했다. 데이터의 가치는 그들의 사용에 있다. 과학 데이터에 대한 완전하고 개방적인 접근은 공적 자금 지원 연구로부터 도출된 과학 데이터의 교환을 위한 국제 규범으로 채택되어야 한다.

(2) **데이터 통제 격차**(Data control divide)

데이터 통제격차는 디지털 격차 개념에서 쉽게 포착되지 않는 핵심적 개념이다. 데이터 격차는 빅데이터 마이닝(Big data mining)시대에 특정 이익과 목적으로 플랫폼의 가용성 데이터 흐름을 결정하는 기술이 집중됨으로써 발생되는 격차이다(Andrejevic, 2017; Cinnamon, 2019). 데이터 통제격차는 플랫폼 회사가 개인행동 데이터를 활용하는 주체가 되고 개인은 이용되는 객체가 되면서 발생하는 비대칭적 격차이다. 이러한 비대칭적인 격차를 알면서도 플랫폼의 개인정보 보호정책과 감시전술에 저항할 수 있는 이용자의 능력은 상당히 제한적이다. 왜냐하면 개인은 시스템의 상호의존성과 복잡성에 대한 통찰력을 얻기 매우 어려운 반면, 플랫폼의 기술과 비즈니스 모델에 의해 운용되기 때문이다(Hartzog & Selinger, 2013; Mager, 2012).

플랫폼은 누가, 어디에, 언제, 어떤 목적으로 데이터를 이용할지에 대한 흐름을 통제함으로써 개인과 공동체, 국가에 부정적 영향을 미칠 수 있다. 데이터 흐름의 주요 통제기제는 알고리즘이다. 알고리즘은 사람이 개입되지 않고 데이터의 흐름을 잡아주기 때문에 통제와는 거리가 멀 것이라고 기대하지만 편향의 사례는 속속히 발견되고 있다. 소셜 미디어 플랫폼 메타데이터는 인간의 현재 행동을 반영하는 것으로 간주하지만, 구글과 트위터 및 다른 사이트들이 채택한 알고리즘은 본질적으로 선택적이고 조작적이다. 디아코포우로스와 샘마(Diakopoulos & Shamma, 2010)는 트위터(Twitter)를 통해 토론 성과를 분석해 정치적 선호를 예측했을 때, 실시간으로 트위터 토론에 영향을 미칠 수 있는 스핀 닥터나 당파적 2인자의 가능성을 무시한다고 했다. 아마존의 추천 알고리즘은 소비자 습관의 특정 패턴을 감지하려 성공적으로 수익을 창출하는 주요 마케팅 전략을 구사한다(Andrejevic, 2011). 소셜 미디어는 인터넷 검색과 마찬가지로 개인화 및 맞춤화가 적용되어 특정 청중이나 개인에게 메시지를 조정한다(Bucher, 2012; Pariser, 2011). 넷플릭스와 같은

OTT서비스는 사용자에 대한 정보를 수집하고 그렇게 수집한 정보를 활용하여 연관이 있거나 관심이 있을 만한 콘텐츠를 추천해 준다. 수집영역에 검색정보와 시청자 평가는 기본이고 시청자가 어느 부분에서 일시정지하고 어떤 부분에서 되감기를 하는지, 심지어 영상의 색감이나 음량까지 집계한다. 더 나은 서비스를 제공한다는 명목 하에 수집된 개인의 정보가 오로지 서비스 제공하는 플랫폼의 통제 하에 있는 것은 매우 위험한 데이터 통제격차를 보여주는 것이다.

그런데 이런 알고리즘 통제가 인종, 성별, 거주지 등으로 차별적이거나 부정확한 판단을 내린다는 사실이 불평등을 야기시킬 수 있다. 예컨대 구글 애드센스(AdSense)는 인종을 드러내는 이름을 사용하여 검색한 흑인 이용자에게 범죄기록 조회광고를 25%나 더 노출했으며, 고임금 직업관련 광고는 여성보다 남성에게 6배나 많이 노출되었다. 또한 의도적으로 가격 담합을 하도록 설계한 알고리즘은 소비자들의 가격 선택권을 박탈하는 도구로 사용되기도 한다(최은창, 2017). 알고리즘이 편향적 사고방식을 강화하여 정치와 공공영역을 양극화할 수 도 있다. 즉 개인행동 데이터는 알고리즘에 의해서 사회적 차이를 세밀하게 구성하여 프로파일링 하는데, 이를 기반으로 편향된 정보를 제공된다. 사람들은 뉴스와 정보를 소수의 플랫폼에 의존함으로써 노출의 다양성은 약화되고 양극화를 심화시킨다(Allcott & Gentzkow, 2017). 플랫폼 노동의 확산으로 데이터와 알고리즘의 통제가 가속화되고 있는데, 알고리즘이 결정하는 기준을 알 수 없다는 점에서 컨베이어벨트와 같은 알고리즘에 더욱 종속될 수도 있다(주영재, 2020.02.15.).

