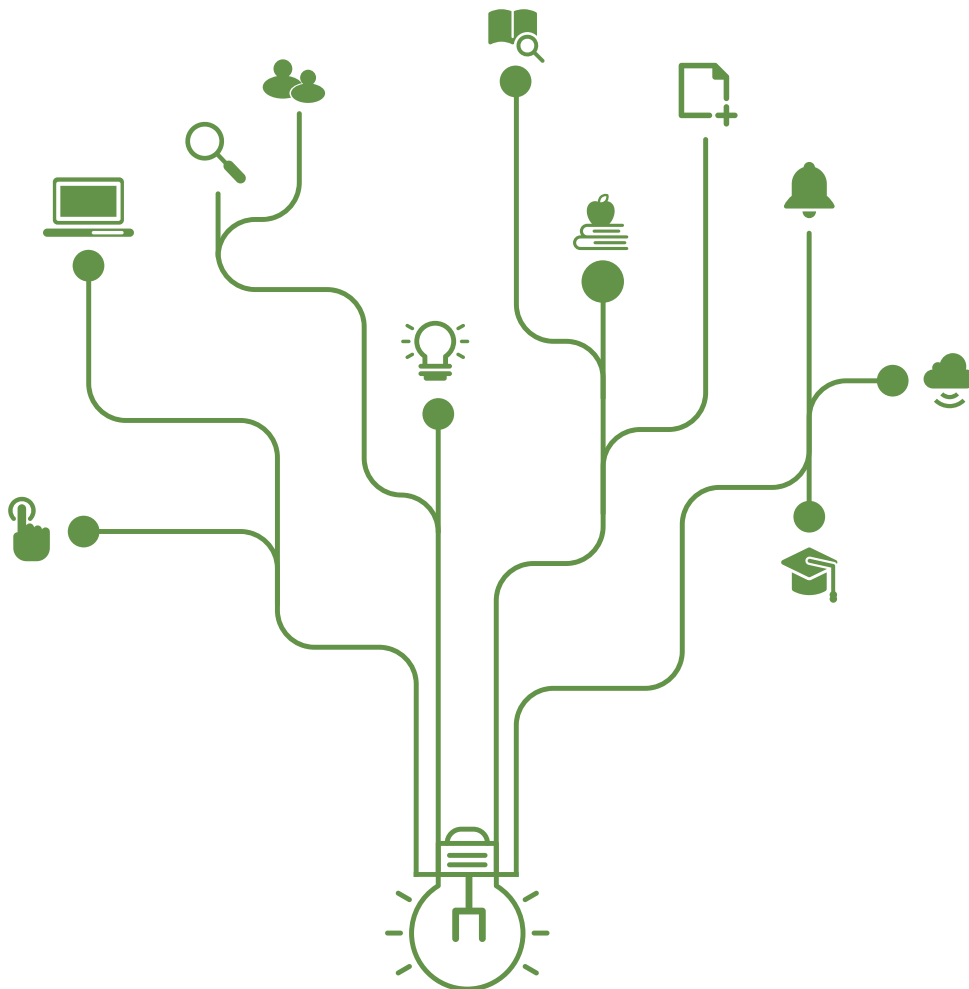


A Sectoral Approach in Assessing the Data Economy Ecosystem: Focusing on the Finance, Real Estate, and Medical Service Sectors in Korea

Man Cho (KDI School of Public Policy and Management)

Sungwuk Moon (Sogang University)

Inbok Rhee (KDI School of Public Policy and Management)



**A Sectoral Approach in Assessing the Data Economy Ecosystem:
Focusing on the Finance, Real Estate, and Medical Service Sectors in Korea**

Abstract

The main objective of this study is to conceptualize “data economy (DE)” and to assess the current state of fostering its ecosystem in three particular service sectors in Korea – finance, real estate, and health care. To that end, we, first, synthesize the key arguments advanced in the recent literature as to the expected socio-economic effects of DE and, second, perform the meta-analyses based on the academic articles of relevancy that are extracted from two international databases: 263,974 articles in the fields of health and medical science from PubMed, established by the National Institute of Health of the U.S.; and, 26,859 articles in the fields of finance and real estate from Web of Science (WOS). Our results are summarized as: for our analytical purpose, DE is defined as an innovation ecosystem in which digital data is collected·shared·analyzed·utilized such that regional or national welfare gain can be realized; Its key elements include DPA(Data·Platform·AI)-driven infrastructure along with various enablers, both market-driven and institution-driven; The expected effects of DE, as discussed in the literature, are categorized into three types – (1) the platform effect, (2) the prediction power effect, and (3) the new analytics effect, whose implications to the three sectors are also elaborated; According to the meta-analyses performed, the DE-related academic studies have been growing in an explosive phase in all three sectors since around 2015, and the trend of increasing co-authorship and inter-country collaboration is also observed; And, those articles written by the Korean authors (according to their affiliations) show a fair standing in terms of the quantity, i.e., number of articles (8th place in both databases) but are being lagged significantly compared to other advanced economies in terms of the quality (e.g., average number of citations, newness of research topics). Using the above findings, we discuss the policy directions and tasks for activating DE and expanding its social effects in the three sectors examined.

**데이터경제 개념화 및 생태계 구축에 관한 연구:
우리나라의 부동산·의료·금융부문을 중심으로**
(KDI국제정책대학원; 2021.8.25)

조만*·문성욱**·이인복*

- 세계 최초의 보험거래소인 영국의 ‘로이드사’는 1687년 Edward Lloyd가 런던항 근처에서 개점한 커피숍에서 타블로이드 “Lloyd’s List”를 발행하여 원양항해에서 수집한 데이터를 공유하면서 시작되었다. [Bernstein (1998)에서 발췌·요약] -

초록

본 연구의 목적은 ‘데이터경제’의 개념화 및 활성화 방안을 우리나라의 3개 서비스부문(금융·부동산·의료)에 초점을 맞추어 분석하는데 있다. 이를 위하여 최근 국내외 데이터경제 관련 문헌의 주요 논거를 정리하였고, 2개의 국제학술논문 데이터베이스를 이용하여 상기 3개 부문의 데이터경제 관련 논문을 추출하여 메타분석을 수행하였다: 미국 국립보건원(NIH)의 ‘PubMed’로부터 추출한 263,974건의 보건·의료 분야 학술논문; 그리고 WOS(Web of Science)로부터 추출한 26,859건의 금융·부동산 분야 학술논문이 포함됨. 분석결과를 요약하면, 데이터경제는 디지털 데이터의 구축·공유·분석·활용을 통하여 지역 또는 국가 차원의 사회적 후생을 증가시키는 혁신생태계로 정의하였고, 이의 핵심 구성요소는 DPA(Data·Platform·AI) 기반의 하부구조와 제반 촉진기제들이 포함된다. 최근 문헌에서 논의되고 있는 데이터경제의 경제·사회적 기대효과는 다음 3개 유형으로 정리하였고 - (1) 플랫폼 효과; (2) 예측력 증대 효과; (3) 새로운 분석기법 효과, 이들이 3개 서비스부문에 주는 시사점을 논의하였다. 메타분석의 결과는 3개 부문의 데이터경제 관련 학술논문이 2015년경부터 폭발적인 증가추세에 있고, 공저자의 숫자 및 다국간 협력논문의 비중이 지속적으로 증가하는 것을 보여 주었다. 우리나라의 경우 논문의 양적인 측면에서는 대체로 양호한 것으로 판단되나 (2개 DB 모두에서 세계 8위를 기록), 질적인 측면에서는 경쟁 상대국에 비하여 상당히 뒤쳐져 있는 것으로 나타났다 (평균 인용횟수, 연구주제의 참신성 측면에서). 상기 분석결과를 기초로 우리나라 3개 부문에서의 데이터경제 활성화를 위한 제반 정책과제를 논의하였다.

* KDI국제정책대학원; ** 서강대학교; 본 논문은 KDI국제정책대학원 연구과제의 결과물을 정리한 것임.

I. 서론

지난 30여년간 인류는 이전에 경험하지 못한 2개의 거대한 플랫폼을 보유하게 된다: 즉, 1989년 영국에서 개발된 인터넷(World Wide Web, www)과 2007년 미국에서 개발되어 확산된 스마트폰. 이들은 현재 개인의 삶에 지대한 영향을 미칠 뿐만 아니라, 기업과 정부의 생존·발전전략의 수립에 있어서도 가일층 영향이 커져가는 상황이다. 특별히, 쓰나미로 표현되는 디지털전환(digital transformation)의 글로벌 추세는 인터넷·모바일 플랫폼을 기반으로 새로운 형태의 데이터를 새로운 방식으로 분석하여 혁신적인 상품·서비스를 개발·보급하는 글로벌기업들의 출현을 가능하게 하였고, 이는 국가 차원에서의 산업생산성 증대 및 일자리 창출과 같은 사회복지 증진으로 이어지고 있다. 예를 들면, 지난 10여년간 글로벌 톱10기업의 대부분이 애플·아마존·구글·알리바바·텐센트 등의 대형 IT기업들로 바뀌었고, 이들은 데이터·플랫폼·인공지능(DPA) 기반의 혁신적 상품을 출시하여 괄목할만한 성과를 이어가고 있다. 또한 세계 주요 국가들은 (미국·중국·영국·일본·EU국가들) 2017년경부터 디지털전환에 대응하기 위한 경제·사회발전전략을 수립하여 추진하고 있다.¹

본 연구에서는 디지털 데이터의 구축·공유·분석·활용이 사회복지 증진으로 이어지는 현상을 ‘데이터경제’로 정의하고, 이의 실현을 위한 연결고리를 우리나라의 3개 서비스부문(금융·부동산·의료)에 초점을 맞추어 분석하고자 한다. 구체적으로, 국가 또는 지역 차원의 사회복지 증진을 위한 데이터경제 생태계는 어떻게 활성화 할 수 있고, 이의 핵심 구성요소는 무엇이며, 또한 이를 촉진하기 위한 공공·민간의 역할은 무엇인지에 대한 개념화를 시도하는 것이 본 연구의 주요 목적이다. 이를 위하여 최근 학술연구의 논거를 기초로 데이터경제의 경제·사회적 기대효과를 정리하였고, 데이터경제 관련 키워드를 이용하여 2개의 국제학술논문 DB로부터 관련 논문을 추출하여 메타분석을 수행하였다: (1) 미국 국립보건원(National Institute of Health, NIH)이 구축한 보건·의료 관련 데이터베이스인 ‘PubMed’를 통하여 총 263,974건의 보건·의료 관련 학술논문 추출; (2) WOS(Web of

¹ 여기에는 2017년 중국의 “빅데이터산업 발전계획” (데이터 개방의 확대, 플랫폼·오픈소스 관련 기술지원, 데이터거래소 설립 등); 2017년 영국의 “영국 디지털전략” (기업·정부의 혁신적·효과적 데이터 기반 구축, 전국민의 데이터 스킬 향상 추진 등); 2017년 일본의 “미래투자전략-Society 5.0 실현을 위한 개혁” (현실데이터 플랫폼 구축, 교육·인재 역량강화, 혁신벤처 선순환 시스템 구축 등); 2018년 EU의 “개인정보 보호에 대한 일반규칙 (GDPR)” (개인정보의 보호 및 활용에 대한 지침) 및 GAIA-X 프로젝트 (데이터경제 활성화를 위한 클라우드 기반 구축); 2019년 미국의 “오픈거버먼트 데이터법” (공공부문의 데이터 플랫폼인 “Data.gov” 구축, 데이터의 전략적 활용을 위한 10년 계획 수립 등) 등이 포함됨.

Science) DB를 통하여 총 26,859건의 금융·부동산 부문(사회과학 중심)의 학술논문을 추출.

미시적 의미에서 ‘데이터(data)’는 정보(information)·사실(fact)·증거(evidence)로 변환되기 이전의 형태를 지칭하고, 따라서 이는 통상 가치중립적인 재화로 인식된다. 데이터가 경제·사회적인 가치를 가지게 되는 것은 이의 공유 또는 거래를 통하여 사용자가 특정 학술적·상업적·정책적 목표와 연계하여 결과를 도출할 때 가능해진다. 즉, 데이터를 특정 맥락에 따라 활용할 때 가치창출이 가능해지고, 데이터경제는 이와 같은 가치창출의 결과가 지역 또는 국가 차원의 사회적 후생 증대로 이어지게 하는 혁신생태계라고 할 수 있다. 이의 실현을 위한 핵심 구성요소는, 첫째, 다양한 정형·비정형 데이터의 효율적인 구축 및 공유, 둘째, 이를 가능하게 하는 인터넷·모바일 기반의 온라인 플랫폼, 그리고 인공지능 등 새로운 분석기법의 사용이라고 할 수 있다. (이를 데이터경제를 위한 DPA – Data-Platform·AI 인프라로 명명할 수 있음.)

이와 같은 DPA 기반의 데이터경제에는 현재 국내외에서 빠르게 형성되고 있는 ‘스마트X’ 산업 대부분이 포함된다고 볼 수 있고, 본 연구의 대상이 되는 3개 서비스부문에 서도 스마트금융 (핀테크, FinTech, 등), 스마트부동산 (프롭테크, PropTech, 등), 스마트의료 (디지털헬스, 원격진료 등) 이에 해당된다. 우리나라에서는 이들 스마트X 부문에 대한 언론과 학계의 관심이 2015년경부터 본격적으로 시작된 것으로 나타났다. 예를 들면, 핀테크에 대한 언론보도의 빈도수는 2015년경부터 급증하였고 (이는 우리나라 최초의 모바일지급결제시스템인 ‘카카오페이’에 대한 인가가 주요 원인인 것으로 보임), 디지털헬스, 프롭테크, 데이터경제에 대한 언론보도는 그 이후에 훨씬 낮은 빈도로 언론에 포착되었다. 반면, 미국의 경우 인터넷이 보급되기 시작한 1990년대 초·중반부터 이를 기반으로 하는 다양한 온라인 금융서비스가 B2B·B2C플랫폼을 통하여 제공되었고, 현재 미국의 핀테크 부문은 1990년대 온라인 금융서비스 부문이 보다 발전·진화된 형태라고 할 수 있다.

최근 문헌에서는 이와 같은 DPA 기반 데이터경제의 사회적 효과를 세 가지 유형으로 분석하고 있다. 첫째, 인터넷·모바일 플랫폼 기반의 비대면서비스는 기존의 대면서비스에 비하여 거래비용을 획기적으로 감소시키고 (따라서 소비자 편익·효용을 획기적으로 증가시킴), 동시에 온라인 거래에서 부수적으로 발생하는 디지털데이터의 (소비자행태 등에 대한) 축적을 촉진한다.² 그러나 이와 같은 플랫폼 기반 서비스의 확대는 관련 부문에서 기존 중개기능의 무용화 (즉, ‘탈중개화’) 및 이로 인한 고용인력 감소를 초래하고, 동

² 핀테크·프롭테크·디지털헬스의 경우 Philippon (2015) and (2016), Baum (2017), Cho (2020), 이상영 (2020), 오상우 (2020), 강건욱 (2020) 등을 참조

시에 비대면서비스에 따르는 사이버 리스크의 증가 (보이스피싱 등), 그리고 데이터를 이용한 국가의 과도한 간섭·통제를 가능하게 하는 등의 부작용을 초래할 수 있다.

둘째, 데이터경제의 또 다른 효과로 데이터의 축적·공유가 개별 기업의 관점에서 최적의 생산기술 또는 비즈니스모델을 예측하는데 있어서 오류의 가능성을 낮추는 것으로 논의되고 있다. (Farboodi and Veldkamp (2021) 등) 즉, 데이터의 축적은 최적의 기술에 대한 예측력을 증가시키고, 이는 개별 기업 및 산업 전반의 생산성 증가로 이어질 수 있다는 논거다. 이와 같은 예측력 증대 효과는 기존에 구축한 데이터의 양이 작은 상황에서, 그리고 스타트업과 같은 소규모 기업의 경우에 한계생산성 증가의 (즉, 규모의 경제 효과가) 나타날 수 있음을 예측하고 있다. 같은 맥락에서, 데이터의 구축이 거시경제의 경기순환성에 대한 예측력 증가 (Ordonez (2013), Fajgelbaum et al. (2017)), 금융시장에서 개별 기업의 신용위험에 대한 예측력 증가 및 이로 인한 기업 자기자본비용의 감소 (Begenau et al. (2018)), 대출시장에서 차입자의 신용리스크에 대한 예측력 증가 및 이로 인한 담보물의 중요성 경감 (Gambacorta et al. (2020)) 등이 최근 문헌에서 논의되었다. 3개 서비스부문에 대한 이의 시사점으로, 데이터의 구축은 금융시장에서 보다 적확(的確)한 리스크의 측정 및 관리, 부동산시장에서 개별 물건에 대한 보다 정확한 가치평가 및 시장 추이의 분석, 그리고 보건·의료 부문에서 최적의 치료법 및 신약개발을 촉진하는 것이 가능하고, 이로 인한 사회복지 증진 또한 기대할 수 있다는 점이다.

셋째, 데이터경제의 또 하나의 중요한 축은 최근 사용이 급증하고 있는 머신러닝(ML)·딥러닝(DL) 등 인공지능(AI) 기반의 분석기법을 통한 정형·비정형 데이터에 대한 의미있는 분석이 가능해 졌다는 점이다. 이는 학술적 실증분석의 영역을 넓히고 있고, 실제로 최근 문헌에서는 다양한 비정형 데이터를 이용한 연구결과가 발표되고 있다: ML방식을 기반으로 인공위성 사진을 통한 경제활동 수준에 대한 예측, 세금보고서(예: 미국의 '10K 보고서')를 통한 기업의 성과 및 유형 분석, 그리고 주택의 외관이 건물가치에 미치는 영향에 대한 실증분석 등이 포함된다.³ 또한 이와 같은 분석을 가능하게 하는 새로운 분석기법의 (regression tree, LASSO, random forest, ensemble 등) 등장과 함께 이를 위한 소프트웨어도 (R, Python 등) 개발·보급되고 있다. 이와 같은 AI 기반의 분석방식은 실증분석의 철학적인 기반에도 변화를 초래하고 있다. 즉, 기존의 이론에서 출발하는 귀납적(top-down) 실증분석 방식이 가설검증에 초점을 맞추고 있다면 (즉, 모수(β)의 부호 및 통계적 유의성에), AI 기반의 실증분석은 사례에서 시작하는 연역적(bottom-up) 방식을 취하고 있고, 이는 β 를 통한 가설검증 보다는 설명변수들 간의 다양한 연계성 패턴을 분석하고,

³ 이에 대한 최근 문헌의 서베이는 Mullainathan and Spiess (2017) 참조

이를 통하여 오차항을 최소화 하는 종속변수의 예측치(\hat{y}) 도출에 초점을 맞추고 있다. 향후 학술연구에서 이와 같은 두 가지 방식을 보완적으로 사용하여 실증분석의 효과를 높이고자 하는 노력이 인문사회계·이공계 모두에서 증가할 것으로 예상된다.

2개 학술논문 DB를 활용한 메타분석에서는 (3개 서비스부문에 대한) 데이터경제 관련 논문이 대부분의 국가에서 2015년경부터 폭발적으로 증가한 것으로 나타났다. 그러나 미국의 경우 2003~08년 기간 중 보건·의료 분야 논문의 증감사이클이 또 한번 관찰된 반면, 우리나라는 2015년 이후의 상승세 시작이 유일한 변곡점인 것으로 관찰되었다. 학술논문의 양적인 측면에 있어서 (연구자의 소속기관 기준) 미국·중국·영국이 다른 나라에 비하여 압도적으로 높은 비중을 차지하고 있고 (중국이 미국의 1/2, 영국이 중국의 1/2), 우리나라는 보건·의료부문 및 금융·부동산부문 모두에서 8위를 기록하였다 (그러나 증가 속도 면에서는 상위권의 다른 나라들에 비하여 빠른 것으로 관찰됨). 또 하나 주목할만한 결과는 공저자 숫자의 지속적 증가와 함께 (금융·부동산 부문의 경우 논문 한 편당 평균 저자수가 1998년 2.5명에서 2020년 4.5명으로 증가), 다국적 협력논문의 비중도 지속적으로 증가하는 추세라는 점이다.

논문의 질적인 측면에서는 우리나라가 경쟁 상대국들에 비하여 상당한 수준으로 뒤져 있는 것으로 나타났다. 금융·부동산 부문 논문의 경우 평균 인용회수가 미국 22.1, 영국 19.6, 네덜란드 15.2에 비하여 우리나라와 중국은 9.8로 미국의 절반에도 미치지 못하는 수준인 것으로 조사되었다. 보건·의료 부문 논문의 경우에는 PubMed에서 제공하는 “MeSH Heading Keywords System (MHKS)”을 통한 개별 연구주제의 참신성을 평가하였고 (키워드가 MHKS에 최초로 등록된 연도와 논문의 발간연도 차이를 구하고, 이의 국가별 평균값을 계산하였으며, 평균값이 낮을수록 연구주제의 참신성이 높음), 분석 결과, 미국·프랑스·독일·일본의 평균값이 약 23.3년인 반면, 우리나라는 25.7년(중국은 25.3년)을 기록하였다. 2.4년의 차이가 일견 크지 않은 것으로 보일 수 있으나, 현재와 같이 기술수준이 매우 빠르게 변화하는 상황에서는 상당한 수준의 격차인 것으로 분석된다.

현재 우리나라는 산업과 학술연구 모두에 있어서 과거의 모방·추격 모드로부터 혁신의 창출·선도로의 대전환이 필요한 상황이다. 이를 위해서는 선진국 수준의 독창적인 연구를 수행할 수 있는 연구자와 연구환경, 원천기술의 개발 및 혁신적 아이디어를 촉진할 수 있는 집단 창조성, 그리고 기술혁신과 경쟁을 통한 기업동학(firm dynamics)의 실현이 (즉, 혁신과 경쟁을 통한 기업의 생멸(生滅), 그리고 이를 통한 산업구조의 변화 및 생산성 증대가) 절실한 것으로 진단하고 있다. (최영락 (2018), 안상훈 (2018)) 데이터경제 생태계의 조성 및 활성화도 이와 같은 학술연구 및 산업구조 개혁의 관점에서 추진할 필요가 있다. 구체적으로, PPAP (Public·Private·Academic· Partnership) 방식의 협력체계를 부문별·지

역별로 도입하고, 이를 현재 진행 중인 스마트X 사업과 데이터경제 생태계를 연계하는 기제로 활용하며, 이의 결과로 지역 일자리 창출 등과 같은 국가·지역 차원의 정책효과를 추구하는 방향으로 관련 사업을 진행해야 할 것으로 생각된다. 또한, 데이터경제 활성화는 매우 다면적인 사업이고, 경제시스템 전반에 큰 영향을 미치는 사안인 만큼, 이의 효과적인 추진을 위해서는 다부처·다부문간의 협력과 정책공조가 필수적인 조건일 것으로 판단된다. 본 고에서는 이와 관련된 정책과제를 다음 3개 유형으로 구분하여 논의하였다: (1) 데이터경제 활성화를 위한 국가 차원의 제도적 기반 마련; (2) 지역별 맞춤형 데이터경제 생태계 조성을 위한 전략 수립; (3) 데이터경제 전문인력 양성 및 다부문·다학제간의 학술연구 증진.

서론을 제외한 본 논문은 다음의 6개 장으로 구성되어 있다: 데이터경제의 개념 및 현황 (제2장); 데이터경제의 경제·사회적 효과에 대한 논거 (제3장); 데이터경제 관련 메타분석 결과 (제4장); 3개 부문의 데이터경제 추진현황 평가 (제5장); 정책 과제 (제6장); 결론 (제7장).

