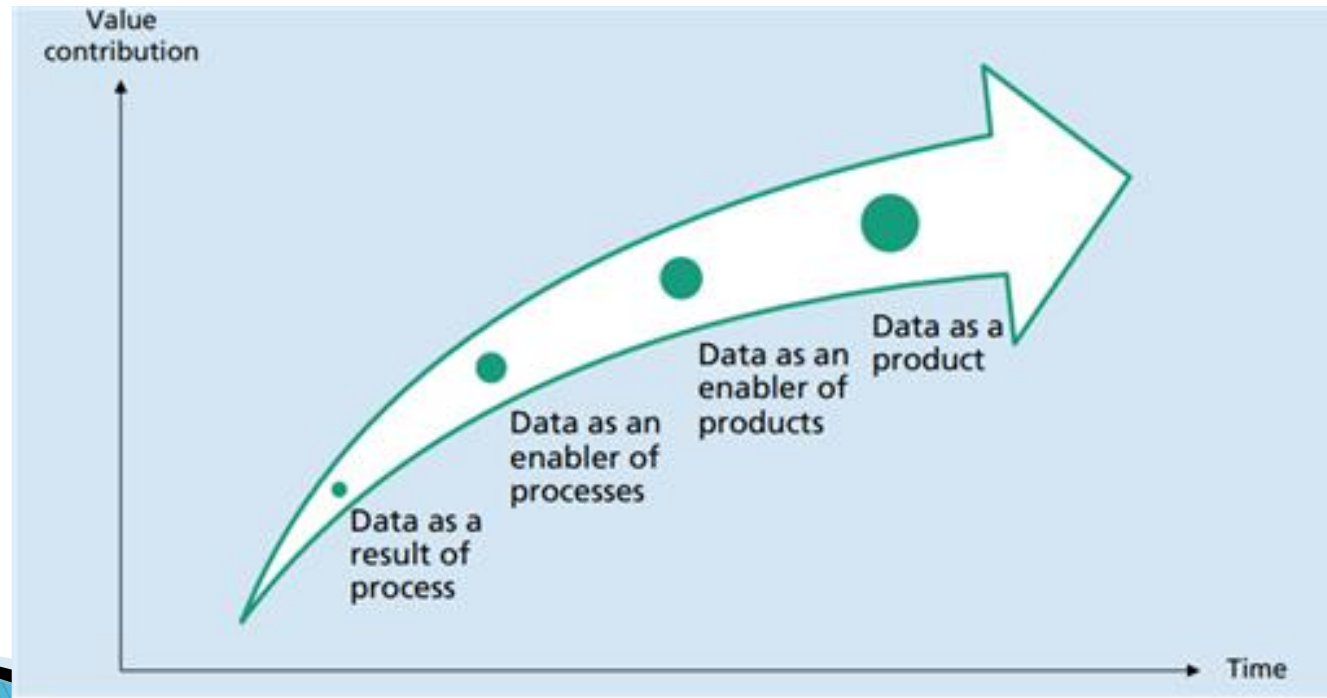


# 데이터 거래의 이론과 실제 및 향후 정책과제

강경훈, 강성호, 전성인  
(동국대) (금융위, UNIDO) (홍익대)

# 상품으로서의 데이터

- 21세기의 원유라고 불리는 데이터의 중요성은 이전에도 잘 알려져 있었는데 시간이 흐름에 따라 데이터가 제공하는 기능이나 역할이 변천
  - 과거에 기업활동의 결과물이었던 데이터는 이후 생산 및 경영과정의 협조, 신제품과 서비스 개발 촉진 역할에서 이제는 상품으로서 시장에서 거래



# 글로벌 데이터 거래 개관 (1)

- Azcoitia et al. (2021)은 데이터거래 산업에 종사하는 데이터 기업을 세 가지로 분류
  - Data Market, Data Provider, Personal Information Management System

