

해외 주요국의 데이터 거래 및 가격설정 사례 분석 및 한국의 데이터 정책 방향 모색

강경훈*, 강성호**, 김영일***

요약

데이터가 경제 내 가장 중요한 부가가치 생산 원천으로 자리매김하는 현 상황에서 데이터의 효율적 생산과 활용을 위해서는 데이터의 생산, 수집, 결합, 가공 및 거래가 원활히 이루어질 수 있는 시스템 혹은 생태계의 정착이 필요하다. 본고는 국내 및 해외 데이터 거래 현황에 대한 소개, 문제점 파악, 데이터 가격설정과 관련한 기존 연구들을 포괄적으로 소개하는 한편 이에 기반하여 우리나라 데이터 거래 활성화를 위한 정책과제를 제시한다. 바람직한 데이터 정책은 데이터 활용이 광범위하게 이루어지고 데이터 집종을 제한함으로써 경쟁을 제고하는 한편 데이터 주체의 사생활 보호에도 만전을 기할 필요가 있다.

* 동국대학교 경영대학 경영학과 교수, Tel: +82-2-2260-8909, E-mail: khkang@dongguk.edu

** 금융위원회 및 UNIDO 수석전문관

*** 나이스평가정보 리서치센터장

I. 머리말

21세기의 원유라고 불리는 데이터의 중요성은 사실 이전에도 잘 알려져 있었다. 다만 시간이 흐름에 따라 데이터가 제공하는 기능이나 역할에는 상당한 변화가 있다. 독일 프라운호퍼 연구소에 따르면 데이터의 기업활동에서의 기능은 다음과 같이 4단계로 구분된다. 우선 기업활동의 결과물로서의 데이터인데 1960-70년대와 같은 데이터 프로세싱의 초기 단계에서 데이터나 정보시스템은 비즈니스를 지원하기 위한 기능으로 활용되었다. 예컨대, 재고 자산 관리 기능은 재고가 어느 지역에 있는지를 체크하고, 실제로 창고 내에 물건이 존재하는지를 실제 현장조사를 하지 않고도 알 수 있도록 도와주는 역할이다. 이 시절에 기업의 가치는 물리적인 제품에 의해 창출되었지 데이터가 창출하지는 않았다.

1980년대와 1990년대의 제조업 자원배분계획(manufacturing resource planing)이나 기업자원 배분계획(enterprise resource planning) 시스템의 유행과 더불어 데이터는 기업전반의 경영과정을 돕는 촉진자로 기능했다. 기업전반을 아우르는 정확한 데이터 없이 경영자는 주문, 현금관리, 물자조달, 결제 등의 과정을 쉽게 만들기 어렵다. 이 단계에서 데이터는 생산, 조달 및 배송, 고객관리 분야에 있어서의 일종의 전략적 자원이 됐다.

한편 21세기에 들어 기업들은 양질의 데이터를 필요로 하는 제품과 서비스를 생산하기 시작했다. 서비스와 제품의 촉진자로서의 데이터라고 할 수 있다. 예컨대, 아디다스의 miCoach, 힐티의 임대사업 모델, 또는 소비자들에게 제공되는 다양한 스마트 서비스들을 들 수 있다.

최근에는 데이터 자체가 마켓플레이스에서 상품으로 거래된다. 이는 데이터에 대한 가격부과를 위해 사용시간과 사용량에 따라 데이터 API를 구축할 필요성을 제기한다. 이런 방식으로 데이터는 기업활동을 촉진시키는 촉매제가 아니라 상품 그 자체가 되었다.

이처럼 상품 자체로서의 데이터가 등장함에 따라 데이터의 거래 메커니즘과 가격 설정 메커니즘이 어떻게 설계되고 작동되어야 하는지는 중요한 문제로 대두되고 있다. 데이터는 경합성, 배제성, 복사가능성 등 여러 면에서 기존 유무형의 상품과 매우 달라 거래 또는 가격설정이 쉽지 않기 때문이다.

이 글은 해외 주요국 및 국내 데이터 거래 현황을 살펴보고 데이터 가격을 산정하는 여러 연구결과를 소개한 후, 이에 기초하여 국내 데이터 거래를 활성화시키는 데 필요한 정책과제를 모색한다. 먼저 다음 절에서는 국내 데이터 거래의 현황과 문제점을 살펴본 뒤 미국, 영국, 중국 등 주요국의 데이터 거래 현황에 대해 알아본다. 3절에서는 데이터 측정과 가치 평가의 어려움과 주요 방법론에 검토한다. 4절에서는 3절까지의 분석을 기초로 향후 정책과제에 대해 논한다.

Ⅱ. 국내외 데이터 거래 현황

1. 국내 데이터 거래 현황 및 문제점

국내 데이터 거래 시장은 초기 단계이지만 규모가 계속 커지고 있다. 하지만 아직 주요국에 비해 성장률이 낮고, 데이터 판매에 대한 일반 기업의 참여율도 낮은 편이다. 최근에는 신용정보회사 등 전통적으로 데이터 사업을 운영해온 기업들 외에도 오픈마켓형 거래소(중개업체)가 등장하고 있다. 해외 기업의 51%가 데이터 판매로 수익을 내고 있으며 성장이 빠른 기업일수록 데이터 판매가 활발한 데 비해, 국내 기업 중 데이터 판매 경험이 있는 기업은 21%에 그치고 있다. 대기업이 주요 판매업체이지만 판매 동인이 부족해 데이터 거래의 활성화를 위해서는 데이터 거래·컨설팅업체(데이터 브로커) 육성이 시급한 상황하다.

국내에서는 데이터 거래가 미미하고 이에 따라 데이터의 경제적 가치 평가도 제대로 이루어지지 못하는 형편이다. 해외 주요국의 경우 기업들이 데이터 거래를 통하여 데이터를 획득하는 비중이 높은 데 비해 국내에서는 기업이 보유한 내부 데이터를 주로 이용하는 것으로 나타나고 있다.

<표 1> 국내외 데이터 획득 경로 비교 (단위: %)

외국 (Q: 다음 각각의 데이터 획득 경로를 이용하는가?)	유형	국내 (Q: 다음 중 주로 이용하는 데이터 획득 경로는?)
<div>72</div> Supplier data <div>69</div> Consumer usage data <div>67</div> Anonymous consumer data	내부데이터	<div>24.9</div> 온라인 회원 및 고객어 이용 동의를 통한 데이터 <div>13.3</div> 자사의 데이터 가공/데이터화/제 의한 데이터 확보 <div>12.4</div> 마케팅 등 고객 커뮤니케이션을 통해 수집 <div>4.0</div> 각종 센서를 통한 자동 데이터 수집 <div>3.5</div> 내방객 등 오프라인을 통한 데이터 수집
<div>97</div> Open data	공공데이터	<div>10.9</div> 공공기관에서 제공되는 데이터(공공데이터)
<div>98</div> Publicly available competitor data <div>77</div> Data from blogs/product reviews <div>77</div> Social media data	웹사이트 데이터 수집	<div>18.6</div> 소셜(SNS), 인터넷 등을 통한 데이터 수집 <div>6.5</div> 수집 솔루션(tool) 등에 의한 웹 데이터 수집
<div>92</div> Proprietary datasets from data aggregators <div>90</div> Data from hyperscalers (i.e., Google, Amazon) <div>84</div> Data from distributors/partners <div>67</div> Data from platform providers	거래소 데이터 거래	<div>5.9</div> 데이터 거래를 통한 데이터 확보
<div>91</div> Analyst/industry reports	기타	

자료: 정준화 · 박소영 (2022)

한국 정부는 디지털 뉴딜 정책을 추진하고 있는데 그 일환으로 2019년부터 16개의 빅데이터 플랫폼을 구축한 바 있다. 민간기업과 공공기관이 빅데이터 플랫폼에 참여해 유·무료 데이터를 거래하고 있다. 2021년 12월 현재 데이터 개방 건수는 6,842건, 활용 사례는 409,440건이다.¹⁾ 또한 민간의 데이터 플랫폼이나 데이터 전문기업도 데이터 거래 및 중개 업무를 하고 있다.²⁾

국내에서 데이터 거래가 활성화되지 못한 것은 관련 규제의 엄격성, 기술 수준 등 여러 요인에 기인하고 있으나 이런 전제조건들 하에서 효율적으로 작동할 수 있는 데이터 거래 메커니즘이 제대로 도입되지 못한 것도 하나의 요인일 가능성이 있다. 이하 해외 주요국의 데이터 거래 현황을 살펴보고 국내 데이터 거래를 활성화할 수 있는 방안에 대한 시사점을 모색한다.

1) 이 가운데 유료 데이터 거래는 7,891건이다.

2) 쿠큰닷넷(Cucon), MY데이터허브(기용정보통신), 금융데이터거래소(금융안전청) 등이 있다.








2. 미국

미국에서는 주로 ‘데이터 브로커’를 중심으로 데이터 거래가 활발히 이루어지는 것으로 파악된다. 그 밖에 데이터 거래소, 직접 판매, 제휴 및 인수·합병 등의 방식도 활용되고 있다.

먼저 데이터 브로커에 대해 살펴보면 판매자·거래소·거래사가 결합되어 운영되는 형태로 볼 수 있으며, 주로 B2B 거래가 중심이다. 데이터 브로커 운영 현황을 보면 여러 기업이나 기관으로부터 데이터를 수집하여 분석·가공한 후 구매자가 원하는 조건의 데이터를 재판매하는 역할까지 하나의 사업체 안에서 수행하는 기관으로 여러 출처로부터 다양한 정보를 수집한다. 먼저 공공기록(public records)은 인구통계, 부동산 기록, 자동차 관련 기록, 결혼증명서, 파산기록, 기업목록, 주 단위 면허기록, 유권자 등록 기록 등이다. 상업적 출처(commercial sources)로터는 소매업체 구매내역, 고용등록정보, 신용정보, 멤버십 데이터, 사용료 카드 데이터, 보증등록, 구독정보 등을 수집한다. 온라인 트래킹(online tracking)을 통해 소셜미디어 프로필, 웹 브라우저 쿠키, 포럼 게시물, 모바일 앱 기록, 웹 브라우징 활동, 디바이스 데이터, 메타 데이터, IP내역 등도 수집한다. 한편 각 출처로부터 수집한 개인에 대한 정보를 기반으로 다른 영역을 추측(예: 온라인 구매 및 검색내역 기반으로 건강 상태를 추측)한 추측성 데이터도 중요한 자원이 된다.

미국의 데이터 브로커 시장 규모는 2018년 기준 1,832억 달러(약 220조 원)에 달한다. 민간에서 운영하는 데이터 브로커 기업의 수도 2,500개를 넘는다. 참고로 2019년 기준 글로벌 데이터 브로커 시장 규모는 2,326억 달러 수준으로 보고되고 있다. 미국 내 대표적 데이터 브로커로 Acxiom, LexisNexis, Nielsen, Experian, Equifax, CoreLogic, Verisk, Qlik, Oracle, Epsilon, People-search sites 등이 있다.

<표 2> 미국의 주요 데이터 브로커

데이터브로커명	구분	주요 서비스	내용
	마케팅 제공 목적의 데이터 브로커	맞춤형 광고정보 제공	• 전세계 7억명 이상의 소비자 명부 보유
		맞춤형 광고정보 제공	• 약 8억건의 부동산 거래정보, 약 1억건의 담보 DB를 자체 가공 및 분석
		맞춤형 광고정보 제공	• 600만개 이상의 상업용 부동산, 14억건 이상의 보험 사기 데이터, 18억개 이상의 소비자 카드 및 저축계좌 정보 제공
		통계 관련 시각화 정보 판매	• 35개국 주식시세정보 등 통계정보 시각화 및 직접 판매
		맞춤형 광고정보 제공	• 2억8,300만개 이상 미국 소비자 프로필 및 7,700만개의 비즈니스 명부 보유
	글로벌 CB사이먼서 데이터 브로커	맞춤형 광고정보 및 고용관련 정보 제공	• 전세계 8억명 이상의 소비자와 9천만개 회사에 대한 정보 보유
		맞춤형 광고정보 제공 및 의료관련 정보 제공	• 2억 4,500만 명 미국 소비자 프로필 및 2,500만 개 이상 미국 회사 정보를 포함하여, 전세계 8억 건 이상의 개인 및 회사의 정보 보유

자료: 각 사 홈페이지

주요 거래 데이터로는 데이터 브로커가 보유하고 있는 원데이터, 구매자가 원하는 그룹(특정 조건에 해당하는 소비자 명부)을 선별하거나 새롭게 생성한 고객군 정보 등이다. 이들은 원데이터도 판매하지만, 구매자의 마케팅 목적으로 개인정보 분석을 통해 특정 상품이나 브랜드의 판매 또는 홍보 방법, 효과 등에 대한 분석 서비스도 거래한다. 다양한 고객(end-user)의 효과와 니즈를 충족시킬 수 있는 여건이 갖춰져 있다.