II. 데이터경제: 개념 및 현황

1. 데이터경제: 개념의 설정

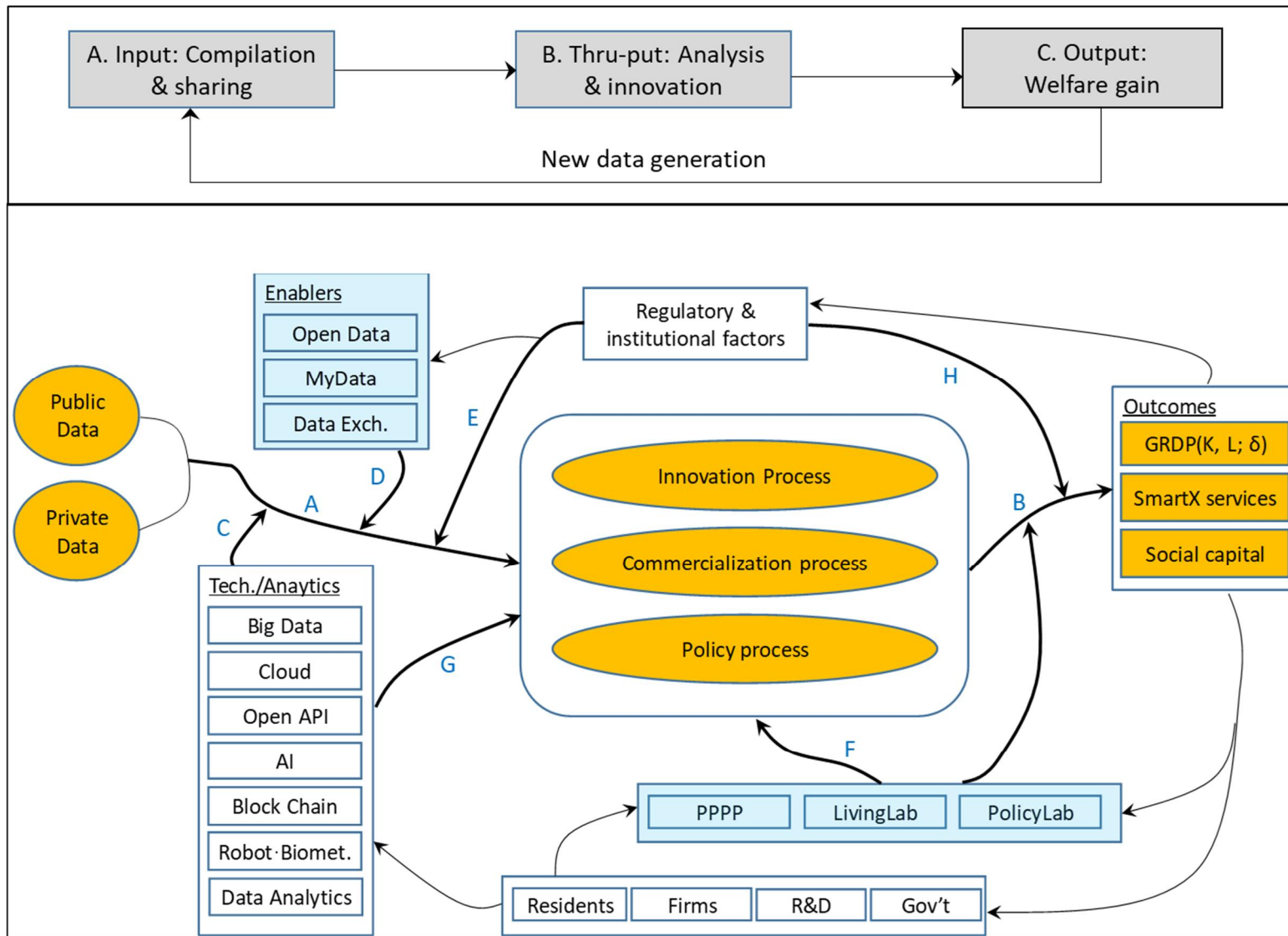
데이터경제는 ‘디지털 형태의 데이터를 유통 또는 거래함으로써 경제·사회적 가치를 창출하는 혁신생태계(innovation ecosystem)’로 정의할 수 있다.⁴ 이의 핵심적인 구성요소는, 첫째, 공공·민간부문에 산재한 다양한 형태의 데이터를 효율적으로, 그리고 안전하게, 수집·축적·공유할 수 있는 제반 기술 요인들 (ICBM – IoT, Cloud, BigData, 그리고 블록체인 등 인터넷·모바일 플랫폼 구축 관련 기술들); 둘째, 축적된 정형·비정형 데이터의 분석이 (특히 AI 기반의 머신러닝·딥러닝 등 새로운 분석기법을 활용한) 사회후생의 증대로 이어지게 하는 가치창출 과정 (소비자효용 증가, 산업생산성 향상, 공공정책의 효과성·투명성 제고 등을 위한); 셋째, 이와 같은 사회적 효과를 촉진하기 위한 제반 제도적 기제들이 포함된다 ([그림 1]의 개념도 참고). 요약하면, 데이터경제는 DPA (Data·Platform·AI) 기반의 혁신생태계라고 정의할 수 있고, 이의 활성화를 위해서는 부문별·지역별 필요에 따라 맞춤형 전략이 필요하다고 하겠다.

데이터경제의 촉진기제와 관련하여 정부는 최근 수년간 다양한 정책들을 추진되고 있다. 여기에는, 개인정보의 보호와 활용을 위한 ‘데이터 3법’의⁵ 국회 통과 및 시행 (2019년 12월), 데이터의 공유 및 거래를 위한 ‘오픈데이터’ (공공부문 데이터의 개방·공유) 및 공공빅데이터 플랫폼 구축, 공공·민간 데이터거래소 설립, 마이데이터(MyData) 사업의 시행, 그리고 규제 샌드박스의 도입 등이 이에 포함된다 (2020~21년 기간 중). 그리고 2020년 7월에는 ‘디지털 뉴딜’ 사업이 발표되었고 (‘한국판 뉴딜 1.0’의 일환으로, 2021년 7월 ‘한국판 뉴딜 2.0’으로 개정), 여기에는 D·N·A (Data·Network·AI) 생태계 강화, 교육·의료 부문에 대한 디지털 인프라 고도화와 비대면 산업 육성 (스마트병원, 원격교육 등을 통한), SOC 디지털화 및 스마트시티·스마트산단 고도화가 포함되었다. 상기 사업들 대부분은 현재 초기 시범사업 단계에 있고, 따라서 이들의 성과에 대한 의미있는 평가는 차후 가능할 것으로 생각된다.

⁴ 유럽연합의 경우 이와 유사하지만 보다 협의의 데이터경제를 정의하고 있다: European Commission, 2018, “A (global) digital ecosystem in which data is gathered, organized, and exchanged by a network of vendors for the purpose of deriving value from the accumulated information.”

⁵ 여기에는 (1) ‘개인정보 보호법’ (주무부처: 행정안전부); (2) ‘정보통신망 이용 촉진 및 정보보호 등에 관한 법률개정안’ (방송통신위원회·과학기술정보통신부); (3) ‘신용정보의 이용 및 보호법’ (금융위원회)이 포함됨.

[그림 1] 데이터경제의 구성요소 및 이들간의 연계채널 (저자 작성)



미시적 의미에서 ‘데이터 (data)’(또는 이의 단수 형태인 ‘datum’)는 일반적으로 정보 (information), 사실(fact), 증거(evidence)로 변환되기 이전의 형태를 지칭하고, 따라서 이는 가치중립적인 재화로 볼 수 있다. 데이터가 경제·사회적인 가치를 가지게 되는 것은 이의 공유 또는 거래를 통하여 사용자가 특정 학술적·상업적·정책적 목표와 연계하여 결과를 도출할 때 가능해진다. 즉, 데이터를 특정 맥락에 따라 (context-specific) 활용할 때 가치창출이 가능해지고, 데이터경제는 이와 같은 가치창출 과정이 지역 또는 국가 차원의 사회적 후생의 증대로 이어지게 하는 혁신생태계라고 이해할 수 있다.

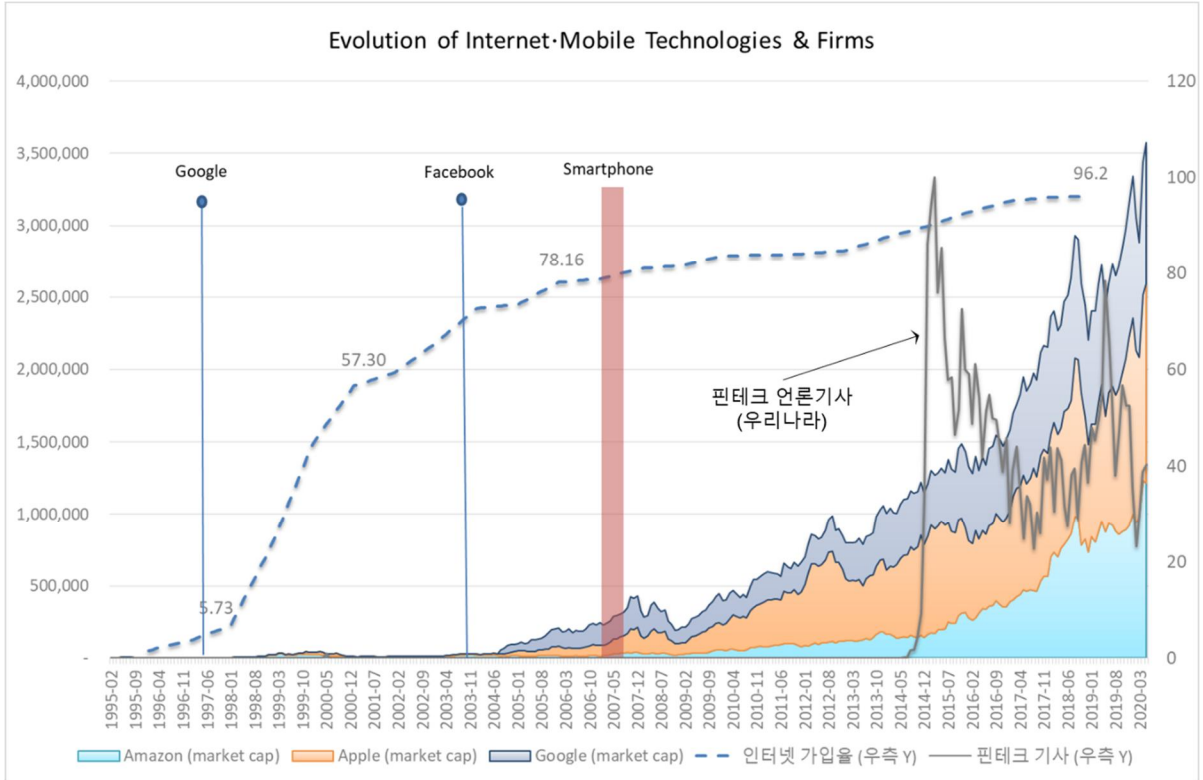
2. 3개 부문 스마트X 산업: 현황

본 연구의 대상인 3개 서비스부문에서는 이와 같은 데이터경제 생태계가 국내외에서 빠르게 형성되고 있고, 여기에는 스마트금융 (핀테크 ~ P2P대출, 모바일지급결제, 로보어드바이저, 가상화폐 포함; 인슈어텍 등), 스마트부동산 (프롭테크, PropTech), 스마트의료가 (디지털헬스, 원격진료) 포함된다.⁶ 우리나라에서 상기 3개 ‘스마트 X’ 부문은 2015년 이후 언론과 학계의 관심을 끌기 시작한 것으로 보인다. [그림 2]에서 보는 바와 같이, 우리나라에서 핀테크에 언론보도는 2015년경부터 급증하였고, 이는 우리나라 최초의 모바일 지급결제시스템인 ‘카카오페이’에 대한 인가가 주요 원인인 것으로 보인다. 반면, 디지털헬스, 프롭테크, 데이터경제에 대한 언론보도는 그 이후에 훨씬 낮은 빈도로 언론에 포착된 것으로 조사되었다.

이와 관련된 추세로, 미국의 대표적인 인터넷 기반의 기업인 아마존·구글은 1990년대에 창업하여, 이들의 시가총액이 2003년경부터 증가하기 시작하였고, 2008년 글로벌금융위기 중 단기간의 하락세가 있었으나, 그 이후 폭발적인 증가세가 이어지고 있다. 반면, 우리나라의 인터넷 가입율은 1998년 6%의 미미한 수준에서 2000년대 초반 급증하였다 (2001년 57%, 2006년 78%로 상승하였고, 현재 96% 수준). 이는 우리나라가 1990년대 인터넷 기반의 혁신사이클에는 참여하지 못하였으나, 2000년대 초반 이후 인터넷·모바일 기반의 혁신사이클에는 매우 빠르게 진입하고 있음을 시사한다.

⁶ 데이터경제 기반의 혁신생태계 구축은 제조업의 부문의 스마트팩토리 (SmartFactory), 스마트팜 (SmartFarm), 자율주행차 (Driverless Car), 그리고 공공부문의 스마트시티 (SmartCity)를 등 수 있다. 스마트팩토리, 스마트금융, 스마트시티의 개요에 대해서 [부록 1]을 참조하고, 최근의 스마트시티 관련 논문으로는 황종성 (2017), 이태준·조만 (2020) 참조.

[그림 2] 핀테크 관련 언론보도 및 관련 추이



데이터경제 기반의 혁신적 금융부문으로 볼 수 있는 핀테크산업은 기존의 대면서비스와 지점망 중심의 금융에 대한 대안적 산업으로 성장하고 있다. 여기에는 온라인 자금 조달 서비스 (P2P 대출 및 크라우드펀딩), 대안적 지급결제 서비스 (cryptocurrency & mobile payment services), 로보-어드바이저 (robo-advisor) 기반의 투자자문 서비스, 보험 서비스 (InsurTech), 규제 준수 서비스 (RegTech) 등이 포함된다. (Cho (2020)) 공급 측면에서의 핀테크 산업은 3개 유형의 서비스 제공자가 주를 이루고 있다: (1) 스타업을 포함한 핀테크 전문 중소기업, (2) 빅테크 기업 (통상 우리나라의 카카오, 중국의 알리바바, 미국의 구글과 같은 비금융 플랫폼 기반의 대기업), 그리고 (3) 기존 금융기관 (은행, 증권회사 등). (핀테크에 대한 보다 자세한 설명은 [부록 1]과 Cho (2020) 참조)

핀테크 부문은 지난 10년간 전 세계적으로 폭발적인 성장세를 보여 왔다. 이의 한 예로, P2P대출 및 크라우드펀딩의 경우 글로벌시장 규모가 2013년 \$117억달러(USD)에서 2018년 3,017억달러로 5년간 25배 성장하였고, 최근까지도 중국을 제외한 세계 전 지역에서 빠른 성장세가 관찰되고 있다. (Cho (2020)) 국내에서도 3개의 인터넷전문은행의 출현

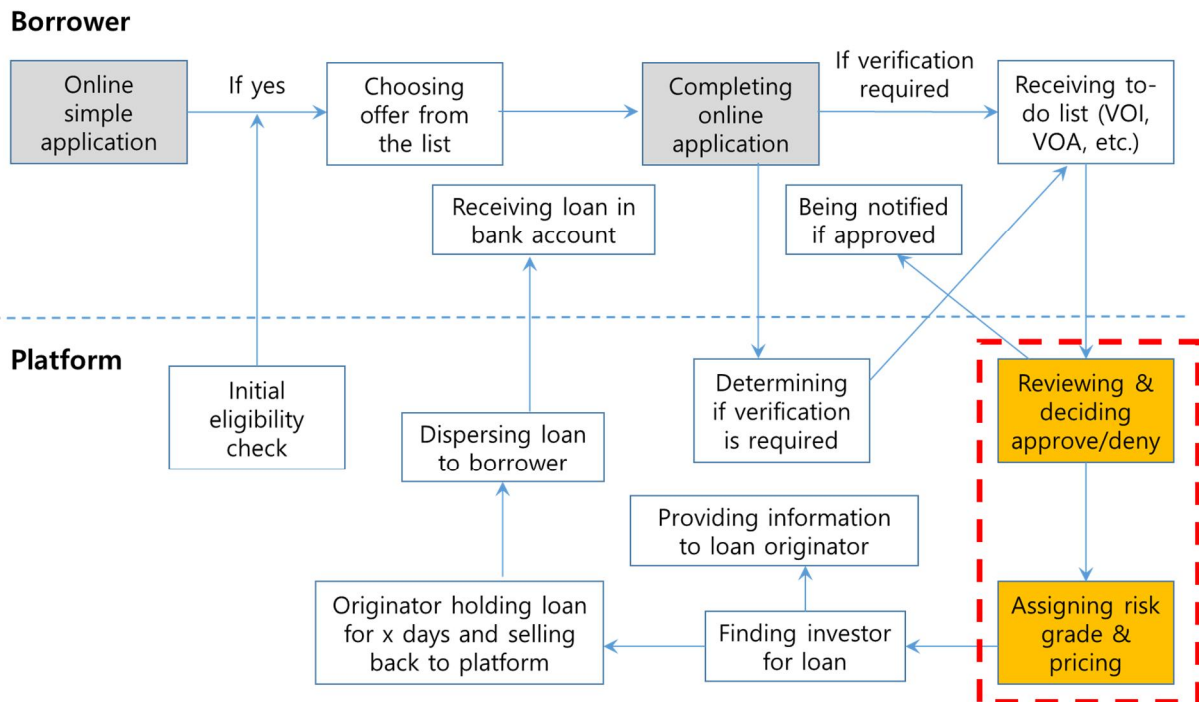
하였고 (케이뱅크 (2016), 카카오뱅크 (2017), 토스 (2019 예비인가)), 다수의 P2P대출업체 현재 영업 중이다. 우리나라의 P2P대출잔액은 2017년 1.7조원에서 2020년 3월 기준 9.6조원으로 크게 증가하였으나, 기존 금융기관의 대출규모에 비해서는 아직 미미한 수준이다.⁷ 현재 우리나라 금융부문에 디지털화를 통하여 효율적으로 공유할 필요가 있는 데이터로는 소비자·기업 신용이력 데이터, 대출성과 (대출실행 이후 연체·조기상환 여부 포함) 데이터, 금융소비자 행태 데이터를 (e.g., digital footprint 등 대안적 데이터 포함) 들 수 있다.

핀테크에 대한 관심이 2015년경부터 고조된 우리나라와 달리, 미국에서는 인터넷이 보급되기 시작한 1990년대 초·중반에 이를 플랫폼으로 하는 다양한 온라인 금융서비스가 등장하였고, 현재 미국의 핀테크 산업은 이들 인터넷 기반 온라인 금융서비스의 발전된 형태라고 할 수 있다. 예를 들면, 1990년대 중반 미국의 주택담보대출 부문은 자동화대출 심사시스템(AUS 또는 Automated Underwriting System)을 소비자(대출신청인)·대출기관·유통화기관을 연결하는 인터넷 기반의 B2B체계로 구축하였고, 이를 통하여 개별 대출신청에 대하여 대출상품·담보물·차입자와 관련된 신용리스크를 미시 데이터와 통계모형(즉, 알고리즘)을 통하여 평가하게 되었고, 이는 금융소비자 효용의 관점에서 편리성·효율성을 획기적으로 높이는 계기가 되었다.⁸ 현재 미국의 핀테크모기지대출 부문은 1990년대 AUS의 연장선에서 대출신청·신용평가·대출승인의 전 과정을 B2C 또는 C2C 온라인시스템에서 (인터넷/모바일 기반) 실행하고 있고, 이 과정에서 정형·비정형 데이터를 사용한 실시간 신용평가, 그리고 자동화가치산정모형(AVM, Automated Valuation Model)을 통한 담보물 평가를 온라인으로 실행하고 있다. ([그림 3]에서 보는 바와 같이 대출실행 여부 및 리스크를 반영한 가산금리 결정이 온라인 대출의 핵심 결정사항임.) 최근 연구에 따르면 미국의 핀테크 모기지대출은 기존의 대출 채널에 비하여 보다 신속한 대출실행과 함께 신용리스크의 관리 측면에서도 보다 우월한 결과를 도출하는 것으로 나타났다 (대출 이후 부실채권의 비율 측면에서). (Fuster et al. (2018))

[그림 3] 미국 P2P대출업체의 전형적인 온라인거래 과정

⁷ 미국·영국 등 외국의 사례와 마찬가지로 우리나라의 P2P대출 및 크라우드펀딩 부문은 기본적으로 고위험 고객군에 대한 대출업무가 위주 (즉, “bottom fishing” 사업모델로 분류; 우리나라의 P2P대출 연체율은 2017년 5.5%에서 2020년 3월 15.8%로 급증).

⁸ 이에 대한 논의는 Cho (2007) 참조



출처: Cho (2020)

DPA 기반의 부동산서비스를 제공하는 프롭테크 부문은 2020년 12월 현재 우리나라에서 총 130개의 업체가 영업 중인 것으로 파악되었다 (‘프롭테크포럼’의 회원사 기준). 이들은 대부분 2013년 이후 설립되었으며, 2019년 이후에는 신규 회사의 진입이 위축되었으나, 이들은 현재 다양한 유형의 플랫폼 기반 서비스를 제공하고 있다: 건물에 대한 공유서비스가 34개사 (26%); 부동산마케팅 플랫폼이 32개사 (24%); 기타 건설솔루션, 스마트홈, 벨류에이션 (AVM 기반의) 전문 업체들이 있고; 금융 관련 프롭테크(‘부동산 핀테크’)도 4개사가 존재함. (이상영 (2020)) 프롭테크 부문은 데이터경제 활성화를 통하여 우리나라 부동산시장의 효율성·투명성을 크게 높일 수 있는 잠재력을 보유하고 있는 것으로 보이지만, 이를 위해서는 주요 부동산 데이터베이스의 보다 효율적인 축적 및 공유, 그리고 기존의 라이선스 기반 서비스제공자와의 (중개업·감정평가업 등) 갈등요인 해소가 해결되어야 할 과제인 것으로 판단된다 (‘타다’와 유사한 문제가 존재). 부동산 부문에서 디지털화를 통해서 공유가 필요한 주요 데이터로는 주택매매자료, 주택전월세신고자료 (신고가격, 동·호수에 대한 비식별화를 통한 개인정보 보호), 상업용 부동산 거래데이터 등이고, 이에 추가적으로 물건 특성 데이터 (내부·외부 특성 포함), 거래에 대한 보다 상세한 사항을 (예: 매매에 걸린 시간, “market in time” 등에) 추가적으로 고려할 수 있다.

마지막으로, 우리나라의 의료부문에서 데이터경제 기반의 혁신을 촉진하기 위해서는 3개 유형의 의료데이터에 대한 디지털화 및 활용도 증대가 주요 과제인 것으로 파악되었

다 (오상우 (2020), 강건욱 (2020)): (1) 개인 의무기록의 활용도 증대 (현재 우리나라는 본인에 대한 의무기록 제공이 매우 제한적이고 활용도 또한 매우 낮은 상황이어서, 향후 소비자 참여형·맞춤형·예방형·예측형⁹ 의료체계를 구축하는데 있어서 장애요인으로 작용); (2) 유전체 데이터를 이용한 보건·의료부문 혁신 제고 (본 데이터는 질병분석 → 서비스/비즈니스모델 개발 → 경제적 가치창출로 이어질 수 있는 잠재력이 매우 높은 분야이고, 의료부문의 데이터경제 활성화를 통한 사회후생 증대가 가능한 중요한 분야임); (3) 생활·환경 데이터의 활용도 제고 (DPA 기반으로 질병의 사전 탐지 및 예방, 그리고 이를 기초로 한 조기 치료의 확대를 위한 중요한 분야고, 현재 많은 혁신이 발생하고 있는 분야이기도 함). 우리나라에서는 상기 데이터의 DB화에 대한 다양한 사업들이 현재 진행되고 있고, 향후 이를 의료부문의 데이터경제 활성화를 위해서는 데이터의 효율적인 (그리고 안전한) 구축·공유·활용과 관련된 제반 기준 및 지침의 마련이 중요한 과제인 것으로 인식되고 있다.