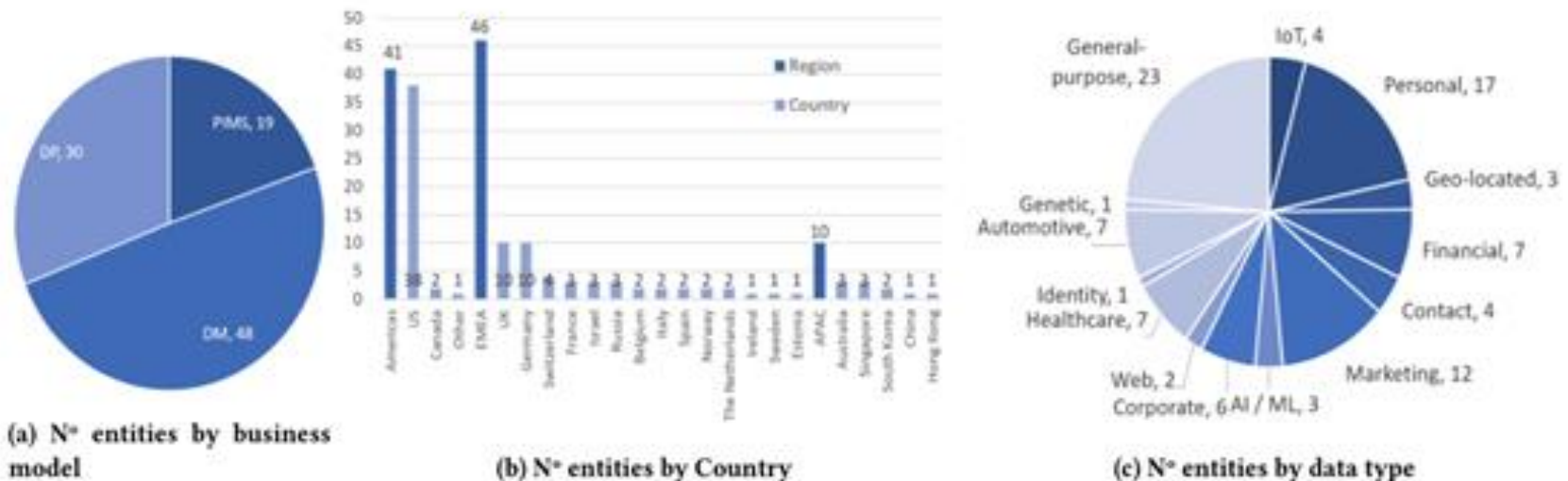


Figure 2: Scope of our survey on entities trading with data

## 글로벌 데이터 거래 개관 (2)

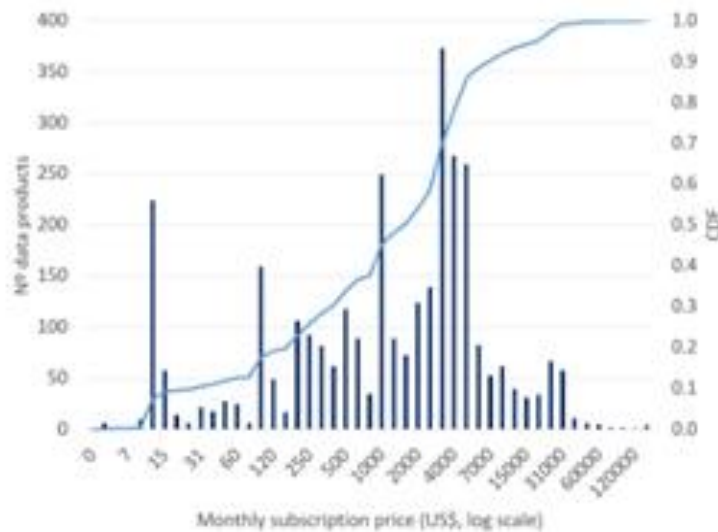
- Azcoitia et al. (2021)에 따르면 데이터 거래가 가장 활발한 나라는 미국
  - 미국은 글로벌 데이터 거래의 30%를 차지하며 캐나다(9.3%), UK(9.2%), 독일(7.6%), 프랑스(7.4%), 스페인(7.1%)이 그 다음 순서