알고리즘 설계에 의한 데이터의 통제는 사회적 약자를 차별하고 경제적 불평등을 악화시킬 수 있는 위험이 있다. 알고리즘 데이터 통제는 구체적으로 알고리즘 규칙을 설계하는 자와 그것에 의해 영향을 받는 자의 문제이다. '주인-대리인(principle-agent)'으로 알고리즘 개발 문제가 발생할 수 있다(주영재, 2020.02.15.). 알고리즘을 이용하는 자가 플

랫폼이고 알고리즘 개발사가 대리인이라면, 대리인은 플랫폼의 선점과 이익을 위해서 복무할 수밖에 없을 것이다. 대리인이 이기적인 동기에 따라 행동하고 도덕적 해이를 보인다면, 데이터로 표현되는 정보의 편향은 가속화되며 정보의 비대칭은 심화될 수밖에 없다. 알고리즘이 데이터를 기반으로 정확한 예측을 하더라도 결정적 판단이 도출된 원인과 이유를 말해주지는 못한다. 문제가 발견되어도 그 원인을 쉽게 알아내거나 수정하기 어렵기 때문에 현실적으로 위험을 직면하게 한다. 알고리즘의 불투명한 설계 또는 알고리즘을 훈련시키는 데이터 셋(data set)에 숨어있는 편향과 오류는 사회적 차별과 배제를 조장할 수 있다. 이로 인해 알고리즘의 주인인 플랫폼은 데이터의 게이트키퍼로서의 권한을 갖게 됨으로써 (Taylor & Broeders, 2015) 사회적 제도를 잠식할 수도 있다.

데이터 통제 격차에서 우위권을 점유한 알고리즘 권력(algorithm power)이 시스템과 제도를 잠식할 가능성에 대해 진술하였지만 그 격차를 해결하기는 쉽지 않아 보인다. 베두치 외(Beduschi et al., 2017)는 구조화되지 않은 개인 데이터를 대량으로 운용하는 복잡한 분석 프로세스가 불투명하기 때문에 규제가 적용되기 어렵다고 했다. 그럼에도 불구하고 알고리즘 작동에 대한 투명성과 책무성을 구현하고 통제 가능성을 확보하는 것은 데이터 통제 격차에서 파생될 위험을 해결하기 위한 필수적인 조치이다. 이에 최근 유럽에서 새롭게 추진되는 법안엔 추천 알고리즘을 공개하고, 소비자의 알고리즘 선택권을 보장하는 내용이 담겨져 있다(정원엽, 2020, 11.18).

2) 데이터 표현 격차(Data representation divide)

데이터 표현 격차(data representation divide)는 데이터를 능동적으로 표현하는 자와 표현하지 않는 자의 격차이다. 데이터 표현은 이용자 개인의 영역이다. 이용자 개인의 데이터 표현 역량의 차이에 따라 발생하는 격차이다. 데이터 표현(data representation)은 데이터를 이용하고

사용하는데 가장 기본이 되는 첫 과정이다. 데이터를 다루고 실제적으로 사용하기 위해서는 데이터를 어떠한 형식으로 표현할 수 있어야 한다. 사회과학적 의미로서 데이터 표현은 이용자 개인의 데이터 관여(data engagement)와 데이터 경험(data experience)과 관련된 것으로, 적극적으로 표현하는 자와 표현하지 않는 자의 격차이다. 적극적으로 데이터를 표현하는 자의 정보는 드러나고, 표현하지 않는 자의 정보는 드러나는 않는다. 따라서 능동적 사용자가 제공한 데이터에 의해 지배적인 관점이 재현되고 수동적인 사용자의 데이터는 정보로 전환되지 못하고 사장될 수 있다.

디지털 데이터에서 개인 데이터가 차지하는 비중이 약 75%라고 한다(IDC Digital Universe study, 2011 Jan.). 개인 데이터는 인터넷, 소셜 미디어, 휴대폰, 스마트기기, 인프라(대중교통, 건물), 신용카드나 직불카드를 이용하는 모든 사람들은 소비자 활동, 선호도 및 욕구, 사는 장소, 일하는 장소, 노는 장소에 대해 상호작용함으로써 데이터를 생성한다(Cinnamon, 2019). 보통, 데이터를 표현하는 사람들은 자신들의 '주제'의 윤곽에 맞추어 그것을 형성하고 세계와 사회에 실질적인 영향력을 미친다. 데이터 표현은 데이터를 둘러싼 조직, 기술 및 문화세계를 재편성한다. 따라서 데이터를 적극적으로 표현하는 사람들은 그들의 관점을 드러낼 수 있지만 데이터를 소비만 하는 사람들은 데이터 표현하는 적극적인 이용자의 영향 아래 놓일 수밖에 없다(Ribes & Jackson, 2013, p.147). 예컨대 '크라우드소싱(crowd sourcing)에 의해 생산된 '이용자 생산 데이터'의 경우를 살펴보면 이용자 데이터 표현의 격차를 쉽게 이해할 수 있다. 구글 맵 메이커(Google MapMaker)와 오픈스트리트맵(OpenStreetMap)의 협업 매핑 프로젝트에서 크라우드소싱 데이터를 생산했다. 그런데, 대부분 남성 지원자에 의해 데이터가 제공되어서 남성 중심의 시설(스트립클럽)의 데이터는 풍부하게 제공되었지만 여성 중심의 시설(돌봄)등은 데이터화되지 않았다. 특정 남성 중심적 관점이