생활·환경 데이터의 활용과 관련하여 개인부착용 장비(wearable devices)를 통한 혈압·혈당의 실시간 체크, 스마트홈(IoT 기반) 내부에서의 생활패턴 데이터의 추적 및 분석을 통한 치매환자의 조기 탐지 및 치료, 디지털 디바이스를 사용한 복약 및 영양 관리 등 다양한 사례들이 현재 보고되고 있다. (오상우 (2020)) [그림 4]에서 보는 바와 같이, 자택에서의 생활패턴을 매시간 측정하고, 이의 변화를 실시간으로 장기간 모니터링하여 치매환자에 대한 발병 초기의 발견 및 치료를 가능하게 한 사례가 DPA 기반의 혁신적 의료서비스의 좋은 예라고 하겠다. 또한 생명보험의 경우에도 생활습관, 부착용 장비를 통하여 소비자 관련 데이터를 수집하고, 이를 상품개발 및 보험료 사정에 반영하는 사례가 보고되고 있고 ('All Life'의 경우 HIV 및 당뇨병 환자들의 상태를 부착용 장비를 통하여 실시간으로 수집·분석하고, 이를 보험상품 개발 및 보험료 산정에 반영), 이는 기본적으로 보험시장에서의 금융포용 확대로 이어지고 있다.¹⁰ 현재 일부 보험회사는 데이터의 구축과 행동과학(nudge 등)에 기반하여 위험의 사전 예방까지를 포함하는 보험서비스를 제

⁹ 미래의학은 향후 4P – Predictive-Preventive-Personalized-Participatory medical service 방식으로 발전할 것으로 예상됨. (강건욱 (2020))

¹⁰ 이와 관련된 인슈어테크 부문의 혁신 사례로, GPS 기반의 앱/디바이스를 사용하여 운전자의 운전행태 및 주행거리를 보험료에 반영 ('Root', 'metromile' 등의 사례; 운전자 행태의 안전성, 주행거리 등을 후불제로 보험료에 반영); 공유경제에 있어서도 인슈어텍업체들이 리스크의 보다 정교한 측정을 통하여 보험상품을 개발·보급하는 사례 존재 ('metromile'이 'Uber'와 협력하여 승용차를 개인적인 용도와 상업용 용도로 사용하는 시간을 분리측정하여 자동차보험 서비스를 제공); 우리나라에서도 '배민커넥트'와 'smallticket'이 협력하여 유사한 배달 관련 보험서비스를 제공

공하고 있고, 이를 위하여 헬스케어산업에 진입하는 사례도 존재하고 있다 ('Clover Health'가 'Roche'와 협력하여 'Clover Therapeutics'를 설립하여 헬스케어산업에 진입). (박소정 (2020))

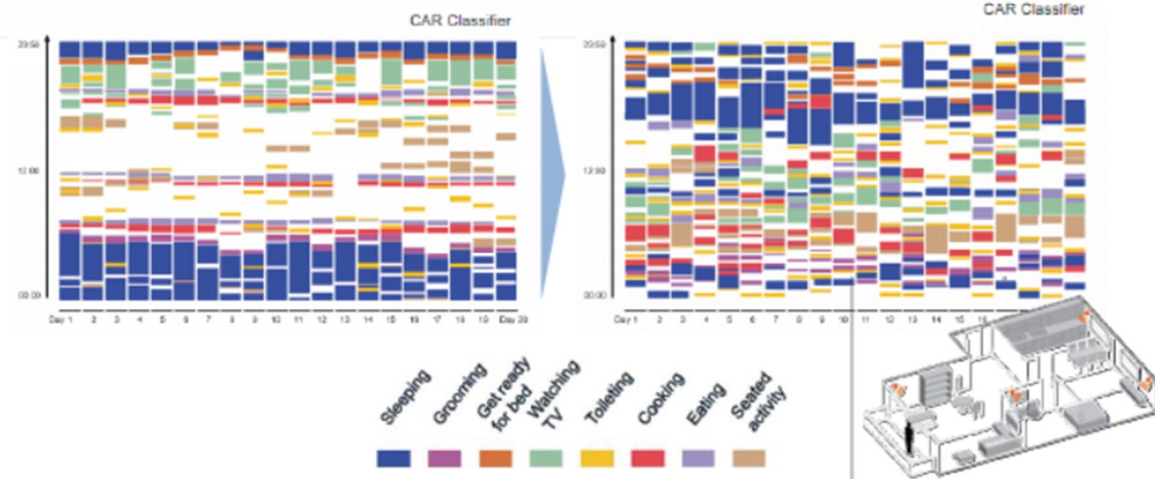
[그림 4] 스마트홈(IoT) 기반의 생활정보 활용 사례 (치매환자 조기 발견 사례)

Health Care Device_낙상방지

DOMO Safety

Behavioral pattern of a healthy 80-year-old woman

Behavioral pattern of a 81-year-old female Alzheimer patient (MoCa 23/30)



출처: 오상우 (2020)

III. 데이터경제의 기대 효과

데이터경제 활성화에 따르는 기대효과는 무엇인가? 이 질문에 대하여 최근 경제학·경영학 분야에서의 관련 연구가 증가하는 추세이고, 본 장에서는 이들 문헌에서 논의한 데이터경제의 경제·사회적 기대효과를 3개 유형으로 구분하여 정리하였다: (1) 플랫폼 (platform) 효과; (2) 예측력 (prediction power) 증대 효과; (3) 새로운 분석기법 (new analytics) 효과.

3.1. 플랫폼 (platform) 효과

금융·부동산·의료 부문에서 인터넷·모바일 플랫폼 기반의 서비스는 기존의 대면서비스에 비하여 거래비용을 획기적으로 감소시키고 (이로 인한 소비자효용의 획기적 증대와 함께), 또한 온라인 거래에서 부수적으로 발생하는 다양한 디지털데이터 (소비자행태 등에 대한) 축적의 증가 효과가 있는 것으로 논의되고 있다. 핀테크 부문에서 이와 같은 플랫폼 효과에 대해서는 이미 다수의 논문들이 발표되었다 (Philippon (2015) and (2016), Cho (2020), 이성복 (2021) 등 참조). 프롭테크 부문에서도 DPA 기반의 서비스 제공이 소비자의 편의 증대 및 거래비용의 감소로 이어질 것으로 예상하고 있고 (Baum (2017), 이상영 (2020)), 디지털헬스 부문에서도 유사한 중개비용의 감소 및 소비자효용 증대 효과가 기대되고 있다. (오상우 (2020), 강건욱 (2020))

그러나 상기 문헌에서는 이와 같은 플랫폼 기반 서비스의 확대는 각 부문에서 기존 중개기능의 무용화 (즉, 탈중개화- disintermediation) 현상을 초래하여 고용인력 감소 효과 또한 나타나는 것으로 논의되고 있다. 그러나 우리나라의 핀테크·프롭테크·디지털헬스 부문에서는 이와 같은 탈중개화 현상이 아직 관찰되지 않는 것으로 파악되지만, 카카오뱅크와 같이 빅테크와 연계된 서비스제공자의 금융산업 진입은 향후 본 부문에서 기존 금융기관들과의 경쟁·경합이 향후 격화될 것으로 예상된다. 또한 이와 같은 플랫폼 기반의 비대면서비스 확대는 사이버 리스크의 증가 (보이스피싱 등), 디지털 데이터를 이용한 국가의 과도한 간섭·통제의 가능성을 높이는 부작용을 초래할 수 있다.

보다 구체적으로, 데이터경제 관점에서 ‘플랫폼’은 어떤 의미를 가지는가? 이에 대하여 문헌에서는 혁신생태계의 플랫폼을 거래플랫폼(Transaction Platform, TP)과 혁신플랫폼

(Innovation Platform, IP)으로 구분한다 (Evans and Gawer (2016)): 전자의 경우 다수의 수요자와 다수의 공급자가 특정 플랫폼을 통하여 온라인으로 연계되어서 상품·서비스의 거래를 함 (예: 미국의 경매 플랫폼 eBay, 영국의 P2P대출 플랫폼 ZOPA – Zone of Possible Agreement 등); 반면, 후자의 경우 플랫폼 제공자는 기본 인프라(하드웨어 및 소프트웨어)를 제공하고, 참여하는 다수의 개인·기업들이 자신의 혁신적 아이디어를 플랫폼을 통하여 공유하는 방식이고 (예: 애플의 “App Ecosystem,” 아마존의 “Market Place Vendors” 등), 이와 같은 혁신플랫폼의 주요 성공요인은 충분한 참여자의 확보와 (“critical mass” 구성을 위한) 함께 이들 간의 상호 보완성인 것으로 (아이디어·자산·기술의 공유를 통한 가치창출에 있어서) 논의되고 있다. 또한 2개 유형 모두에 있어서 플랫폼 참여와 데이터 공유에 대한 준칙의 마련이 (즉, 플랫폼 사용과 관련된 기본적인 규범과 원칙에 대한) 필수적인 과제라고 할 수 있다.

이와 관련하여 다면적 플랫폼(“Multi-Sided Platform, MSP”) 비즈니스 이론은¹¹ 거래비용이 큰 시장일수록 (“high-friction markets”), 그리고 보다 기술 집약적인 시장일수록 (혁신적 ICT, Cloud 등을 사용하는), 지속가능하고 성공적인 MSP의 출현이 가능한 것으로 논의하고 있다. (Rochet/Tirole (2003) & (2006)) 이는 데이터경제 기반의 금융·부동산·의료 서비스 모두에 적용된다고 볼 수 있다. 또한 각 부문에서의 플랫폼의 구축에 있어서 민간참여자는 TP 구축에 초점을 맞추어 혁신적인 상품·서비스의 개발·거래에 치중해야 하는 반면, 공공부문은 대·중·소기업들과 협력하여 다양한 IP를 구축하여 데이터경제의 촉진기제를 조성하는 역할을 담당해야 할 것으로 생각된다.

3.2. 예측력 (predictive power) 효과

데이터경제의 또 다른 효과로 데이터의 축적·공유가 개별 기업의 관점에서 최적의 생산기술 또는 비즈니스 모델을 예측하는데 있어서 오류의 가능성을 낮추는 것으로 논의

¹¹ MSP 이론에 의하면 동일 플랫폼에 다수의 경제주체가 참여하는 경우 이들 간의 간접적 네트워크 효과(indirect network effects)로 인하여 거래되는 상품/서비스의 가격이 기존의 방식과 달리 책정될 수 있음 (즉, 가격이 한계비용(MC)보다 낮게 책정되거나, 심지어는 음의 가격이 나타날 수 있음). 즉, 플랫폼 참여자를 “money-side”와 “subsidy-side”로 구분하고, 후자에 대해서는 플랫폼의 “critical mass”의 구축을 위하여 (즉, 참여를 유도하기 위하여) 할인가격 또는 음의 가격을 (즉, 쿠폰/상품권 제공) 설정하는 것이 가능함.

되고 있다. Farboodi and Veldkamp (2021)의 이론적 모형에 따르면, 식 (1)에서 보는 바와 같이 생산기술의 질을 반영한 기업의 생산량(Quality-adjusted Production, QP)은 질적인 요인(QL)과 투입요소(FA, 즉 자본 및 노동)의 양에 의해서 결정되고, QL은 다시 산업 전반의 최적 기술(식 (2)의 \overline{QL} , 또는 기술의 프론티어)과 개별 기업이 예측을 통하여 선택·적용하는 생산기술로 구분한다¹²:

$$(1) QP_{i,t} = QL_{i,t} \cdot FA_{i,t}$$

$$(2) QL_{i,t} = \overline{QL} - (q_{i,t} - [\theta_t + \varepsilon_{a,i,t}])^2$$

(상기 식에서 i 는 개별 기업, q 는 실제 적용한 생산기술, 대괄호 안의 $[\theta + \varepsilon]$ 은 오차항을 포함한 최적 기술에 대한 예측치를 나타냄.) 식 (2)에서 데이터의 축적은 기업의 최적 생산기술(θ^*)을 예측하는데 있어서 오차($[\varepsilon \cdot \varepsilon]$ 또는 Mean Squared Error, MSE)을 줄이는 효과가 있고, 이는 다시 질적인 생산요인(QL) 및 생산량(QP)의 증가를 초래하는 것으로 분석하고 있다. 또한 생산의 증가는 이에 부수적인 산물인 데이터의 양을 증가시키고, 기업들은 이를 공유·거래함으로써 추가적인 이윤이 가능한 것으로 논의하고 있다.

데이터의 축적으로 인한 생산량 증가는 (즉, MSE를 감소시키는 최적 기술에 대한 보다 정확한 예측으로 인한) 일반적으로 규모의 비경제 효과를 가지는 것으로 상기 논문은 논의하고 있다. 즉, 새로운 아이디어 없이 데이터의 양을 늘리는 것은 MSE 감소효과가 축적된 데이터의 양이 증가함에 따라서 감소하고, 이는 MSE의 하한선이 영이라는 사실에 기인한다. 그러나 이와 같은 규모의 비경제는 통상 축적된 데이터의 양이 이미 충분할 때 나타나는 현상이고, 스타트업과 같은 소규모 기업, 그리고 기존에 구축한 데이터의 양이 작은 상황에서는 데이터의 축적에 따른 생산량의 증가와 이의 증가 속도가 같이 커지는 규모의 경제 (increasing returns to scale) 효과가 나타날 수 있는 것으로 논의하고 있다. (Farboodi and Veldkamp (2021)) 상기 모형으로부터 도출된 또 하나의 논거는 MSE 감소에 기반한 개별기업의 생산성 향상과는 별도로, 기술 프론티어(\overline{QL}) 자체의 발전은 산업 전반의 생산성을 증가시킨다는 것이다 (즉, “envelope-pushing game changer” 유형이 혁신을 의

¹² 최근 일부 연구에서는 데이터를 자본의 성격을 가지는 무형자산으로 보는 시각이 있으나, 상기 분석은 이를 자본이라기 보다는 생산의 질을 향상시키는 정보의 요소로 해석

미함).

같은 맥락에서 다양한 논문들이 데이터경제의 예측력 증대 효과를 분석하고 있다: 거시경제적인 측면에서 데이터의 축적으로 인한 경기순환성에 대한 예측력 증가 (Ordonez (2013), Fajgelbaum et al. (2017)); 금융시장에서 개별 기업의 신용위험에 대한 예측력 증가로 인한 기업 자기자본비용의 감소 (Begenau et al. (2018)), 대출시장에서 차입자의 신용리스크에 대한 예측력 증가로 인한 담보물의 중요성 경감 (Gambacorta et al. (2020)).

금융시장에서의 데이터 축적은 대출·투자·보험의 모든 금융서비스 분야에서 보다 적확(的確)한 리스크의 측정이 가능해 지고, 이로 인해서 서비스제공자와 소비자 간의 정보 비대칭성이 경감될 수 있다고 하는 것이 주요 시사점이다. 실제로 최근 문헌에서는 소비자 신용리스크의 측정에 있어서 온라인 거래로부터 수집한 대안적 데이터의 사용이 (이를 “digital footprint”로 명명) 부도확률 추정 모형의 정확도를 높이고, 이의 결과로 금융소비자의 신용등급의 추정 및 적용의 효율성을 높임으로서 저신용자 및 신용이력 부족 소비자(“thin filers”)에 대한 금융포용이 확대되는 효과가 있는 것으로 논의하고 있다. (Lin et al. (2013), Puri et al. (2017), Hildebrand et al. (2017), Freedman and Jin (2018), Berg et al. (2018) 등) 인슈어테크의 경우에도 데이터 및 테크놀로지의 활용을 통하여 제반 리스크를 보다 정교하게 측정하고, 이를 상품개발에 반영하여 위험을 효율적으로 공유·분산하는 것을 가능하게 하며, 이의 결과로 역선택 방지, 보험사기 감소와 같은 사회적 순기능을 높이는 결과가 도출되는 것으로 논의하고 있다. (박소정 (2020))

부동산시장의 경우에도 데이터의 축적 및 공유는 개별 물건에 대한 가치평가, 그리고 부동산시장의 추이 분석에 있어서 (가격지수를 이용한) 예측력을 높이는 것으로 논의되고 있다. (Glaezer et al. (2018), Wan and Lindenthal (2021) 등) 의료서비스의 경우에도 데이터의 축적과 플랫폼을 통한 공유·분석은 최적의 치료법 또는 신약 개발에 따르는 오차 수준을 감소시키는 효과가 있을 수 있고 (관련 R&D 활동을 통하여), 이로 인한 사회복지의 증진과 이어지는 결과를 도출할 수 있을 것으로 논의하고 있다. (오상우 (2020), 강건욱 (2020))

3.3. 새로운 분석기법 (new analytics) 효과

데이터경제 또 하나의 중요한 축은 머신러닝·딥러닝 등 인공지능 기반의 분석기법을

통하여 다양한 정형·비정형 자료에 대한 분석이 가능해 졌다는 점이다. 예를 들면, 인공 위성 사진을 기초로 한 경제활동 수준에 대한 예측, 세금보고서(예: 미국의 ‘10K 보고서’)를 통한 기업의 성과 및 유형 분석 등이 최근 행해지고 있고, 분석방식에 있어서도 AI 기반의 새로운 방식이 등장하고 있으며 (예: regression trees, LASSO, random forests, ensemble 등), 이를 용이하게 적용할 수 있는 소프트웨어 또한 개발·보급되고 있다 (R, Python 등). 이와 관련하여 PubMed 및 WOS에서 추출한 의료·금융·부동산 부문에서 데이터경제 관련 학술논문의 키워드 분석에서 최근 (2016~20년 기간 중) AI 관련 단어가 최빈명사인 것으로 조사되었다 ([그림 5], [그림 6] 참조).

- 보건·의료 학술논문의 2016~20년 기간 최빈 키워드 (출처: PubMed): 1. Machine Learning (ML); 2. Neural Networks; 3. Algorithms; 4. Deep Learning (DL, 2016~20년 기간 중 새로 출현한 키워드); 5. Artificial Intelligence
- 금융·부동산 학술논문 2016~20년 기간 최빈 키워드 (출처: WOS): 1. BigData; 2. ML; 3. Artificial Intelligence; 4. Digitalization; 5. Social Media; 9. DL (2016~20년 기간 중 새로 출현)

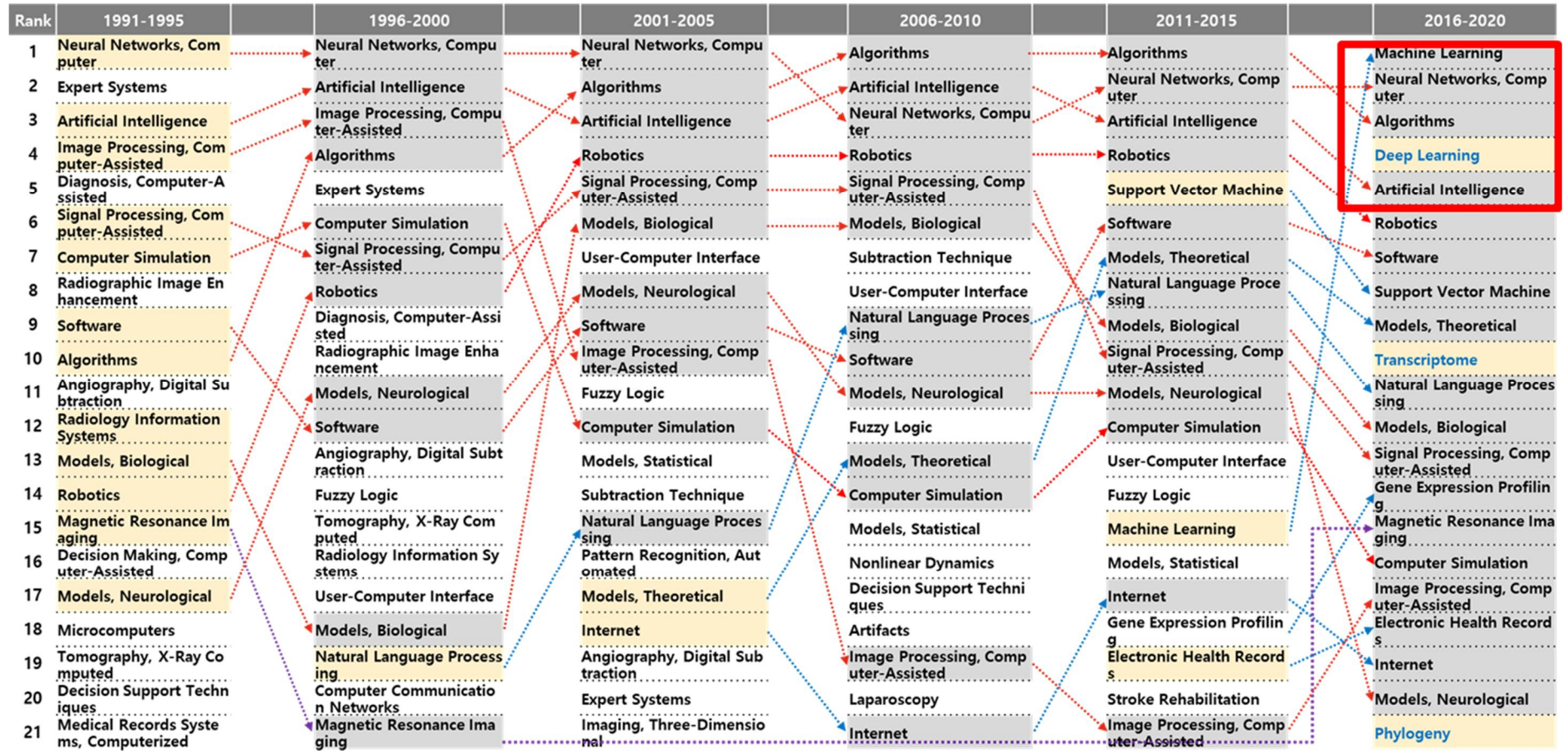
ML·DL 등 AI 기반의 데이터 분석 방식은 이미 ‘일반목적의 기술(General Purpose Technology, GPT)’에 근접한 것으로 평가받고 있다: 즉, 현재 널리 사용되고 있고 (pervasive), 진행 중인 기술진보 뿐만 아니라 미래의 혁신에도 적용 가능하며, 경제 전반에 의미있는 영향을 미친다는 의미. (Teece (2018)) AI 기반의 데이터분석은 기본적으로 다양한 알고리즘을 컴퓨터에 입력하여 인간의 인지·해석·소통·결정 능력을 대신하는 방식이고, 최근에는 신경망(neural network) 방식을 기반으로 하는 딥러닝(DL) 분석의 적용범위가 확대되고 있는 상황이다. 금융부문에서의 로보 어드바이저, 의료부문에서의 AI를 활용한 신약개발, 부동산부문에서 ML방식으로 건물 외관이 자산가격에 미치는 영향에 대한 분석 등에 대한 실증분석 결과가 최근 문헌에서 보고되고 있다.¹³ 이들 논문들은 AI 기반의 분석방식이 기존의 방식에 비하여 (예: 경제이론의 검증방식인 “econometrics”) 예측오차 (MSE)를 대체로 감소시키는 것으로 보고하고 있다.