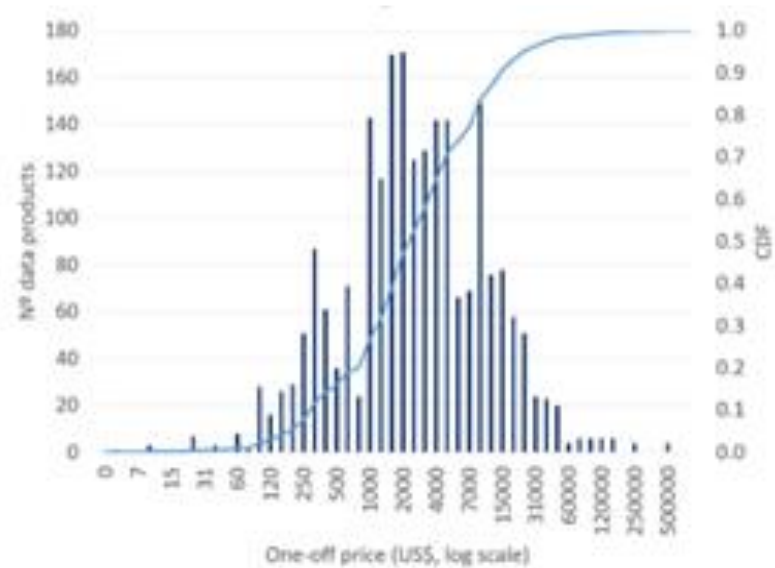
Figure 3: Data products by country

# 글로벌 데이터 거래 개관 (3)

- Azcoitia et al. (2021)에 따르면 데이터 상품의 과금 방식은 크게 정기구독 (Subscription) 방식과 일회성 구매(One-off) 방식으로 구분
  - 구독방식과 일회성 구매방식의 중위값은 각각 월 1,417 USD 및 2,176 USD로서 일회성 방식이 구독방식보다 비싸게 거래되는 것으로 평가



(a) Subscription-based



(b) Fixed price (one-off)

# 미국의 데이터 거래 현황 (1)

- 미국에서는 주로 '데이터 브로커'를 중심으로 데이터 거래가 활발히 이루어짐
  - 그 밖에 데이터 거래소, 직접 판매, 제휴 및 인수·합병 등의 방식도 활용
  - 데이터 브로커는 판매자, 거래소, 거래사가 결합되어 운영되는 형태로 볼 수 있으며 주로 B2B 거래 중심

기업명	주요 서비스	비고
Acxiom	맞춤형 광고정보 제공	· 전세계 9억6,500만 개 이상의 소비자 명부를 보유
CoreLogic	맞춤형 광고정보 제공	· 약 8억 건의 부동산 거래정보, 약 1억 건의 담보 DB를 자체 가공 및 분석
Verisk	맞춤형 광고정보 제공	· 600만 개 이상의 상업용 부동산, 14억 건 이상의 보험 사기 데이터, 18억 개 이상 소비자 카드·저축계좌 정보 제공
Qlik	통계 관련 시각화 정보 판매	· 35개국 주식시세정보 등 통계정보 시각화 및 직접 판매
LexisNexis	맞춤형 광고정보 제공	· 2억8,300만 개 이상 미국 소비자 프로필, 7,700만 개의 비즈니스 명부 보유




# 미국의 데이터 거래 현황 (2)

- 데이터 브로커가 거래하는 데이터는 개인정보를 포함하여 매우 광범위하며 원데이터뿐 아니라 새롭게 생성한 고객 정보 및 분석 서비스도 제공
  - 거래 데이터는 이름, 현재 및 이전 주소, 생일, 성별, 혼인여부, 가족상태, 사회보장번호, 교육수준, 정치성향, 자산, 직업, 전화번호, 이메일 주소, 구매습관, 의료정보 등 다양한 정보를 포괄
  - 미국의 데이터 유통시장이 성장한 것은 사후거부권(Opt-Out) 방식을 채택함에 따라 개인정보 수집이 상대적으로 용이했기 때문으로 설명
  - 최근 데이터 브로커를 통한 개인정보 유출 사건이 발생하면서 데이터 브로커에 대한 규제 입법이 추진 (캘리포니아주 및 버몬트주에서는 데이터 브로커의 정의 및 등록을 요구하는 주법을 신설)



# 미국의 데이터 거래 현황 (3)

- 미국의 데이터 거래소는 민간 위주로 운영되며, 2019년 말 기준 650여개의 각종 데이터 거래소가 운영되고 있음

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



# 미국의 데이터 거래 현황 (4)

- 미국의 데이터 거래소에서는 주로 공개된 자료에 기반한 정보, 판매자가 보유하고 있는 정보, 판매자가 가공한 정보 등이 거래되며 개인식별정보 등은 거래되지 않는 것으로 보임
  - 개인정보를 제외한 정보의 원데이터 또는 특정 목적을 위해 가공된 정보 등으로 구성
- 미국에서는 데이터가 주요 자산인 기업의 인수·합병 등을 통해 데이터 획득 및 시너지 효과를 활용하는 사례도 많음

인수 기업	인수대상 기업	연도	인수금액
Microsoft	깃허브(오픈소스 코드 공유 플랫폼)	2018년	75억 달러
Microsoft	LinkedIn(구인구직 SNS 앱)	2016년	262억 달러
Oracle	DataLogix(데이터 브로커)	2014년	12억 달러
LexisNexis	ID Analytics(데이터 브로커)	2020년	3.75억 달러