크라우드소싱 데이터베이스를 통해 더 넓은 지역으로 전파되기도 한다(Stephens, 2013). 위키피디아(Wikipedia)의 예에서도 이용자의 참여에 의한 데이터의 표현은 편향적이라는 지적이 일고 있다. 적은 비율의 열성적인 편집자(슈퍼편집자)들이 위키의 편집에 영향력을 끼치고 이들의 편집은 편향적인 지식정보를 낳고 있다. 예컨대 위키의 많은 정보는 북반구(Global North)에 한정되어 지역적 편향성을 보이고 있다. 또한 젠더 편향을 보여주는 사례도 나타난다. 위키피디아의 자발적 편집자의 빈도를 조사한 결과, 여성은 20% 미만으로 나타났고, 사용자 모임의 공헌자도 여성은 9%에 불과한 것으로 나타났다. 또한 위키백과의 인물정보 중에서 여성 인물정보는 15%에 불과한 것으로 나타났다. 이렇듯 남성중심의 편향된 디지털 표현은 세계에 대한 우리의 이해에 영향을 미친다. 일부의 경우에는 새로운 정보 엘리트가 정보의 세상을 지배할 수도 있음을 경고하기도 했다(Shaw & Graham, 2017, p.913). 자발적인 편집 참여자의 대부분이 북반구에 있고 남성으로 편향되면서 지식정보의 관점의 편향화될 수도 있다.

데이터 표현 격차는 권력구조를 재현하고 있다. 데이터를 표현하는 사람들은 대부분 남성이고 일반적으로 부유하고, 교육받고, 사회에서 영향력 있는 집단으로 나타났다. 인종과 경제적 수준에 따라 사회적인 격차를 강화한다(Stephens, 2013). 이러한 데이터 표현의 격차를 극복하기 위해서 크라우드소싱(crowdsourcing)된 데이터의 편향을 식별하고 수량화하기 위한 통계방법을 개발하여 '관리된 군중(managed crowd)'의 접근방식을 제안하기도 한다. '관리된 군중(managed crowd)'은 '사회적 관계가 중심이 되는 네트워크 기반 자본주의'에서 데이터 표현의 편익과 위험을 교육시키고 구조적 인식을 증진시킴으로써 데이터 표현 리터러시를 해결할 개념이 될 수 있다. 예컨대 우리나라의 '마이데이터(My data)'는 정보주체인 개인이 본인의 정보를 적극적으로 관리, 통제하고 이를 신용관리, 자산관리, 나아가 건강관리까지 개인생활에 능동적으로

활용하는 일련의 과정이다. 개인 데이터의 표현이 위험한 수준에서만이 아니라 편익적 차원에서도 활용될 수 있도록 뒷받침하는 제도이다. 이렇게 개인의 데이터 표현에 대한 구조적 문제를 인식시켜 동의를 구축하는 것은 데이터 표현과 수집에 긍정적인 역할을 할 것이며 데이터 표현격차를 완화시킬 수 있을 것이다.

3) 데이터 참여행위의 단계적 흐름(cascade)과 성과의 격차

반다익(van Dijk)의 자원전용이론 모델(〈그림 3〉 참고)의 '정보커뮤니케이션 접근'에서 새로운 '정보커뮤니케이션의 특징'에 따른 격차의 단계적 흐름이 선형적으로 제시되고 있다. 데이터 커뮤니케이션 기술의 특징은 사회적, 심리적, 경제적 요인 이외에도 실용주의적이고 복합적인 이유로 행위성과에 영향을 미칠 수 있다(Selwyn, 2004, p.348). 데이터 격차는 단순히 인터넷 보유자와 비보유자의 이분법적인 차이를 넘어서 인터넷 접근뿐만 아니라 기술사용 및 결과를 통합하는 다면적인 개념이다. 플랫폼에서 데이터 참여의 행동 패턴이 경제, 정치, 문화, 공간 이동성, 사회 제도, 사회 네트워크, 지역 사회등의 이익에 영향을 미치고 있다.