¹³ 이에 대한 서베이는 Mullainathan and Spiess (2017) 참조

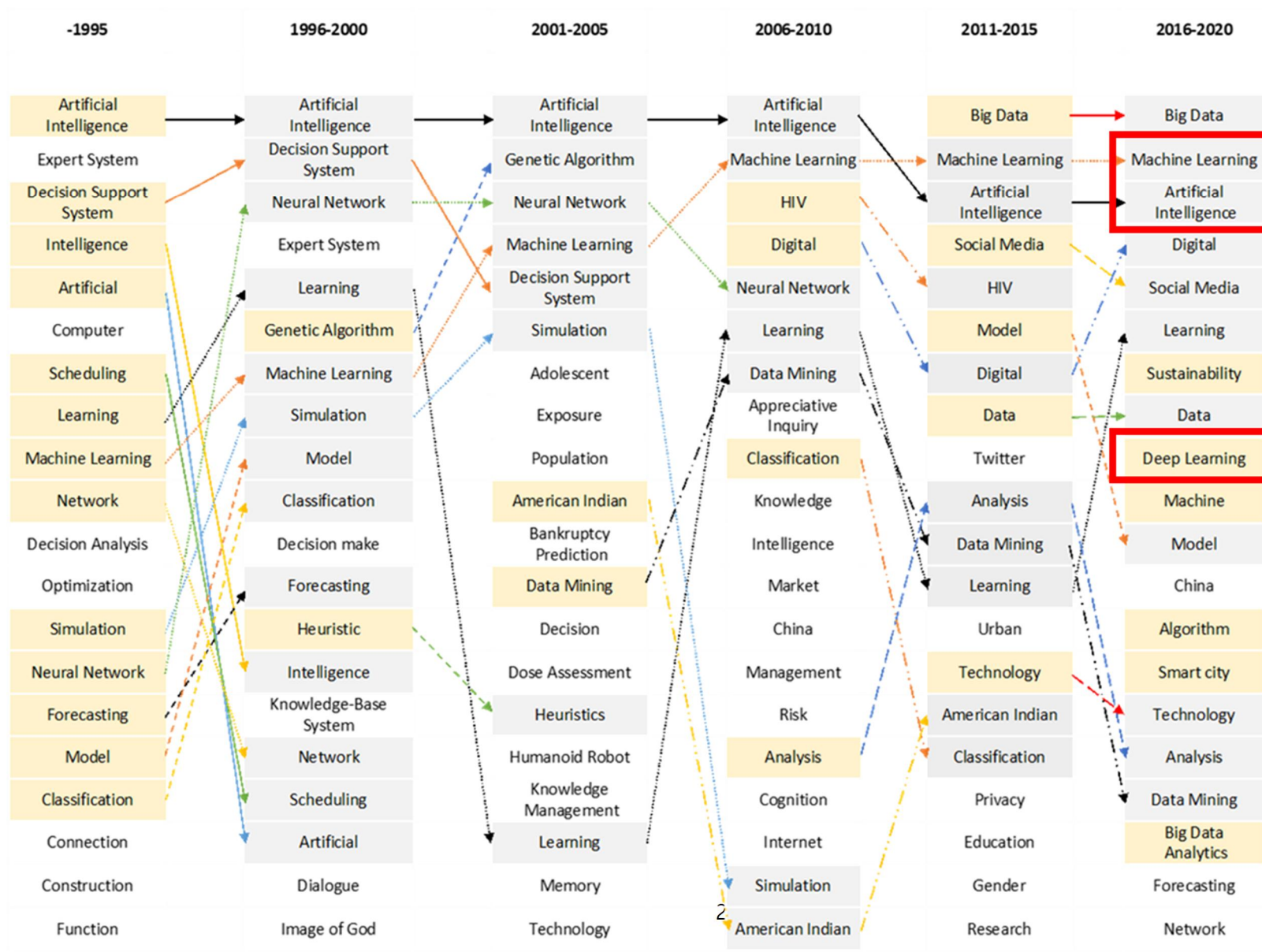
$$(3) y = \beta x + \varepsilon \rightarrow \hat{y} = \hat{\beta} x + \hat{\varepsilon}$$

이와 같은 새로운 분석방식은 학술연구 방식에 있어서도 변화를 초래하고 있다. 즉, 모형의 추정을 통하여 이론으로부터 도출된 가설을 검증하는 기존의 귀납적(top-down) 연구에서는 통상 설명변수(x)들 간의 독립성, 그리고 모형의 적절성을 가정하고, 추정한 모수($\hat{\beta}$)의 부호 및 통계적 유의성에 초점을 맞추는 방식이다. 반면, AI 기반의 분석방식은 사례 중심의 연역적(bottom-up) 방식에서 출발하여, 설명변수(x)들 간의 다양한 연계성 패턴에 대한 분석을 기초로 오차항($\hat{\varepsilon}$)을 최소화 하는 종속변수의 예측치(\hat{y}) 도출에 초점을 맞추는 방식이다. (Mullainathan and Spiess (2017)) 따라서 기존의 실증분석이 통상 이론화·가설검증(이론에 대한)·법·진리로 이어지는 귀납적 방식 위주라고 한다면, ML·DL 기반의 분석은 사례의 관찰·일반화·이론화 과정을 거쳐서 진실을 추구하는 대안적 방식이라고 할 수 있다. 향후 이공계·인문사회계 모두에서 상기 두 가지 방식을 혼합하여 보다 의미있는 연구결과를 도출하고자 하는 노력이 증대될 것으로 예상된다.

[그림 5] 보건·의료 부문 키워드 추이 (PubMed의 MeSH Headings 기준; 1991~2020)



[그림 6] 금융·부동산 부문 키워드 추이 (WOS 논문 대상; 1991~2020)



IV. 데이터경제 관련 메타분석

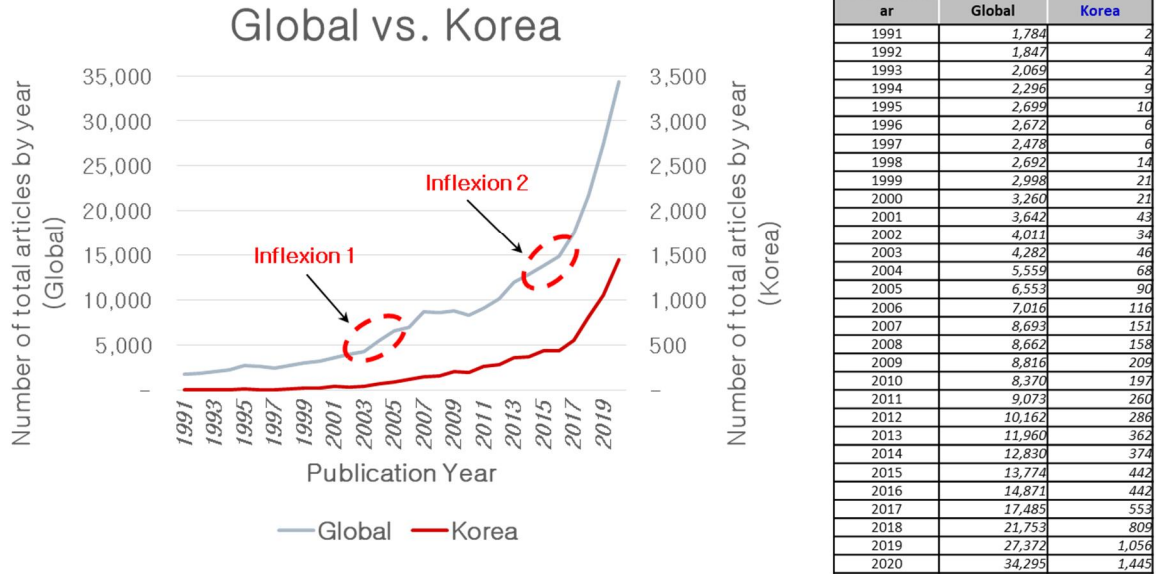
금융·부동산·의료 부문에서의 데이터경제 관련 우리나라의 학술연구 활동에 대한 평가를 위하여 2개 국제학술논문 데이터베이스를 이용하여 관련 논문을 추출하였다. 먼저 보건·의료 관련 논문은 “PubMed Article Bibliography Data”에서 4개의 주요 키워드를 (AI, ML, Digitalization, Digital Health; “or”로 연결하고, 이들의 변형 관련어 포함) 사용하여 1862-2020년까지의 모든 논문을 추출하였고, 이 중 1991년 이후로 한정하여 총 263,974건의 학술논문을 메타분석에 포함하였다. 금융·부동산 관련 학술논문은 Web of Science (WOS) DB에서 관련 키워드를 (FinTech, PropTech, Data Economy, Big Data, AI, ML, DL, Digitalization 및 이의 변형어) 사용하여 사회과학 학술지 중심으로 (일부 다학제적 저널 포함) 논문을 추출하였고 (1978~2022년 기간 중), 총 26,859 건의 학술논문을 분석에 포함하였다.

결과 (1): 2015년 이후 관련 학술논문의 폭발적 증가

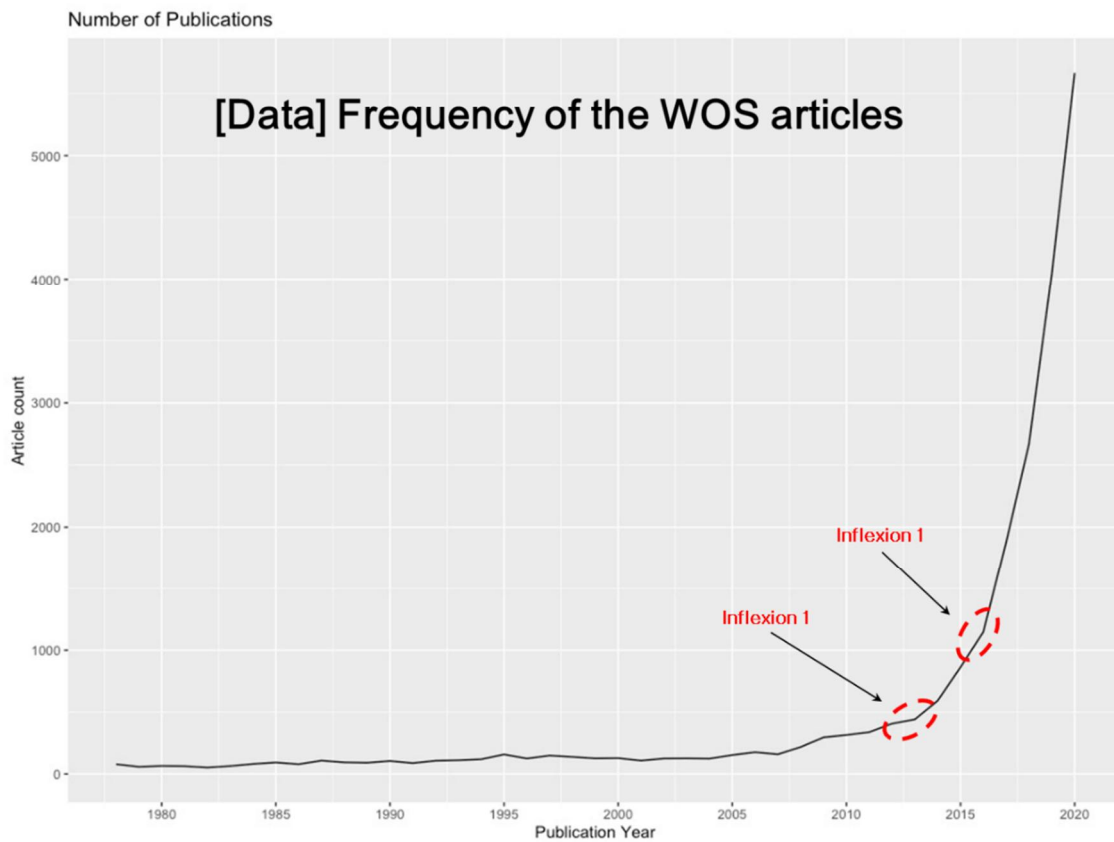
보건·의료 분야의 데이터경제 관련 논문의 숫자가 2015년경부터 폭발적으로 증가하는 것으로 관찰되었다. 금융·부동산 분야 역시 2015년경부터 학술논문의 폭발적인 증가 추세가 관찰되었고, 이의 상승속도는 의료 부문에 비하여 더욱 가파른 것으로 나타났다. 그러나 미국의 경우 2015년 이전 2003~05년 기간 중 보건·의료 분야의 논문이 증가하는 시기가 한번 더 관찰된 반면, 우리나라는 2015년부터의 급증이 유일한 상승 추세인 것으로 나타났다. 이는 인터넷 기반의 플랫폼 산업이 미국에서 우리나라 보다 더 빠른 시기에 (1990년대) 시작되었고, 이의 결과로 아마존·구글·마이크로소프트 등의 글로벌 IT·플랫폼 기업들의 약진과 함께, 데이터경제 관련 학술연구에 있어서도 여타 국가들과 달리 2000년대 초반 또 한번의 상승기가 있었던 것으로 해석할 수 있다 (즉, 우리나라는 2003~05년 관찰된 1차 상승 웨이브에 편승하지 못했던 것으로 해석됨).

[그림 7] 보건·의료 분야 학술논문 추이: Global vs. Korea

Digitalization in Biotechnology and Medicine



[그림 8] 금융·부동산 분야 학술논문 추이

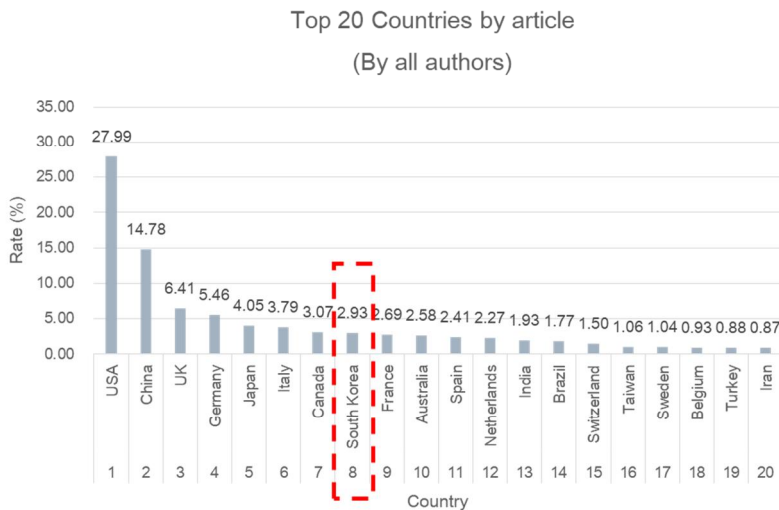


결과 (2): 미국·중국 저자의 지배적인 논문 발간 (영국이 3위)

학술논문의 양적인 측면에서 미국·중국·영국의 (소속기관 기준) 비중이 압도적으로 높은 것으로 나타났다 (의료 분야의 경우 미국·중국의 비중이 전체 논문의 약 50%에 달하고, 중국이 미국의 1/2, 영국이 중국의 1/2에 해당함) ([그림 9] 참조). 영국 다음으로는 독일, 일본, 이탈리아 순이고, 우리나라는 8위를 기록하였다. 의료·보건 부문 논문의 증가 속도를 살펴보면, 미국은 표본 시작년 (1991년) 다른 나라들에 비해서 압도적으로 앞서 있는 상황이었으며, 측정기간 동안의 증가율도 높았으며, 유럽 국가들은 표본 시작년 논문 개수가 미국보다는 낮으나 아시아 국가들보다 많았으며 증가율은 미국보다 높은 상황이고, 한국, 중국은 표본 시작년 논문개수가 매우 낮았으나 증가율은 상기 국가들에 비하여 높은 상황이다 ([그림 10] 참조). (일본은 표본 시작년 논문 개수는 많으나 증가율이 낮은 상황임.) 금융·부동산의 경우에도 의료와 유사한 결과가 관찰되었다: 미국·중국·영국 순서이고, 다음은 스페인·이탈리아·독일이고, 우리나라는 8위에 위치하고 있음 ([그림 11], [그림 12] 참조).

[그림 9] 보건·의료 분야 학술논문: Top 20 Countries

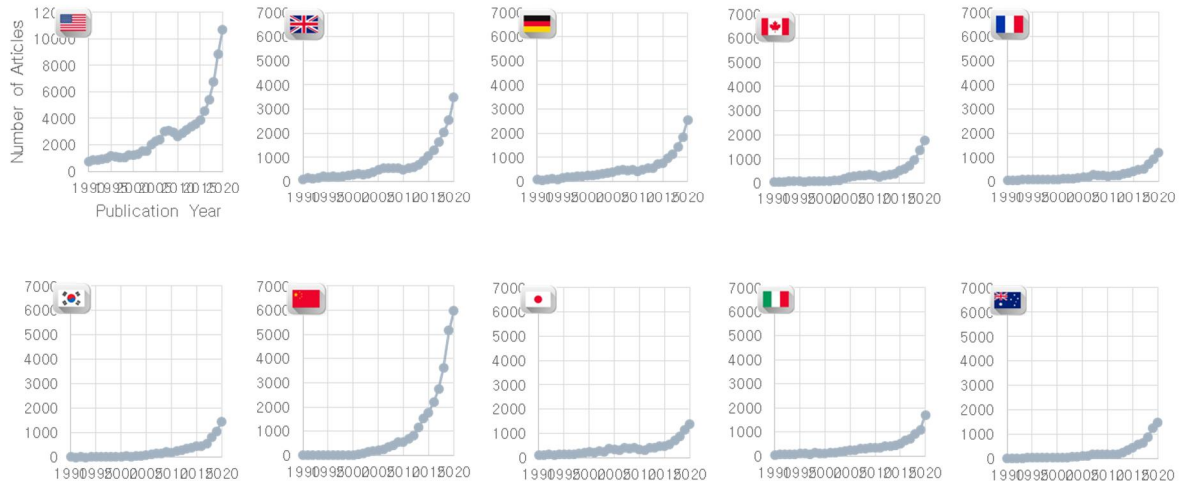
Top 20 Countries



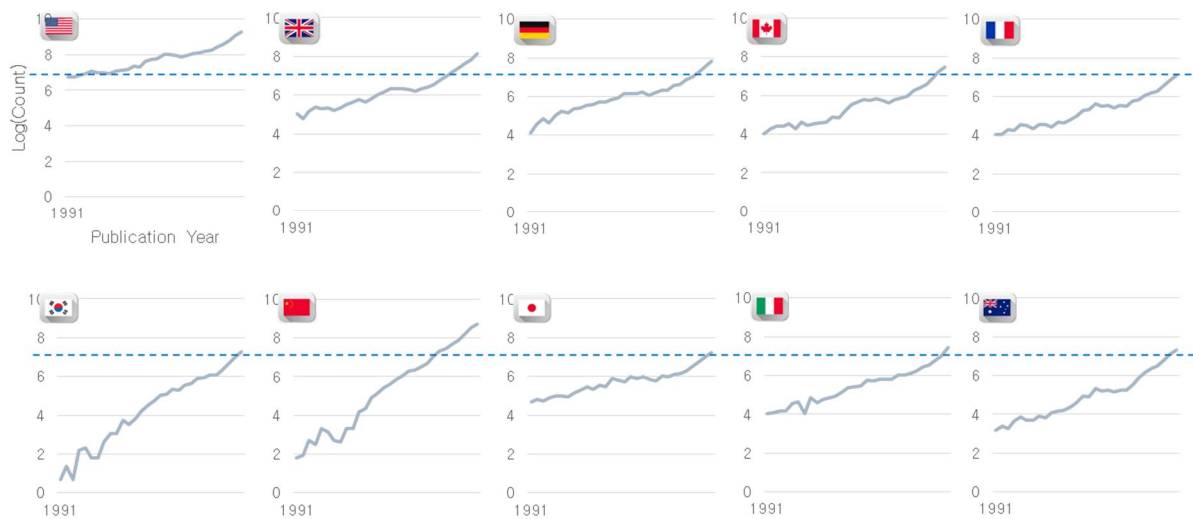
	Country	Count	Rate(%)
1	USA	245,185	27.99
2	China	129,437	14.78
3	UK	56,105	6.41
4	Germany	47,804	5.46
5	Japan	35,446	4.05
6	Italy	33,185	3.79
7	Canada	26,878	3.07
8	South Korea	25,676	2.93
9	France	23,591	2.69
10	Australia	22,612	2.58
11	Spain	21,102	2.41
12	Netherlands	19,854	2.27
13	India	16,865	1.93
14	Brazil	15,478	1.77
15	Switzerland	13,123	1.50
16	Taiwan	9,327	1.06
17	Sweden	9,089	1.04
18	Belgium	8,156	0.93
19	Turkey	7,706	0.88
20	Iran	7,599	0.87

[그림 10] 보건·의료 분야 학술논문: 국가별 추이 (단순 증가 vs. 로그 변환 증가)

Growth in Numbers: 10 Countries



Growth in Numbers: Log Transformation



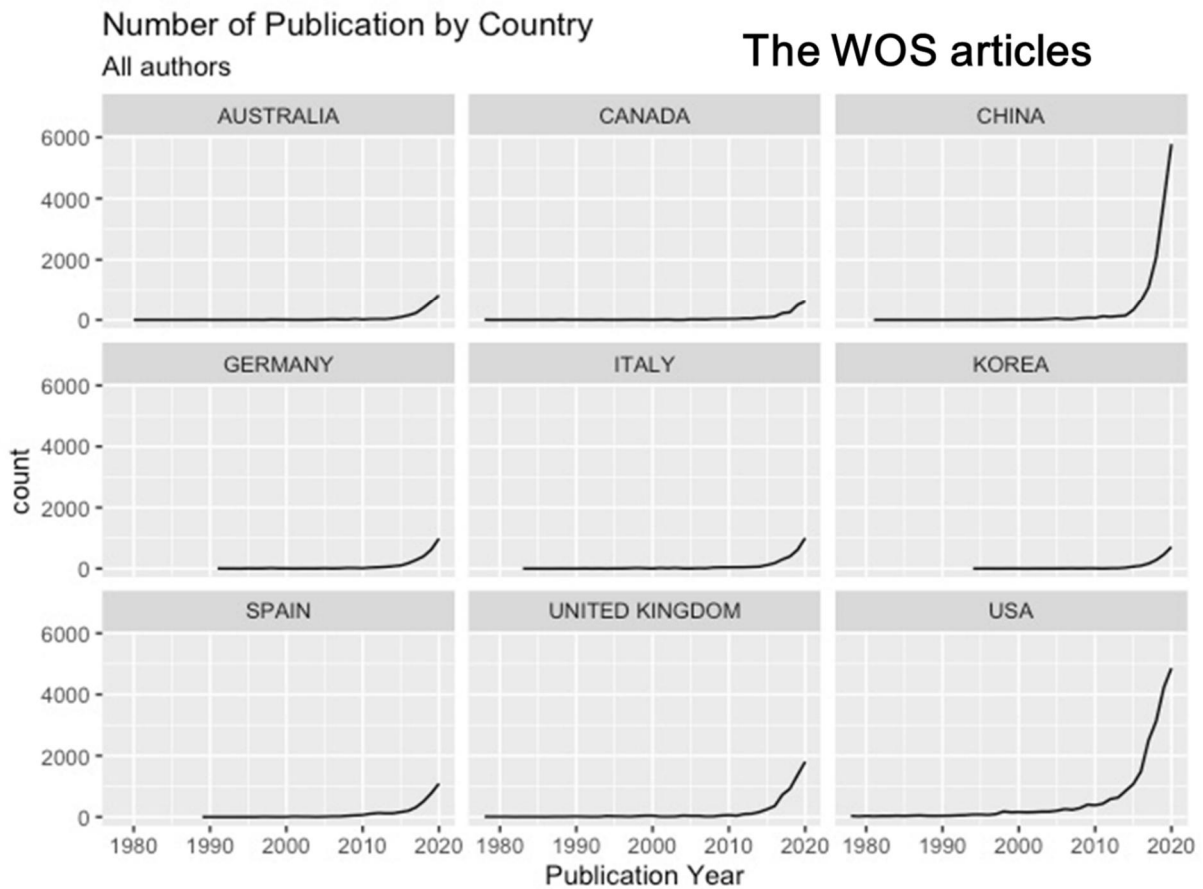
[그림 11] 금융·부동산 분야 학술논문: Top 10 Countries

Distribution of the WOS articles

Corresponding Author's Countries							
	Country	Articles	Freq	SCP	MCP	MCP_Ratio	
1	USA	7,010	0.27	5,657	1,353	0.19	
2	CHINA	4,314	0.17	2,869	1,445	0.34	
3	UNITED KINGDOM	2,132	0.08	1,343	789	0.37	
4	SPAIN	1,127	0.04	826	301	0.27	
5	ITALY	1,074	0.04	686	388	0.36	
6	GERMANY	972	0.04	623	349	0.36	
7	AUSTRALIA	779	0.03	471	308	0.40	
8	KOREA	704	0.03	526	178	0.25	
9	CANADA	691	0.03	464	227	0.33	
10	NETHERLANDS	570	0.02	340	230	0.40	