# 중국의 데이터 거래 현황

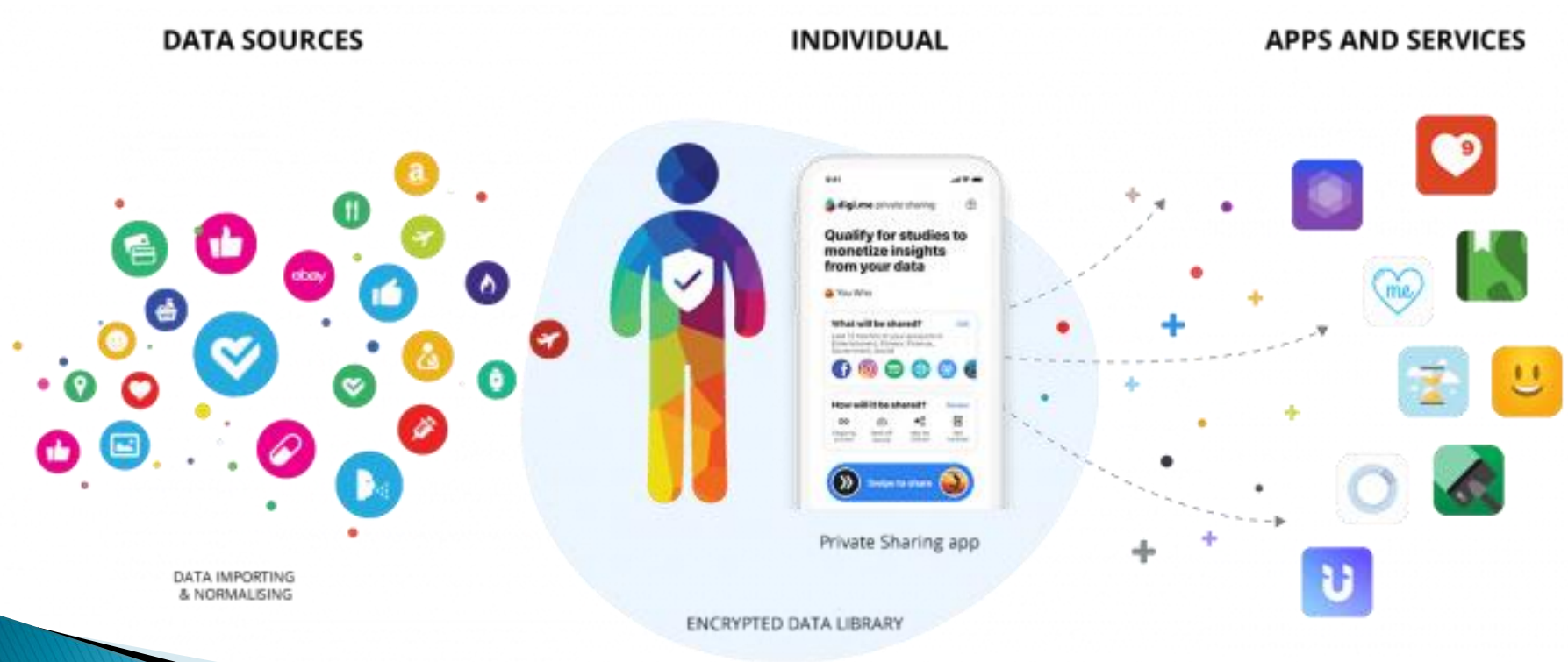
- 중국의 데이터 시장은 정부 주도 하에 빅데이터 종합 시범구를 지정하고 세계 최초 데이터 거래소인 귀양 빅데이터거래소(2014)를 설립하며 빠르게 성장
  - 중국 정부는 '국유자본, 정부주도, 기업참여, 시장운영'을 원칙으로 하여 거래소에서의 데이터 권리를 보증하고 다양한 거래 주체의 적극성을 유도하며 데이터의 공신력과 신뢰성을 확보
  - 회원의 약 15%가 원 데이터(raw data)를 제공하고 있는데 거래소 내에서 유통되고 있는 데이터의 80%가 정부 데이터로 이루어짐
  - 데이터 거래소의 서비스 비용으로 데이터 매각 금액의 10%를 수수료로 책정하며, 데이터 가격은 협의 가격, 경매, 콜옵션 등 세 가지 방식으로 책정
  - 중국 정부는 데이터 거래소에서 거래되는 데이터가 각종 가격지수를 통해 결정되며 상품 유형에 따라 자동으로 가격이 결정되는 표준안을 구축