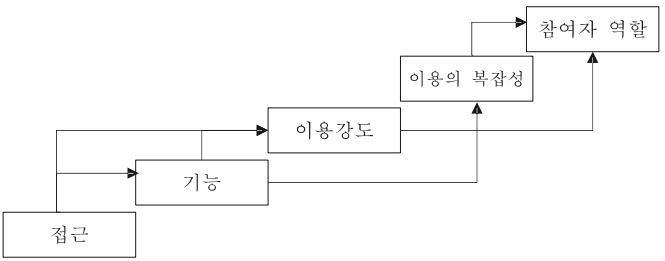

그림 4. 데이터 참여 행위의 단계적 흐름

출처: Andreotti et al., (2020, van Dijk, 2005에서 기반).

반다익은 정보기술의 특징에 따른 물리적, 동기유발적, 기능적, 이용적 접근이 후속혁신에 영향을 미치는 순환관계를 통해 참여가 촉진되는 것으로 설명하였다. 정보기술의 특징에 따른 격차의 차원은 접근격차, 이용격차, 활용격차로 구분되기도 하였다. 〈그림 4〉에서 제시한 플랫폼 데이터 생태계에서 데이터 기술에 의한 격차의 단계적 흐름은 이용과 활용뿐만 아니라 이용자의 적극적인 참여도 강조하고 있다(Andreotti et al., 2020). 플랫폼 데이터 생태계는 소비자, 공급자, 소유자의 다중행위자의 환경이기 때문에 행위자들의 적극적인 참여는 플랫폼을 활성화시킨다.

플랫폼의 소비자, 공급자, 소유자의 생태계의 이용자가 플랫폼 접근성을 확보한다는 것은 디지털 매체의 소유, 인터넷 근접성이 높고 테크노포비아등의 동기성이 낮지 않았을 때 가능할 것이다. '기능'은 플랫폼을 이용할 수 있는 기능을 갖추었는지에 대한 것이다. 이용강도와 이용의 복잡성, 참여자의 역할등은 활용적 측면으로 이해할 수 있다. 플랫폼 데이터의 이용강도와 이용의 복잡성은 플랫폼의 '활용'의 격차를 낳을 것이다. 참여자의 역할은 데이터의 생성과 표현에 영향을 미칠 수 있다.

데이터 격차의 단계적 흐름에서 볼 때 물리적이며 동기유발적 접근성은 기능에 영향을 미친다. 어떻게 이용할 것이며 얼마나 이용할 것인가에 대한 이용강도에 영향을 미치기도 한다. 플랫폼 데이터를 이용할 수 있는 기능은 얼마나 이용할 것인가를 결정한다. 또한 기술의 복잡성에 대응한 전문적인 활용에 영향을 미친다. 플랫폼 데이터의 이용 강도와 이용의 복잡성은 참여자의 역할에 영향을 미치는 것으로 모델화되어 있다. 플랫폼 데이터의 참여자인 소비자, 공급자, 소유자의 네트워크 효과가 플랫폼 파워를 강화한다고 할 때 행위자들의 참여의 역할은 중요하다. 참여자의 역할에 따라 데이터의 품질과, 데이터의 생성과 데이터의 통제에 영향을 미치는 점을 감안할 때, 참여자의 역할은 플랫폼 데이터 격차의 단계의 최종의 심급에서 매우 중요하다.

데이터 격차의 단계적 흐름의 모델에서 보면, 플랫폼 데이터 격차를

해소하기 위해선 최초에 물리적이고 동기유발적 접근 격차를 극복하기 위한 리터러시를 개발해야 할 것이다. 그러나 이미 디지털 기기와 네트워크 접근 역량은 극복된 것으로 이해하기 때문에 접근 리터러시 보다는 플랫폼 데이터의 기능에 대한 기능적 리터러시에 집중할 필요가 있다. 플랫폼 데이터 기술은 새로운 차원의 이용강도와 복잡성을 포함하기 때문에 기술적 기능 리터러시를 모색해야 할 것이다. 예컨대, 5G, VR, AR, MR, AI, IoT를 이해하고 데이터 코딩처럼 실질적 데이터 만드는 것도 전문 교육시킬 필요가 있다. 이용강도와 이용의 복잡성에 따라 참여자의 역할 격차가 발생할 수 있기 때문에 능동적 참여자와 소극적 참여자를 구별하여 전문적인 데이터 전문 교육과 보편적 시민 데이터 교육을 실시할 필요가 있다. 보편적 시민 데이터에게는 데이터 접근과, 데이터 생성 및 데이터 통제에 대한 편익과 위험을 인식시켜야 한다. 또한 알고리즘 리터러시를 통해서 맞춤형 정보와 왜곡과 편향의 위험성을 제고함으로써 플랫폼 데이터 생태계의 참여적이며 '관리된 대중'을 육성해야할 것이다.