데이터 브로커가 보유하고 있는 데이터는 개인정보를 포함하여 매우 광범위한 범위에 걸쳐 있다. 거래 데이터 범위는 이름, 현재 및 이전 주소, 생일, 성별, 혼인여부, 가족상태, 사회보장번호, 교육수준, 정치성향, 자산, 직업, 전화번호, 이메일 주소, 구매습관, 의료정보 등 다양한 정보를 포괄한다.

데이터 구매자는 주로 기업체 또는 단체이다. 민간기업, 공공기관(연방정부, 주정부, 지방정부 등), 비영리조직(대학교, 자선단체 등) 등 다양한 조직이 데이터를 구매하고 있다. 예컨대 페이스북이 2012년 DataLogix로부터 페이스북 이용자들이 SNS 상품광고를 조회하거나 오프라인 상점에서 상품을 구매한 것과 관련된 데이터를 매입한 사례를 들 수 있다.

<표 3> 고객 산업군별 판매 서비스 유형

구분	Direct Marketing	Online Marketing	Marketing Analytics	Identity Verification	Fraud Detection	People Search
간편결제제공업체				○	○	
변호사/수사관	○					
자동차회사	○	○	○			
교육기관	○			○	○	
에너지/유틸리티	○					
정부기관	○		○	○	○	○
병원/여행/엔터	○		○	○		
개인 소비자						○
보험회사	○		○	○	○	
금융회사	○	○	○	○	○	○
마케팅/광고회사	○	○	○	○	○	○
미디어회사	○		○			○
비영리단체	○	○		○	○	
제약회사	○		○			○
부동산서비스업체	○				○	○
소매업체	○	○	○	○	○	○
테크회사	○	○	○			○
통신사	○		○	○	○	

자료: Data Brokers A call for Transparency and Accountability, Federal Trade Commission, 5.2014.

미국의 주요 데이터 브로커 가운데 Acxiom은 1969년 여론조사 자료 및 분석업체로 설립되었는데 현재 전 세계 7억 명 이상의 소비자 정보 DB를 보유한 미국의 대표적 데이터 브로커로 성장했다. 1990년대 후반 IT시장이 발전하면서 사람들이 남기는 개인정보의 흔적을 수집하기 시작했다. 백화점·피트니스센터 회원 등록정보, 신용카드 사용, 보험 가입 등 금융정보를 수집하였으며 인터넷의 보편화, 전자상거래 등 파생산업의 확산, 스마트폰의 대중화로 수집하는 소비자정보의 규모가 급속도로 확장했다. 현재 미국 아칸소주에 본사가 있으며 영국, 중국, 독일, 폴란드 등에 사무소를 운영하고 있다. 2022년 기준 임직원 수 3,380명, 매출액 6억 1,700만 달러를 기록했다.

현재 Acxiom은 IPG의 자회사로 데이터베이스 기반 컨설팅 분석 서비스를

제공하고 있으며 기존 Acxiom은 LiveRamp로 사명을 변경하고 온보딩 및 데이터연계기술 서비스를 제공하고 있다. 2014년 LiveRamp(데이터온보딩, 데이터연계기술회사)를 인수했으며 2018년 2월 LiveRamp와 Acxiom Marketing Solution으로 사업 분할 뒤 2018년 7월 당시 매출의 70%를 차지하는 Acxiom Marketing Solution을 광고회사인 IPG에 매각된 바 있다.

Acxiom은 특정 데이터를 보유하고 있는 파트너사와 협력 또는 직접 온라인·오프라인을 통해 다양한 데이터를 수집한다. 2억 6천만 명 이상의 미국인을 포함하여 전 세계 여러 나라의 소비자 7억 명 이상에 대한 정보를 보유하고 있으며 저장된 개인 한 명에 대한 개인정보 관리 항목이 1,500여 종에 달한다. Acxiom은 특정 개인의 인적사항을 수집한 후 13자리 수로 구성된 고유번호를 부여해 관리하고 있다.

<표 4> Acxiom의 주요 데이터 상품

DB상품명	내용
INFOBASE® 전세계 7억명 이상의 개인 정보 DB 개인 및 가구에 관리항목이 약 1,500개	<ul style="list-style-type: none"> 개인정보: 이름, 전화번호, 주소, 연령, 성별, 종교, 교육수준, 직업, 이메일 주소 가구정보: 구성원 수, 자녀 수, 결혼/이혼, 출산 금융정보: 소득, 순자산, 경제적 안정성 라이프 관련: 주택구입, 이사, 자동차 구입 관심사: 스포츠, 레저, 펫, 엔터테인먼트, 커뮤니티 활동 구매정보: 구매 이력, 지불 방법 지리정보: 위도/경도, 인구통계 (패키지DB: Block, DMA, ZIP+4 등)
PERSONICX® 마케팅 전략에 활용할 수 있는 소비자그룹에 대한 다차원 세분화 솔루션	<ul style="list-style-type: none"> 인구통계, 사회 경제 및 필수 구매 행동 패턴 등을 기반으로 기존고객 및 잠재고객군을 그룹화한 DB PERSONICX와 다른 데이터베이스를 결합하여 타겟 마케팅 등에 활용 다양한 패키지 DB를 제공: <ol style="list-style-type: none"> PERSONICX PRIME: 개인을 79가지 세그먼트 및 15개 그룹화 PERSONICX LIFESTAGE: 인구통계, 사회경제적 행동을 기반으로 가구를 70가지 세그먼트 및 21개 그룹화 PERSONICX LIFESTAGE INSURANCE: 인구통계, 보험성향 기반으로 개인을 70가지 세그먼트 및 13개 그룹화 PERSONICX FINANCIAL: 인구통계학적 특성과 관계없이 금융활동에 대한 성향 기반으로 13개 그룹화 PERSONICX HISPANIC: 히스패닉 소비자의 인구통계, 사회경제적 행동 기반으로 15개 그룹화
Audience PROPENSITIES® 소비자 행동 예측 솔루션	<ul style="list-style-type: none"> 3,500개 이상의 시장 선도 지표를 활용하여 브랜드 친밀도, 선호도 및 행동을 예측 행동 및 태도 지표는 실행 가능한 정보를 기반으로 생성되며, 쇼핑의도, 태도 데이터, 미디어 행동 등과 연결됨 고급 분석 툴은 소비자 브랜드 친화도 및 선호도, 시장 내 관심사 및 시기, 미디어 시청 습관 등을 예측

자료: Acxiom data Catalog (2022)

Acxiom은 주로 데이터베이스 기반의 마케팅 사업을 영위하고 있으며 익명 처리된 데이터 자체를 직접 판매하기도 한다. 마케팅에 활용할 수 있는 소비자 정보를 수집·분석하여 고객사에 최적화된 잠재고객 발굴 및 마케팅 방안 등을 제공하고 수수료를 취득한다.

<표 5> Acxiom의 주요 서비스와 활용 마케팅 예시

Acxiom을 통한 주요 서비스		Acxiom의 데이터 활용 마케팅 예시	
서비스 종류	설명	기업명	마케팅 내용
Data Management	• 타겟 마케팅 목적으로 고객사에 맞는 소비자 데이터를 선별할 수 있게 지원	 미국 항공사	• Southwest 항공사의 실시간 타겟마케팅 • Southwest 비행기 좌석 스크린에는 승객별로 다른 광고 방송을 함 • 비행기 승객 명단이 확보되면, Acxiom DB와 매칭시켜 승객의 쇼핑 습관, 구매 패턴을 분석한 후 승객별 최적 광고를 방송함
Identity Resolution	• Real Identity™ 솔루션 제공 • 여러 채널을 통해 방문하는 고객의 신원 정보를 동기화하여 식별자를 구성할 수 있는 서비스		
Digital Transformation	• 디지털 마케팅 솔루션에 대한 구현, 관리 및 최적화 등 일련의 과정을 지원	 개인용 잔디 및 정원 등 원예 장비 제조 및 판매	• Acxiom의 INFOBASE 데이터베이스 기반의 모델링 기술을 활용 • 3개의 소비자 그룹을 광고 타겟으로 분류 : Loyalists, Switchers, single-family homeowners • 2021.3~8월 중 2,680만 명의 잠재고객에게 광고 노출, 노출되지 않은 대조군에 비해 2.25배 높게 구매로 연결
Analytics Services	• 데이터 마이닝, 모델링, 머신러닝 등의 기술을 활용, 고객사의 마케팅 전략 분석 서비스 제공		
Data Privacy	• 고객사가 데이터 규정 및 개인정보 보호지침을 준수하고 정책 수립을 준비할 수 있도록 지원		

자료: Acxiom 홈페이지

미국의 대표적인 데이터 브로커 중 하나인 Equifax는 1899년 설립되었으며 미국 애틀랜타에 위치한 대표적인 글로벌 CB이기도 하다. 미국 3대 신용정보회사³⁾ 중 가장 오래된 회사이며 데이터브로커 등록의무가 있는 캘리포니아 및 버몬트 주에 데이터브로커로 등록한 상태이다. 2022년 기준 임직원 수 14,000명, 매출액 51.2억 달러이며 미국 본사 외에 북미, 중남미, 유럽, 아시아 태평양 지역의 24개국에서 운영하고 있다. 전통적으로는 신용정보회사이나 데이터 다양화를 추진하며 데이터베이스 마케팅 사업 및 신상품을 출시하며 서비스를 확장하였다. 주로 특정 분야의 기업인수를 통해 서비스 분야를 확장한 것으로 알려져 있다.

3) 미국 3대 신용정보회사는 Equifax, Experian, TransUnion이다.

<표 6> Equifax의 주요 서비스

Equifax의 주요 서비스		데이터 브로커로서의 주요 서비스
구분	세부 서비스	주요 서비스 내용
Workforce Management	• Manage HR Compliance Activities	• Workforce Solutions - 2007년 TALX를 인수하면서 workforce solutions 시장에 진출 - 고용 이전 단계~고용 이후까지 인적자원 관리를 위한 모든 서비스 제공 - 미국 성인 5.4천만 명에 대한 실명인증(Verification)과 HR정보(소득, 업무경력 등) 기반으로 인력검증서비스/급여관리솔루션 등을 제공 - '22.9월 Reemployment Service 출시: 재취업 솔루션
	• Manage Unemployment Activities	
	• Onboard New Hires	
	• Pre-Screen Employment Applications	
	• Simplify & Manage Tax Forms	
Data-Driven Marketing	• Analyze Market Potential	• 다양한 보유 다양한 데이터 기반으로 고객사 타겟층 분석, 확장 및 마케팅 전략 제공 • 고객사 위험 및 기회요인 분석
	• Deliver a Better Customer Experience	
	• Marketing Strategies, Tactics & Results	
	• Retain & Grow Customer Relationships	
Credit Risk	• Target Hight Right Audiences Online	• 신규고객/마켓의 신용위험, 포트폴리오 퍼포먼스 최적화 등 서비스 제공
	• Access Market & Business Intelligence Tools	
	• Analyze Multi-Data Assets for Credit Risk	
	• Identify New Customer Credit Risk	
	• Monitor and Manage Accounts	
Identity & Fraud	• Optimize Portfolio Performance	• Kount 360: 계정생성 및 보안결제 플랫폼 - 2021년 전자상거래 사기 및 보안서비스 제공 전문업체인 Kount 인수 • 사기 위험 탐지 및 손실 방지, 규정 준수 지원
	• Assist Credit Risk Compliance	
	• Prevent Fraud Risk & Loss	
	• Support Regulatory & Security Compliance	
Verification	• Verify, Authenticate & Link Identities	• 교육 이력, 정부 프로그램 자격, 소득 및 고용 정보 등 신원 확인
	• Verify Education History	
	• Verify Eligibility for Government Programs	
	• Verify Income & Employment	

자료: Equifax 홈페이지 (강경훈, 강성호, 김영일 (2023)에서 재인용)

한편 미국의 데이터 유통시장이 성장할 수 있었던 이유는 사후거부권 (Opt-Out) 방식을 채택하여 자유롭게 개인정보를 수집할 수 있기 때문이다. 데이터 브로커 산업을 직접적으로 규제하는 연방정부 차원의 별도 법률은 존재하지 않으며, 다만 일부 주에서 간접 또는 직접적으로 데이터 브로커를 규제하는 사례가 있다. 데이터 브로커를 통한 개인정보 유출 사건이 발생하면서 데이터 브로커에 대한 규제 입법이 추진되고 있으나, 이미 활발하게 운영되고 있는 데이터 브로커 산업을 침체시킬 우려 및 데이터 활용의 중요성이 부각되면서 제안된 법안 통과가 쉽지 않다.