SCP: Single Country Publications
MCP: Multiple Country Publications

[그림 12] 금융·부동산 분야 학술논문: 국가별 추이

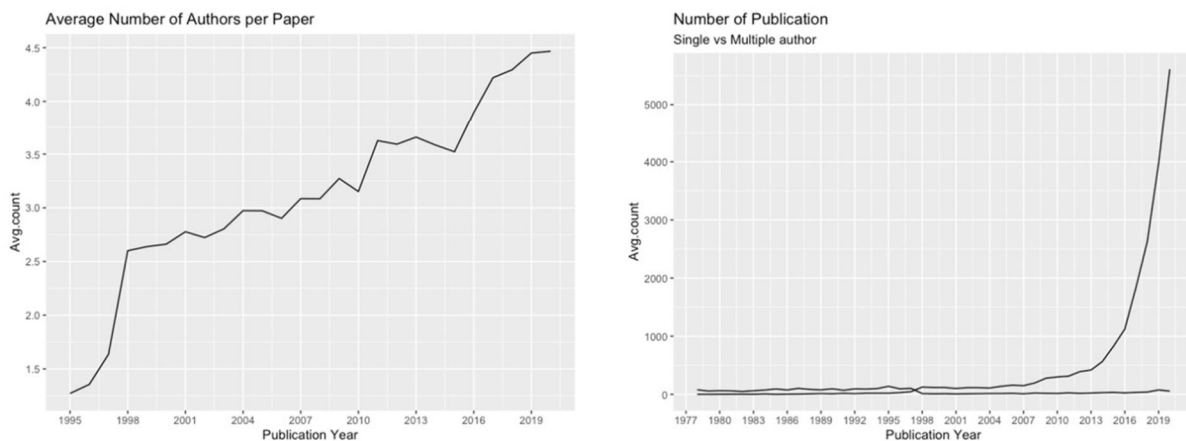


결과 (3): 공저자 숫자 및 국제 학술협력 증가

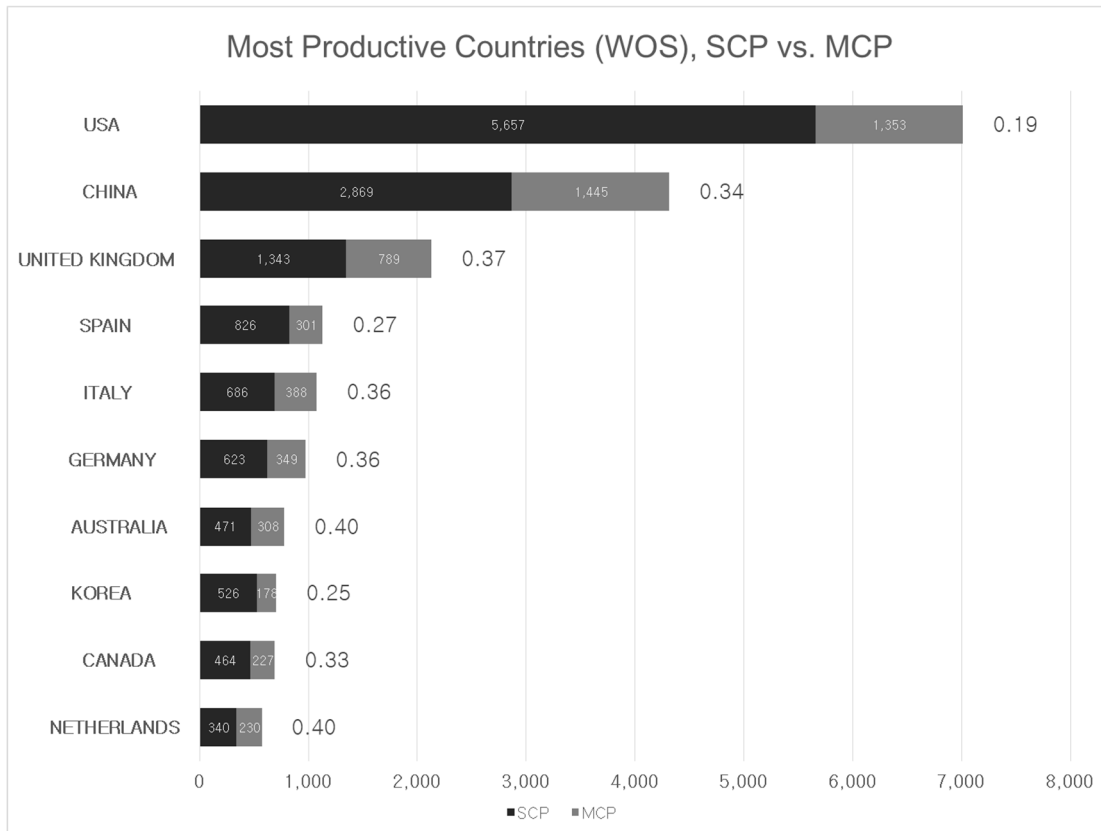
메타분석에 나타난 또 하나의 중요한 추이는 학술논문의 작성에 있어서 공저자 수 및 국제협력 논문의 비중이 지속적인 증가 추세다. 미국에서 발간된 금융·부동산 분야 논문의 80%가 다국가 저자에 의한 논문이며, 논문 한 편당 평균 저자수도 1998년 2.5명에서 2020년 4.5명으로 증가하였고, 대부분의 국가에서 다국간 논문이 최근 급증하는 추세다 ([그림 13] ~ [그림 15] 참조). 의료 분야의 연구에 있어서도 우리나라의 연구자들은 전체 논문의 약 35%를 미국 연구자들과 공저하였고, 미국 이외에 중국·독일·영국·일본·호주의 순서로 협력하는 것으로 나타났다 (흥미로운 것은 이란과의 협력 비중이 높은 점임). 한 가지 특이한 사항은 우리나라 의료부문의 연구에서 2006년 이전에는 국제협력이 매우 미미한 것으로 관찰되었고, 그 이후 빠르게 증가하고 있다는 점이다 (특히 2009년 경부터 증가율이 크게 상승함). 저자의 소속과 관련하여 한국의 경우 병원의 참여 비중이 상대적으로 높으며, 미국의 경우 연구 주체별 비중이 다양하다는 특징이 관찰되었다.

[그림 13] 금융·부동산 분야 학술논문: 공저자 수 및 국제협력 논문 추이

Trends of co-authorship & cross-country collaborations

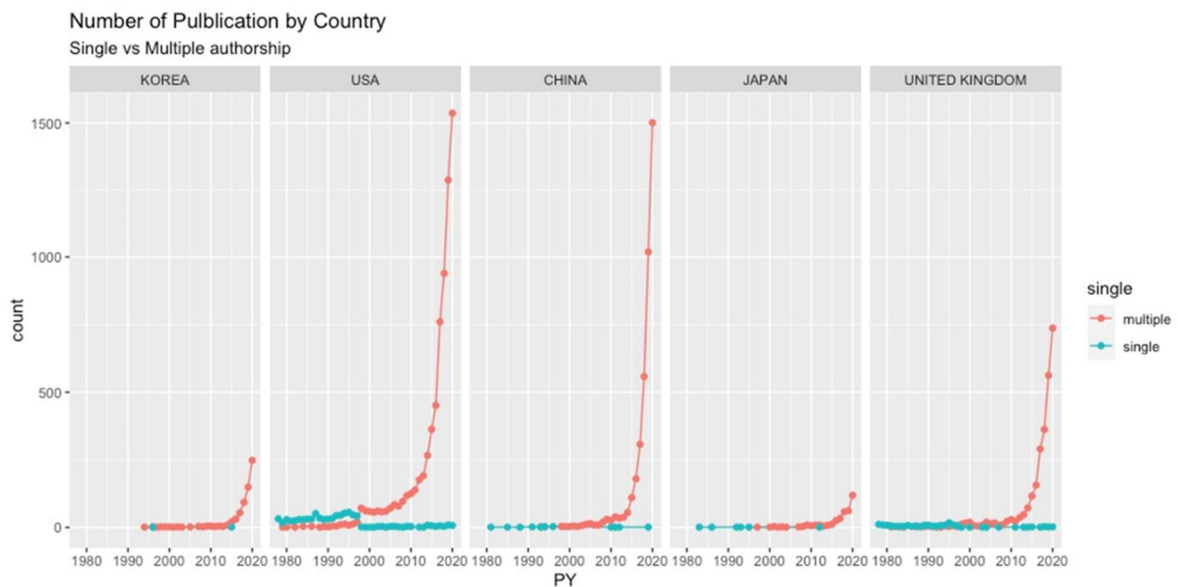


[그림 14] 금융·부동산 분야 학술논문: 국제협력 논문의 국가별 추이 (1)

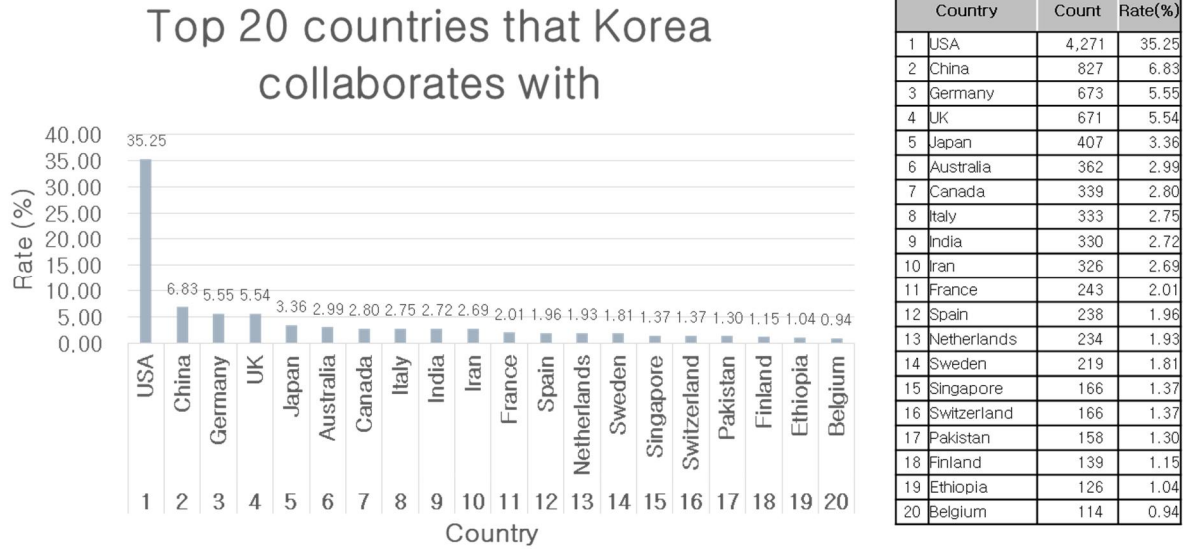


[그림 15] 금융·부동산 분야 학술논문: 국제협력 논문의 국가별 추이 (2)

Distribution of the WOS articles



[그림 16] 금융·부동산 분야 학술논문: 국내 연구자들의 국제협력 추이 (국가별)



[그림 17] 금융·부동산 분야 학술논문: 국내 연구자들의 국제협력 추이 (시기별)

1991 ~ 1995	2011 ~ 2015	2016 ~ 2020
1 South Korea 27	1 USA 389	1 USA 3,830
	2 China 109	2 China 688
	3 UK 51	3 Germany 621
	4 Japan 40	4 UK 618
	5 Germany 39	5 Japan 354
	6 Italy 34	6 Australia 341
	7 Spain 32	7 Iran 324
	8 Australia 21	8 Canada 323
	9 Canada 20	9 India 308
	10 France 12	10 Italy 278
	11 Belgium 11	11 France 230
	12 Czech Republic 9	12 Netherlands 227
	13 India 9	13 Sweden 215
	14 Malaysia 9	14 Spain 202
	15 Pakistan 9	15 Switzerland 161
	16 Netherlands 7	16 Singapore 154
	17 Taiwan 7	17 Pakistan 149
	18 Austria 6	18 Finland 132
	19 Greece 6	19 Ethiopia 126
	20 Poland 6	20 Belgium 102
	21 Singapore 6	21 Poland 102
	22 Luxembourg 4	22 Portugal 102
	23 New Zealand 4	23 Brazil 100
	24 Portugal 4	24 Taiwan 82
	25 Sweden 4	25 Denmark 79

결과 (4): 연구의 질적인 측면에서의 국가 간 차이

학술논문의 질적인 측면에서 주요 국가들간에 상이한 패턴이 관찰되었고, 우리나라의 경우 보건·의료 분야와 금융·부동산 분야가 이 면에서 유사한 것으로 나타났다. 금융·부동산의 경우 논문 한 편당 평균 인용회수가 미국 22.1, 영국 19.6, 네덜란드 15.2에 비하여 우리나라와 중국은 9.8인 (미국의 절반에도 미치지 못하는 수준인) 것으로 관찰되었다. 보건·의료 분야의 경우 PubMed에서 제공하는 “MeSH Heading Keywords System (MHKS)”을 사용하여 개별 논문의 핵심 키워드가 MHKS에 최초로 등록된 연도와 논문의 발간연도 차이를 구하고, 이를 국가별로 평균을 계산하여 연구주제의 참신성을 비교하였다 (평균값이 작을수록 연구주제가 참신한 것으로 평가함). 이 면에서 미국·프랑스·독일·일본이 약 23.3년을 기록하여 비교 국가들을 앞서는 것으로 나타난 반면, 우리나라는 25.7년으로 (중국은 25.3년) 상기 선진국들에 비하여 연구주제의 참신성이 떨어지는 것으로 관찰되었다. 상기 결과인 2.4년의 차이가 크지 않은 것으로 보일 수 있으나, 현재와 같이 기술수준이 빠르게 변하는 상황에서는 상당한 수준의 격차일 수 있다. 또한 미국의 경우 2003-2006년과 2016년 이후 등 두 번에 걸친 관련 MHKS 토픽의 급상승기(“waves”)가 있었으나, 우리나라의 경우 첫번째 웨이브에 편승하지 못했고 두번째 웨이브에는 편승한 것으로 나타났다. 마지막으로, COVID19과 같은 사회적 문제에 대한 디지털 헬스의 반응이 국가별로 차이가 있다: (1) 미국의 경우 상위 20개 토픽에 6개의 COVID관련 토픽이 바로 반영되나, (2) 한국의 경우 상위 20개 토픽에 2개의 COVID관련 토픽만 반영.

[그림 18] 금융·부동산 분야 학술논문: 평균 인용회수 국제비교

Total and average citations (WOS)

Total Citations per Country				
	Country	Total Citations (A)	Number of articles (B)	Average Article Citations (A/B)
1	USA	154,979	7,010	22.11
2	CHINA	42,604	4,314	9.88
3	UNITED KINGDOM	41,984	2,132	19.69
4	GERMANY	13,585	972	13.98
5	ITALY	13,582	1,074	12.65
6	AUSTRALIA	10,951	779	14.06
7	SPAIN	10,621	1,127	9.42
8	CANADA	8,913	691	12.90
9	NETHERLANDS	8,684	570	15.24
10	KOREA	6,897	704	9.80

[그림 19] 보건·의료 분야 학술논문: 연구주제의 참신성 국제비교

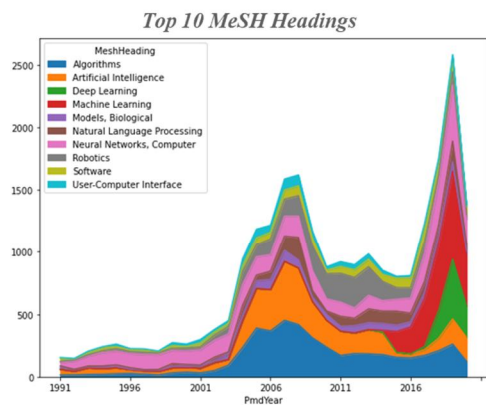
“Newness” of Topics by MeSH Age

Total MeSH Headings				Major topic of MeSH Headings			
Country	Sum. (MeSH Age)	Total Count (Article)	Avg. (MeSH Age)	Country	Sum. (MeSH Age)	Total Count (Article)	Avg. (MeSH Age)
1 Germany	4,067,471	155,373	26.2	1 USA	2,323,299	101,181	23.0
2 France	1,854,449	70,681	26.2	2 France	219,370	9,372	23.4
3 USA	19,282,407	734,135	26.3	3 Germany	470,789	20,003	23.5
4 Japan	2,998,968	112,909	26.6	4 Japan	303,831	12,863	23.6
5 UK	4,113,648	152,001	27.1	5 Italy	258,077	10,711	24.1
6 Canada	2,069,033	75,999	27.2	6 Switzerland	119,539	4,967	24.1
7 Switzerland	971,324	35,758	27.2	7 Canada	255,851	10,507	24.4
8 Netherlands	1,507,057	54,966	27.4	8 UK	529,376	21,623	24.5
9 Italy	2,373,947	85,842	27.7	9 Spain	184,541	7,450	24.8
10 Spain	1,413,584	49,963	28.3	10 Netherlands	169,766	6,806	24.9
11 Australia	1,525,376	52,812	28.9	11 India	137,596	5,459	25.2
12 India	1,147,221	39,614	29.0	12 China	881,420	34,838	25.3
13 China	7,153,615	245,107	29.2	13 South Korea	175,603	6,836	25.7
14 South Korea	1,682,656	55,677	30.2	14 Australia	188,267	7,269	25.9
15 Brazil	1,292,673	42,256	30.6	15 Brazil	139,553	5,060	27.6

[그림 20] 보건·의료 분야 학술논문: 미국의 연구주제 추이

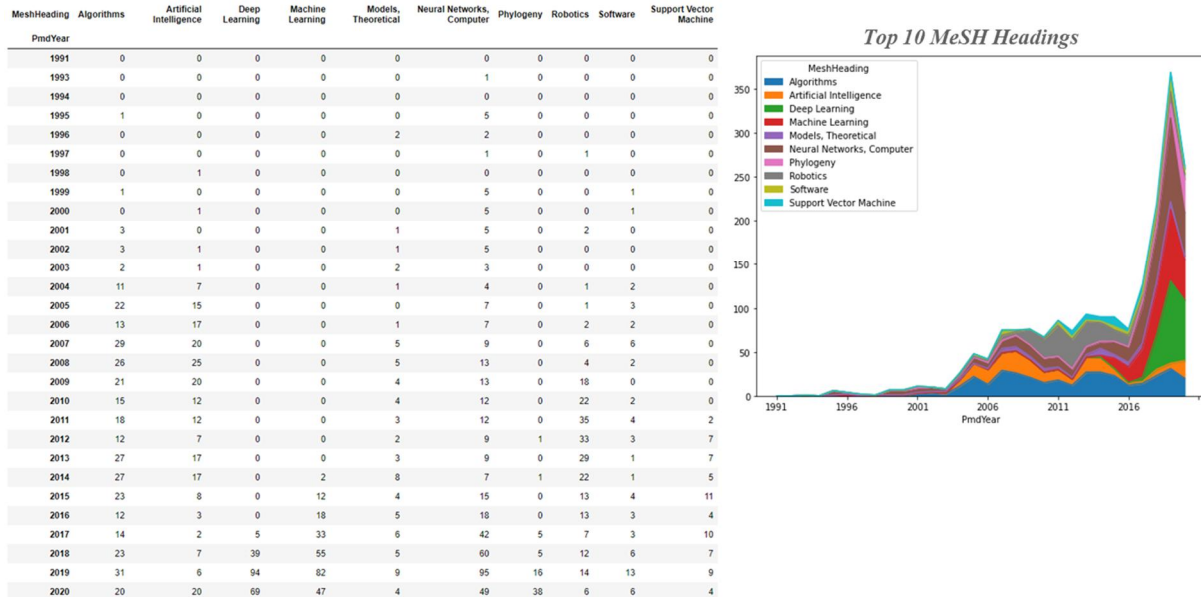
□ Digital health research trends: USA (Cont.)

PmdYear	Algorithms	Artificial Intelligence	Deep Learning	Machine Learning	Models, Theoretical	Neural Networks, Computer	Phylogeny	Robotics	Software	Support Vector Machine
1991	0	0	0	0	0	0	0	0	0	0
1993	0	0	0	0	0	1	0	0	0	0
1994	0	0	0	0	0	0	0	0	0	0
1995	1	0	0	0	0	5	0	0	0	0
1996	0	0	0	0	2	2	0	0	0	0
1997	0	0	0	0	0	1	0	1	0	0
1998	0	1	0	0	0	0	0	0	0	0
1999	1	0	0	0	0	5	0	0	1	0
2000	0	1	0	0	0	5	0	0	1	0
2001	3	0	0	0	1	5	0	2	0	0
2002	3	1	0	0	1	5	0	0	0	0
2003	2	1	0	0	2	3	0	0	0	0
2004	11	7	0	0	1	4	0	1	2	0
2005	22	15	0	0	0	7	0	1	3	0
2006	13	17	0	0	1	7	0	2	2	0
2007	29	20	0	0	5	9	0	6	6	0
2008	26	25	0	0	5	13	0	4	2	0
2009	21	20	0	0	4	13	0	18	0	0
2010	15	12	0	0	4	12	0	22	2	0
2011	18	12	0	0	3	12	0	35	4	2
2012	12	7	0	0	2	9	1	33	3	7
2013	27	17	0	0	3	9	0	29	1	7
2014	27	17	0	2	8	7	1	22	1	5
2015	23	8	0	12	4	15	0	13	4	11
2016	12	3	0	18	5	18	0	13	3	4
2017	14	2	5	33	6	42	5	7	3	10
2018	23	7	39	55	5	60	5	12	6	7
2019	31	6	94	82	9	95	16	14	13	9
2020	20	20	69	47	4	49	38	6	6	4



[그림 21] 보건·의료 분야 학술논문: 한국의 연구주제 추이

□ Digital health research trends: Korea (Cont.)



[그림 22] 보건·의료 분야 학술논문: 연구주제 (2019 & 2020)

□ COVID-19 shock: Overall

2019		
Rank	Mesh Headings	Count
1	Machine Learning	1,773
2	Neural Networks, Computer	1,697
3	Deep Learning	1,211
4	Algorithms	811
5	Artificial Intelligence	523
6	Robotics	399
7	Software	334
8	Natural Language Processing	260
9	Support Vector Machine	250
10	Electronic Health Records	212
11	Models, Biological	208
12	Magnetic Resonance Imaging	205
13	Models, Theoretical	198
14	Image Processing, Computer-Assisted	186
15	Phylogeny	180
16	Gene Expression Profiling	174
17	Transcriptome	169
18	Signal Processing, Computer-Assisted	164
19	Computer Simulation	162
20	Computational Biology	150

2020		
Rank	Mesh Headings	Count
1	Machine Learning	1,088
2	Neural Networks, Computer	1,007
3	Deep Learning	840
4	Artificial Intelligence	499
5	Algorithms	444
6	Pandemics	236
7	Robotics	221
8	Phylogeny	219
9	Betacoronavirus	206
10	Software	186
11	Coronavirus Infections	144
12	Pneumonia, Viral	140
13	Telemedicine	136
14	Support Vector Machine	127
15	Image Processing, Computer-Assisted	112
16	Dental Implants	109
17	Magnetic Resonance Imaging	102
18	Natural Language Processing	101
19	COVID-19	97
20	Mobile Applications	93

V. 부문별 데이터경제 평가

1. 금융부문 데이터경제 평가

우리나라 금융부문의 데이터경제 생태계 조성은 현재 타 부문에 비하여 보다 원활하게 진행되고 있는 것으로 보인다. 구체적으로, 2019.12월 데이터3법의 국회 통과 이후 P2P대출법 및 전자상거래법에 입법논의가 진행되고 있고, 디지털데이터의 유통·거래를 위한 금융데이터거래소 설립, 금융소비자의 개인정보를 이용한 서비스업인 본인신용정보관리업(마이데이터) 육성, 금융규제샌드박스 도입 등이 현재 금융부문에서 데이터경제 생태계 구축과 관련하여 진행되고 있다. 특별히 마이데이터의 경우 2021년 40여개의 업체가 금융위원회로부터 허가를 받을 예정이고, 이들은 고유업무인 개인 신용정보의 수집을 통한 조회·열람서비스 이외에도 다양한 부수업무를 수행할 수 있다 (로보어드바이저를 통한 투자자문, 금융상품자문업, 대출의 중개·주선, 신용정보업, 데이터 분석 및 컨설팅 서비스 등).¹⁴

우리나라 금융부문에서의 데이터경제 생태계 조성은 주로 기존의 금융기관, 그리고 빅테크와 연계된 인터넷전문은행에 의해서 이루어지고 있고, 핀테크 관련 스타트업 등의 중소기업의 역할은 아직 크지 않은 것으로 보인다. 개별 금융기관의 경우 자체의 인터넷·모바일 플랫폼을 기반으로 하는 온라인 금융서비스 이외에도, 은행권 전반의 ‘오픈뱅킹’과 같은 통합 플랫폼을 통한 서비스도 확대되는 추세다. 이에 대한 보다 미시적인 분석이 필요하겠으나, 데이터경제의 플랫폼 효과 측면에서 우리나라의 대출·보험·투자 부문은 국제 경쟁력이 있는 온라인 서비스를 이미 제공하고 있는 것으로 보인다.