5. 결론

데이터의 부자와 빈자, 데이터 가진자와 가지지 못한자 사이에 격차가 현실화되고 있다. 데이터 불평등에 의한 위험도 커지고 있다. 이때 행동하지 않으면 완전히 새로운 불평등 전선이 열리며, 아는 사람과 그렇지 않은 사람 사이에 세상을 갈라놓을 것이다. 언어, 빈곤, 교육 부족, 기술 인프라 부족, 원격성 또는 편견과 차별에 의해 데이터와 정보의 새로운 세계에서 많은 사람들이 배제될 것이다(UN Data Revolution Group, 2014, p.7). 반다익의 정보격차의 개념은 정보커뮤니케이션 기술의 특징에 의해서 새로운 차원의 격차로 진화하고 있다. 플랫폼 데이터 생태계가 구축되고 데이터 표현과 데이터 생성이 플랫폼에서 발생하면서 새로운

차원의 격차와 불평등의 문제가 발생하고 있다. 플랫폼 생태계는 소유자, 공급자, 소비자의 상호작용에 의해서 구축된다. 그 과정에서 데이터가 발생하기 때문에 플랫폼에서의 데이터는 자원(resources)이면서 자산(asset)이기도 하다. 플랫폼 데이터의 격차로 인한 불평등은 참여를 촉진함으로써 해소될 수도 있다. 자발적 데이터 표현의 참여와 플랫폼에서 데이터 생성과 알고리즘에 의한 데이터 통제의 문제에 대한 감시등의 적극적인 참여가 새로운 불평등을 해소할 수 있는 방안일 수 도 있다.

본 연구는 새로운 불평등의 발생 원인을 데이터 생성격차(데이터 접근격차와 데이터 통제격차)와 이용자 간의 데이터 표현격차로 제시하였다. 또한 이러한 격차에 의한 사회, 정치, 경제적인 영향을 분석하였고, 장기적인 불평등의 위험에 대해서 진술하였다.

첫째, 플랫폼과 데이터 생성 역량이 압축적인 확장 유형화하면서 네트워크 효과를 만들고 있다. 플랫폼의 데이터 생성역량이 강화됨으로써 데이터 접근격차와 데이터 통제격차가 격화되고 있다.

우선, 데이터 접근은 단순히 데이터를 확인하는 것이 아니라 데이터를 처리하는 능력이 있는가에 대한 것이다. 데이터 접근격차는 데이터를 처리할 능력이 있는 자와 없는 자의 격차이다. 플랫폼의 데이터 처리 능력의 소유와 비소유의 격차로도 볼 수 있기 때문에 플랫폼에 의해 개인의 데이터가 처리되는 과정에서 발생할 수 있는 위험과 불평등 가능성을 진단해야 할 것이다. 플랫폼의 데이터 집중에 의한 데이터 접근의 격차와 데이터화 자체의 접근이 불가능함으로써 발생하는 격차로 인해 '개인데이터 자기결정권' 또는 '자기정보관리통제권'등의 데이터 접근권의 불평등의 시정하기 위한 조치가 마련되기도 하였다. 또한 데이터 접근격차를 줄이기 위한 방안으로 '개방형 데이터'가 제안되기도 한다. 즉 개방성이 낮고 통제력이 높은 플랫폼으로 인해 데이터 접근의 격차가 발생함으로써 격차를 완화하기 위한 방안인 것이다. 자본화된 플랫폼의 힘이 데이터 접근격차를 가속화할 새로운 시스템을 구축할 수 있기 때문에 장기적인 해결

책에 대해 고민할 필요가 있다.

다음은 데이터 통제에 대한 것이다. 플랫폼의 데이터 처리에는 통제의 기능이 강화되고 있다. 데이터 통제는 알고리즘 체계에 의해서 형성된다. 데이터 통제격차는 반다익의 정보격차에서는 다루어지지 않는 차별적인 개념이다. 플랫폼 사업자가 데이터의 흐름에서 중립적일 것이라 간주하지만 플랫폼 사업자의 편의에 따라 알고리즘 규칙이 생성된다. 이로 인한 데이터 흐름의 왜곡은 정보의 편향성을 강화시킬 수 있다. 알고리즘에 의한 프로파일링은 표현의 자유 측면에서 우려된다. 왜냐하면 만약 알고리즘 설계에 누군가 관여한다면 그 본래 의도는 감춰지고 알고리즘 권력(algorithm power)을 활용하여 정보를 통제 왜곡하고 검색순위를 정하고 정보에 대한 수용자의 관점을 조작할 수 있기 때문이다. 이 경우에는 이용자들은 알고리즘에 의한 판단에 의하여 좌우되는 객체로 전락할 수 밖에 없다. 알고리즘에 의한 데이터 통제는 정치적으로 마케팅 측면에서도 정보의 다양성 측면에서 영향을 미치고 있다. 플랫폼의 자본화가 가속화하고 알고리즘 개발의 주인과 대리인의 문제가 발생화한다면 플랫폼 자본화에 복무하기 위한 알고리즘의 편향성과 비대칭성은 심화될 수 밖에 없다. 알고리즘의 불투명한 설계 또는 알고리즘을 훈련시키는 데이터셋에 숨어있는 편향과 오류는 사회적 차별과 배제를 조장할 수 있다. 2018년 유럽연합은 데이터 통제 격차에 대한 규범으로 개인 데이터의 수집, 분석 및 통제에 대한 최신의 법적 프레임인 GDPR(General Data Protection Regulation)을 시행했다. 이 규범은 동의(agreement) 문제를 강화했으며 현재와 미래의 데이터 이용에 대해 이용자 동의는 '자유롭게 부여되어야 하며 구체적이고 정보를 알려줘야 하고 모호하면 안된다' 강조하고 있다(European Union, 2016a). 이 외에도 데이터 주체는 데이터 접근권(사용된 분석과정과 목적에 대한 정보 포함), 부정확성에 대한 수정, 특정한 상황에서 처리 제한(restriction of processing), 데이터 삭제(잊혀질 권리), 자동화된 데이터 처리만으로 의사결정을 받