<표 7> 주요 데이터 브로커의 개인정보 유출사건

데이터 브로커	유출사건 내용
Equifax	2017년 9월 Equifax에서 1억 4,550만 명의 미국 소비자의 개인정보가 유출. 개인정보에는 이름, 사회보장번호, 생년월일, 주소, 운전면허증번호 등이 포함
Experian	4회의 개인정보 유출 사고가 발생. 최근에는 2023년 1월, 브라질 자회사인 Serasa Experian에서 2억 2천만 명의 개인정보 및 1억 4백만 건 차량 기록 데이터가 다크웹에서 거래. 개인정보에는 이름, 세금 식별자, 얼굴 이미지, 주소, 전화번호, 이메일, 신용점수, 급여 등이 포함
LexisNexis	모회사인 Relx에서 2005년 사회보장번호, 운전면허증 정보 및 31만 명의 주소가 유출. 2009년에는 3만 3천 명의 정보가 유출
Leap Lab	은행계좌정보와 사회보장번호가 포함되어 있는 수십만 건의 대출 상환신청서를 사기꾼에게 재판매

자료: 기사 참고 및 미국 ‘데이터 브로커’ 제도의 국내법적 함의(경제규제와 법, Vol.11, No.2.)

캘리포니아 주 및 버몬트 주에서는 데이터 브로커의 정의 및 등록을 요구하는 주법을 신설했다. 캘리포니아 주에서는 소비자 프라이버시 전반에 관한 개인정보보호법 내에 데이터 브로커에 대한 내용이 포함된 반면, 버몬트 주에서는 데이터브로커 자체에 관한 법령을 신설하였다. 데이터 브로커의 개인정보 처리와 관련하여 일정한 의무를 부여하고 있으며 사업등록을 통해 정부의 관리감독을 도모하고 있다. 그러나 데이터 브로커들은 여전히 개인정보 수집으로부터 자유로운 편이며 2023.8월 현재 캘리포니아, 버몬트, 텍사스, 오리건 4개 주에서 입법을 통해 데이터 브로커를 허용하고 있다.

<표 8> 캘리포니아 주 및 버몬트 주 데이터 브로커 관련 법 비교

	California	Vermont
법률명	<ul style="list-style-type: none"> 캘리포니아 소비자 프라이버시법 California Consumer Privacy Act, CCPA 	<ul style="list-style-type: none"> 데이터 브로커와 소비자보호에 관한 법 An Act Relating to Data Brokers and Consumer Protection
제정일	<ul style="list-style-type: none"> 2018.6.18 제정, 2020.1.1 시행 	<ul style="list-style-type: none"> 2018.5.22 제정, 2019 발표
특징	<ul style="list-style-type: none"> 미국 최초의 개인정보보호에 관한 일반법 & 미국 역사상 가장 강력한 개인정보보호 법으로 평가⁶ 	<ul style="list-style-type: none"> 2017.9월 Equifax에서 버몬트주 시민 247,607명(버몬트 인구의 약 40%)의 개인 정보가 유출된 사건을 배경으로 제정
데이터 브로커 정의	<ul style="list-style-type: none"> 1798.99.80.(d) "Data broker" means a business that knowingly collects and sells to third parties the personal information of a consumer with whom the business does not have a direct relationship. 	<ul style="list-style-type: none"> §2430. DEFINITIONS(4)(A) "Data broker" means a business, or unit or units of a business, separately or together, that knowingly collects and sells or licenses to third parties the brokered personal information of a consumer with whom the business does not have a direct relationship.
등록 의무	<ul style="list-style-type: none"> 매년 차기년도 1월 31일까지 연 400달러의 비용을 지불, 캘리포니아주 법무장관에게 등록 요구 주정부 홈페이지에 회사 이름, 웹사이트 주소, 물리적 주소 등 정보 공개 등록의무 위반 시 하루 100달러, 전년도 수수료, 조사·기소 과정에서 소요된 비용 등의 벌금 청구 	<ul style="list-style-type: none"> 매년 차기년도 1월 31일까지 연 100달러의 비용을 지불, 버몬트주 국무장관에게 등록 요구 주정부 홈페이지에 회사 이름, 웹사이트 주소, 물리적 주소 등 정보 공개 등록의무 위반 시 하루 50달러(최대 10,000달러), 전년도 수수료, 조사·기소 과정에서 소요된 비용 등 벌금 청구
기타	<ul style="list-style-type: none"> 사업자는 개인정보보호를 위해 정보 성격에 적절한 합리적인 보안절차와 실무관행을 이행하고 유지해야 하는 의무가 있으며, 이를 위반함으로써 발생한 소비자의 피해와 관련해 민사소송 대상이 될 수 있음(§ 1798.150(a)) 이를 위반하는 사업체, 서비스 제공자 또는 기타 사람은 각 위반사항에 대한 가처분 및 2,500달러 이하 또는 7,500만 달러의 벌금이 부과(§1798.155) 	<ul style="list-style-type: none"> 새로운 보안 프로그램 요구 사항을 충족시키지 못하면 "불공정하고 기만적인 상업 행위(unfair and deceptive acts in commerce)"로 선언(commits)될 수 있음(§ 2433(b)(1)) 한편, 벌칙규정과 관련하여 법무부 장관이 시행 조치를 취하게 되지만 민간 부문에서 시민들이 신용보고 관련 법률(credit reporting laws)에 따라 민사 소송을 제기할 수 있음(§2433(b)(2))




자료: 각 법안 원문 (강경훈, 강성호, 김영일(2023)에서 재인용)

미국의 데이터 거래를 주도하는 것이 데이터 브로커이지만 데이터 거래소를 통한 거래도 활발하며 빠르게 증가하고 있다. 데이터 거래소는 데이터를

유통하는 장소를 제공하며 거래를 희망하는 자는 거래 플랫폼에서 데이터를 제공 또는 취득한다. 데이터 거래소는 데이터 중개 및 가격 책정 역할을 수행한다. 주로 판매자 위주의 운영 방식을 택하고 있으며 판매자가 보유한 데이터를 원하는 가격 또는 거래소에 의해 책정된 가격으로 제시하면 해당 데이터를 필요로 하는 소비자가 해당 거래소를 통해 구매하는 방식이다. 판매자(개인·기업)가 수집·가공한 데이터 셋을 거래소에 업로드하면, 구매자가 원하는 데이터를 검색하여 구입하는 형태로 “중개” 기능에 초점을 두고 있다. 데이터 가격은 판매자가 결정하거나 거래소의 가치평가 시스템을 통해 책정되고 있다.

미국의 데이터 거래소는 민간 위주로 운영되며, 2019년 말 기준 각종 데이터 거래소의 수는 650개가 넘는다. 미국 내 대표적 데이터 거래소로 Factual Data, Microsoft Azure Data market, Propublica Data Store 등이 있다.

<표 9> 미국의 주요 데이터 거래소

거래소명 (설립년도)	주요 거래 데이터 유형	데이터 가격 책정
 (2010)	<ul style="list-style-type: none"> 인구통계, 환경, 금융, 소매, 날씨, 스포츠 등에 관한 정보 Excel, PowerPivot, Tableau 등 다양한 인터페이스 제공 	<ul style="list-style-type: none"> 판매자가 사용 조건, 가격 책정 모델 및 가격을 포함한 제품의 세부 정보를 설정 판매자 대부분이 기업 구매자가 데이터를 구매할 때 수수료 20% 부과
 (2007)	<ul style="list-style-type: none"> 위치정보(지리적 위치, 위성, 관심지역 데이터 등) 	<ul style="list-style-type: none"> 모든 데이터셋이 유료
 (2014)	<ul style="list-style-type: none"> 건강, 형사·사법, 교육, 정책, 비즈니스, 교통, 국방, 환경, 금융, 지역 등 다양한 종류 	<ul style="list-style-type: none"> 무료·유료 혼재 거래소가 데이터 가격 책정 (개인/비상업적/상업적 용도에 따라 가격 상이)

출처: 각 사 홈페이지 및 데이터 거래 활용 활성화를 위한 법제도 개선방안 (입법조사처, 2020.9.30.)

데이터 거래소에서 거래되는 데이터는 주로 공개된 자료에 기반한 정보, 판매자가 보유하고 있는 정보, 판매자가 가공한 정보 등이 거래되며 개인식

별정보 등은 거래되지 않는 것으로 보인다. 개인정보를 제외한 정보의 원데이터 또는 특정 목적을 위해 가공된 정보 등으로 구성되어 있다. 인구통계, 금융, 소매, 스포츠, 환경, 날씨 등에 관한 기록이나 이미지, 공개기록 요청을 통해 수집한 데이터, 개인이 가공한 정보 등 범위는 다양하다. 다만 거래소가 취급하는 데이터 유형별로 형태는 상이한데, Microsoft Azure의 경우 기업간의 거래가 대부분으로 거래 데이터 규모 등이 방대한 반면, Propublica Data Store의 경우 개인 거래가 활발한 것으로 추정된다. 판매자가 판매하고자 하는 데이터 리스트를 사용자가 검색 또는 문의하여 구매하는 방식 위주이며, 거래소를 통해 필요한 데이터를 요청할 수도 있다.

<표 10> Propublica Data Store의 거래 데이터 유형 예시

데이터 유형	예시
거래소가 수집	<ul style="list-style-type: none"> • New Mexico School Discipline <ul style="list-style-type: none"> - 공립학교 데이터베이스를 통해 해당 데이터 수집
판매자가 수집 및 가공	<ul style="list-style-type: none"> • Home Price Impact of Tax Cuts and Jobs Act of 2017 <ul style="list-style-type: none"> - Moody's Analytics의 수석 경제학자 Mark Zandi가 계산한 전국 약 3,000개 카운티의 예상 주택가치 감소에 대한 카운티별 목록

자료: Propublica Data Store 홈페이지 (강경훈, 강성호, 김영일(2023)에서 재인용)

<그림 1> Propublica Data Store에서 거래되는 데이터의 가격 책정 예시

PREMIUM

Police Officer and State Trooper Earnings (New Jersey)

Source	463 municipal police departments, New Jersey State Police
Date Released	January 2022
Dates Covered	2019
Rows	23890
Related Content	The Pay Check
Methodology	How we tracked the pay of 24,000 cops

This data set, created as part of a two-year investigation by NJ Advance Media for NJ.com, includes the 2019 earnings of 21,000 local officers and 2,900 state troopers in New Jersey, including overtime, off-duty jobs and contractual perks that can add tens of thousands of dollars to the average officer's paycheck. It is a comprehensive look at police compensation that researchers said has few parallels in the nation.

The effort required more than 700 public records requests and hundreds of hours of data inputting, since many municipalities provided only paper records that had to be broken down by hand. The resulting interactive database can be found at <https://projects.nj.com/paycheck/>. For more details on how we cleaned and analyzed the data, see our methodology.

PURCHASE THE DATA

Individual License — \$200.00

Individual License — \$200.00

Non-Commercial License — \$600.00

Commercial License — \$2,000.00

SAMPLE

FREE SAMPLE DATA

ADD TO CART

자료: Propublica Data Store 홈페이지

데이터 구매자와 관련하여 거래소 플랫폼에 판매 데이터 목록 등이 공개되어 있는 경우 개인의 접근이 용이한 구조이며, 구매는 개인뿐만 아니라 기업·단체도 가능하다.

3. 영국

영국은 데이터 관련 정책을 정부에서 적극 수립하고 있다. 특히 정보 주체의 자기정보 결정권을 확립함으로써 PDS 플랫폼을 통하여 능동적이고 적극적으로 자기 정보를 관리 및 거래할 수 있도록 있다. 영국은 유럽의 여러 나라 중에서 데이터 산업이 가장 활성화된 국가이다. 영국 정부는 이미 MyData (2011), Open Banking (2018) 정책을 통해 통신, 금융, 에너지 등 여러 산업의 기업들이 보유한 고객정보에 대한 접근을 허용한 바 있다. Hub of All things (HAT)라는 개인데이터 플랫폼을 통해 개인의 정보계좌를 제공하고 있다. 데이터 주체들은 자신의 SNS, 소비, 검색, 건강 등 데이터를 플랫폼을 통해 직접 수집, 관리, 가공할 수 있다. 데이터거래 플랫폼에서는 다양한 데이터가 수집되고 있으며 이용을 위해 거래된다.

대표적인 거래소인 digi.me는 개인데이터저장소(Personal Data Store: PDS)를 통해 개인이 모든 데이터를 한 곳에서 수집 및 관리하는 공유 플랫폼이다. 개인의 금융, 소셜, 엔터테인먼트, 의료, 헬스케어 등과 관련한 여러 정보를 취합하며 정보 주체인 사용자가 데이터를 관리⁴⁾한다. 개인은 PDS를 통한 자기정보 결정권을 직접 행사하며 금전적 보상을 지급받는다. App에 공유되는 개인데이터 거래 당 0.1달러(사용자 1인당 연간 최대 3달러) 또는 digi.me를 통해 창출되는 App 수익의 7.5%가 정보주체에게 지급된다.