그러나 데이터경제의 예측력 증대 효과 측면에서 우리나라의 금융부문은 아직 데이터 기반의 리스크관리 문화가 정착되지 않은 것으로 판단되고, 이는 향후 본 부문에서의 데이터경제 활성화를 통하여 개선해야 할 핵심적인 사항인 것으로 생각된다. 금융시장에서의 리스크관리는 기본적으로 소비자(차입자)·금융상품·담보물로부터 발생할 수 있는 비체계적(idiosyncratic) 리스크와 금융시장·거시경제로부터 발생할 수 있는 체계적(systematic) 리스크를 제반 미시·거시 데이터를 통하여 측정하고, 이를 세 가지 결정에 반영하는 것

¹⁴ 우리나라 금융부문에서의 데이터경제 촉진기제에 대한 소개는 [부록 2] 참조

이라고 할 수 있다 - 리스크 회피 (예: 대출심사를 통한 ‘risk avoidance’ 여부 결정), 리스크 기반의 가격산정 (데이터 분석을 통한 ‘risk-based pricing’ 실현), 그리고 리스크의 공유·전이 (파생상품시장을 통한 ‘risk transfer & hedging’ 실현). 미국·영국·EU국가들의 경우 다양한 정형·비정형 데이터를 활용한 신용리스크의 측정 및 관리에 대한 실증분석 결과가 지난 수년간 다수 발표되고 있고, 이들에 의하면 대안적 데이터(digital footprint 등)의 사용이 신용리스크 모형의 정확도를 높이고, 이를 반영한 신용등급의 산정은 소비자 계층별 맞춤형 상품개발을 가능하게 하여, 금융포용의 효과를 도출하는 것으로 논의하고 있다. 우리나라에서도 향후 제반 정형·비정형 데이터 및 AI기반의 새로운 분석방식을 활용하여 보다 과학적이고 증거에 기반한 금융리스크의 측정과 관리를 촉진할 필요가 있고, 이를 위한 공공·산업·학계의 협력이 금융부문에서의 데이터경제 생태계의 중요한 과제인 것으로 생각된다.

금융규제 측면에서도 데이터경제 활성화는 금융기관에 대한 건전성 규제 뿐만 아니라 금융소비자 보호를 위한 행위규제에 있어서도 사회적 효과가 있는 규제방안의 도출에 기여할 수 있을 것으로 생각된다. 즉, 다양한 정형·비정형 데이터의 활용과 함께, AI기반의 분석기법을 사용하여 보다 효과적인 (그리고 데이터·증거 기반의) 미시·거시 건전성 규제 방안을 도출할 수 있을 것으로 보인다 (개별 금융기관 및 금융시장 차원의 신용리스크 관리 차원에서). 금융소비자 보호 측면에서, 독일의 경우 금융기관이 금융소비자에게 제공하는 핵심정보문건(Key Information Document, KID)의 일부로 금융상품의 신용·시장리스크를 등급화 하여 소비자에게 제공하게 하고 있으며 (Elsen (2021)), 미국의 소비자보호청(Consumer Financial Protection Bureau, CFPB)에서는 소비자에게 제공하는 요약정보를 “적시에 이해가능한 (timely and understandable)” 형태로 고안하기 위하여 다양한 행태과학(behavioral science) 기반의 실증적 연구를 학계와 협력하여 수행하고 있다 (Knoll (2021)). 마지막으로, 플랫폼을 기반으로 하는 빅테크 금융서비스 제공자의 독과점 방지 등에 대한 보다 의미있는 규제를 위해서도 (동시에 이들에 의한 금융혁신을 촉진하기 위해서) 데이터·증거 기반의 ‘스마트규제’가 향후 우리나라의 금융부문에서 추진해야 할 과제인 것으로 판단된다.

2. 의료부문 데이터경제 평가

우리나라 보건·의료부문에서의 데이터경제 활성화는 3개 유형의 데이터를 효율적으로

로 구축·공유하여 이의 활용도를 높이고 - (1) 개인 진료기록 데이터, (2) 유전체 데이터, (3) 생활·환경 데이터, 동시에 AI 기반의 분석기법을 활용한 신약개발 및 새로운 치료법 개발의 촉진이 핵심사항인 것으로 보인다. 산업생산성 측면에서 의료부문은 금융·부동산 부문과 달리 DPA 기반의 가치창출을 통하여 이에 직접적으로 기여할 수 있는 잠재력이 매우 큰 분야라고 하겠다. 이를 지역·국가 차원의 사회적 효과로 연결하기 위해서는 상기 데이터를 효율적이고 안전하게 구축·공유하고, 이를 통한 R&D활동을 촉진하는 것이 의료부문 데이터경제 생태계 조성을 중요한 과제인 보인다. 특별히 우리나라에는 양질의 의료데이터가 대량으로 존재하고 있고, 이를 DPA 기반으로 구축하여 효율적으로 활용할 경우 (연구용·상업용 용도로) 이의 부가가치가 매우 높을 것으로 예상하고 있다. (오상우 (2020))

본 부문에서는 현재 의료 빅데이터 사업 (“국가 바이오 빅데이터 구축 사업”), 의료 마이데이터 사업 (“MyHealthway”)에 대한 시범사업이 진행되고 있고, 향후 이들 사업의 진행에 있어서 민간부문의 참여를 보다 적극적으로 유도할 필요가 있다 (민간 스타트업에 대한 투자 유치 강화 등을 통한). 또한 마이데이터 사업의 추진에 있어서 금융·부동산 등 타 부문과의 연계를 통하여 다부문·다학제적 연구 및 협력을 강화할 필요가 있다. 그러나 의료부문에서의 데이터경제 플랫폼 효과는 (즉, 인터넷·모바일 기반의 플랫폼을 통한 거래비용 감소, 그리고 이로 인한 의료소비자의 효용 증대는) 현재 미미한 것으로 판단되고, 이의 주된 이유는 이와 같은 혁신적 의료서비스에 대한 기존 업계의 반발과 함께 개인 의료정보의 오남용에 대한 과도한 우려인 것으로 논의되고 있다.

유전체 데이터의 구축을 통한 의료부문의 R&D 제고를 위해서 현재 K-DNA 사업이 사업이 진행되고 있고, 이는 2018년 이후 미국에서 실행 중인 “All of US” 사업을 (100만 명의 유전체 정보를 디지털 데이터로 구축하여 연구용으로 활용하고, 참여자에게 자료의 접근성 부여) 벤치마킹 하여 설계되었다. 미국 사업의 경우 사업기간인 2018~27년 기간에 대한 중·장기 계획을 시기별로 구체적인 목표를 설정하여 진행하고 있고 ([그림 23] 참조), 우리나라에서도 이와 같이 가시적인 목표를 정하여 성과위주(result-oriented)로 사업을 진행할 필요가 있는 것으로 보인다 (이는 의료부문 뿐만이 아니라 타 부문의 데이터경제 관련 사업의 추진에도 적용할 필요가 있음).

[그림 23] 미국 “All of US” 사업의 시기별 성과에 대한 목표

Table 2. Scientific Goals of the All of Us Program and Expected Timelines.*

Goal	Years				
	End of 2018 (N=94,000)	End of 2019 (N=>200,000)	2020–2022 (N=<650,000)	2023–2027 (N=>1 million)	After 2027 (N=>1 million)
Return data to participants	+	+	+++	+++	++++
Establish demonstration projects†		+	+++	+	+
Discover genetic and environmental correlates with disease			++	+++	++++
Improve predictions of therapeutic safety and efficacy			++	+++	+++
Discover disease biomarkers			++	+++	+++
Connect mobile health, digital health, and sensor data with clinical outcomes			++	+++	+++
Develop new disease classifications			+	+++	++++
Support clinical trials			+	+++	+++
Enable machine-learning applications			++	+++	++++
Improve understanding of health disparities			++	+++	+++
Develop and test new therapeutic agents					++

* The expected number of participants in the cohort is shown for each time period. The number of plus signs in each cell indicates the anticipated relative degree to which each goal may be accomplished during the estimated timeline for focused research.
 † Demonstration projects are scientific studies implemented by the All of Us program to show the quality, usefulness, validity, and diversity of the All of Us research data set and platform. In these projects, the population and data are further characterized, and the data are evaluated with a view to determining whether known associations can be replicated.

N Engl J Med 2019; 381:668-676

3. 부동산부문 데이터경제 평가

우리나라의 부동산부문은 금융시장 및 실물시장 모두에서 데이터경제 생태계의 조성이 상대적으로 더딘 것으로 평가되고 있다. 이의 주된 이유는 주거용·비주거용 부동산 거래데이터에 대한 구축 및 공유가 아직 활성화 되지 않았고, DPA 기반의 혁신에 대한 기존 라이선스 업계의 반발 및 정부 차원의 촉진기제 형성이 더딘 상황이며, 투기과열지구 지정 및 LTV·DTI 규제로 인하여 부동산금융 부문에서 효율적인 대출리스크의 측정 및 관리가 어려운 환경이라는 점 등이 포함된다. 본 부문에서의 데이터경제 생태계 조성 과 관련하여, 지난 10년간 프롭테크 관련 언론기사에 등장한 5개 최빈명사(사람이름) 중 1위가 프롭테크 업체인 직방의 사장이고 대부분이 민간부문 인사인 반면, ‘데이터경제,’ ‘핀테크,’ ‘디지털헬스’의 경우 대부분이 정부 관련 인사인 것으로 나타났다. ([부록 3] 참조) 또한 핀테크 부문의 경우 ‘핀테크 지원센터’와 같은 정부주도의 촉진기제가 현재 작동하

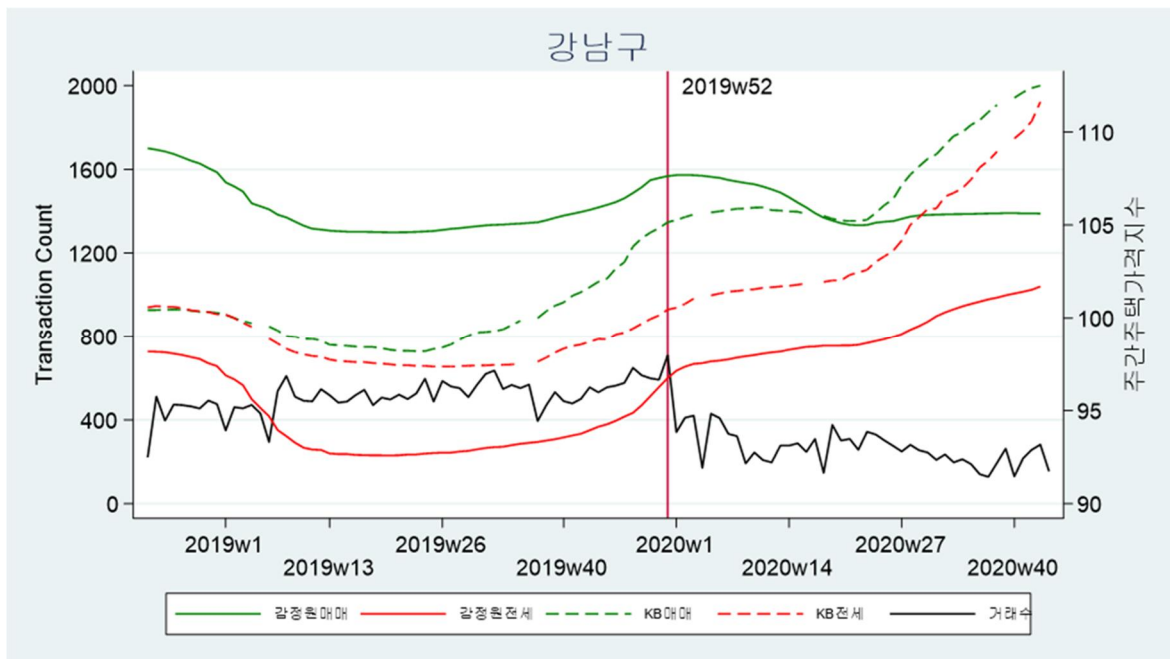
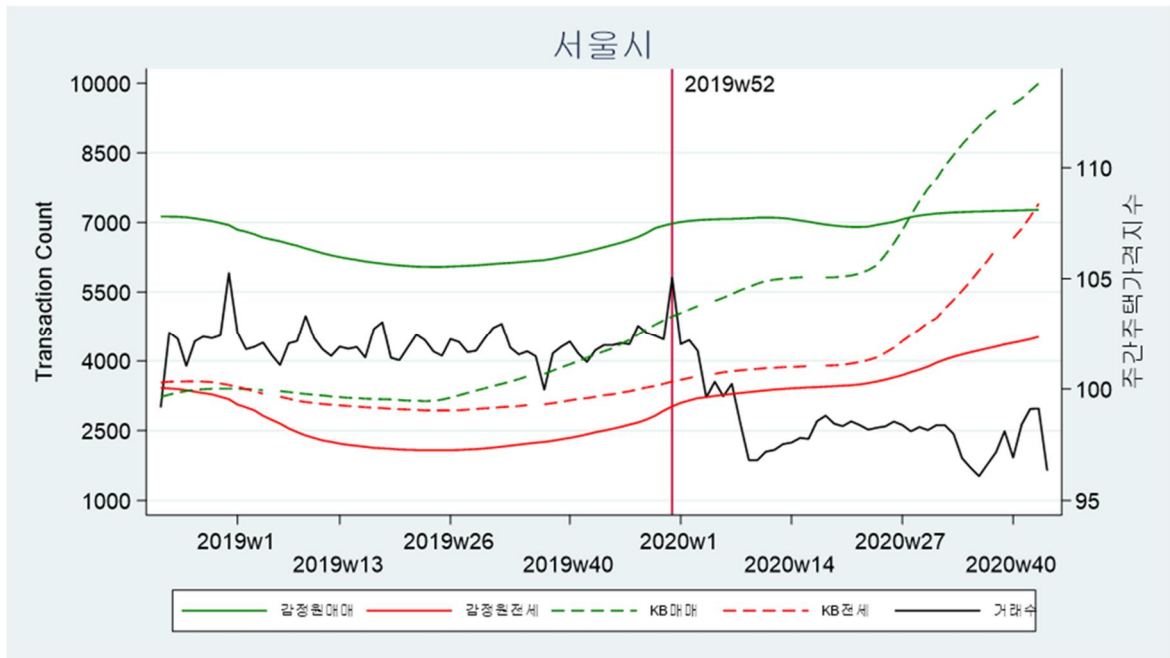
고 있는 반면, 프롭테크 부문은 이와 같은 기제가 현재 없는 것으로 파악되었다.

부동산부문의 데이터 활용과 관련하여 미국에서는 1990년대 AUS(Automated Underwriting Service)의 일환으로 AVM(Automated Valuation Model)을 통한 주택에 대한 자동화가치평가서비스가 확산되었고, 이의 결과로 Zillow.com과 같은 플랫폼 기반의 성공적인 프롭테크 업체들이 등장하였다. 우리나라의 경우에도 최근 주택에 대한 데이터·모형 기반의 가치평가를 수행하는 프롭테크 업체들이 등장하였고, 이들 중에는 규제 샌드박스를 통과하여 혁신금융서비스 업체로 지정되기도 하였다. 그러나 현재 이들 국내 AVM업체들은 사업영역이 비아파트로 한정되어 있고 (연립주택, 다세대주택, 오피스텔, 단독주택 등),¹⁵ 향후 AVM서비스를 모든 유형의 주택으로 확대하여 이들에 의한 효율적이고 혁신적인 서비스를 촉진할 필요가 있다. 또한 다양한 데이터와 모형을 사용하여 주택 등 부동산물건에 대한 적절한 가치평가가와 관련된 R&D 활동을 PPAP 체계를 활용하여 촉진할 필요가 있는 것으로 보인다.

부동산시장의 추세를 투명하고 정확하게 측정하여 모니터링 하는 것은 관련 경제주체 (소비자·정부·기업) 모두에게 매우 중요한 공공재의 성격을 가진다. 이와 관련하여 우리나라의 주택시장 추이는 현재 다양한 가격지수를 통하여 파악되고 있다 (부동산원 아파트 실거래가격지수, REB 주택가격동향조사, KB국민은행 주택가격동향, 부동산114 주택가격지수 등), 최근 언론에 보도된 바와 같이 2020년 이후 이들 지수들 간에 상당한 수준의 괴리가 관찰되고 있다 ([그림 24] 참조). 부동산가격지수는 사용된 데이터(샘플), 지수산정 방식, 발표주기, 지역단위에 따라서 큰 차이가 발생할 수 있는 실증연구 분야이고, 향후 우리나라의 주거용·비주거용 부동산시장의 적정 가격지수 산정방식과 심도있는 연구를 학계와 유관기관들이 협력하여 진행될 필요가 있는 것으로 보인다. 이는 우리나라 부동산시장에서의 투명성 제고를 위한 매우 중요한 과제일 뿐만 아니라, 데이터경제의 예측력 증대 효과 측면에서도 중요한 과제인 것으로 판단된다. 또한 이와 관련된 부동산 부문 데이터의 (주택 매매 및 전월세 거래, 건물 및 공간 정보, 부동산금융 – 대출 및 지분투자 – 데이터) 디지털화 및 효율적인 공유가 본 부문에서의 데이터경제 생태계 조성을 위한 중요한 과제인 것으로 보인다.

¹⁵ 이는 아파트의 경우에는 감정평가 결과, 부동산원 시세, KB국민은행 시세만을 사용하도록 되어 있는 규제가 원인임.

[그림 24] 우리나라 주택 매매·전세가격지수의 괴리 현상



VI. 정책 과제

정책과제 (1): 데이터경제 활성화를 위한 국가 차원의 제도적 기반 마련

데이터경제 활성화를 위한 국가 차원의 법적·제도적 기반을 구축하기 위하여 중앙정부는 다음 사항을 지속적·효율적으로 추진할 필요가 있다. 첫째, 개인정보의 적절한 보호와 함께 이의 연구용·상업용 활용을 위한 범정부 차원의 기준 및 지침을 마련할 필요가 있다 (개인정보 비식별화, 데이터결합, 데이터 클리닝, 데이터 분석기법 공유 등). 또한 부문별·부처별로 공유 대상 데이터의 식별 및 디지털화 추진, 기존 기득권 산업과의 갈등조정 방안 도출, 그리고 플랫폼을 통한 온라인 거래의 안전성 확보 등이 데이터경제 활성화와 관련된 국가 차원의 정책과제라고 하겠다.

둘째, 공공·민간의 공유 대상 데이터를 다음 3가지 유형별로 분류하여 각각에 대한 차별화된 데이터 구축·공유 전략을 수립할 필요가 있다: (1) 공공재 (비식별화를 통하여 일반에 무료로 제공, 예를 들면 R&D 용도의 대출 데이터 샘플, 유전체 데이터 등), (2) 공유재¹⁶ (비식별화 이외에 추가적인 보호장치를 통하여 사용자 네트워크 내에서 공유, 예를 들면 P2P대출 플랫폼에서 블록체인을 통하여 개인정보 공유); (3) 사유재 (‘데이터거래소’를 통하여 유료로 거래하고, 사적 혜택이 발생하는 데이터).

셋째, 국가 차원의 데이터경제 혁신생태계 조성을 지속적·효과적으로 추진할 수 있는 중앙정부 내의 컨트롤타워 구축이 필요할 것으로 생각된다 (예: 디지털혁신부). 현재 우리나라는 ‘4차산업혁명위원회’와 산하의 특위/소위 중심으로 데이터경제·스마트X 관련 정책에 대한 논의 및 전략을 수립하고 있으나, 향후 이를 보다 지속적·효과적으로 추진하기 위한 정부내 상설조직이 필요해 보이고, 이를 중심으로 상기 논의한 국가 차원 정책

¹⁶ 데이터의 일반적 속성으로 (1) 비경합성 (non-rivalrous, 특정 경제주체의 소비가 다른 주체의 소비에 영향을 주지 않음), (2) 비배타성 (non-excludable, 상품·서비스의 소비에 있어서 특정 주체를 제외하는 것이 가능하지 않거나 매우 높은 비용을 촉발하게 됨), (3) 한계비용 제로 (zero marginal cost, 추가적 경제주체에 대한 서비스 제공의 비용이 발생하지 않음)가 논의됨. (Floridi (2010), Sonderegger (2016)) 최근에는 ICT를 활용하여 데이터를 비경합적이지만 배타적인 (non-rivalrous but excludable) 재화로 변환하여 활용하는 추세이고 (예: 블록체인을 활용하여 P2P 네트워크 간의 개인정보 공유), 이 경우에는 데이터를 공유재(common good)로 인식할 수 있다.

과제들의 추진과 함께, 관련 부처들간의 정책공조, 중장기 계획의 수립·집행·모니터링, 그리고 데이터경제 관련 공공부문의 주요 사업들에 대한 타당성 평가 및 사회적 비용·편익 분석을 정기적으로 실행하는 과제도 본 기구의 업무영역으로 볼 수 있다.

정책과제 (2): 지역별 맞춤형 데이터경제 생태계 조성을 위한 전략 수립

데이터경제 생태계는 기본적으로 지역별 특성을 고려하여 (산업구조 등에 대한) 차별화된 방식으로 진행되어야 하고, 이를 촉진하기 위한 지역단위의 데이터경제 활성화와 관련된 다음의 정책과제를 추진할 필요가 있는 것으로 판단된다. 첫째, 지역 데이터경제 생태계 조성을 스마트X 사업과 연계하여 지역 일자리 창출에 초점을 맞추어 추진할 필요가 있다 (광역 및 기초 자치단체별에 대한 차별화된 전략 필요). 또한 이를 주도할 청년층 중심의 전문인력을 지역에 투입하여 지역경제 디지털전환을 촉진하는 방안을 고려해 볼 수 있다. 예를 들면, 읍·면·동 단위의 ‘데이터경제 청년지도자’를 육성하고, 이들이 주민 삶의 질 향상에 직접적인 영향을 주는 스마트시티·스마트홈·스마트금융·스마트의료·스마트팜 등의 스마트X 사업을 종합적으로 선도하고, 또한 이들의 정주여건을 지원하기 위한 귀농·귀촌 임대주택 리츠 등의 사업을 이에 연계할 수 있다. 이들의 선발을 위한 국가 자격시험 및 연수과정의 시행 또한 관련 정책사항으로 추진할 수 있다.

둘째, 지역별 데이터경제 생태계 조성을 위하여 공공·민간·학계의 구체적이고 효과적인 데이터경제 사업을 지역별로 발굴하여 추진할 필요가 있다. 이와 관련하여 현재 유럽의 스마트시티 사업의 일환으로 추진 중인 PPAP (Public·Private·Academic·Partnership) 제도를 도입하고, 이의 성공사례를 확산하는 방식으로 지역 데이터경제 활성화를 촉진할 필요가 있다. 예를 들면, 유럽연합의 스마트시티 사업인 “EU Replicate,” 영국의 “Bristol is Open” 등이 지역별 PPAP 사업의 벤치마킹 사례가 될 수 있고, 이와 같은 추진체계를 통하여 학계와 관련 산업의 참여를 활성화 시키고, 이를 통하여 혁신적 아이디어가 R&D를 통하여 상용화·기업화되는 지역 맞춤형 혁신생태계를 촉진할 필요가 있다. 이를 위해서 대기업(빅텍 포함)·중소기업·스타트업 간의 협력체계를 구축하고, 이를 통하여 지역 일자리 창출 및 지역경제 활성화에도 기여하는 방향으로 데이터경제 촉진 사업을 기획·추진할 필요가 있다.

셋째, 지역 데이터경제 사업의 추진에 있어서 이를 성과위주(result-oriented)로 진행할 필요가 있고, 이를 위해서는 구체적이고 실질적인 평가지표(KPI)를 구축하여 사업을

평가하고 모니터링할 필요가 있다. 또한 KPI의 구성은 주민 삶의 질 향상, 역내 일자리 창출, 산업생산성 제고 등과 같은 지역경제 활성화와 직접적인 상관관계가 있는 지표를 포함해야 하고, 사업의 사전 타당성 평가와 사후 정기적인 모니터링을 수행하여 공공부문 사업의 성과를 판단할 필요가 있다.

정책과제 (3): 데이터경제 전문인력 양성 및 다부문·다학제 간의 학술연구 증진

데이터경제 활성화의 기대효과를 (III장에서 논의한 플랫폼 효과, 예측력 증대 효과, 분석기법 효과) 극대화 하기 위해서는 다부문·다학제 간의 기초연구 활동을 증진하고, 또한 데이터경제 관련 전문인력을 지속적으로 양성할 필요가 있다. 이와 관련된 추진과제는 아래와 같다.

첫째, 현재 전공별·부문별로 파편화 되어 있는 대학·대학원 교육을 디지털전환 시대에 부합하는 수요자 중심의 체계로 변환시킬 필요가 있다. 이를 위하여 데이터과학 (Data Science) 전공의 다학제적인 단과대학을 지역별로 설립하고 (수도권 대학의 참여 유도), 이를 중심으로 지역 맞춤형 데이터경제 혁신생태계의 조성 및 활성화에 기여하도록 하는 방안이 필요한 것으로 보인다. 현재 데이터경제는 컴퓨터공학 등 이공계 중심으로 편성되고 있으나, 향후 이의 커리큘럼을 보다 다학제적 편성하여 (경제학·정치학·사회학 등의 사회과학 및 인문학을 포함한) 데이터경제 관련 산업의 수요에 부합하는 새로운 분석 방식 및 이의 사회과학·자연과학에서의 적용사례를 교육할 필요가 있다.