지 않을 권리(데이터 주체에 관한 법적 영향이 발생할 경우)등에 대해 명시하고 있다(European Union, 2016b). 특히 유럽연합의 GDPR의 데이터 주체에 관한 법적 영향일 발생할 경우 알고리즘 데이터 처리만으로 의사결정 되지 않고 인간이 개입하여 의사결정을 도모해야한다는 조항은 데이터가 개인의 정체성과 인격을 담고 있음을 인정하는 대목이다.

둘째, 데이터 표현 격차를 제시하였다. 주도적으로 데이터를 표현하는 이용자와 데이터를 표현하지 않는 이용자 간의 격차는 세계를 표현할 존재론적 차이를 보이고 있다. 이용자의 자발적인 데이터 표현은 정보제공에 유익한데, 자발적이고 적극적인 데이터 표현 역량의 차이에 따라 정보의 지향성과 품질이 달라질 수 있다. 따라서 이용자의 자발적 데이터 표현은 사회적 성과의 차이를 발생시킨다는 면에서 그 중요성을 인식할 필요가 있다. 집단지성 대표 사례인 '위키피디아'에 자발적이고 적극적인 이용자가 만들어낸 데이터 표현의 결과는 북반구의 남성중심의 편향적인 정보를 만들고 있으며, 검색을 통해서 편향적 정보를 재생산할 수 있다. 이러한 사례는 이용자들에게 데이터 표현의 자발적 행위 참여의 유익함과 위험함에 대해 교육시킬 필요성을 시사한다. 이용자 간의 격차를 줄임으로써 바로 대칭적이며 균형적 데이터를 표현할 수 있을 것이다. 따라서 시민들에게 데이터로 표현하는 사회적 영향력을 알리고 윤리와 규범 기반의 데이터 표현 역량에 대한 리터러시가 필요하다.

플랫폼 생태계에서 데이터 격차는 데이터 접근, 데이터 통제, 데이터 표현의 비대칭성에 의해 발생한다. 새로운 차원의 격차에서 '참여'는 데이터 관련 일련의 행위이기도 하지만 사회적 불평등을 이해하고 해결하려는 모든 활동이기도 하다. 데이터 격차의 불평등은 시민의 주체적이고 자발적인 참여를 통해 해결될 것이다. 데이터 격차의 차원에서 데이터 주권, 마이데이터를 둘러싼 데이터 소유권 의제 등을 다루고 있지만, 이용자들에게 데이터의 편익과 위험에 대한 데이터 리터러시를 통해서 플랫폼 데이터 생태계에 대한 인식을 제고할 필요가 있겠다.

이에 본 연구는 다음의 데이터 리터러시를 제안하고자 한다. 데이터 접근격차에서 플랫폼의 데이터 집중에 의한 데이터 처리와 정보의 편향의 가능성에 대한 리터러시가 필요할 것이다. 데이터 통제격차에 알고리즘 리터러시를 통해서 알고리즘 규칙의 투명성, 감시, 알고리즘 선택권 제공 및 개인정보 보호와 같은 리터러시를 실시해야 할 것이다. 데이터 표현격차에 대해서 데이터 표현으로 인한 사회경제적인 긍정적 효과와 부정적 효과에 대한 리터러시를 하고 자발적인 참여를 통한 유의미한 데이터 생산능력을 키울 필요가 있다. 결국 이러한 리터러시 교육은 '데이터 주권(Data Sovereignty)' 교육과 맥을 같이 하면서 이루어질 필요가 있다.

참고문헌

공정거래위원회 (2017). 〈해외경쟁정책동향〉, 제 124호.