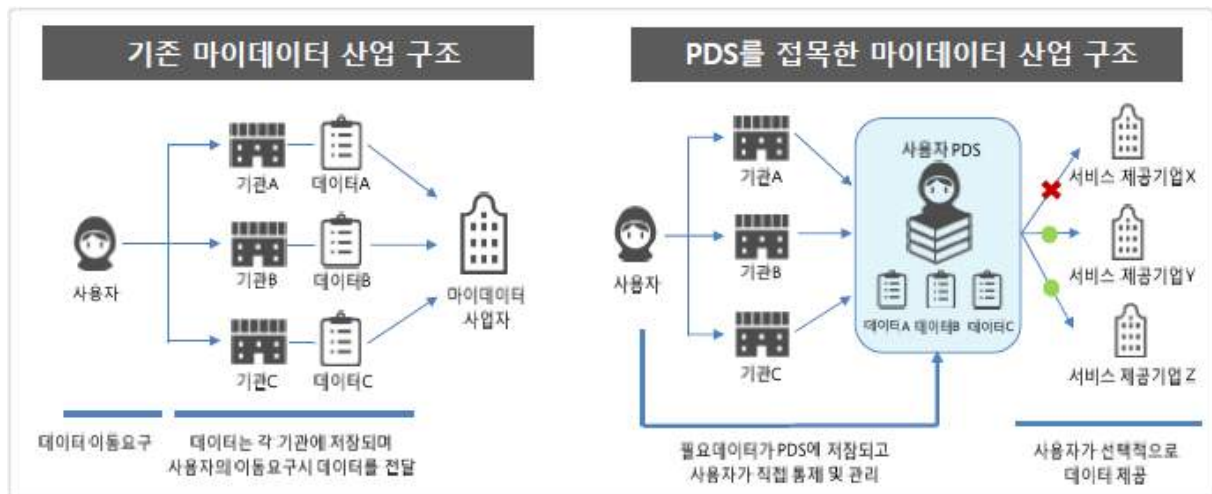
한국에서도 이미 신용정보법을 통한 금융분야 마이데이터 사업이 시작했으며 개인정보보호법 개정을 통해 국내 본인신용정보관리업(MyData) 사업이 시행되면 영국과 같은 PDS 정보수집이 가능할 것으로 보이나, 데이터 유통까지는 허용하지 않고 있는 차이가 존재한다.

영국 정부는 마이데이터 서비스와 개인의 데이터 거래를 하나의 플랫폼

4) 공유 허용, 삭제, 공유 중지, 재공유 등을 포함한다.

안에서 가능하도록 함으로써 정보주체가 주도하는 데이터 유통시장을 만들었다. 개인데이터저장소(Personal Data Store: PDS)가 마이데이터 사업의 또 다른 수익모델로 각광받고 있다. PDS란 개인 동의를 기반으로 정보주체가 자신의 데이터를 수집·저장·사용·공유하는 일을 플랫폼 한 곳에서 관리·통제할 수 있도록 지원하는 서비스이다. 정보주체가 PDS에 접속하여 자신이 무슨 정보를 어느 기업에 제공할지 여부를 직접 결정하는 방식이다. 여러 곳에 흩어져 있는 자신의 데이터를 모아 안전한 장소에 보관하는 한편 어느 데이터를 어느 기업에게, 어느 범위까지 공유할 것인지에 대한 전송요구·제3자 제공 등을 정보 주체의 의지에 따라 관리한다.

<그림 2> 마이데이터 산업 구조의 비교



영국의 주요 PDS 플랫폼에는 digi.me, citizen.me 및 mydex 등이 있다. 먼저 digi.me는 2009년 소셜미디어 백업서비스 업체로 설립되었으며 현재 개인 데이터저장소 운영업체이다. digi.me는 초기 투자금 £120만으로 시작하였는데 2016년 Series A £610만으로 늘어났으며 2021년 early stages of revenue generation으로 진입한 바 있다. digi.me의 자산은 2015년 £120만에서 2018년 £740만으로 성장했다. 2022.10월 프라이버시 및 데이터 공유 기술회사인 World Data Exchange(본사: 시드니)가 인수하였으며 현재 digi.me 앱의 명칭은 Personal Data Explorer로 변경되었으나 아직 digi.me와 혼용하고 있다.

<표 11> 영국의 주요 PDS 플랫폼

PDS명			
수집 정보	<ul style="list-style-type: none"> 금융, 소셜, 헬스케어, 의료, 엔터테인먼트 등 매우 다양 	<ul style="list-style-type: none"> SNS, 금융, 메일, 헬스케어 등 개인정보, 설문조사 분석 결과 등 	<ul style="list-style-type: none"> 자동차, 전기, 세금, 여행, 쇼핑 등 주로 공공기록
공통 특징	<ul style="list-style-type: none"> 정보주체자의 개인정보를 한 곳의 저장소에 모아 공유·사용·관리하는 기능 지원 기업(마이데이터 사업자 등)에게는 개인정보를 안전하게 전달, 사업적으로 활용할 수 있도록 제공 B2B-B2C를 결합한 B2B2C 형태 		
저장 형태	<ul style="list-style-type: none"> 허브-비집중형 : 개인 클라우드 저장 	<ul style="list-style-type: none"> 허브-비집중형 : 스마트폰 개인 클라우드 저장 	<ul style="list-style-type: none"> 허브-집중형 : 정부 클라우드 저장
주요 특징	<ul style="list-style-type: none"> 주로 암호화된 개인데이터를 마이데이터 사업자에게 판매 	<ul style="list-style-type: none"> 주로 개인정보 분석 결과 및 설문조사 결과 등을 기업에게 제공하는 컨설팅 측면의 역할이 큼 암호화된 개인데이터를 판매하기도 함 	<ul style="list-style-type: none"> 정부 인증 기업 (CIC, Community Interest Company) 각 개인 데이터 저장소는 개별적으로 암호화되고, 각 개인은 데이터 저장소에 대한 개인키로 접근 공공서비스 성격이 강해, 민간기업에 비해 데이터 판매 수수료가 낮게 책정

자료: 각 사 홈페이지

digi.me는 개인 동의 하에 여러 곳에 산재해 있는 개인정보를 한 곳에 수집한다. 이러한 개인정보 가운데 필요한 정보를 마이데이터 사업자에게 제공하여 수수료를 취득한다. 개인정보 제공자는 개발된 앱을 바탕으로 다양한 서비스를 제공받으며 마이데이터 사업자는 제공받은 데이터를 활용하여 여러 가지 앱을 개발하고 이를 digi.me 플랫폼에서 상용화할 수 있다. digi.me의 수수료는 데이터 전송(transaction) 당 0.15달러 또는 digi.me를 통해 창출되는 앱 수익의 7.5%인데 프리미엄 서비스 옵션도 존재한다.

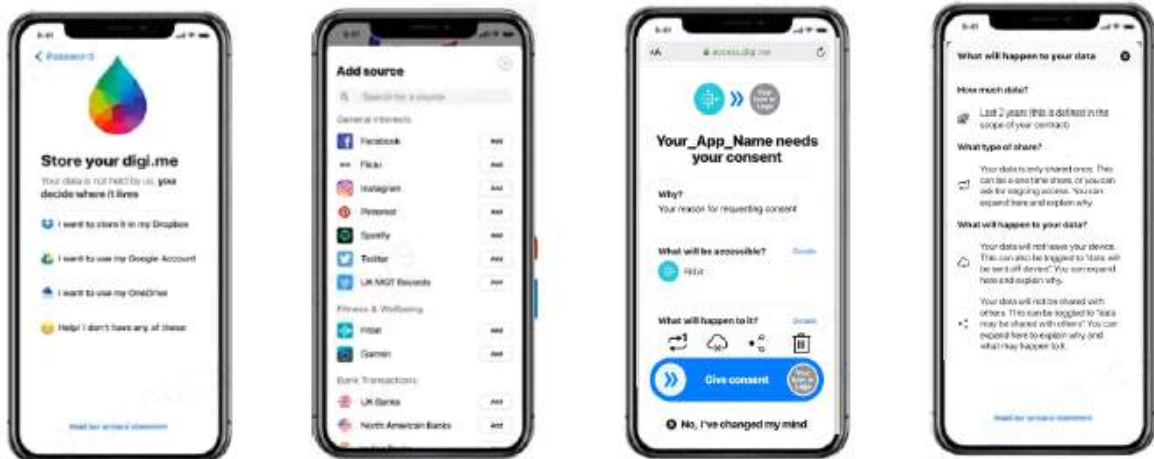
<그림 3> digi.me 앱의 작동 원리



자료: digi.me 홈페이지

<그림 4> digi.me 앱의 이용 절차

1. 저장 스토리지 선택:
Dropbox, OneDrive, Google Drive 중 택1
 2. 수집할 데이터(앱) 선택
→ 각 아이디/패스워드 입력
→ 암호화 방식으로 수집·저장
 3. 마이데이터 사업자의 데이터 공유 요청 시 '동의'를 눌러 진행
- * 어떤 데이터가 활용되는지, 언제 공유를 중단할 수 있는지, 어떤 혜택을 받을 수 있는지 구체적으로 명시



자료: 2020 마이데이터 컨퍼런스 Digi.me 발표

digi.me 사업모델의 특징으로는 동의 기반, 암호화를 통한 뛰어난 보안성 및 데이터 품질 및 최신성 확보 등을 들 수 있다.

<표 12> digi.me 사업모델의 주요 특징

	내용
동의 기반	<ul style="list-style-type: none"> ◆ 플랫폼 내 개인데이터 수집·저장·공유에 이르는 모든 과정을 사용자의 선택에 맡김 ◆ 데이터 공유 조건 등을 명확하게 명시, 데이터 접근이 기록됨 ◆ GDPR·CCPA 등에서 요구하는 '잊혀질 권리' 등의 기능 지원
보안 측면	<ul style="list-style-type: none"> ◆ 수집된 정보는 각 사용자의 클라우드* 저장소에 암호화되어 저장 * Dropbox, Google Drive, OneDrive ◆ 암호화된 정보는 스마트폰에 있는 키 값+사용자가 저장한 암호를 통해서만 해제 가능 ◆ 중개업자로서 중립성을 지키기 위해 데이터를 보거나 저장·변경·활용하지 않음 - 수집된 데이터를 활용하여 자체 서비스를 제공하지 않음
데이터 품질·최신성 확보	<ul style="list-style-type: none"> ◆ 사용자 동의에 근거하여 데이터를 제공받으므로, 품질 및 신뢰성이 우수 ◆ 사용자에 의해 지속적으로 업데이트되는 최신 데이터를 안정적으로 확보 가능 * 데이터 브로커는 온-오프라인을 통해 무작위 수집·가공작업을 거쳐 통합된 형태로 판매되는 경우가 있음

자료: digi.me 홈페이지

digi.me가 수집하는 정보는 금융, 소셜, 헬스케어, 엔터테인먼트, 의료 등 정보 범위 및 확장성이 뛰어나다. 영국, 미국, 네덜란드 3개국의 의료정보도 수집하고 있다.

<표 13> digi.me가 수집하는 주요 데이터

구분	주요내용
금융	• 1,000여개 이상의 은행 및 금융회사의 계좌·금융거래 정보
	• 호주, 뉴질랜드, 영국, 유럽
	• Visa, Mastercard, American Express 등 주요 신용카드 사용 내역
	• 플레이드(Plaid)와 제휴를 통해 데이터 취합 관련 확장성 도모
	• 최대 24개월 거래 기록, 계좌 세부정보 및 잔액 등
소셜미디어	• 페이스북, 플리커(Flickr), 인스타그램, 트위터, 핀터레스트(Pinterest) 등
	• 포스트, 사진, 좋아요 기록, 위치정보, 코멘트 등의 활동내역
의료 기록	• 영국 NHS GP Records, 네덜란드 GP, 미국 Hospital(300개 이상 병원) 등 1,000여개 기관 연계
	• 알리지, 건강상태, 면역기록, 처방 및 약물 기록 등의 데이터 표준화 시스템
헬스케어	• 애플헬스(Apple Health), 핏빗(Fitbit), 가민(Garmin), 구글핏(Google Fit)
	• 최대 24개월 일별 통계
	• 걸음, 거리, 칼로리, 활동수준, 수면활동, 건강데이터 등 실시간 데이터 확보
엔터테인먼트	• 스포티파이(Spotify), 유튜브(YouTube) 등
	• 음악 기록 및 플레이 리스트, 팟로우 채널 및 좋아요 기록 등
자동차	• 영국 등록 차량의 경우, 영국 면허청(Driver and Vehicle Licensing Agency) 연계 데이터 취합
	• 자동차 이력 성능 검사(Motor Ordinance Test, MOT) 상 통과여부, 권고사항, 문제사항 등

자료: digi.me 홈페이지

다음으로 citizen.me는 2013년 설립되었으며 주로 B2B 기반으로 데이터 분석결과를 제공하는 개인데이터저장소 운영업체이다. citizen.me는 초기 투자금 £150만으로 시작하였는데 2021년 £206만 투자 유치에 성공했다. citizen.me가 서비스를 제공하는 국가는 총 158개 국으로 가장 큰 커뮤니티는 영국, 미국, 인도, 브라질 순이며 사용자는 50만 명, 데이터 거래량은 12,00만 건이며 고객사는 약 400여개이다.

citizen.me도 digi.me처럼 여러 곳에 산재해 있는 개인정보를 개인 동의를 받아 한 곳에 수집한다. 다만 데이터 자체의 판매보다는 다양한 데이터를 심층적·다각적으로 분석하여 의미있는 정보를 제공하는 것이 주요 사업모델이다. 개인정보 제공자는 설문조사에 대한 현금 보상을 받으며 앱커뮤니티 활동 및 개인 활동내역 대시보드 등 참여하게 된다. 기업은 마케팅, 신제품 출시에 필요한 소비자 인식 등 설문조사 분석결과를 제공받아 사업에 활용한다. citizen.me는 기본적으로 개인데이터 판매에 대한 수수료를 부과하는데 설문조사 가능 인원, 국가, 범위 등에 따라 월정액 등 다양한 옵션이 있다.