# 영국의 데이터 거래 현황 (1)

- 영국은 데이터 관련 정책을 정부에서 적극 수립하며 PDS 플랫폼을 통해 정보 주체가 능동적으로 자기 정보를 관리 및 거래하는 결정권을 행사하도록 함
  - 영국은 유럽국가 중 데이터 산업이 가장 활성화된 국가로서 Mydata(2011) 및 OpenBanking(2018) 정책을 통해 금융, 통신, 에너지 등의 회사가 보유한 고객 정보 접근을 소비자가 허용
- 대표적인 거래소인 digi.me는 개인데이터저장소(Personal Data Store: PDS)를 통해 개인이 모든 데이터를 한 곳에서 수집 및 관리하는 공유 플랫폼
  - 개인의 금융, 소셜, 의료, 헬스케어, 엔터테인먼트 관련 다양한 정보를 취합하며 정보 주체인 사용자가 데이터를 관리(삭제, 중지, 재공유 등)
  - App에 공유되는 개인데이터 거래당 0.1달러(사용자 1인당 연간 최대 3달러) 또는 digi.me를 통해 창출되는 App 수익의 7.5%를 정보주체에게 지급

# 영국의 데이터 거래 현황 (2)

- 영국은 마이데이터 서비스와 개인의 데이터 거래를 동일 플랫폼 안에서 가능하도록 해 정보주체 주도의 데이터 유통의 장을 마련

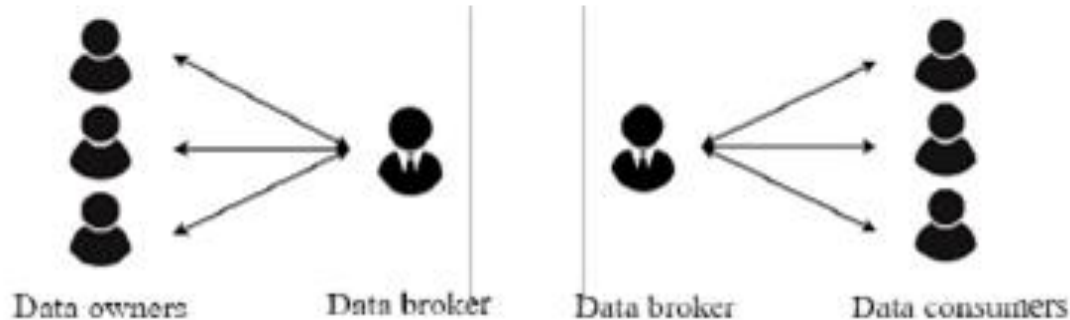


# 데이터 거래시장 유형 (1)

- Zhang et al. (2022)은 데이터 시장을 sell-side, buy-side, two-sided 등으로 유형화
  - sell-side 시장: 여러 소스의 데이터를 통합하고 데이터 소비자에게 데이터를 판매하는 시장으로서 예를 들어 위치 데이터를 주로 판매하는 AggData 등이 있음
  - buy-side 시장: 개인과 조직이 내부 데이터로 수익을 창출할 수 있으므로 데이터 브로커가 데이터 소유자로부터 데이터를 조달
  - 양면(two-sided)시장: 데이터 소유자가 데이터 소비자에게 데이터를 제공할 수 있도록 하여 양면을 결합하는데 대표적으로 Data&Sons 등이 있음

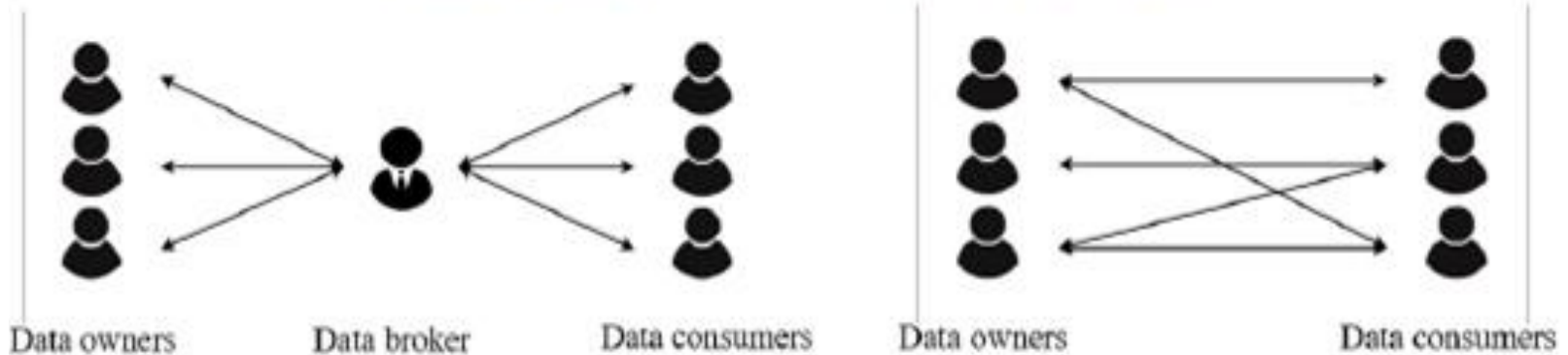
# 데이터 거래시장 유형 (2)