이상기 · 정준희 · 이정훈 · 심우민 (2018). 〈글로벌 기업의 데이터 독점 관련 이슈와 대응방안 연구〉, 방송통신위원회.

정원엽 (2020, 11, 18). 알고리즘 공개, 기업분할, 최강 플랫폼법 예고한 유럽, 〈중앙일보〉. Retrieved from https://mnews.joins.com/article/23923299?fbclid=IwAR0EwX8KQE-026-_WI82hGAYGq9ddXgVMFseWraLxwxYroa79ERhA7IClaQ#home

최은창 (2017). 알고리즘 거버넌스, 〈미래연구포커스〉, summer. 2017, 과학기술정책연구원

주영재 (2020, 2, 15). 알고리즘 이용한 통제도 사용자의 지시통제로 봐야. 〈경향신문〉. Retrieved from http://news.khan.co.kr/kh_news/khan_art_view.html?art_id=202002151423011

Allcott, H. & Gentzkow, M.,(2017). Social Media and Fake News in the 2016 Election, *Journal of Economic Perspectives, 31*(2), 211-236.

Andreotti, A., Anselmi, G., Eichhorn, T., & Hoffmann, C. Pieter(2020), *Participation in the Sharing Economy,* Report from the EU H2020 Research Project Ps2Share: Participation, Privacy, and Power in the Sharing Economy.

Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, *8*, 1673-1689.

Berners-Lee, T. (2010 Feb.). http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide.html. Accessed 8 January 2020.

Beduschi, A, Cinnamon, J & Langford, J. (2017). Building digital identities: The challenges, risks and opportunities of collecting

behavioural attributes for new digital identity systems. http://hdl.handle.net/10871/28297

Boyd, D. & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Cinnamon, J. (2017). Social injustice in surveillance capitalism. *Surveillance & Society, 15*(5), 609-625.

Cinnamon, J. (2019). Data inequalities and why they matter for de velopment, *Information Technology for Development.* Retri eved from https://doi.org/10.1080/02681102.2019.1650244

Commission on Truth, Trust and Technology (2018). *Tackling the informationcrisis: A policy framework for media system resilience*. London: Department of Media and Communications, LSE. https://tinyurl.com/y6rx9dbn

Coyle, D. (2018). Platform dominance: The shortcomings of antitrust policy. In M. Moore & D. Tambini (Eds.), *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple* (pp. 50-70). New York: Oxford University Press.

De Vito, M. (2017). From Editors to Algorithms. *Digital Journalism, 5*(6), 753-773.

Diakopoulos, N. & Shamma, D. A. (2010. Apr). *Characterizing debate performance via aggregated Twitter sentiment.* Paper presented at the CHI Conference. Atlanta, GA. Available at http://dl.acm.org/citation.cfm?id=1753504 .

Dutton, W. H., & Helsper, E. J.,(2007). *Internet in Britain: 2007.* Oxford Internet Surveys. University of Oxford.

European Union. (2016). *General data protection regulation.* Chapter 2, Conditions for consent (Article 7(4), Recital 32). Retrieved from https://gdpr-info.eu/recitals/no-32/

European Commission (2017). The 2017 EU Justice Scoreboard.

Farhad Manjoo. (2016,1,21). Tech's 'Frightful 5' will dominate digital life for foreseeable future, The *New York times*, https://www.nytimes.com/2016/01/21/technology/techs-frightful-5-will-dominate-digital-life-for-foreseeable-future.html

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you.* Penguin Books Limited.

Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., & Schader, M. (2011). *Managing the crowd: Towards a taxonomy of crowdsourcing processes.* Detroit: AMCIS.

Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone?. https://doi.org/10.5210/fm.v16i2.3316.

Geiger, D, Seedorf, S., Schulze, T. & Nickerson R. (2011). Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes. *Americas Conference on Information Systems* (AMCIS), 4-7 August 2011. Detroit, Michigan.

Hargittai, E. (2002). Second-Level Digital Divide: Differences in People's Online Skills. *First Monday, 7*(4). Retrieved from https://doi.org/10.5210/fm.v7i4.942..

Gerlitz, C. & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society, 15*(8). 1348-1365.

Hartzog, W. & Selinger, E. (2013). Surveillance as Loss of Obscurity. *Washington and Lee Law Review, 72*(3), 1342-1386.

IDC. (2011 Jan.). *IDC Digital Universe study, Big Data, Bigger Digital Shadows and Biggest Growth in the Far East*, EMC Corporation.

ID4D (2016). *Identification for Development Strategic Framework*, World Bank Group.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal, 79,* 1-14.

Langley, P. & Leyshon, A. (2016). Platform capitalism: The intermediation and capitalization of digital economic circulation. *Finance and Society, 2*(1). DOI: 10.2218/finsoc.v3i1.1936

Manovich , L. (2012). Trending: the promises and challenges of big social data. In M. Gold (ed.). *Debates in the Digital Humanities* (pp.460-475). Minneapolis: University of Minnesota Press.