수집 정보의 범위와 관련하여 퀴즈, 이벤트 참여, 설문조사 방식 등을 통해 수집한 소비자 의견 정보를 활용하는 점이 특징이다. 앱 활용내역을 통한 개인의 SNS 활동, 헬스케어 정보, 오프라인 관심사 등 다양한 범위를 포괄한다. 또한 설문조사 결과와 개인정보를 결합하여 분석한 정보도 포함한다.

<표 14> citizen.me가 수집하는 주요 데이터

구분	수집 데이터 출처 및 내용
설문조사	• 설문조사 등을 바탕으로 분석된 정보
건강	• Apple Health Kit
금융	• 은행 앱을 통한 금융거래 기록
소셜미디어	• 페이스북, 트위터, 링크드인 활동 내역
인구통계	• 인구통계 데이터 (연령, 거주지역, 가구 관련 정보)
사이코그래픽	• 25개 매개변수 활용
디지털행위	• 모바일 앱 사용/검색 기록 및 활동 내역, 구글맵
미디어소비	• 유튜브, 넷플릭스 시청 내역
소비	• 아마존, 우버 이츠(Uber Eats), 이베이, 부킹 등 앱

자료: citizen.me 홈페이지

citizen.me의 특징은 아래 표와 같다.

<표 15> citizen.me 사업모델의 주요 특징

내용	
동의 기반	<ul style="list-style-type: none"> ◆ 플랫폼 내 개인 데이터 수집·저장·공유에 이르는 모든 과정을 사용자의 선택에 맡김 ◆ 데이터 전송 시 마다 동의가 필요, 데이터 공유 조건을 등을 명확하게 명시
보안 측면	<ul style="list-style-type: none"> ◆ 가입한 앱 정보를 Citizen.Me 앱에 연결하여 개인 데이터를 스마트폰에 로컬로 저장 ◆ 저장된 개인 데이터는 개인 클라우드(Android, iOS)에 백업할 수 있음 ◆ 개인 데이터는 익명으로 집계된 형식 및 암호화되어 기업으로 전송
기업 컨설팅 서비스 용이	<ul style="list-style-type: none"> ◆ 기업 마케팅 서비스 활용에 용이: 사용자 동의에 근거한 최신 개인 데이터와 해당 사용자를 대상으로 실시한 설문조사 결과를 함께 분석하는 서비스를 제공 ◆ 특정 정보 소비자군에 대한 빠른 정보수집이 가능 <ul style="list-style-type: none"> - 앱 사용자(시민) 간 공통점 등을 기반으로 커뮤니티 활동(시민 간 데이터 공유 가능)

자료: citizen.me 홈페이지

citizen.me의 수수료 옵션은 개인 데이터 자체를 제공하거나 데이터 분석 결과를 제공하고 수수료를 받는 구조로 되어 있다. 상업적 용도로 공유된 데이터에 대한 수수료 및 기업 서비스 마케팅 및 제품 출시 등에 활용되는 기업 컨설팅 서비스에 대한 수수료를 취득한다.

<그림 5> citizen.me의 서비스 이용료 예시

Data Snapshots	Deeper Insight	Tailor-Made
Best for exploring the platform and quick access to small audiences with simple data requirements.	Best for businesses wanting deep data and insight from thousands of global respondents regularly.	Best for businesses who need a customisable platform, regular training, and top level analysis.
<ul style="list-style-type: none"> ✓ Up to 10 surveys /month ✓ US & UK audiences of up to 250 people ✓ Detailed survey data ✓ No minimum contract ✓ Pay-as-you-go survey costs 	<ul style="list-style-type: none"> ✓ Unlimited surveys ✓ Global audiences of up to 2000 people ✓ All types of CitizenData ✓ Full access to survey analysis ✓ 3 month minimum contract ✓ Pay-as-you-go survey costs 	<ul style="list-style-type: none"> ✓ Unlimited surveys ✓ Global, unlimited audiences ✓ All types of CitizenData ✓ Quarterly platform training ✓ 12 month minimum contract ✓ Pay-as-you-go survey costs
Free Pay-as-you-go	£299 Service fee per month	Let's Talk Tailored to suit your organisation's needs

자료: citizen.me 홈페이지

4. 중국

중국은 정부 주도 하에 데이터 거래시장을 빠르게 발전시키고 있는 대표적인 사례이다. 중국 정부는 빅데이터 종합 시범 구를 지정하고 세계 최초의 빅데이터 거래소인 귀양 빅데이터거래소를 2014년에 설립하였다. 중국 데이터시장의 빠른 성장세는 데이터 거래와 플랫폼 시장을 먼저 개척하고 이후 법적·제도적 기반을 마련하여 절차상의 효율성을 추구한 데 힘입은 바 큰 것으로 보인다. 국유자본과 정부주도이면서도 기업 참여와 시장 운영을 원칙으로 함으로써 데이터의 공신력과 신뢰성을 확보하고, 여러 거래 주체의 데이터 권리를 보증하는 한편 거래의 적극성을 유도했다.

특히 귀양에 빅데이터 거래소를 2014년 설립하여 데이터 산업의 빠른 성장을 주도했으며, 빅데이터 표준체계를 구축함으로써 데이터 거래의 활성화를 추진했다. 중국의 데이터 거래소는 산업 데이터가 많지 않고 주로 중국 정부가 보유한 데이터 위주로 구성되어 있다. 이에 따라 데이터 공급자보다 데이터를 구매하려는 수요자 중심의 시장이라고 할 수 있다. 데이터 공급자인 기업이 자발적으로 보유 데이터를 거래소의 구매자들에게 판매하는 것이 주된 운영방식인 국내 또는 해외 주요국의 데이터 거래소와는 차이가 있다.

데이터 거래소는 데이터 매각금액의 10%를 수수료로 책정하여 데이터 거래소의 서비스 비용에 충당한다. 데이터 거래방식은 크게 세 가지인데 가격을 협의하거나, 경매 또는 콜옵션 등이다. 중국정부는 빅데이터 거래와 관련하여 세부적인 표준안⁵⁾을 수립하고 데이터 가격도 표준화한 바 있다. 데이터 거래소에서 거래되는 데이터의 가격은 각종 가격지수를 반영하여 산정되며, 데이터 상품의 유형에 따라 자동으로 가격이 매겨지는 공식을 갖추고 있다.

Ⅲ. 데이터 가치평가 및 규모 측정의 주요 방법론

데이터 거래가 활성화되기 위해서는 데이터 거래 메커니즘뿐 아니라 가치

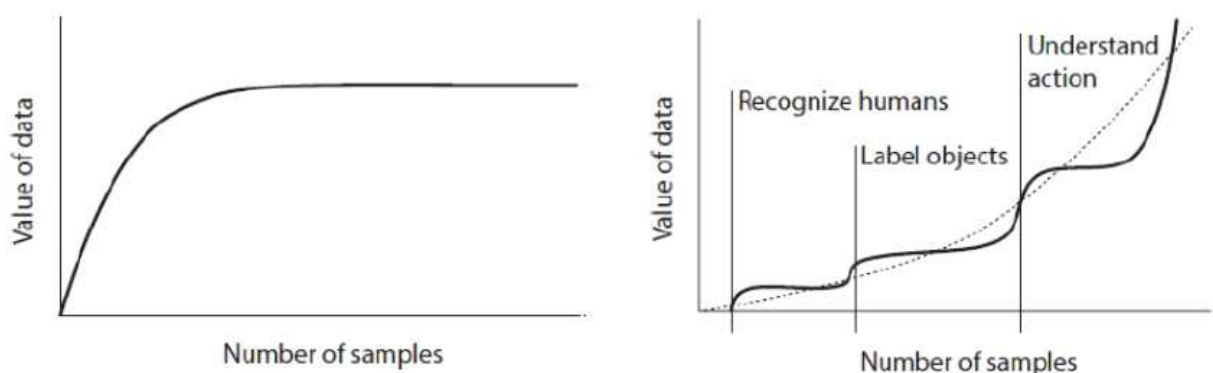
5) 중국 정부는 빅데이터 거래표준, 기술표준, 안전표준, 응용표준을 핵심내용으로 한 1계획 4표준을 발표한 바 있다.

평가 및 가격산정이 정확해야 한다. 그런데 데이터는 일반적인 재화나 서비스와 다른 여러 가지 특이한 성질을 가지고 있어 가치평가 및 프라이싱이 어렵다. 데이터의 특징으로 비경합성(non-rivalry), 복제 가능성(replicability), 결합에 따라 가치 변동폭이 크다는 성질, 사전적인 가치평가가 어려움(usage value not easy to estimate a priori), 재산권을 명확히 규정하기 어려움, 개인 사생활(privacy)과 직결될 수 있음, 플랫폼 및 네트워크 특성 등을 들 수 있다.

데이터 상품의 특이성과 더불어 데이터 프라이싱을 어렵게 만드는 요인도 많다. (Liang et al, 2017) 우선 데이터 원천의 다양성을 들 수 있는데 이는 데이터를 일관적으로 분류하기 어려우며, 데이터를 수집하는 원천이 매우 다양하기 때문에 데이터를 수집하는 데에 소요된 원가계산이 어렵다는 문제이다. 데이터 관리의 복잡성도 있는데 데이터를 관리하고 분석하는 역량과 방법은 매우 다양하며 동일한 데이터라고 하더라도 그 데이터의 가치는 이를 어떻게 관리, 분석, 처리하느냐에 따라 달라진다는 것이다. 데이터의 다양성은 데이터에 대한 수요가 다양하기 때문에 데이터를 어떠한 수준(레벨), 정확성, 유형(타입)으로 나누느냐에 따라 그 가치가 달라진다는 문제이다.

한편 관측치의 수와 데이터의 가치에 대하여 전통적인 통계학과 머신러닝에서의 차이가 있을 수 있다. 데이터의 한계가치를 결정하는 주된 요인은 개별 머신러닝의 복잡도가 아니라 ‘여러 문제에 걸친 복잡도의 분포이기 때문에 전통적인 통계학에서 이야기하는 한계가치 감소 현상이 나타나지 않을 수도 있다.

<그림 6> 문제의 복잡도와 데이터의 한계가치



자료: Posner and Weyl, 2018, Radical Markets

데이터의 가치평가뿐 아니라 데이터의 양을 측정하는 것도 쉽지 않다. 데이터와 관련된 정책을 수립하기 위해서는 데이터의 가치 및 양을 파악할 필요가 있다. 즉 측정이 정책에 앞서야 한다는 것이 상식인데 실제로는 매우 어렵다. 데이터 측정과 관련한 어려움은 다음의 세 가지로 요약할 수 있다.

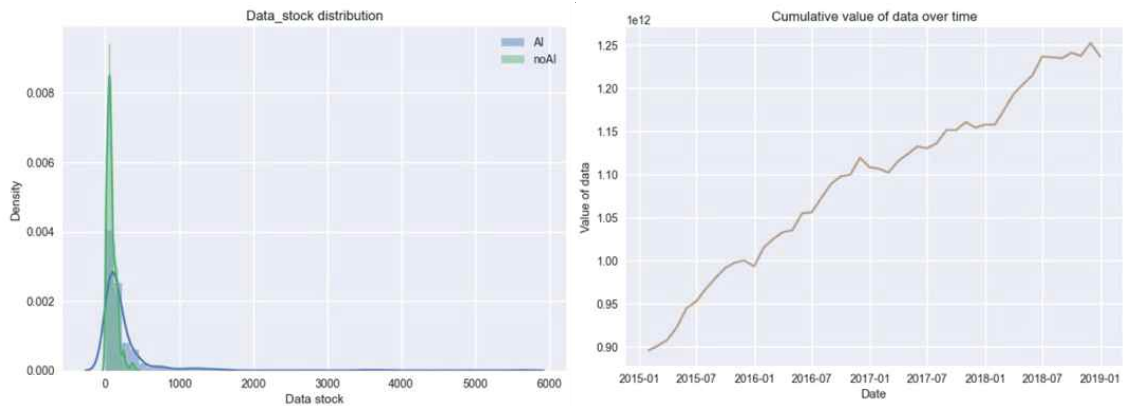
첫째, 대부분의 기업 데이터는 쉽게 관찰할 수 없다. 데이터는 건물처럼 셀 수 있지 않으며, 노동자처럼 세금을 신고하지도 않는다. 데이터를 측정하려면 기업 저장장치에 있는 디지털 기록을 바탕으로 추정해야 하는데 이 기록들은 각 기업의 중요한 비밀인 경우가 많다. 회사가 가치를 창출하는 고도로 보호된 비밀인 경우가 많다.