둘째, 데이터경제 생태계의 조성 및 관련된 다학제적인 기초연구를 이공계·인문사회계 연구자들이 협력하여 증진할 필요가 있고, 특별히 양적인 기준에 (논문 편 수에) 의한 연구성과의 평가방식을 질적인 측면을 (논문의 인용 등) 보다 적절하게 반영하는 방식으로 변화시킬 필요가 있다. 이를 통하여 우리나라의 데이터경제 관련 연구가 국제 학술연구에 보다 보다 선도적으로 기여할 수 있는 환경을 조성할 필요가 있고, 이를 위하여 학술연구의 지원 및 평가체계의 재정비, 그리고 대학·연구기관(공공 및 민간)·산업 간의 협업체계를 강화할 필요가 있다.

셋째, 현재 진행 중인 데이터경제 관련 공공부문 사업들을 다부문·다부문간의 협력을 통하여 시너지 효과를 높일 필요가 있고, 여기에는 오픈데이터, 공공빅데이터 플랫폼, 마이데이터, 규제 샌드박스 등이 포함된다. 이들 사업들은 현재 대체로 부처별·부문별로 구

분되어 (“silo”화 되어) 진행되고 있다.¹⁷ 향후 다부문 간의 융복합을 증진시키는 차원에서 이들 사업을 부처간 공조를 강화하여 추진할 필요가 있고, 여기에는 성공사례의 공유·확산, 표준화 및 적정 규제 방안의 마련, 그리고 빅데이터·블록체인·인공지능 등 데이터 활용과 관련된 신기술에 대한 공동 실험 등이 포함된다.

¹⁷ 이와 관련하여 최영락(2018)은 우리나라 과학기술정책의 접근방식이 여전히 후진성을 보이고 있고, 그 이유로, 첫째, 기존 산업화 과정의 성공 패턴인 선진국 추격·모방에 안주하고 있고, 둘째 집단이익으로 왜곡된 연구비 배분, 권위주의 등의 후진적 연구 문화가 지배하고 있으며, 셋째, 대학, 정부출연연구기관, 기업 등 과학기술 주체들이 파편화된 채 각자의 이익에만 집착하고 있고, 넷째, 연구비 연구 성과 등에서 이룩한 작은 성취에 만족한 채 치열한 글로벌 경쟁의 거친 파도를 헤쳐 나가려는 의지가 상실되었으며, 다섯째, 국가연구개발사업의 관료화 및 선진국 유행에 따라 자주 변하는 연구정책, 그리고 연구사업의 미시적 관리와 부처 간 칸막이 등을 들고 있다.

VII. 결론

본 연구는 우리나라의 3개 서비스부문(금융·부동산·의료)을 대상으로 데이터경제의 개념화 및 활성화 방안을 다학제·다부문간의 연구를 통하여 논의하였다. 이를 위하여 국내외 데이터경제 관련 문헌의 주요 논거를 정리하였고, 2개의 국제학술논문 데이터베이스를 이용하여 상기 3개 부문의 데이터경제 관련 논문을 추출하여 메타분석을 수행하였으며,¹⁸ 상기 분석의 결과를 바탕으로 3개 서비스 부문에서의 데이터경제 활성화를 위한 제반 정책과제를 아래와 같이 제시하였다.

• 데이터경제 활성화를 위한 국가 차원의 제도적 기반 마련

- 개인정보의 보호와 활용을 위한 범정부 차원의 기준 및 지침의 마련, 또한 부문별·부처별로 공유 대상의 데이터 식별, 갈등조정 및 적정 규제 방안 마련 등
- 공공·민간 부문의 공유 대상 데이터를 공공재·공유재·사유재로 구분하여 차별화된 데이터 구축 및 공유 전략 수립
- 국가 차원의 데이터경제 혁신생태계 조성을 지속적·효과적으로 추진할 수 있는 중앙정부 내의 컨트롤타워 구축 (예: 디지털혁신부)

• 지역별 맞춤형 데이터경제 생태계 조성을 위한 전략 수립

- 지역 데이터경제 생태계를 지역 일자리 창출과 명시적으로 연계하고, 이의 일환으로 지역 스마트X 사업과 연계하여 데이터경제 관련 사업을 추진
- 지역별 데이터경제 생태계 조성을 위하여 공공·민간·학계의 구체적이고 효과적인 협력 사업 추진 (유럽의 스마트시티 관련 PPAP 사례 벤치마킹)
- 지역 데이터경제 사업의 추진에 있어서 이를 성과위주(result-oriented)의 평가방식 도입, 그리고 이를 위한 구체적이고 실질적인 평가지표(KPI) 구축

¹⁸ 본 연구를 위하여 2020년 12월 KDI국제정책대학원이 주최로 관련 세미나를 개최하였고, “데이터경제 활성화를 위한 연구 및 정책 과제: 금융·의료·부동산 부문을 중심으로,” 세미나에서 발표된 내용이 본 논문에 다수 활용되었음.

- **데이터경제 전문인력 양성 및 다부문·다학제 간의 학술연구 증진**

- 현재 전공별·부문별로 파편화 되어 있는 대학·대학원 교육을 수요자 중심으로 전환하고, 지역별로 다학제·다부문의 데이터과학 단과대학 설립
- 선도적 학술연구 증진을 위한 연구성과의 평가방식 개선 (질적인 측면 반영) 및 학술연구의 지원 및 평가 체계 재정비
- 현재 진행 중인 데이터경제 관련 공공부문 사업들을 다부문·다부문간의 협력을 통하여 시너지 효과 증진 (오픈데이터, 공공빅데이터 플랫폼, 마이데이터, 규제 샌드박스 등이 포함)

서론에서 언급한 바와 같이, 현재 우리나라는 과거의 모방·추격형 경제시스템을 혁신을 창출·선도할 수 있는 체계로 대전환 하는 것이 필요한 상황이다. 학술연구의 관점에서 이를 실현하기 위해서는 글로벌 톱 수준의 연구자들이 글로벌 톱 수준의 연구주제를 글로벌 톱 수준의 연구지원시스템 하에서 추진할 때 가능해질 것으로 생각된다. 또한 데이터경제 활성화와 같은 혁신생태계의 구축에 있어서 이를 통하여 도달하고자 하는 궁극적인 목표가 어디에 있는지를 분명히 하고, 이를 기반으로 관련사업들을 기획·추진할 필요가 있다.¹⁹ 즉, 향후 3개 부문의 데이터경제 혁신생태계 구성에 있어서 이를 통하여 궁극적으로 도달하고자 하는 목표를 시기별·부문별로 구체화 하고, 이에 부합하는 제도적 기반을 단계적으로 구축하여야 할 것으로 생각된다. 이에 포함되어야 할 필수적인 사항으로, 보다 종합적인 다학제·다부문간의 연구의 증대, 글로벌 톱 수준 연구자들의 장기적 육성 및 이들을 통한 글로벌 톱 수준의 연구결과를 도출, 그리고 이를 기반으로 하는 산업부문의 혁신 및 정책사례 증대를 들 수 있다. 이를 위해서는 현재 단기적 성과 (논문 편 수) 위주로 진행되는 학술연구의 성과평가 체계를 보다 장기적으로 - 그리고 연구자 중심으로 - 개선하고, 대학 및 공공·민간연구소들이 보다 유기적으로 협력하여 장기적인 연구결과를 촉진할 수 있는 제도적 기반을 마련하며, 또한 이를 위한 국제협력을 강화할 수 있는 실질적인 방안을 마련하여야 할 것으로 사료된다.

¹⁹ 이와 관련하여 최영탁(2018)은 우리나라 과학기술 부문 대전환의 (등정) 목표가 뒷동산이나, 북한산이나, 백두산이나, 에베레스트이냐에 따라 정책프레임 및 제도적 기반 마련을 위한 전략도 달라져야 할 것이고, 현재 우리나라의 관련 연구자들이 백두산 또는 에베레스트를 목표로 삼아야 함에도 불구하고 뒷동산 내지 북한산에 만족하고 있는 상황을 개탄하고 있음.

References

<국내>

강건욱, 2020, “의료 부문의 혁신 및 기업화,” 2020.12월 KDI국제정책대학원 ‘데이터경제 세미나’에서 발표.

박소정, 2020, “인슈어테크와 데이터 경제,” 2020.12월 KDI국제정책대학원 ‘데이터경제 세미나’에서 발표.

오상우, 2020, “디지털 의료와 데이터 경제,” 2020.12월 KDI국제정책대학원 ‘데이터경제 세미나’에서 발표.

안상훈, 「혁신과 경제성장: 생산성 및 기업동학을 중심으로」 과학기술정책 제1권 제1호, 2018, pp. 35~61.

이상영, 2020, “프롭테크와 데이터경제,” 2020.12월 KDI국제정책대학원 ‘데이터경제 세미나’에서 발표.

이태준·조만, 2020, 「국내 스마트시티 사업에 대한 평가 및 정책제언」, KDI GKED 연구보고서, 2020.2.

정성훈·김민호, 2020, 「공장의 스마트화와 사업체 성과」, KDI연구보고서, 2020.5.

조만·이태준, 2019, 「우리나라의 스마트시티 정책: 데이터경제 활성화 관점에서의 평가 및 제언」, 연구논문, 2019.12.

최영락, 「한국의 과학기술정책 회고와 전망」 과학기술정책 제1권 제1호, 2018, pp. 7~33.

황종성, 「스마트시티 발전동향과 쟁점을 통해 본 국가전략 연구과제」 한국통신학회지, 제34권 제8호, 2017, pp. 14-18.

<해외>

Bartel, Ann, Casey Ichniowski, and Kathryn Shaw, “How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills,” *The Quarterly Journal of Economics*, 122(4), 2007, pp.1721~58.

Begenau, Juliane, Maryam Farboodi, and Laura Veldkamp, “Big Data in Finance and the Growth of Large Firms,” Working Paper 24550, National Bureau of Economic Research April 2018.

Berg, T., V. Burg, A. Gombović, and M. Puri, 2018, “On the Rise of FinTechs – Credit Scoring using Digital Footprints,” *Presented at the 2018 CCAF Conference*.

Bernstein, Peter, 1998, “Against the Gods: The Remarkable Story of Risk,” John Wiley and Sons.

Cambridge Center for Alternative Finance (CCAF), 2019, “RegTech in Financial Services: Technology Solutions for Compliance and Reporting.”

Cho, Man, 2020, “FinTech Megatrends: An Assessment of Their Industrial and Welfare Implications,” KDIS Working Paper, October 2020.

Cho, Man, 2007, “180 Years’ Evolution of the US Mortgage Banking system: Lessons for Emerging Mortgage Markets, 2007, *International Real Estate Review*, Vol. 10(1), pp. 171-212.

Evans, P.C., Gawer, A., 2016. *The Rise of the Platform Enterprise: A Global Survey*. Center for Global Enterprise, New York.

Fajgelbaum, Pablo D., Edouard Schaal, and Mathieu Taschereau-Dumouchel, “Uncertainty Traps,” *The Quarterly Journal of Economics*, 2017, 132 (4), 1641–1692.

Farboodi, Maryam, and Laura Veldkamp, 2021, “A GROWTH MODEL OF THE DATA ECONOMY,” NBER Working Paper 28427.

Frost, J., L. Gambacorta, Y. Huang, H.S. Shin, and P. Zbinden, 2019, “BigTech and the changing structure of financial intermediation,” *BIS Working Papers No 779*.

Fuster, A., M. Plosser, P. Schnabl, and J. Vickery, 2018, “The Role of Technology in Mortgage Lending,” NBER Working Paper 24500.

Gambacorta, Leonardo, Yiping Huang, Zhenhua Li, Han Qiu, and Shu Chen, 2020, “Data vs collateral,” *BIS Working Papers No 881*.

Gans, J. S., & Stern, S. (2003). The product market and the market for "ideas": commercialization strategies for technology entrepreneurs. *Research Policy*, 32(2), 333-350.

Huawei and Navigant, “UK smart cities index 2017. Assessment of strategy and execution for the UK's leading smart cities, Navigant Consulting, Inc, 2017.

International Monetary Fund (IMF), 2016, “Robots, Growth, and Inequality,” *Finance and Development*, September 2016.

Ismagilova E, Hughes L, Dwivedi YK et al, “Smart cities: Advances in research—An information, systems perspective, *International Journal of Information Management*, 47, 2019, p88-100.

Jagtiani, J., and C. Lemieux, 2019, “The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform,” *Philadelphia Fed Working Paper*.

Kagermann, Henning, Wolfgang Wahlster, and Johannes Helbig, 2013, “Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Final Report of the Industrie 4.0 Working Group,” *Forschungsunions and acatech*, 2013.

Kim, H.S., 2019, “5G and Its Applications,” *Korea Telecom, presentation material*.

Lin, M., N. Prabhala, and S. Viswanathan, 2013, “Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending,” *Management*

Science 59(1), 17-35.

McKinsey Global Institute, 『Smart Cities: Digital Solutions for A More Livable Future,』 2018.

kim, J.K., 2020, “Korean Economic Development: Keys to Unlocking “Shared Growth,” KDIS Lecture Material.

Mullainathan, Sendhil, and Jann Spiess, 2017, “Machine Learning: An Applied Econometric Approach,” *Journal of Economic Perspectives* , Vol. 31(2), pp. 87–106.

Ordonez, Guillermo, “The Asymmetric Effects of Financial Frictions,” *Journal of Political Economy*, 2013, 121 (5), 84.

Park, S., Choi, G. J., & Ko, H. (2020). Information technology–based tracing strategy in response to COVID-19 in South Korea—privacy controversies. *JAMA*, View Online. <https://jamanetwork.com/journals/jama/fullarticle/2765252>.

Philippon, T., 2016, “The FinTech Opportunity,” NBER Working Paper Series.

Philippon, T., 2015, Has the us finance industry become less efficient? on the theory and measurement of financial intermediation. *The American Economic Review* 105(4), 1408–38.

Puri, M., J. Rocholl, and S. Steffen, 2017, “ What do a million observations have to say about loan defaults? Opening the black box of relationships, *Journal of Financial Intermediation* 31, 1-15.

Sonn, J.W., M. Kang, and Y. Choi, 2020, “Smart city technologies for pandemic control without lockdown,” *International Journal of Urban Sciences*.

Teece, D. J. (2018). Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world. *Research Policy*, 47(8), 1367-1387.

The Economist, 2020.3.2, “Special Report: The Data Economy.”

The Government of the Republic of Korea. (2020). Flattening the curve on COVID-19 (15 April ed). Seoul: The Government of the Republic of Korea.

Wan, Wayne, and Thies Lindenthal, 2021, “Towards Accountability in Machine Learning Applications: A system testing approach,” mimeo.

[부록 1] 스마트팩토리, 스마트금융, 스마트시티: 개요

1. 제조업: 스마트팩토리 (SmartFactory)

최근 선진국을 중심으로 제조업의 기존 생산방식이 가지는 한계를 극복하고, 이를 통하여 새로운 성장동력을 얻고자 하는 제조업 혁신의 개념으로 스마트공장(smart factory) 또는 공장의 스마트화(smartzation)에 대한 관심이 고조되고 있다. 스마트공장 개념의 발상지는 독일이고, 이는 자국의 민간 기업들이 최근의 급격한 기술 진보와 소비자 선호의 다양화, 그리고 인구고령화 및 근로자의 숙련도 감소 등 복합적인 사회·경제적 변화에 적응하기 위한 산업 전반에서의 패러다임 전환(paradigm shift)이 필요함을 인식하여 추진하기 시작한 정책과제라고 하겠다. (정성훈·김민호 (2020)) 구체적으로, 2011년 하노버 산업박람회(Hannover Messe)에서 ‘Industrie 4.0’이라는 산업정책을 제안하였고, 본 계획은 이후 산업협회와 정부, 그리고 학계의 공동협력을 거쳐 수정·보완되었으며, 마침내 독일의 공식적인 국가전략으로까지 확산된다.

‘Industrie 4.0’ 정책의 핵심에는 바로 스마트공장이 있고, 이는 모든 공정과 생산품, 그리고 생산요소들이 디지털(정보)화되고 서로 네트워킹하는 공장이라는 뜻이다 (“digitization and networking of all processes, products and resources”). 즉, 사물인터넷(IoT), 센서(sensor), RFID(Radio-Frequency ID 또는 전자태그) 등의 새로운 기술들을 활용하여 제품의 설계와 개발, 생산 등에서 발생하는 정보, 생산된 제품 자체에 대한 정보, 공구, 부품, 원부자재 등 투입요소에 대한 정보 모두를 디지털화(digitization) 하여 표준화하고, 그 정보를 경제주체들 간에 서로 주고받을 수 있도록 연결(networking)하여 통합된 관리를 수행할 뿐만 아니라, 머신러닝(machine learning)과 같은 분석기법을 통하여 축적된 자료를 처리·분석함으로써 생산성 향상, 비용 절감, 새로운 비즈니스 모델의 개발 등을 가능하게 하자는 개념이다.²⁰

스마트공장의 효과에 대하여 Bartel et al.(2007)은 공장의 스마트화 수준이 높아지면 공장 전체의 생산성이나 이윤을 높이기 전에 특정 단계에서부터 먼저 효율성이 개선될 수 있음을 논의하고 있다. 예를 들면, 생산 단계에서 시스템을 통합하고 공정을 최적화하면 제품 생산에 소요되는 단위 시간을 줄일 수 있고, 각 공정에서 발생한 데이터를 분석하

²⁰ Industrie 4.0은 스마트공장 외에도 스마트 물류(smart logistics), 스마트 그리드(smart grid), 스마트 서비스(smart service) 등을 포함하고 있으나, 이들은 제조를 중심으로 확장된 개념이라고 할 수 있다. (Kagermann et al., 2013)

면서 공정의 어느 부분에서 특히 불량품이 많이 생기지는지를 파악하여 그 부분을 개선할 수 있다. 정성훈·김민호(2020)도 실증분석을 통하여 우리나라에서의 공장스마트화가 생산품의 1일 생산량, 리드타임, 제품불량률에서는 개선효과를 보였으나, 공장의 전반적인 생산성 증가에는 그 영향이 미미한 것으로 보고하고 있다. 스마트공장과 관련하여 주목할 사항은 향후 5G기술의 보급이 확대됨에 따라서 데이터의 축적 및 공유에 있어서 초고속화, 초연결성, 초저지연성이 가속화될 전망이고, 이는 스마트공장을 포함한 데이터 기반의 생산성 향상에 큰 역할을 할 것으로 예상된다. (Kim (2019))

2. 스마트금융 (핀테크)

핀텍은 금융서비스 부문에서 데이터와 플랫폼을 활용한 혁신적인 금융중개서비스 방식이라고 정의할 수 있고, 여기에는 가상화폐 (블록체인 기반의), 크라우드펀딩 (Crowdfunding, Market Place Lending 등의 다양한 P2P대출 포함), 모바일 지급결제, 로보-어드바이저(Robo-advisor), 레그텍(RegTech) 등이 포함된다. 지난 수년간 미국·유럽을 중심으로 핀테크의 사회적 효과에 대한 다양한 연구결과가 발표되었고²¹, 이를 요약하면 핀텍산업은 데이터-플랫폼 기반의 금융중개서비스 모델을 제시함으로써 기존의 지점망 중심의 금융중개서비스에 비하여 효율성 및 금융소비자의 복지를 획기적으로 높이고 있다는 것이다.

이는 구체적으로 금융중개의 초과수익(Excess Yield, EY) 개념을 통하여 이를 설명할 수 있다: 즉, 식 (1)에서 보는 바와 같이 금융중개서비스에서의 초과수익(EY)은 통상 대출금리에서, r^l , 무위험이자율과, r^f , 리스크스프레드(또는 가산금리)를, RP, 뺀 나머지 부분으로 정의한다. 따라서 EY는 금융중개서비스의 제공으로부터 발생하는 이윤 및 제반 비용으로 해석할 수 있다.

$$(1) \quad EY = r^l - r^f - RP$$

문헌에서 보고하는 핀텍의 사회적 효과는 기존의 지점망 중심의 금융중개 모델에 비하여 거래비용·시간·편의성 측면에서의 효율성을 획기적으로 높인다는 것이다. (Philippon (2016), IMF (2017), Frost et al. (2019), Jagtiani and Lemieux (2019), OECD (2019), FSB (2019)) 특별히 Philippon은 미국의 경우 1980년대 초반 이후 금융서비스 부문의 EY가 지속적으로

²¹ 이에 대한 요약은 Cho (2019) 참조

과도한 수준을 보이고 있고 (장기균형 수준에 비하여), 이로 인하여 금융서비스 비용이 지속적으로 지나치게 높았으며 (“Finance has been too expensive!”), 이는 1990년대 초반부터 진행된 미국 은행들의 M&A를 통한 대형화의 결과인 것으로 해석하고 있다. 이와 같은 상황에서 미국의 핀텍산업은 기존 금융부문에 대한 경쟁을 심화시킴으로써 금융중개의 EY를 낮추고 효율성을 높일 수 있는 기제인 것으로 논의하고 있다. 또한, 핀텍산업의 활성화는 개도국 및 선진국 모두에서 금융서비스의 대상을 확대시킴으로써 금융포용의 효과가 상당한 수준으로 발생하고 있는 사실 또한 문헌에서 논의하고 있다. (Jagtiani et al. (2018), De Roure et al. (2108))

데이터경제의 관점에서 핀텍산업은 대안적인 데이터(alternative data)를 사용하여 기존의 금융개서비스에서 발생하는 정보의 비칭성(information asymmetry) 문제를 경감시키는 것으로 논의되고 있다. (Lin et al. (2013), Puri et al. (2017), Hildebrand et al. (2017), and Freedman and Jin (2018), Berg, Burg, Gombović, and Puri (2018)) 아래에서는 이와 관련된 3가지 사례를 살펴 보기로 한다: (1) 신용평가(credit evaluation)에서의 효율성 제고; (2) 로보-어드바이저 (Robo-Advisor)를 활용한 투자자문서비스의 효율성 및 형평성 제고; (3) 레그텍(RegTech)을 활용한 규제준수(regulatory compliance) 업무의 효율성 제고.

첫째, 핀텍산업은 온라인 상에서 수집한 금융소비자의 행태와 관련된 제반 데이터를 활용하여 (이를 “soft data” 또는 “digital footprint”로 지칭) 신용평가를 수행하고, 이와 같은 방식은 기존의 신용평가 모형에 비하여 정확도를 상당한 수준으로 높이는 것으로 보고되고 있다. (Lin et al. (2013), Puri et al. (2017), Hildebrand et al. (2017), and Freedman and Jin (2018), Berg et al. (2018)) 대안 자료의 예로는 차입자의 소셜네트워크(social/friend network)와 이들의 신용도, 지역 경기 (local economy) 및 기타 정보, 온라인에서의 상품구매 패턴 등이 포함되고, 이와 같은 데이터는 블록체인 등을 활용하여 통상 네트워크 내부에서 공유된다. 이와 같은 대안적 자료의 사용은 기본적으로 정적인 형태의 소비자-기업 신용평가를 보다 동적인 형태로 전환하고, 이 과정에서 현재 시점의 소비자 행태를 보여주는 다양한 데이터와 함께 새로운 분석기법을 통하여 (예: 머신러닝, AI, text analytics, cognitive computing 등) 신용평가 모형을 구축한다. 이와 관련하여 Berg et al.은 신용평가모형의 정확도에 대한 실증분석을 AUC (Area Under Curve) 방식을 통하여 수행하였고 ([그림 7] 참조),²² 기존의 소비자신용등급 만을 모형에 포함하는 경우 AUC 68.3%였으나, 신용등급과

²² AUC는 신용평가모형의 정확도 분석에 사용되는 통상적인 지표이고, 이는 기본적으로 지니계수와 같은 원리로 추정한다 (AUC 확률이 높을수록 신용평가모형의 정확도가 높은 것으로 해석). 예를 들면, 10,000건의 대출기록 중 100건의 실제 연체가 발생한 샘플을 기초로 2개의 연체확률모형 A & B를 추정하고, 각각의 모형에 기초해서 10,000건 모두의 연체확률을 추정하여 높은 확률 (high-risk) 부터 정렬하여 (rank-ordering from the high-risk cases) X축으로 하고, Y축은 각각의 모

함께 대안적 데이터를 모형에 포함하는 경우 AUC가 73.6%로 5.3%p 증가한 것으로 보고한다 (5%p의 AUC 증가는 통상 매우 높은 수준의 정확도 증가로 해석됨).