- sell-side, buy-side, two-sided 시장의 구조



(a) Buy-side market

(b) Sell-side market



(c) Centralised two-sided market

(d) Decentralised two-sided market

# 데이터 상품의 특이성과 프라이싱 (1)

## ➤ 데이터의 특이성

- 비경합성(non-rivalry)
- 복제 가능성(replicability)
- 결합에 따라 가치 변동폭이 큼
- 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

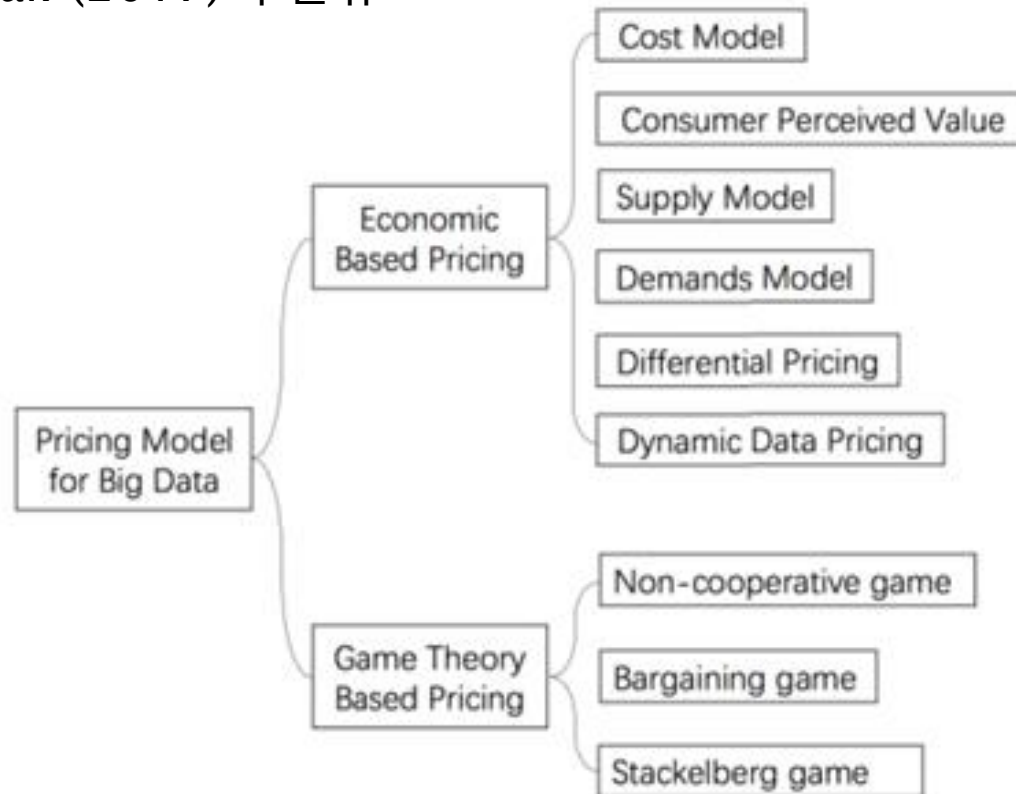


# 데이터 상품의 특이성과 프라이싱 (2)

- 데이터 상품의 특이성과 더불어 데이터 프라이싱을 어렵게 만드는 요인
  - 데이터 원천의 다양성: 데이터를 일관적으로 분류하기 어려우며, 데이터를 수집하는 원천이 매우 다양하기 때문에 데이터를 수집하는 데에 소요된 원가계산이 어려움
  - 데이터 관리의 복잡성: 데이터를 관리하고 분석하는 역량과 방법은 매우 다양하며 동일한 데이터라고 하더라도 그 데이터의 가치는 이를 어떻게 관리, 분석, 처리하느냐에 따라 달라짐
  - 데이터의 다양성: 데이터에 대한 수요는 다양하기 때문에 데이터를 어떠한 수준(레벨), 정확성, 유형(타입)으로 나누느냐에 따라 그 가치는 달라짐