둘째, 만일 기업의 모든 서버와 드라이브에 접근하여 수집한 모든 데이터를 볼 수 있다고 하더라도 각 데이터의 품질이 다르다는 문제가 있다. 서버나 드라이브의 저장 용량을 비트의 척도로 사용할 수 있는데 해당 데이터의 대부분은 회사와 전략적으로 관련이 없는 경우가 많다. 예를 들어 삭제된 이메일이나 정크 이메일을 모두 저장하고 있는 기업의 가치가 더 높다고 평가하기는 어려울 것이다. 기업은 불확실한 결과를 효과적으로 예측할 수 있는 데이터를 보유함으로써 가치를 얻기 때문이다.

셋째, 동일한 데이터라도 사용자나 기업에 따라 다른 가치를 가질 수 있다. 많은 경우 금융자산은 공통가치(common value)를 가지고 있다고 할 수 있으나 자산으로서의 데이터는 사적가치(private value)의 구성 요소가 크다. 한 기업의 매출을 예측하는 데 유용한 데이터가 다른 기업의 매출을 예측하는 데는 도움이 되지 않을 수 있다. 어느 투자자의 이익에 도움이 될 수 있는 데이터가 다른 거래자에게는 무용할 수도 있다. 기업 수준에서 정확한 데이터 평가를 위해서는 데이터의 사적가치 구성 요소를 고려해야 한다.

이러한 어려움에도 불구하고 경제학자들은 기업 및 경제 전체의 수준에서 데이터를 측정하기 위해 다양한 방법론을 고안하고 있다. 예를 들어 기업 활동이나 성과와 데이터 간의 공분산에 주목하거나 기업의 데이터 관련 노동자 수 및 급여를 분석하기도 한다. 기업의 이윤을 데이터 가치와 연결시키는 연구도 있다. 아래 그림은 금융기관의 데이터 분포와 데이터 관련 노동 자료로부터 추정한 데이터 가치에 대한 연구결과이다.

<그림 7> 금융회사의 데이터 보유 분포 및 데이터 가치 추이



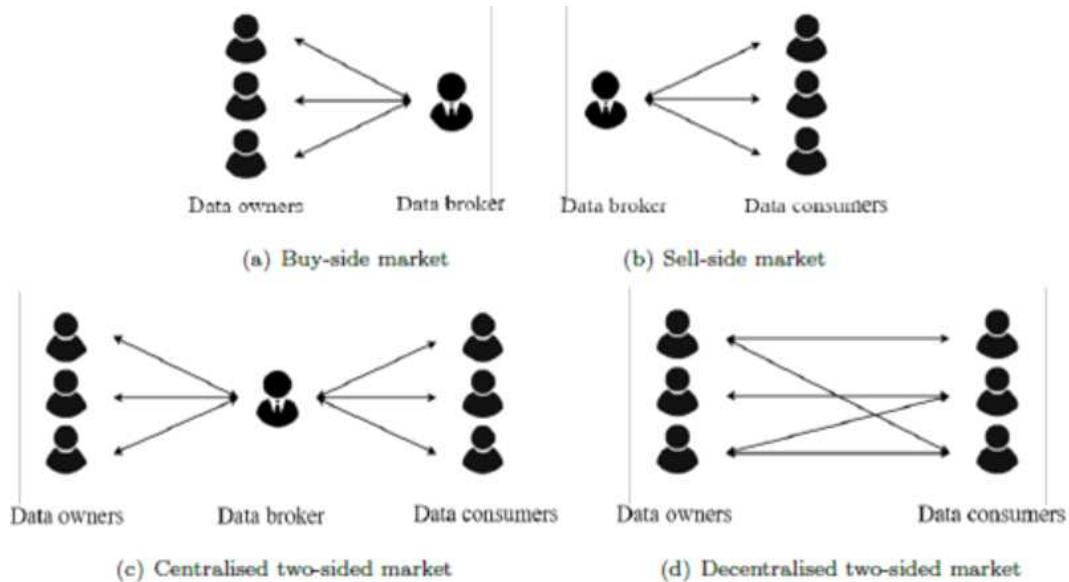
자료: Abis and Veldkamp (2020)

또한 데이터의 가치를 평가하기 위한 연구도 매우 활발하다. 이러한 방법론은 모든 경우에 적용되는 일률적인 접근법이 아니고 다양한 상황에 따라 달라지는 다양한 접근 방식이다. 이하에서는 여러 학자들의 데이터 측정 및 가치평가 방법론을 개략적으로 개관한다.

Zhang et al. (2022)은 데이터 시장을 sell-side, buy-side, two-sided 등으로 유형화하고 유형에 따른 프라이싱 방법론을 소개한 바 있다. 우선 sell-side 시장은 여러 소스의 데이터를 통합하고 데이터 소비자에게 데이터를 판매하는 시장으로서 예를 들어 위치 데이터를 주로 판매하는 AggData 등이 있다. query type에 따라 더욱 세분화하여 null query, one-off query, general query 등에 따라 프라이싱이 달라진다.

buy-side 시장에서는 개인과 조직이 내부 데이터로 수익을 창출할 수 있으므로 데이터 브로커가 데이터 소유자로부터 데이터를 조달한다. 프라이버시 이슈가 중요한데 differential privacy 및 기타 ad hoc privacy로 구분된다. 양면(two-sided)시장에서는 데이터 소유자가 데이터 소비자에게 데이터를 제공할 수 있도록 하여 양면을 결합하는데 대표적으로 Data&Sons 등이 있다. 양면(two-sided)시장은 위 구분들에 따라 더욱 다양하게 세분화된다.

<그림 8> 데이터 시장의 주요 유형

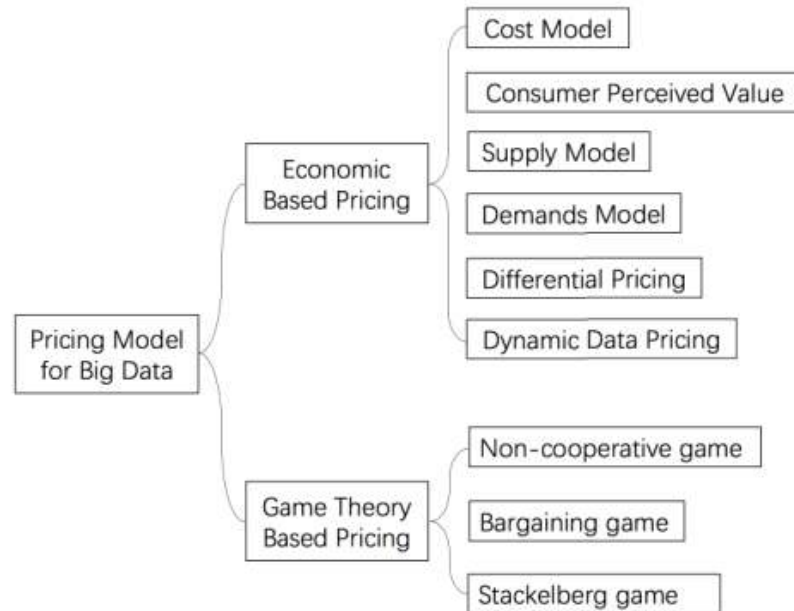


자료: Zhang et al. (2022)

현실의 데이터 마켓·데이터 플랫폼에서는 다양한 가격책정 방식을 사용하고 있으나, 가장 흔히 관찰할 수 있는 데이터 프라이싱 방법은 아래와 같다. (Liang et al, 2017) 우선 Free Data Strategy인데 이는 주로 공공데이터에 적용되는 방식이며 데이터의 판매를 목적으로 하지 않는 데이터베이스들이 사용한다. Usage-Based Pricing은 데이터 사용시간과 사용량, 스트리밍 량에 따라 과금하는 방식이다. Package Pricing Strategy는 기본적으로 사용량에 따라 데이터 가격을 부과한다는 점에서 Usage-Based와 동일하나, 시간이나 사용량에 따라 가격을 차별화한다. Flat Pricing이나 Two-Part Tariff도 많이 사용되고 있다. Freemium의 경우 무료로 데이터를 제공하나, 완전한 기능을 모두 사용하기 위해서는 추가적인 사용료를 납부해야 한다.

다양한 어려움에도 불구하고 그간 학계에서는 데이터의 올바른 가격을 추정하기 위한 많은 연구들이 이루어졌는데 Liang et al (2017)은 기존의 데이터 프라이싱 모델들을 “방법론”에 따라 크게 “경제적 기반”과 “게임이론 기반”의 모델로 구분하고 있다.

<그림 9> 데이터 프라이싱 모델의 구분 (Liang et al., 2017)



자료: Liang et al (2017)

Economic Based Model의 사례로서 아래 모형들을 들 수 있다.

- Cost Model: $\text{Desired Income} = \text{Cost} \times (1 + \text{profit percentage})$
(Rocoroni, 2010)
- Consumer Perceived Model: 소비자들이 느끼는 가치에 따라 가격을 책정하는 방식으로 Harmon et al(2009)은 $P = (\text{Performance}, \text{Environment}, \text{Motivation}, \text{WTP}, \text{Supplier's value}, \text{Economic value})$ 함수를 사용
- 수요-공급모델
- Differential Pricing Model: 다른 데이터와의 비교를 통해 산정
- Dynamic Data Pricing: 시스템의 혼잡도에 따라 데이터 가치를 산정. 피크 타임에는 데이터의 사용료를 경감해주는 모델

한편 Game Theory Based Model의 사례로는 아래 연구들이 있다.

- 비협조게임: Luoing et al. (2016), Yaiche et al. (2000)
- 슈타켈버그 게임: Haddadi and Ghasemi (2016)
- 협상게임 모형: Mao et al. (2017)

Majumdar and Gurtoo (2023)은 데이터 프라이싱 방법론들을 “요소 (Factor)” 에 따라 분류하는데 어떠한 요소가 가격을 결정하는 주된 Factor인지에 따라 기존 연구들을 아래와 같이 5가지 유형으로 구분한다.

<표 16> 데이터 프라이싱 모델의 구분 (Majumdar and Gurtoo, 2023)

Theme	Label	Publication
Theme 1	General Pricing	Heckman et al., (2015), Muschalle et al., (2013), Kushal et al., (2011), Stahl and Vossen, (2013), Zhang& Beltran (2020), Liang et al., (2018), Fricker and Maksimov, (2017)
Theme 2	Quality-Based Pricing	Yang et al., (2019), Yu and Zhang, (2017), Stahl and Vossen, (2016), Tang et al., (2016), Bataineha et al., (2016)
Theme 3	Query-Based Pricing	Koutris et al., (2015), Li and Miklau., (2012), Balazinska et al., (2013), Bergemann and Bonatti, (2015)
Theme 4	Privacy-Based Pricing	Mehta et al., (2019), Li et al., (2014), Lia and Raghunathan, (2013), Shen et al., (2016), Yang and Xing, (2019), Jaising et al., (2008), Gkatzelis et al., (2015)
Theme 5	Special Cases	Harris, (2000), Ye et al., (2018), Li et al., (2017), Golrezaei and Nazerzadeh., (2014), Zheng et al., (2017)

자료: Majumdar and Gurtoo (2023)

Quality Based Pricing은 데이터의 품질을 측정하고 이를 토대로 데이터의 적정가격을 추론하는 방식인데 주로 데이터의 수량, 적시성, 업데이트 여부 등을 데이터의 품질을 결정하는 요인변수로 두고 데이터의 점수와 등급을 산정한다. Query Based Pricing은 수요자들이 필요한 데이터를 검색, 분류하기 위한 Query를 바탕으로 가격을 부과하는 모델이다. Quality Based 방식이 데이터의 공급측면(품질)에 기반한다면, 이는 수요자들의 데이터 사용(수요, 요청)에 의거해 과금하는 방식이다. 주로 데이터베이스를 연구하는 컴퓨터 공학에서 주된 주제로 다루어지고 있다. Privacy Based Pricing은 개인정보

공급자가 상실하게 되는 프라이버시의 가치를 보상하는 방식으로 데이터의 가격을 산정하는 모델이다. 이외에 데이터에 대한 설문조사를 통해 가치를 산정하거나, 데이터 시장의 정보비대칭 등에 중점을 두고 있는 연구들도 존재한다.