Mahapatra, P. & Rao, P.V.C. (2001). Cause of death reporting systems in India: a performance analysis. *National Medical Journal of India 14,* 154-162.

McCarthy, M. T. (2016). The big data divide and its consequences. *Sociology Compass, 10*(12). 1131-1140.

Napoli, P. M., & Caplan, R. (2017). Why media companies insist they are not media companies, why they're wrong, and why it matters. *First Monday, 22*(5). Retrieved from https://firstmonday.org/ojs/index.php/fm/article/view/7051/6124.

O'Neil, C. (2016). *Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Books.

Opher, Albert, Chou, Alex, Onda, Andrew & Sounderrajan. (2016). *The Rise of the Data Economy: Driving Value through Internet of Things Data Monetization: A Perspective for Chief Digital Officers and Chief Technology Officers,* IBM. https://www.ibm.com/downloads/cas/4JROLDQ7

Ozanne, L. & Ballantine, P.(2010). Sharing as a form of anti-consumption?: An examination of toy library users. *Journal of Consumer Behaviour, 9*(6), 485-498.

Ribes, D., & Jackson, S. J. (2013). Data bite man: The work of sustaining a long-term study. In *"Raw data" is an oxymoron* (pp. 147-166). Cambridge, MA: MIT Press.

Seidel, C. E., Thapa, B. E., Plattfaut, R., & Niehaves, B. (2013). *Selective crowdsourcing for open process innovation in the public sector: Are expert citizens really willing to participate?* Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance.

Selwyn, N. (2006). Digital division or digital decision? A study of non-users and low-users of computers. *Poetics 34*(4), 273-292.

Shaw, J. & Graham, M. (2017). An Informational Right to the City? Code, Content, Control, and the Urbanization of Information. DOI: 10.1111/anti.12312

Stephens, M. (2013). Gender and the GeoWeb: Divisions in the production of user-generated cartographic information. *GeoJournal, 78*(6), 981-996.

Sundararajan, A. (2016). *The Sharing economy: The end of employment and the rise of crowd-based capitalism,* The MIT Press.

Taylor, L. & Broeders, D. (2015). In the name of development: Power, profit and the datafication of the global South. *Geoforum,* 64, 229-237.

The Economist (2017). The world's most valuable resource is no longer oil, but data. Retrieved from https://tinyurl.com/ybsfte99

The Economist. (2014). Off the map: Rich countries are deluged with data; developing ones are suffering from drought. Retrieved from http://www.economist.com/news/international/21632520-rich-countries-are-deluged-data developing-ones-are-suffering-drought?fsrc=scn/tw/te/pe/offthemap

UN Data Revolution Group. (2014). *A world that counts: Mobilising the data revolution for sustainable development.* Geneva: United Nations Secretary General's Independent Expert Advisory Group.

van Dijk, J. (2005). *The deepening divide: Inequality in the information society.* London, UK: Sage.

van Dijk, J. (2013). *The Culture of connectivity: A Critical History of Social Media.* New York: Oxford University Press.

van Deursen, Helsper, E.J. & Eyon, R. (2014). Development and validation of the internet skills scale. *Journal Information, Communication & Society. 19*, 804-823.

van Dijk, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society, 12*(2), 197-208.

van Dijk, J., Nieborg, D. & Poell, T. (2019), Reframing platform power. Retrieved from https://www.researchgate.net/publication/334824790

van Dijk, J. (2017). Digital divide: Impact of access, The International Encyclopedia of Media Effects published 08 March 2017. Retrieved from https://doi.org/10.1002/9781118783764.wbieme0043

van Dijk, J. Poell, T. & De Waal, M. (2018). *The Platform Society Public Values in a Connective World,* Oxford University Press.

투 고 일 자: 2020년 07월 05일

심 사 일 자: 2020년 11월 17일

게재확정일자: 2020년 11월 25일

Abstract

Data Divide in the Platform Data Ecosystem:

Beyond digital inequality

Mi Kyung Kim
Professor, Department of Media Communication, Chungwoon University

This study reviewed the divide that would be caused by platform data concentration in the platform data ecosystem, reviewing the analysis of platform data production structure and data transaction structure. With the resourceization of data and the compress or expand the stack of platform and data, the rich and poor of platform data can occur and it causes structural inequality. This paper presented the data access gap, data control gap according to the platform's data production. It also pointed out the problem of information bias reproduced by the data representation divide between users. It is proposed to devise a data literacy strategy from a large framework of data sovereignty, as the network effect of the platform data ecosystem causes performance and inequality in both aspects of benefits and risks.

KEYWORDS Platform Data Ecosystem, Data Access Divide, Data Control Divide, Data Representation Divide, Data Literacy