# 데이터 프라이싱 이론 (1)

- 데이터가 다양하고 가격 책정이 어려운 만큼 데이터 프라이싱 이론 및 방법론도 다양하게 제기
  - Liang et al. (2017)의 분류



# 데이터 프라이싱 이론 (2)

## ➤ Majumdar et al. (2023)의 분류

- Quality Based: 데이터의 품질을 토대로 적정가격을 추론
  - Yang et al. (2019), Yu and Zhang (2017), Stahl and Vossen (2016), Tang et al. (2016), Bataineha et al. (2016)
- Query Based: 수요자들이 필요한 데이터를 검색, 분류하기 위한 Query를 바탕으로 가격을 부과하는 모델
  - Koutris et al. (2015), Li and Miklau. (2012), Balazinska et al. (2013), Bergemann and Bonatti (2015)
- Privacy Based: 개인정보 공급자가 상실하는 프라이버시의 가치를 기반
  - 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)
- 기타 다수의 연구들

# 데이터 프라이싱 이론 (3)

## ➤ 김수진 등 (2021)의 분류

- 자산으로서의 데이터 가치 평가: 데이터를 기업의 무형자산으로 보고 전통적인 자산평가방법인 원가기반, 수익기반, 시장기반의 접근법을 적용
- 상품으로서의 데이터 가치 평가: 원가기준, 시장기준, 소득기준, 혜택기준, 영향기준 접근법으로 세분

## ➤ 데이터 거래시장 유형별 분류

- sell-side 시장: query type에 따라 더욱 세분화하여 null query, one-off query, general query 등에 따라 프라이싱이 달라짐
- buy-side 시장: 프라이버시 이슈가 중요한데 differential privacy 및 기타 ad hoc privacy로 구분
- 양면(two-sided)시장은 위 구분들에 따라 더욱 다양하게 세분화

# 데이터 거래 구조 설계시 고려해야 할 논점 I: Possibility of Price Discrimination (1)

- Inherent Large Sample Property: 임요한(2023)
  - 동등한 확률적 특성을 보유한 데이터라고 하더라도 gain function의 기울기는 보유 데이터의 규모에 따라 체감함
  - 따라서 기 보유 데이터의 규모가 큰 구매자의 지불 용의는 크지 않음
  - 기 보유(혹은 기 구입) 데이터의 규모가 일종의 교섭력(bargaining power)을 부여하는 효과 야기함
  - 만일 데이터를 sequential 하게 구입할 수밖에 없다면 지불하려는 가격은 점차 낮아질 가능성
  - 따라서 second-round seller에 대해서는 (비록 데이터의 확률적 특성이 동등하다고 하더라도) 낮은 가격을 지불하려는 성향 발생 => 가격 차별?

# 데이터 거래 구조 설계시 고려해야 할 논점 I : Possibility of Price Discrimination (2)

## ➤ Data Externality a la Acemoglu: 임종호(2023)

- 구입한 데이터와 미구입한 데이터 간에 통계적 상관관계가 있다면, 구입 데이터를 통해 미구입 데이터의 특성에 관한 부분적인 추론이 가능
- First-round seller는 자신의 데이터 판매 행위를 통해 second-round seller가 보유한 데이터의 가치를 감소시킴 => 음(-)의 외부효과 발생

## ➤ Ignorability와 데이터의 품질

- Ignorability 고려할 경우 구입한 데이터는 양면성을 보임
- 구입한 데이터는 원 데이터의 확률 분포를 왜곡없이 반영하지 못함 => 구매 행위를 통한 데이터 확보 의욕의 일반적 저하
- Acemoglu 모형에서 ignorability를 추가로 고려할 경우 First-round seller는 사생활 보호 성향이 낮은 판매자가 될 가능성 큼 => 음의 외부효과로 인해 구매자는 ignorability를 인식하면서도 거래에 임할 가능성