IV. 맺음말 및 국내 데이터 거래 활성화를 위한 정책과제

최근 국내외를 막론하고 데이터 경제로의 전환이 빠르게 진행 중에 있으며 이제는 데이터 자체가 상품으로서 시장에서 거래되는 단계에 이르렀다. 그러나 지금까지 살펴본 것처럼 데이터는 일반 재화나 서비스와 성격이 매우 달라 기존의 시장이나 거래소, 나아가 가격기구가 제대로 역할을 하기 어려운 실정이다.

세계 각국에서는 다양한 형태의 데이터 거래 시장이 등장하고 있는데 거래 메커니즘이나 가격 산정 방법론도 가지각색이다. 이는 데이터가 다양하고 데이터의 원천이 다양할 뿐 아니라 상품으로서의 데이터의 특성이 다른 재화와 매우 다르기 때문이다. 반면 국내의 경우 아직까지 데이터 거래 규모가 작을 뿐 아니라 관련 제도 정비도 미흡한 부분이 많다. 데이터 거래 활성화를 위한 정책과제는 매우 다양한데 아래에서는 데이터의 소유권 확립 문제, 데이터의 집중 문제 및 데이터 거버넌스의 세 가지 문제를 중심으로 검토한다.

우선 시장에서 데이터를 거래하도록 하여 데이터의 효율적 배분 또는 이용을 도모하기 위한 첫 번째 과제는 데이터의 소유권을 명확히 하는 것이다. 다만 데이터의 비배제성과 복사 가능성 등을 감안할 때 데이터의 소유권을 정하는 문제가 간단하지 않을 수 있다.

특히 데이터가 일단 만들어지거나 수집되었다면 이를 광범위하게 사용하는 것이 바람직하다. 의료 분야에서 기계 학습을 이용하는 사례를 생각해 보자. MRI, CT 및 초음파 검사에서 검사결과를 판독하고 진단을 내리도록 알고리즘을 훈련할 수 있다. 알고리즘을 훈련하는 데 더 많은 스캔과 보고서를 사용할수록 진단이 더 정확해지는 것은 분명하다. 한 병원의 환자에게만 훈

련된 알고리즘은 국내 모든 환자의 스캔을 사용하여 훈련된 알고리즘보다 열등하다. 모든 병원이 세계에서 가장 재능 있는 방사선 전문의를 동시에 고용할 수 없지만 잠재적으로 최고의 데이터를 사용할 수는 있다.

이러한 점에 주목하여 Jones and Tonetti (2020)는 소비자 개인에게 데이터 소유권을 부여하는 경우 경제 전체의 후생이 증가하고 경제성장에도 도움이 된다는 연구 결과를 제시했다. 소비자들은 자신들의 데이터를 여러 플랫폼, 기업들에 판매함으로써 더 우월한 알고리즘을 만들 뿐 아니라 경제 내 경쟁을 제고한다. 더욱이 자신의 프라이버시 보호 문제를 함께 고려하여 내리는 결정이란 점에서도 바람직한 해법이 될 수 있다.

앞에서 살펴본 영국의 PDS는 현재로서 Jones and Tonetti (2020)가 제시하는 정책 방향에 가장 들어맞는 해결책으로 보인다. 또한 일본의 정보은행제도 역시 개인 정보주체에게 개인정보 활용에 따른 수익을 환원한다는 점에서 유사하다. 다만 아직까지 실적이 많지 않아 digi.me 또는 citizen.me 등처럼 효과적으로 작동하는지 판단하기 어렵다. 또한 일본 정보은행의 경우 개인의 어느 정보가 어디에 제공되었는지 정보은행을 통해 확인하고 통제하기 때문에 정보은행의 업무 부담이 훨씬 더 클 수 있다.

데이터의 소유권을 법적으로 정의하는 것은 또 다른 난제이다. 최근 데이터 거래가 활성화됨에 따라 법학계, 국회 등에서는 데이터의 소유권을 인정할 것인지 또한 어떤 개념으로 어떤 개념에 포섭해야 할지 등에 대해 논의가 진행하고 있으며 다양한 법률 개정안이 발의되기도 하였다. 이러한 논의는 국내외를 막론하고 활발하게 이루어지고 있으며 여러 가지 방안이 백가쟁명으로 제시되고 있어 아직 뚜렷한 방향성이 제시되지 못한 상태로 보인다.

이러한 법률적 문제는 이 보고서의 범위를 넘어서는 것이어서 자세히 다루기는 어려운데 한 가지 강조되어야 할 점은 데이터의 소유권을 정의하고 확립하는 문제는 기존 법률체계를 감안하여 법학자들이 다루어야 하지만 그 큰 그림은 데이터 경제의 생태계를 활성화시키는 방향이어야 한다는 것이다. Jones and Tonetti (2020)는 법률적 문제를 추상하고 경제학 이론에 입각해서 데이터 소유권의 문제를 검토했는데 데이터의 종류가 다양하고 데이터 거래 역시 매우 세분화될 수 있는 만큼 데이터 소유권에 대한 경제적 분석도 계

속 진행될 필요가 있다.

한 가지 더 강조할 점은 데이터의 소유권을 정의하고 확립하는 문제는 관련 기술 발전에 따라 끊임없이 변모할 수 있다는 점이다. 경합성이나 비배제성의 정도가 달라질 수도 있으며 이전에 생각지 못하던 전혀 새로운 해법이 제시될 수도 있다. 나아가 데이터 경제의 제도를 설계할 때 이처럼 새롭고 창의적인 해법이 튀어나올 수 있도록 유인을 마련해 놓는 것도 중요하다.

한편 개인 정보주체의 데이터 소유권을 확립하는 것이 데이터 경제 생태계를 활성화시키는 완벽한 해결책에 미흡할 수도 있다. 개인 모두가 자신의 데이터 관리에 얼마나 시간을 들일 수 있는지에 대해서 낙관적인 답변이 나오기 어려울 수 있기 때문이다. 영국의 digi.me 등에서 개인에게 데이터 이용에 대한 대가를 지불하는 모델을 보여주고 있지만 금액이 매우 작기 때문에 많은 개인들은 데이터 관리 자체를 무시하거나 경시할 수 있다. 특히 소득이 높거나 바쁜 개인들이 이 경향을 보일 가능성이 더 높는데 데이터 이용 업체의 입장에서는 이 개인들의 데이터가 더 가치가 있을 수 있다.

이처럼 개인의 데이터 관리 및 거래가 이상적인 방법일 수는 있지만 현실성이 부족할 수 있다. 이에 대한 해법으로 제시되는 데이터 유니언(data union), 데이터 트러스트(data trust) 또는 마이데이터의 메커니즘을 개선하고 PDS의 주요 장점을 접목시킬 필요가 있는데 쉽지 않은 과제로 보인다.

또한 식별정보를 없앤 가명정보 또는 익명정보의 거래에 대해서는 이러한 메커니즘과 다른 방법론을 사용할 수 있다. 다만 여기서도 가명 또는 익명정보를 만들기 위한 재료로서의 개인정보에 대해서 어떻게 보상을 해줄 것인지 고려가 필요하다. 가명정보 및 익명정보의 거래 메커니즘도 일반 재화나 증권의 거래 메커니즘과 크게 다르며 식별이 가능한 개인정보의 거래 메커니즘과도 달라질 수 있음에 유의할 필요가 있다.

한편 데이터의 배분 문제를 시장기구에만 맡길 수 있는지에 대해서도 검토가 필요하다. 최근 데이터의 중요성이 커지고 이를 둘러싼 경쟁이 격화되면서 데이터 집중에 대한 우려도 늘어나고 있다. 예를 들어 데이터를 많이 확보한 기업이 독점하는 승자독식 시장이 되면서 많은 기업이 퇴출될 수 있다는 것이다. 데이터 기반 산업에 진입하는 신규 기업들이 데이터 부족으로 인해 부족한 경쟁력을 쉽게 끌어올리지 못하는 이른바 콜드스타트(cold start)

문제도 심심치 않게 들린다.

그리고 데이터를 많이 보유하고 있는 플랫폼들이 폐쇄형 생태계로 발전하려는 경향을 보이는 것도 문제다. 플랫폼들은 소비자가 되도록 많은 활동에 동일한 플랫폼을 사용하도록 유도함으로써 더 많고 더 다양한 데이터를 보유하려고 한다. 이들 입장에서는 경쟁 우위를 제공하는 사용자 데이터를 경쟁상대와 공유하거나 판매하고 싶어 하지 않을 수 있다. 더 많은 데이터를 독점적으로 사용하는 경우 경쟁력을 최대로 키울 수 있을 것이다.

EU 경쟁당국의 수장인 Margrethe Vestager에 따르면 “경쟁에 긴요한 중요 자원을 소수의 회사만이 통제하고 이를 공유하기를 거부한다면 경쟁은 성공할 수 없는데, 데이터가 이러한 중요자원 중 하나가 되어가고 있다.⁶⁾” 이러한 맥락에서 최근 미국, EU 등 해외 주요국들은 데이터의 집중 또는 독점적 지배를 규제하려는 움직임을 보이고 있다. EU는 데이터 공유(pooling or sharing)와 관련하여 다양한 제안을 내놓고 있으며 미국도 플랫폼의 경제력 집중을 억제하는 규제 법안을 발의하면서 데이터 관련 규제도 포함시킨 바 있다.

여기서 한 가지 유의할 점은 대부분의 해외 주요국에서 데이터 집중을 해소하는 문제를 필수설비원칙(essential facilities doctrine)을 통해 접근하지는 않는다는 것이다. 최근 데이터 경제가 진전됨에 따라 데이터의 중요성이 날로 커지고 있으며 데이터를 필수설비로 간주해도 무방할 것으로 보이는 것이 사실이다. 다만 필수설비원칙을 이용하여 데이터 독과점 기업을 규율하기 위해서는 어떤 종류, 어느 범위의 데이터가 필수설비인지 보여야 하는데 이 문제가 매우 난해하다. 또한 경쟁업체가 필수설비에 해당하는 데이터를 생산할 수 없다는 것을 입증해야 하는데 이 역시 쉽지 않다. 규모는 작더라도 어느 정도의 데이터를 만드는 것이 어렵지는 않기 때문이다. 결국 필수설비원칙을 이용한 법적 다툼에서 승소한 적이 거의 없는 것으로 알려져 있다. EU에서 데이터 공유 의무제 등을 도입하는 것은 기업 간 경쟁에서 데이터가 중요하다는 사실에 따른 것이지만 필수설비원칙에 직접 기반하지는 않은 것으로 알려져 있다.

6) Competition can't work if just a few companies control a vital resource that you need to be able to compete - and if they refuse to share it with you. Right now, it looks as though data is becoming one of those vital resources.

한국의 경우 데이터 거래가 활성화되지 못한 상태여서 대부분의 기업들이 주로 내부 데이터에 의존하는 상태이므로 데이터 집종의 문제는 해외 주요국에 비해 더욱 심각할 가능성이 있다. 이에 따라 EU에서와 같이 데이터 의무 공유제의 도입을 검토할 필요가 있다. 데이터 공유를 의무화하지 않고도 미국에서처럼 데이터 거래 활성화를 통하여 데이터가 널리 이용될 수 있다면 좋겠으나 현실성이 높지 않기 때문이다.

다만 데이터 의무 공유제 등을 도입하는 경우에도 개인에 대한 데이터의 경우에는 데이터 보호 규칙이 준수되어야 한다. 또한 데이터를 수집하거나 생산하는 기업이나 기관의 입장에서 계속 데이터를 수집·생산할 유인이 사라지지 않도록 해야 한다. 데이터를 수집하고 생산하는 데에도 노력과 시간이 필요한데 기업이 적절한 보상 없이 데이터를 외부 기업이나 기관과 공유해야 한다면 향후 데이터 수집 및 생산에 필요한 노력을 줄일 수 있기 때문이다.

결국 데이터를 둘러싼 경제생태계는 시장기구는 물론이고 데이터 집중 문제에 대응한 규제체계까지 포함하여 설계되어야 한다. 데이터 소유권을 확립하고 거래규칙을 잘 만들어 시장 거래가 활발히 이루어지도록 하는 한편 데이터 집중에 따른 폐해를 방지하기 위해 데이터 공유제 등도 포함되어야 한다. 이와 함께 프라이버시 보호가 일상적으로 끊임없이 이루어져야 하므로 제대로 잘 작동하는 데이터 경제생태계를 구축하는 일은 매우 어려운 과제이다. 하지만 목표가 멀고 어렵더라도 어차피 갈 수밖에 없는 길이며 이 길에서 낙오하는 나라나 경제는 도태될 것이 분명하다. 데이터 주체와 데이터 사용자 간, 데이터를 수집하거나 이용하는 기업들 간 견제, 균형과 협력이 이루어지는 데이터 경제생태계는 향후 경제발전에 필수적인 요소이기 때문이다.