이와 같은 신용평가모형의 정확도 제고는 리스크관리의 효율성을 높일 뿐만 아니라, 핀텍 기업들이 다양한 온라인-오프라인, 그리고 금융-비금융 부문으로 업역을 확대하여 생산성 제고 및 고용창출에 기여하는 것으로 보고되고 있다. 예를 들면, Alibaba의 자회사인 Ant Financial은 (2018년 1월 기준 세계 10위의 금융서비스 회사) ‘Sesame Credit’이라는 신용평가회사를 통하여 소비자신용등급을 산정하고, 이를 위해서 계열사인 AliPay와 같은 모바일지급결제시스템을 통하여 수많은 소비자에 대한 정보를 실시간으로 수집하여 빅데이터화 하고, 이를 신용평가에 반영하고 있다. 또한 산출된 신용등급은 계열사들이 운영하는 보험, 여행, 레저, 의료 등의 다양한 서비스업권에 사용되고 있다. 이와 같은 빅텍(BigTech) 기반의 핀텍업체들은 보유하고 있는 방대한 소비자 네트워크를 활용하고, 다양한 업역간의 범위의 경제 효과를 증대시킴으로써 고용창출 및 산업생산성 제고에 기여하는 것으로 논의되고 있다. (Citi GPS (2018), Frost et al. (2019))

둘째, 로보-어드바이저(RA)는 이미 투자자문 분야에서 기존의 사람 중심의 서비스를 빠르게 대체하고 있다 (로보-어드바이저의 통상적인 서비스 절차는 [그림 7] 참조). 예를 들면, 바둑의 AlphaGo 경우와 마찬가지로, S&P사가 보유한 RA인 Kensho는 15명의 월스트리트 주식분석가들과의 대결에서 이들이 10여일에 걸쳐서 분석한 정형·비정형 데이터를 단 몇 시간만에 분석을 완료하였다. 또한 Kensho와 같은 RA는 주식수익률 등의 숫자 기반 데이터 뿐만 아니라 기업의 연차보고서, 신문기사 등의 텍스트 데이터도 빠른 속도로 분석하여 최적의 투자 포트폴리오를 제시하고, 이를 동적으로 조정(rebalancing) 하는 것으로 알려져 있다. (Citi GPS, 2018)

소비자 복지 차원에서는 기존의 사람 중심의 투자자문서비스는 고객층이 주로 고소득·고자산계층에 집중되어 있는 반면 (높은 수수료 및 최소계정잔액 요구 등의 이유로), 로보-어드바이저는 투자자문서비스를 저소득층을 포함하여 보다 확대한 것으로 알려져 있고 (이를 “democratization of investment consultancy”로 표현), 이는 추가 소비자에 대한 RA서비스 제공의 한계비용이 기본적으로 영에 가깝다는 점에서 기인한다 (물론 개발단계에서의 투자비용은 큰 수준으로 발생). RA는 일반적인 금융이론에 입각하여 (예: MPT, Modern Portfolio Theory, & CAPM, Capital Asset Pricing Model 등) 다양한 정형·비정형 데이터를 통하여 지속적으로 학습하는 머신러닝의 방식이 AI-알고리즘의 핵심이라고 할 수 있

형에서 high-risk 구간에서 (예: top 25%) 실제 연체건수의 비율을 (전체 100건의 연체 중) 나타냄. A & B의 AUC가 각각 65% & 70%면 B모형의 정확도가 더 높은 것으로 해석.

다. 그러나 이의 결과로 제시되는 포드폴리오에 대한 설명이 쉽지 않다는 “blackbox effect”가 결점으로 지적되고 있다.

셋째, 레그텍(RegTech)을 활용한 사기성·위법성 금융거래의 사전적 방지가 핀테크 부문에서 데이터를 활용한 또 하나의 사례로 들 수 있다. 글로벌금융위기 이후 자금세탁 방지 (Anti-Money Laundering, AML), 테러리스트 자금조달 차단 (Anti-Terrorist Funding) 등에 대한 규정이 강화되었고, 이를 위반하는 경우 엄청난 규모의 벌금이 부과될 수 있다는 점이 (실제로 수십억 달러의 벌금이 부과된 사례 존재) 레그텍산업이 빠른 성장을 이룰 수 있었던 배경으로 알려져 있다. (주요 레그텍 서비스 제공자로는 Ayasdi (2008, US), RedOwl Analytics (2011, US), Elliptic (2013, UK), Compliance Advantage (2014, UK), Saude (2014, UK) 포함) [그림 8]에서 보는 바와 같이, 레그텍은 AI, 블록체인, 바이오메트릭스, Digital Twin 등의 다양한 기술을 사용하여 금융기관 내부·외부의 데이터를 분석하고, 이를 통하여 위법적·사기성 대출 탐색, 전사적 리스크관리, 규제보고서 작성의 자동화, 동적 시장 분석 등의 기능을 담당하고 있다. (통상적인 레그텍의 영역에 대하여 CCAF (2019)는 다음을 포함 - KYC - Know-Your-Customer, AML, ERM - Enterprise Risk Management, fraud detection, compliance to privacy laws, and various other applications) 향후 데이터경제가 더욱 활성화 되면 레그텍 부문의 성장도 더욱 가속화될 전망이고, 이를 통하여 새로운 금융서비스 모델의 동반 성장도 (예: 마이데이터, 데이터익스체인지 등) 예상해 볼 수 있다.

3. 공공부문: 스마트시티 (SmartCity)

공공부문에서의 데이터경제 사례로 ‘스마트시티’(SmartCity, SC)를 들 수 있다. SC는 기존의 도시 개발·운영에 대한 일반적 원칙을 공유하지만 (즉, 주민의 안전과 편의를 위한 제반 서비스를 안정적·효율적으로 제공하고, 인구·고용·주택·상업용부동산·도시인프라 등의 하부시장의 여건을 파악하여 예산제약 하에서 최적의 계획을 수립·집행), 다음의 몇 가지 측면에서 차별화되는 특성이 존재하는 것으로 파악되고 있다: (1) 혁신적 정보통신 기술(ICT) 및 공공·민간의 데이터를 활용한 차세대 도시서비스(“advanced urban services”)의 개발 및 제공; (2) 정책수립 및 서비스제공에 있어서 주민들이 프로슈머(prosumer)로서 수요·공급 양 측면 모두에서 보다 적극적으로 참여; (3) 이의 효과적인 추진을 위해서 관련 산업·R&D부문·공공부문 간의 혁신생태계를 조성하여 이들 간의 협업을 통하여 정책결정 및 공공서비스 제공. ([그림 9] 참조) (조만·이태준 (2019), 이태준·조만 (2020))

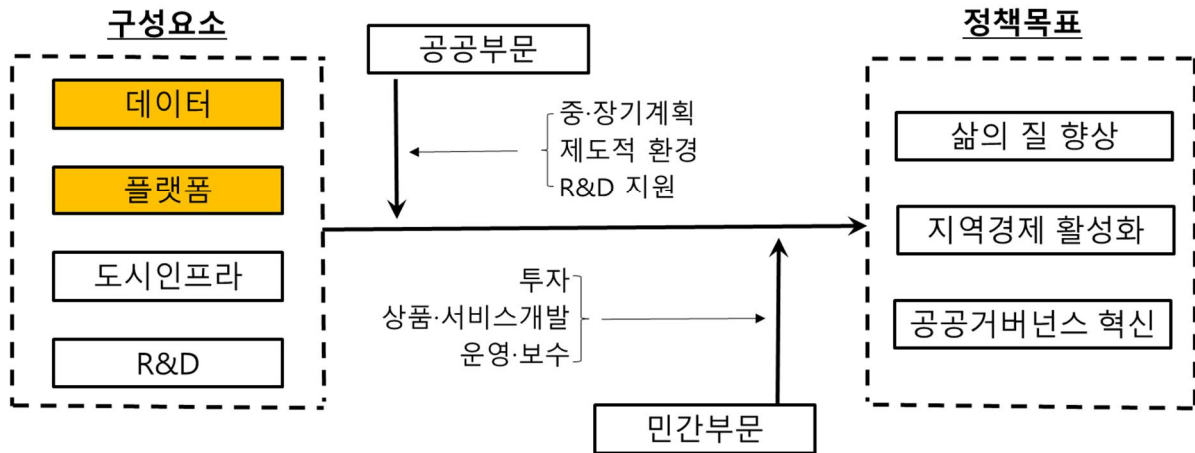
SC는 한국·중국·싱가포르 등의 아시아 국가들 뿐만 아니라 영국·네덜란드·핀란드 등의 유럽 국가에서도 국가 차원의 경제·사회개발 수단으로 추진하고 있고, 이의 실현을 위해서 다양한 형태의 디지털트윈 방식을 사용하여 데이터의 효용을 높이고 있다. 스마

트시티는 기본적으로 스마트X 산업의 집합체로 볼 수 있고, 이태준·조만(2020)에서는 Ismagilova et al. (2019)의 논의를 기초로 아래 8개 SC 도메인에 대한 우리나라 광역 지방자치단체의 스마트시티 사업 추진현황을 분석하였다: 스마트모빌리티, 스마트 리빙, 스마트 환경, 스마트 시티즌, 스마트 (지역)경제, 스마트 아키텍춰, 스마트 금융. 현재 우리나라의 스마트시티 사업은 매우 초기 단계인 것으로 파악되고 있고, 조만·이태준 (2019)는 이의 활성화를 위하여 제반 분야의 정책방향을 제시하고 있다.²³ 스마트시티 추진과 관련된 현재 파악된 국내외 우수사례는 아래와 같고, 향후 이에 대한 보다 심도있는 분석이 필요한 상황이다.

- Sitra, and the Kalasatama-Otniemi cluster (Helsinki, Finland); “Fair Data Economy”
- Innovate UK by FCA (Financial Conduct Adm.); “Consumer-oriented financial innovation” focusing on the FinTech hub in London (& the PolicyLab initiative)
- “Bristol is Open,” Bristol, UK; For PPAP
- Copenhagen Data Exchange, Denmark
- Amsterdam LivingLab case, Netherland
- Barcelona, data-driven SC, Spain
- MyData by PingAn Insurance Group, Shenzhen, China
- Dallas and Austin, Open Data, the U.S.
- Two SC experiments in Korea, in Sejong City and Busan
- CLC (Center for Livable Cities) and LARC (Living Analytics Research Center), Singapore
- Fujisawa township, Japan; For resident participation for increasing QOL

²³ 첫째, 스마트시티에 대한 범정부적 비전 및 중장기 전략을 제시 및 이의 효과적인 실행을 위한 관련 부처들 간의 정책공조; 둘째, 미국·영국·일본의 사례를 참조하여 지방정부 주도의 스마트시티 모델을 추진하고, 이 경우 대·중·소기업 및 스타트업 간의 협업, 그리고 연구개발 부문과의 공조 강화; 셋째, 데이터·플랫폼 기반의 스마트시티를 구현하기 위해서 공공·민간 소유의 데이터가 구축·공유·분석하고, 이를 위해서는 개인정보의 균형있는 보호와 활용을 위한 법적·제도적 장치 마련; 넷째, 공공거버넌스 개혁을 실현하기 위해서 지역별로 차별화된 경제·사회문제를 식별하고, 이의 해결을 위한 정책의 계획 및 실행 단계에서 리빙랩 방식을 통하여 경제주체들의 참여를 확대하고, 또한 정책의 수립·실행·평가의 전 단계에서 정책랩 방식 도입.

스마트시티: 구성요소, 정책목표, & 공공 민간의 역할



[그림 9] 스마트시티: 구성요소, 정책목표, & 공공 민간의 역할 (출처: 조만·이태준 (2019))

최근의 COVID19 사태와 관련하여 SC는 추가 확진자 방지를 위한 데이터 분석에 활용된 것으로 알려져 있다. (Sonn et al. (2020)) 본 사태에 대한 대응과 관련하여 한국은 국제적 모범사례로 평가받고 있고, 여기에는 3T(Testing·Tracing·Treatment)의 모든 측면에서 빠르고 효과적인 대응이 있었던 결과인 것으로 논의된다. (Kim (2020)) 구체적으로, 확진자의 동선 추적을 위하여 다양한 데이터가 빠르게 수집·분석되었고, 여기에는 다음 사항이 포함된다: (1) 추가 확진자에 대한 인터뷰와 CC-TV 등을 통한 동선 파악; (2) ‘정보화진흥원’ 등에서 구축한 ‘Smart City Data Hub’(대구 등의 도시에 존재)을 활용하여 경찰·신용거래연합(the Credit Finance Association)·스마트폰서비스제공자·신용카드사(22개 업체)들이 보유한 데이터를 신속하고 일괄적인 방식으로 취합·분석; (3) 필요한 경우 추가적으로 교통카드 데이터를 이용하여 확진자의 동선에 대한 분석 (대중교통 수단의 탑승 및 하차 정보 등 사용); (4) 파악된 개별 확진자의 동선을 기초자치단체별로 주민들과 공유. (Sonn et al. (2020), Park (2020), Choi, 2020, The Government of the Republic of Korea (2020)) 향후 본 사례를 데이터경제 관점에서 더욱 심도있게 분석할 필요가 있다. 이 경우 SC가 전염병 대응행에 대한 대응과 같은 재난리스크 관리의 차원 뿐만 아니라, 지역경제의 복원(“urban resilience”)과 같은 보다 복합적인 차원에서의 연구도 고려해 볼 만 하다.

[부록 2] 금융부문 데이터경제 촉진기제: 현황

금융부문 데이터경제의 촉진기제 (1): 데이터 거래소

데이터 거래소는 공공·민간에 산재한 각종 데이터를 수집하여 빅데이터화 하고, 이를 기초로 수요자의 필요에 부합하는 다양한 형태의 데이터를 생성하여 판매하는 플랫폼이라고 할 수 있다. 세계 최초의 데이터거래소는 2014년 12월에 설립된 중국의 구이양빅데이터거래소(GBDE)로 알려져 있고, 현재 중국에는 이 외에도 20여 곳의 데이터거래소가 운영 중이다. GBDE는 자본금의 36%가 국유자본이고, 중국의 대표 IT기업인 텐센트, 알리바바 등을 포함해 약 2,000개 회원사가 참여하고 있으며, 이 중 225개 기업은 직접 데이터를 판매하고 있다. 유럽과 미국 또한 각각 917개, 650개에 달하는 데이터 거래소가 운영 중에 있고, 일본은 2018년 최초의 데이터거래소가 민영으로 출범하였다.

우리나라 최초의 데이터 거래소는 한국정보화진흥원(NIA)과 매일방송(MBN)이 공동으로 구축하여 2019년 12월 출범한 한국데이터거래소(KDX)이고, 여기에서는 경제·산업, 금융, 보건의료, 소비, 유통, 통신, 물류, 부동산, 상권·동선, 시세, 정보·검색, SNS, 공공 데이터, 미디어 등 14개 카테고리의 데이터를 거래할 수 있다. 공공부문의 데이터거래소는 현재 금융데이터거래소(금융위원회), 빅데이터 플랫폼(과기부), 데이터스토어(데이터산업진흥원) 등 10개가 운영 중이고, 중소벤처기업부, 서울시 등에서도 자체적으로 준비하고 있다. 민간에서도 네이버, SKT, LGCNS 등의 플랫폼이 10개에 달하는 것으로 알려져 있다. 최근 정부는 각종 데이터 거래소 및 거래 플랫폼을 통합한 데이터 거래소의 구축을 추진하고 있으며, 한 걸음 더 나아가서 ‘데이터청’의 신설까지도 검토하고 있다.²⁴

금융데이터거래소는 2020년 5월에 설립되었고, 9월 12일 현재 거래건수는 477건에 달한다 (판매액은 3억 5700만원으로 집계).²⁵ 본 거래소에서는 데이터 결합 전문기관인 신용정보원과 금융보안원의 개인정보의 비식별화 및 데이터 결합과정을 거친 자료를 거래하고 있다. 최근에는 비금융사 최초로 ‘네이버’가 AI 기반의 ‘네이버 클라우드 데이터 샌드박스’를 금융데이터거래소에 공개할 예정임을 발표했는데, 이는 스타트업·대학연구진·공공기관을 대상으로 네이버가 구축한 2가지 유형의 데이터와 (온라인 쇼핑 트렌드 데이

²⁴ 조선일보 (2020.8.7)

²⁵ 한국경제 (2020.9.12)

터 및 지역 비즈니스 데이터) 분석 도구를 제공할 예정인 것으로 알려져 있다.²⁶ 향후 본 거래소에는 신용카드사를 포함한 금융기관, 편의점 체인, 부동산 관련 회사 등 다양한 사업주체가 참여할 수 있을 것으로 보인다 (예: 주택정보회사인 ‘공감랩’은 주소기반 시세정보와 거래 사례 등의 데이터를 실시간으로 업데이트해서 받아볼 수 있도록 하는 서비스 제공).²⁷

금융부문 데이터경제의 촉진기제 (2): 마이데이터 (MyData)

마이데이터는 본인 정보의 주체인 개인의 동의에 따라 본인 데이터를 제3자에게 개방할 수 있도록 하고, 마이데이터 사업자는 분산되어 있는 개인 금융정보를 통합 관리하고, 데이터 분석에 기반한 개인의 자산운영에 제언을 하는 제도를 의미한다. 해외에서는 이미 이 부문의 선도기업들이 출현하고 있고 (미국의 Mint, 영국의 Digi.Me, 중국의 핑안보험 등), 금융혁신 차원에서 본 부문의 활성화를 추진하고 있다. 우리나라에서도 현재 금융위원회가 마이데이터 업체에 대한 평가를 진행 중이고, 2021년 상반기에 40개의 마이데이터 사업체를 인가할 계획으로 있다.

우리나라의 마이데이터 사업은 금융·통신 분야에서 우선적으로 시행하고, 의료·유통·에너지·학술 등의 분야로 확대할 계획이다. 금융부문의 마이데이터 사업인 ‘본인신용정보관리업’의 고유업무는 신용정보 제공·이용자 또는 공공기관이 보유한 개인신용정보를 수집하고, 수집된 정보의 전부 또는 일부를 신용정보주체가 조회·열람할 수 있게 하는 업무로 규정되어 있다 (『신용정보의 이용 및 보호에 관한 법률』제2조 제9호의 2, 시행령 제2조 제21항). 그러나 본 사업자는 이 외에도 경영업무를 수행할 수 있고, 여기에는 로보어드바이저를 통한 투자자문·일임서비스, 금융상품자문업, 대출의 중개·주선, 전자금융업, 신용정보업 등이 포함된다. 또한 데이터 분석·컨설팅, 정보계좌 업무, 정보관련 권리(예:프로파일링 대응권) 대리행사 업무, 본인신용정보관리업 관련 연수·교육·출판, 금

²⁶ 중앙일보 (2020.11.22)

²⁷ 의료부문 빅데이터와 관련하여 4차산업혁명위원회 산하 헬스케어특별위원회는 미래 헬스케어 산업을 선도하기 위한 6대 핵심 추진과제를 설정하고, 이의 일환으로 본인 동의를 기반으로 헬스케어 데이터를 종합 분석할 수 있는 스마트 임상센터 구축을 제안하고 있음. 이는 임상시험센터 간 통합 네트워크를 구축하고 첨단 기술을 임상시험에 접목하는 방안을 포함하고 있음. 또한 신약 개발에 소요되는 시간과 비용을 단축하기 위해 인공지능 활용 방안을 적극 추진하고, 신개념 융복합 의료기기에 대한 표준·규격 마련 등 신속한 인허가 진행을 위한 제도 개선을 계획 중임.

용상품 광고·홍보, 본인인증 및 식별확인 업무 등의 매우 다양한 부수업무를 수행할 수 있다.

》 2019년 마이데이터 실증서비스 선정 과제

분야	과제명	수행기관		서비스
		주관	참여	
의료	의료 마이데이터 플랫폼 및 검진 데이터를 활용한 건강관리 서비스	강남세브란스병원 (연세대학교 산학협력단)	아름정보기술, CJ프레시웨이, 에스푸드	건강 검진·처방전 등의 데이터를 이용한 영양 건강식단 추천
	응급상황을 위한 개인건강지갑 서비스	브이티더블유	삼성서울병원, 서울아산병원, 동아대학교병원	응급환자의 응급 진료기록 및 일상 생활 속 건강기록 보관, 개인건강지갑 제공
	MyHealth Data 플랫폼 및 서비스 실증	서울대학교병원	차의과학대학교 산학협력단, 메디블록, 웰트, 삼성화재	환자 동의에 따른 개인 의료정보 기반의 건강정보 교류 플랫폼 개발, 라이프로그 데이터를 활용하여 개인 건강진진 코칭
금융	본인정보 통합조회 및 생애주기별 맞춤형 금융상품 추천 서비스	NHN페이코	KEB하나은행, 한화생명보험, 한화투자증권, 한화손해보험, 신한금융투자, 웰컴저축은행	금융·비금융 빅데이터 기반 맞춤형 금융상품 추천
에너지	사용자 맞춤형 에너지 절감 서비스	다음소프트	세종시, 에이엠에이닷컴, 유디아이	세종시 주민 대상 에너지 데이터 수집 및 에너지 수급 예측, 맞춤형 서비스 구현
유통	개인데이터 저장소 기반 소상공인 마케팅 관리 서비스	한국신용데이터	신한카드	사업장 정보를 활용한 소상공인 경영 개선, 개인 대상 할인·이벤트 정보 제공
	소상공인 성장 지원 문서·자금 플랫폼 서비스	한국기업데이터	한국전자세제협회, 비즈니스온 커뮤니케이션, 기용정보통신	소상공인(개인사업자)을 위한 정책자금 매칭 등 종합경영관리 서비스 제공
기타	연구자를 위한 매칭 및 분석 서비스	코난테크놀로지	국회도서관	국회도서관으로부터 수집된 학술연구자 정보를 활용하여 저작물 통합관리

Source : 과학기술정보통신부(2019.05), '의료·금융·에너지 등 마이데이터 서비스 8개 과제 선정', 삼성KPMG 경제연구원 재구성

금융부문 데이터경제의 촉진기제 (3): 오픈뱅킹 (Open Banking)

오픈뱅킹은 은행권과 핀테크 기업이 스마트폰에 설치한 응용프로그램(앱)을 통하여 공동으로 이용할 수 있는 결제시스템으로, 참여하는 모든 금융기관의 계좌에서 결제를 비롯해 잔액 조회, 거래내역 조회, 계좌실명 조회, 송금인 정보조회, 입금입체, 출금이체 등의 금융서비스를 실시간으로 이용할 수 있는 시스템을 의미한다. 금융소비자 입장에서는 여러 앱을 설치할 필요 없이 하나의 은행 앱 또는 핀테크 앱만으로 모든 은행의 계좌를 조회·이체할 수 있는 서비스다. 따라서 소비자의 편익이 제고될 수 있고, 참여하는 금융기관들 간의 서비스 경쟁이 심화되어 금융산업 전반의 혁신이 촉진되는 효과를 기대할 수 있다.

우리나라에서는 2019년 12월부터 운영되고 있고, 2020년 9월 현재 누적 사용자가

5,185만명인 것으로 집계되고 있다. 기능별로는 총 이용건수의 약 80%(3,972만 건)가 잔액조회에 이용되고 있고, 2019년 11월 28일 기준 성인 1인당 2.3개의 계좌를 오픈뱅킹에 등록하고, 일 평균 165만 건이 이용되고 있다. (영국의 오픈뱅킹은 2018년 1월 시행된 이후 이를 통한 API 호출 건수는 2018년 6월 190만 건에서 2019년 8월 1억 1,050만 건으로 약 58배 증가하였으며, 총 180개의 기관이 이에 참여하고 있다.) 현재 오픈뱅킹 참가기관을 은행·핀테크사에서 타 금융업권으로 확대하는 것을 추진하고 있고, 여기에는 상호금융 중앙회 (농협, 수협, 신협, 새마을금고, 저축은행, 산림조합), 우정사업본부 및 17개 증권사가 확대 대상이다.

[부록 3] 데이터경제, 핀테크, 프롭테크, 디지털헬스 관련 언론보도 분석