# 다양한 가격 차별 가능성에 대한 정책 대응(1)

- Large sample property에 의한 가격 체감 현상은 시장에서 거래 당사자의 교섭에 의해 해결 가능하다고 판단됨
    - 일회성 거래 형태에서 데이터 규모를 당사자간 교섭에 의해 결정하거나,
    - 일회성 거래(one-off) 형태 대신 데이터 구독형 거래(subscription) 형태를 선택하고 최소 구독 기간을 설정하는 등 교섭이 가능
    - 다만 데이터를 표준화하고 동일한 가격으로 불특정 다수에게 판매하는 통상적인 거래소 형태의 거래는 불가능할 수도 있음
- => 데이터의 판매자(혹은 그 대리인)와 데이터의 구매자 간 교섭에 의한 거래가 가능하도록 기초적 제도만을 정비해 주면 될 듯함(임요한(2023))



# 다양한 가격 차별 가능성에 대한 정책 대응(2)

- Negative Externality와 Ignorability는 데이터 거래 구조 설계에서 중요한 정책적 대응 요소를 발생시킴
  - Negative Externality를 internalize 할 수 있는 메커니즘 필요 => first-round seller와 second-round seller의 이해관계 통합 필요
  - Ignorability 에 따른 데이터 품질 저하를 해결할 수 있는 메커니즘 필요  
=> 데이터 판매 여부에 대한 의사결정을 배제할 수 있는 구조?
- 다음과 같은 정책 대응 방향을 모색할 필요
  - 데이터의 원초적 보유자가 최종 데이터 구매자 또는 (거래만을 중개하는) 거래소를 개별적으로 직접 접촉하는 시장 구조는 바람직하지 않음
  - 데이터의 원초적 보유자는 자신의 이해관계를 대변할 수 있는 “통합적” 대리자에게 데이터를 집적한 후 그 대리자가 집적된 데이터를 구매자에게 판매하는 구조가 더 바람직 (강경훈·전성인(2018))  
=> 다만 대리자를 통한 데이터 집적의 경우에도 대리자의 배당이 데이터 제공자의 privacy cost를 올바르게 반영하지 못할 경우 ignorability 문제의 해결은 여전히 쉽지 않음 => 정확한 compensation mechanism 개발 필요

# 데이터 거래 구조 설계시 고려해야 할 논점 II: 공공 정보의 (무료) 개방과 데이터 (유료) 거래 (1)

## ➤ 공공 정보 (무료) 개방이 초래하는 긍정적 효과

- 데이터 구매자가 데이터 유료 거래에 참여하기 이전에 보유 가능한 데이터 규모를 확대함으로써 유료 데이터의 gain function의 기울기를 작게 함으로써, 데이터 구매자의 유료 데이터 확보 비용을 절감시킴
- 시장에서 거래되는 유료 데이터가 보완적 성격일 경우 공공 정보의 무료 개방은 시장 거래 데이터의 효능을 증가시켜 생산성 향상에 기여
- 공공 정보가 “강제 수집” 정보인 경우 ignorability 논점에서 자유스러움
- 공공 정보는 개인의 privacy 제공 동의를 받지 못한 정보이므로, 다양한 익명화 조치가 정보 제공에 선행해야 하는데 이를 통해 데이터 품질이 저하될 수 있으므로, 역설적으로 시장 거래 데이터의 필요성을 창출할 수 있음

# 데이터 거래 구조 설계시 고려해야 할 논점 II: 공공 정보의 (무료) 개방과 데이터 (유료) 거래 (2)

- 공공 정보 (무료) 개방이 초래하는 부정적 효과
  - 구매자가 데이터 유료 거래에 참여하기 이전에 보유 가능한 데이터 규모를 확대함으로써 유료 데이터의 gain function의 기울기를 작게 하는 것은 시장에서 거래되는 유료 데이터의 매력을 감소시켜 전체적인 시장 거래를 위축시킬 가능성
  - 공공 정보가 유료 데이터와 상관 관계가 있을 경우 Acemoglu 형태의 음(-)의 외부 효과를 발생시켜 데이터 보유자의 gain을 감소시키고 시장 거래 자체를 위축시킬 가능성
- 정보의 (무료) 개방이 초래하는 일반적인 부정적 효과
  - 데이터의 원초적 제공자의 효용을 반영하지 않은 수집 과정 때문에 공공 정보의 제공시에도 원초적 제공자의 효용을 무시하게 됨 => 과다 제공 가능성
- 공공 정보 (무료) 개방의 일반적 원칙?
  - 공공 정보의 제공 대상을 “원초적 정보 소유자로부터 사후 보상을 약속하고 데이터를 집적한 중간 대리자”으로 한정하여 정보 제공? (현재 단계 conjecture)

감사합니다!