이처럼 다양한 데이터 시장과 제도들을 아우르는 데이터 경제생태계를 어떻게 구축하느냐의 문제는 이 보고서에 다 담기 어려운 난제이다. 다만 그 출발점 중의 하나는 효율적인 데이터 거버넌스 체계를 만드는 데 있다. 정용찬(2017)에 따르면 “데이터 거버넌스는 공공과 민간 데이터를 포괄한 국가 전체 데이터 자산 관리에 대한 권한, 통제, 의사 결정으로 정의된다.” 그런데 현재 한국에서는 현재 데이터와 관련된 정부의 정책기능이 여러 부처에

흩어져 있으므로 이를 조정할 필요가 있다. 한국행정연구원(2020)도 “공공데이터뿐만 아니라 민간데이터를 포함하여 국민과 밀접한 데이터를 수집·관리·공개·활용 등 통합 관리하고 활성화하는 주도적 역할을 책임질 조직이 필요하다”고 지적한 바 있다.

현재 한국에서는 공공데이터 관련 정책은 행정안전부가 맡고 있으며 민간데이터 관련 정책기능은 과학기술정보통신부가 담당하고 있다. 여기에 더해 개인정보보호를 전담하는 개인정보보호위원회가 있으며 금융부문의 개인정보 등 데이터는 금융위원회가 맡고 있어 데이터를 활용하는 데 있어 여러 가지 애로가 발생하고 있는 실정이다.

먼저 금융, 비금융 등 데이터나 정보의 종류에 따라 정책당국이 달라지는 경우 효과적인 정책대응에 애로가 생기기 쉽다. 것이다. 현재 우리나라의 경우 개인정보보호위원회를 중심으로 데이터 보호 거버넌스 체계가 구축됐지만 산업 분야별 제정된 개별법 규제가 따로 있어 공백 혹은 중첩 문제가 발생하고 있다.⁷⁾ 향후 데이터 경제를 활성화하는 한편 개인정보 보호원칙도 충실히 수행하기 위해서는 거버넌스 체계를 일관화가 시급하다는 지적이 많다.

데이터 정책의 거버넌스를 통일하는 문제는 되도록 넓은 범위의 데이터가 결합될 때 그 가치가 커진다는 점에서 중요하다. 금융데이터와 비금융데이터의 결합으로부터 부가가치 창출이 크다는 점은 잘 알려진 사실이다. 예를 들어 금융 데이터와 비금융 데이터를 결합하여 생성한 빅데이터를 활용하면 신용등급 평가가 훨씬 더 정확해진다는 것은 잘 알려진 사실이다. 플랫폼에 속하거나 제휴 맺은 기업의 금융, 비금융 거래 데이터를 통해 기업의 상황을 완벽하게 파악하는 것도 가능하다. 또한 플랫폼을 통해 운영되는 기업이 대출금을 상환하지 않으면 다른 비재무적 운영에도 악영향을 미칠 수 있다. 소비자나 동료 기업들이 대출 원리금 상환을 연체한 기업과 거래를 하지 않는 것은 일반적인 현상이기 때문이다. 그런데 동일한 플랫폼 내에서는 연체정보가 더욱 빠르고 폭넓게 전파될 수 있으므로 이러한 압박을 받는 기업은 성실하게 대출금을 상환하려고 노력할 것이다. 이뿐 아니라 이질적인 데이터들을 결합하여 부가가치를 높이는 사례는 무수히 많이 있다.

7) 금융 분야는 신용정보법, 지리정보 분야는 위치정보법, 통신 분야는 전기통신사업법 등에서 개인정보와 관련한 규정을 두고 있으며 일반법인 개인정보보호법과 상충되거나 중첩되는 문제가 종종 발생한다.

만일 데이터 정책의 거버넌스가 데이터의 종류에 따라 분산되어 있는 경우 데이터 결합에 방해가 될 것은 쉽게 알 수 있다. 예를 들어 과거 신용정보법에 기반한 마이데이터 사업자들 간에 데이터 공유 범위에 대해 논의하는 과정에서 고객이 구매한 신발의 브랜드가 신용정보인지 아닌지를 두고 논란이 된 적 있다. 이커머스에 기반한 마이데이터 업체들은 신용정보가 아니니 공유하기 어렵다는 입장이었고 금융회사에 기반한 마이데이터 업체들은 필요한 정보라는 입장이었다. 만약 개인정보보호법에 기초하여 마이데이터 업이 도입되었다면 하등의 필요가 없는 논쟁이었다고 할 수 있다.

결국 금융·비금융을 구별하지 않고 민간 및 공공 데이터 관련 정책을 총괄 추진할 수 있는 체계와 법제도적 기반을 마련할 필요가 있다. 다만 데이터 정책을 담당하는 기존의 부처·기관과의 관계를 충분히 고려할 필요가 있다. 만일 새로운 부처나 기구에서 데이터 정책을 총괄한다면 구체적인 설립 형태, 업무범위 및 역할 등에 대해 충분히 논의할 필요가 있다. 데이터 정책의 거버넌스 체계를 통일하는 문제와 관련하여 앞에서 언급한 것처럼 현재의 개인정보보호위원회로 통일하는 방안이 유일한 수단은 아니다. 영국의 Digital Markets Unit 또는 미국의 Data Authority와 같은 정부기구를 신설하는 방안도 검토해볼 수 있다. 데이터 경제생태계는 개인정보만 취급하는 것이 아니라 가명, 익명정보 및 비개인데이터 등 매우 다양한 데이터들을 취급하기 때문이다.

<참고 문헌>

- 강경훈, 강성호, 김영일, 2023, 해외 주요국의 데이터 거래 및 가격설정 사례 분석 및 한국의 데이터 정책 방향 모색, Working Paper
- 김수봉, 권현영, 2022, “금융데이터거래 정보보호 강화방안: 데이터브로커 보안이슈를 중심으로,” 정보보호학회논문지 제32권 제1호, 141-154
- 김수진, 이정현, 박천웅, 2021, “데이터거래 활성화를 위한 데이터상품가치 평가모델 연구,” 한국콘텐츠학회논문지, Vol. 21 No. 12, 34-46
- 신용우, 2020, 데이터 경제 활성화를 위한 입법정책 방안, 국회입법조사처 입법·정책보고서 Vol. 69
- 정용찬, 4차 산업혁명 시대의 데이터 경제 활성화 전략, KISDI Premium Report, 2017. 6.
- 정준화, 박소영, 2022, “데이터 거래 활성화를 위한 거래소·거래사·크롤링의 현황과 개선과제,” 국회입법조사처, NARS 현안분석, 제254호
- 한국행정연구원, 2020, “데이터 거버넌스의 현안 및 쟁점”, 정부디자인 ISSUE, 제5호
- Abis, S. and L. Veldkamp, 2020, “The changing economics of knowledge production”, Working Paper
- Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian and Asu Ozdaglar, 2019, Too Much Data: Prices and Inefficiencies in Data Markets, Working Paper
- Azcoitia, Santiago Andrés and Nikolaos Laoutaris, 2022, “A Survey of Data Marketplaces and Their Business Models,” arXiv:2201.04561 [cs.DB], <https://doi.org/10.48550/arXiv.2201.04561>
- Bataineha, Ahmed Saleh., Mizounib, Rabeb., Barachic, May El., and Bentahar, Jamal., (2016). Monetizing Personal Data: A Two-Sided Market Approach. Procedia Computer Science, Volume.83: 472-479.
- Bergemann, Dirk., and Bonatti, Alessandro., (2015). Selling Cookies. Cowles

Foundation Discussion Paper No. 1920R.

- Fricker, Samuel. A., and Maksimov, Yuliy V., (2017). Pricing of Data Products in Data Marketplaces. International Conference on Software Business.
- Golrezaei, Negin., and Nazerzadeh, Hamid., (2014). Pricing Scheme for Metropolitan Traffic Data Markets. In Proceedings of 3rd International Conference on Data Management Technologies and Applications: 266-271
- Haddadi, S. and A. Ghasemi, 2016, "Pricing-based stackelberg game for spectrum trading in self-organised heterogeneous networks," IET Commun., vol. 10, no. 11, pp. 13741383, Nov. 2016.
- Harmon, R., H. Demirkan, B. Heey, and N. Auseklis, 2009, "Pricing strategies for information technology services: A value-based approach," in Proc. 42nd Hawaii Int. Conf. Syst. Sci. (HICSS), 1-10.
- Harris, Ray., (2000). Data Pricing Policy. International Archives of Photogrammetry and Remote Sensing. Vol. XXXIII, Part B6:341-350
- Heckman, J.R., Boehmer, E.L., Peters, E.H., Davaloo, M., Kurup, N.G. (2015). A Pricing Model for Data Markets. In iConference 2015 Proceedings.
- Koutris, Paraschos., Upadhayaya, Prasang., Balazinska, Magdalena., Howe, Bill., and Suciu, Dan., (2015). Query-Based Data Pricing. Journal of the ACM. Vol. 62, Issue 5, 1-44
- Kushal, Avanish., Moorthy, Sharmadha., and Kumar, Vikash, 2011, Pricing for Data Markets, University of Washington Working Paper
- Li, Chao., and Miklau, Gerome., (2012). Pricing Aggregate Queries in a Data Marketplace. Journal of the ACM.
- Li, Chao., Li, Yang Daniel., Miklau, Gerome., and Suciu, Dan., (2014). A Theory of Pricing Private Data. ACM Transactions on Database Systems. Volume.39, Issue. 4:1-28.
- Li, Xijun., Yao, Jianguo., Liu, Xue., and Guan, Haibing., (2017). A First

- Look at Information Entropy- Based Data Pricing. IEEE 37th International Conference on Distributed Computing Systems.
- Lia Bai, Xiao., and Raghunathan, Srinivasan., (2013). Pricing and Disseminating Customer Data with Privacy Awareness. *Decis Support Syst.*59: 63 - 73.
- Liang, Fan., Yu, Wei., An, Dou., Qingyu, Yang., Fu, Xinwen., and Zhao, Wei., (2017). A Survey on Big Data Market- Pricing, Trading and Protection. IEEE; Special Section on Privacy Preservation for Large-Scale User Data in Social Networks, Vol.6:15132-15154
- Majumdar, Rupsa and Anjula Gurtoo, 2023, Data Pricing Models for Data Exchange: Review Analysis, Working Paper
- Mao, Y., T. Cheng, H. Zhao, and N. Shen, 2017, "A strategic bargaining game for a spectrum sharing scheme in cognitive radio-based heterogeneous wireless sensor networks," *Sensors*, vol. 17, no. 12, p. 2737
- Mehta, Sameer and Dawande, Milind and Janakiraman, Ganesh and Mookerjee, Vijay. (2019). How to Sell a Dataset? Pricing Policies for Data Monetization. *Information Systems Research*, Forthcoming
- Muschalle, Alexander., Stahl, Florian., and Vossen, Gottfried. (2013.) Pricing Approaches for Data Markets. M. Castellanos, U.Dayal, and E.A. Rundensteiner (Eds.): *BIRTE 2012*, LNBIP 154: 129 - 144.
- Pei, Jian, 2022, "A Survey on Data Pricing: From Economics to Data Science," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 10, October 2022
- Posner, Eric A. and Eric Glen Weyl, 2018, *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*, Princeton University Press
- Roncoroni, A., 2010, "Commodity price models," in *Encyclopedia of Quantitative Finance*. Hoboken, NJ, USA: Wiley
- Shen, Yuncheng., Guo, Bing., Shen, Yan., Duan, Xuliang., Dong,

- Xiangquian., and Hong, Zhang., (2016).A Pricing Model for Big Personal Data. Tsinghua Science and Technology.Vol,21; No 5 :482-490.
- Stahl, Florian., and Vossen, Gottfried.,(2016). Fair Knapsack Pricing for Data Marketplaces. J. Pokorn' et al. (Eds.): ADBIS 2016, LNCS 9809: 46 - 59,
- Tang, Ruiming., Amarilli, Antoine., Senellart, Pierre., and Bressan, Stephane., (2016).A Framework for Sampling-Based XML Data Pricing. A. Hameurlain et al. (Eds.): TLDKS XXIV, LNCS 9510: 116 - 138
- Yang, Jian., Zhao, Chongchong., and Xing, Chunxiao. (2019). Big Data Market Optimization Pricing Model Based on Data Quality. Wiley
- Yang, Jian., and Xing, Chunxiao., (2019). Personal Data Market Optimization Pricing Model Based on Privacy Level. Information.10(4):123.
- Ye, Peng., Qian, Julian., Chen, Jieying., Wu, Chen-Hung., Zhou, Yitong., Mars, Spencers De., (2018). Customized Regression Model for Airbnb Dynamic Pricing. Applied Data Science Track Paper.
- Zhang, Mengxiao., Fernando Beltran and Jiamou Liu, 2022, A survey of Data Pricing for Data Marketplaces, Working Paper
- Zheng, Zhenzhe., Peng, Yanquing., Wu, Fan., Tang, Shaojie., and Chen, Guihai. (2017). An Online Pricing Mechanism for Mobile Crowdsensing Data Markets. Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing. Article.6(July): 1-10